E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.41;TP18　　文献标识码:A　　文章编号:1006-8961(2026)07-2408-26

论文引用格式:Xue X Q, Wen J, Liu X W and Zhang J P. 2026. Prior-driven multi-view clustering: a perspective from geometry and semantics to LMM cognition. Journal of Image and Graphics, 31(7):2408-2433(薛旭倩,文杰,刘新旺,张军平. 2026. 先验驱动的多视图聚类:从几何与语义空间到多模态大模型认知前瞻. 中国图象图形学报, 31(7):2408-2433)[DOI:10.11834/jig.260125]

先验驱动的多视图聚类:从几何与语义空间到多模态大模型认知前瞻

薛旭倩[1],文杰[2],刘新旺[3],张军平[1*]

1. 复旦大学计算与智能创新学院,上海　200433; 2. 哈尔滨工业大学(深圳)计算机科学与技术学院,深圳　518055;
3. 国防科技大学计算机学院,长沙　410000

摘　要: 随着多模态大模型与海量异构数据的涌现,多视图聚类(multi-view clustering, MVC)作为无监督知识发现与数据底层关联挖掘的核心技术,其研究范式正经历深刻变革。现有综述多局限于底层算法网络结构的横向归纳,难以揭示该领域在不同技术时代的内在发展逻辑。为此,本文打破传统分类框架,首创性地提出先验驱动的理论视角,对多视图聚类20年来的发展脉络进行了系统性重构。首先,梳理了基于几何先验的浅层结构挖掘,分析了欧氏原型、仿射子空间与流形邻域中的显式数学约束机理;其次,归纳了基于语义协同先验的深层空间建模,揭示了模型如何在嵌入、隐空间、增强及拓扑空间中捕获非线性的跨视图一致性;最后,前瞻性地探讨了基于多模态大模型认知先验的深度对齐,阐述了聚类技术赋能海量数据治理、混合专家(mixture-of-experts, MoE)路由及检索增强生成(retrieval-augmented generation, RAG)的基础设施作用,并分析了多模态大模型逻辑推理反哺聚类任务的潜在机遇。本文通过构建几何—语义—认知的跨范式分析框架,深刻揭示了多视图聚类由底层数据驱动向高阶知识驱动转型的内在逻辑。在此基础上,分析了类别分布长尾失衡、视图严重缺失及评估体系滞后等开放环境下的核心挑战,并探讨了相应的解决思路,旨在为多视图聚类在多模态大模型时代的理论创新与工程实践提供新的研究路线。
关键词: 多视图聚类(MVC);先验驱动学习;多模态大模型(LMMs);几何结构;语义协同;认知对齐;综述

Prior-driven multi-view clustering: a perspective from geometry and semantics to LMM cognition

Xue Xuqian[1], Wen Jie[2], Liu Xinwang[3], Zhang Junping[1*]

1. College of Computer Science and Artificial Intelligence, Fudan University, Shanghai　200433, China;
2. College of Computer Science and Technology, Harbin Institute of Technology, Shenzhen　518055, China;
3. College of Computer Science, National University of Defense Technology, Changsha　410000, China

Abstract: With the rapid emergence of large multimodal models (LMMs), massive heterogeneous data are proliferating across various industrial and scientific domains. In this context, multi-view clustering (MVC) serves as a cornerstone technology for unsupervised knowledge discovery and latent correlation mining. At present, MVC is undergoing a profound and historical paradigm shift. Traditional surveys predominantly focus on the horizontal categorization of algorithmic network structures. However, this approach often fails to reveal the intrinsic evolutionary logic across different technological eras.

收稿日期:2026-03-09;**修回日期:**2026-04-04;**预印本日期:**2026-04-11
*通信作者:张军平　jpzhang@fudan.edu.cn
基金项目:国家自然科学基金项目(62576103)
Supported by: National Natural Science Foundation of China(62576103)

Departing from these conventions, this study proposes a pioneering, prior-driven theoretical perspective to reconstruct systematically the developmental trajectory of MVC over the past two decades. This goal is achieved through a trans-paradigm analytical framework of geometry-semantics-cognition. During the initial stage of shallow structural mining, the research paradigm focused on explicit mathematical constraints within original or kernel-induced feature spaces. Euclidean space methods, such as multi-view K-Means and nonnegative matrix factorization, identify global prototypes by minimizing squared error or Frobenius norm reconstruction loss. By contrast, affine space methods leverage self-representation properties to model data as a union of low-dimensional subspaces, while manifold space techniques utilize spectral graph theory to transform clustering into optimal graph cut problems by preserving local topological correlations. As the field transitioned into deep spatial modeling based on semantic collaborative priors, researchers utilized the powerful nonlinear mapping capabilities of deep neural networks to project heterogeneous data into high-order semantic spaces. This evolution encompasses several distinct research paradigms. 1) Embedding space research focuses on deep subspace clustering, employing autoencoders to learn discriminative features while maintaining cross-view consistency. 2) Latent space methods utilize probabilistic generative models, such as variational autoencoders and generative adversarial networks, to align latent distributions and infer missing view information through adversarial games. 3) Augmented space paradigms introduce contrastive learning to maximize mutual information between views via InfoNCE-like losses, enhancing representation robustness. 4) Topological space studies leverage graph neural networks to mine intra-view geometric structures and inter-view complementary semantics synergistically. Moving into the current era of LMMs, this study progressively explores deep alignment based on cognitive priors, where LMMs are viewed not merely as data sources but as knowledge bases that contain human-level common sense and logical reasoning capability. We systematically elucidate the infrastructural role of MVC in empowering massive data governance. Specifically, MVC facilitates semantic deduplication to enhance data quality and employs token-level clustering to optimize mixture-of-experts (MoE) routing for expert specialization. Furthermore, MVC enables hierarchical semantic chunking, which is critical for precise document retrieval within retrieval-augmented generation (RAG) frameworks. Beyond these applications, we analyze the potential for LMM logical reasoning to back-propagate into clustering tasks. This synergy elevates MVC from pure statistical feature alignment to a new dimension of knowledge-driven cognitive logic consistency. Beyond theoretical frameworks, the review highlights the transformative effect of MVC in diverse real-world scenarios, ranging from multi-omics fusion in smart healthcare for cancer subtyping and spatial-spectral fusion in remote sensing for urban functional zone identification to cross-perspective pedestrian recognition in public security and unsupervised anomaly detection in industrial internet of things networks. Despite these advancements, achieving a transition from laboratory benchmarks to robust industrial infrastructure requires addressing several core challenges. First, the scaling bottleneck where in the $O(N^2)$ or $O(N^3)$ computational complexity of traditional methods must be reduced to linear levels to handle million-scale web data. Second, the extreme robustness required in open environments to handle not only severely incomplete views but also the pervasive long-tailed category imbalance where minority abnormal samples are frequently overwhelmed by dominant patterns. Third, the urgent need for evaluation system reconstruction to move away from legacy small-scale datasets with shallow features toward native heterogeneous multimodal benchmarks that reflect real-world weak alignment and complex noise distributions. Ultimately, this survey aims to provide a novel research road map for theoretical innovation and engineering practice, advocating for a transition toward intelligent decision systems characterized by high-level semantic decoupling and cognitive logic alignment in the era of LMMs.

Key words: multi-view clustering (MVC); prior-driven learning; large multimodal models (LMMs); geometric structure; semantic collaboration; cognitive alignment; review

0 引 言

随着信息采集技术的发展,多视图数据已成为大数据环境下的普遍存在形式(Breit 等, 2020;Gao 等,2020a)。多视图数据是指从不同维度、模态或来源对同一事物进行刻画的数据集合,例如同一事件的文字、图像与视频描述,或同一物体的不同角度拍摄。多视图聚类(multi-view clustering,MVC)作为无监督学习领域的核心任务,旨在通过挖掘多个视图间

的共性一致性信息与特性互补性信息,将样本划分为若干具有语义意义的类簇。相比单视图聚类,多视图聚类能够有效克服单一数据源的不确定性与片面性,为复杂对象的关联分析提供更具鲁棒性的表征。

近年来,多视图聚类的研究重心已发生显著的范式转移,从依赖浅层矩阵分解与图投影的统计特征对齐,发展至基于深度表示学习的高维语义空间建模。当前研究的核心诉求在于攻克多模态数据在真实场景下的复杂性(Liu等,2026)。具体而言,这些挑战集中在三大维度:首先是结构一致性与视图特异性的权衡,即如何在诱导全局共识表征的同时,精准保留各视图底层的局部流形和拓扑,避免在深度网络中发生表征塌陷;其次是隐空间的解耦与高阶关联挖掘,旨在从高维冗余与强噪声干扰中学习出视图间的本质语义,并捕获超越简单成对关系的复杂高阶交互(如借助张量或超图机制);最后是开放环境下的鲁棒性建模,旨在应对现实中普遍存在的多视图数据缺失(incomplete multi-view clustering, IMVC)(Lin等,2021a)与类别分布不平衡(Xue等,2025)问题。尽管深度学习凭借极强的非线性拟合能力提升了MVC的性能基线,但如何构建具备可解释性且符合人类认知先验的聚类框架,仍是该领域待解决的核心问题。

现有的多视图聚类综述已对传统几何方法、多核学习及深度表示学习等方法进行了系统的梳理与归纳。例如,Li等人(2019a)侧重于多视图表示的对齐与融合机制。Fang等人(2023)与Zhou等人(2024)详细评估了传统方法与深度学习方法在各类基准上的性能表现。Liu等人(2026)全面总结了过去20年基于图、矩阵分解及张量等算法的技术谱系。然而,这些工作大多侧重于算法网络结构的横向分类,倾向于将聚类视为独立于下游应用的纯数学优化与模式识别问题。在当前由多模态大模型(large multimodal model, LMM)驱动的技术范式下,现有综述普遍缺乏对聚类与多模态大模型协同发展视角的审视,忽视了两个关键维度:一是自下而上的基座支撑维度,即聚类技术如何作为基础组件,服务于多模态大模型预训练的数据治理、混合专家架构(mixture-of-experts, MoE)中的动态表征路由,以及检索增强生成(retrieval-augmented generation, RAG)中的多模态语义分块与聚合;二是自上而下的认知反哺维度,即现有的视图融合多局限于浅层统计与几何层面的对齐,忽视了如何利用大模型的泛化认知与推理能力为聚类任务注入逻辑先验,从而将多视图聚类从单纯的数据驱动升维至知识驱动的认知任务。

鉴于传统以算法网络结构(如图学习、矩阵分解等)为导向的分类法容易割裂聚类任务在不同技术时代的内在逻辑发展,本文首创性地提出了一种先验驱动的全新多视图聚类分类体系,如图1所示。

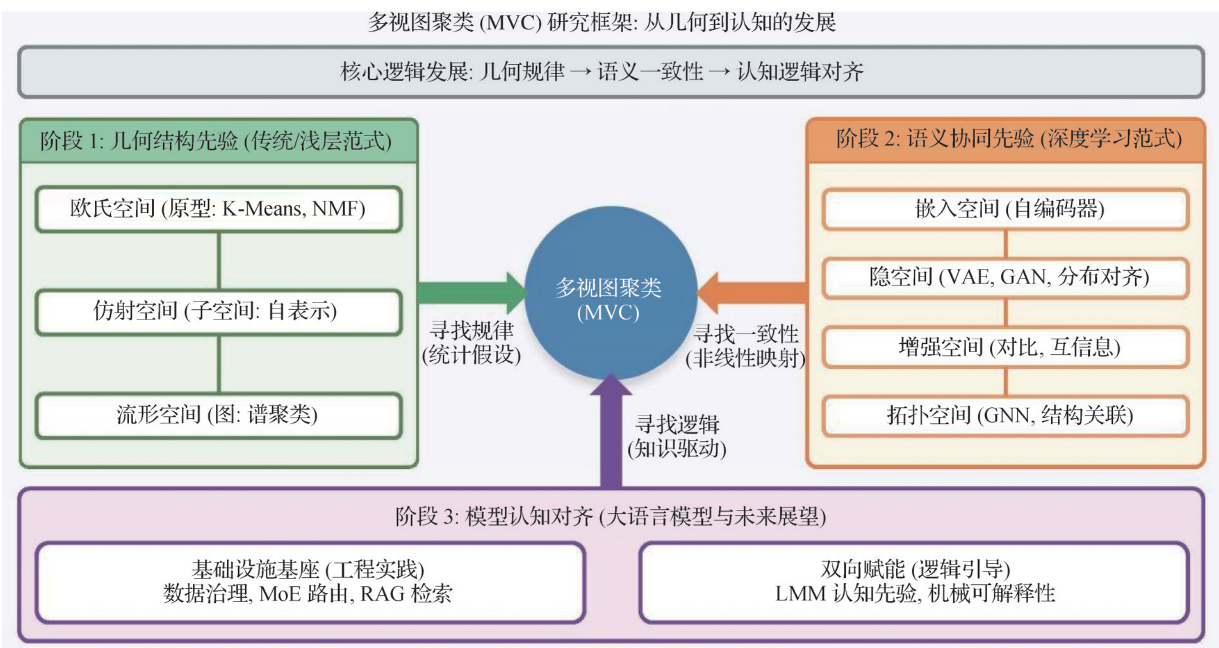


图1　多视图聚类研究框架

Fig. 1　Research framework of multi-view clustering

具体而言，通过构建发展脉络，将现有的多视图聚类研究重构为三大核心范式：1）基于几何先验的浅层结构挖掘（强调多流形假设与统计分布的对齐）；2）基于语义先验的深层空间建模（聚焦非线性空间的特征解耦与高阶关联）；3）基于认知先验的多模态大模型对齐（探索多模态大模型逻辑增强、知识引导及基础组件）。

为直观呈现多视图聚类领域的发展历程，本文梳理了相关文献的研究全景与技术脉络，如图2所示。图2涵盖了近20年来该领域的系列代表性研究，不仅梳理了三大核心范式下各个空间分支的学术谱系，更直观呈现了研究重心从底层特征拟合、非线性表征学习向高阶认知对齐发展的历史轨迹。与现有综述相比，这一分类框架不仅清晰界定了多视图聚类在现代人工智能生态中的位置，更为推动该领域从单纯的无监督数据分析工具，向具备常识认知与逻辑引导能力的智能决策系统转型，提供了新的研究思路。

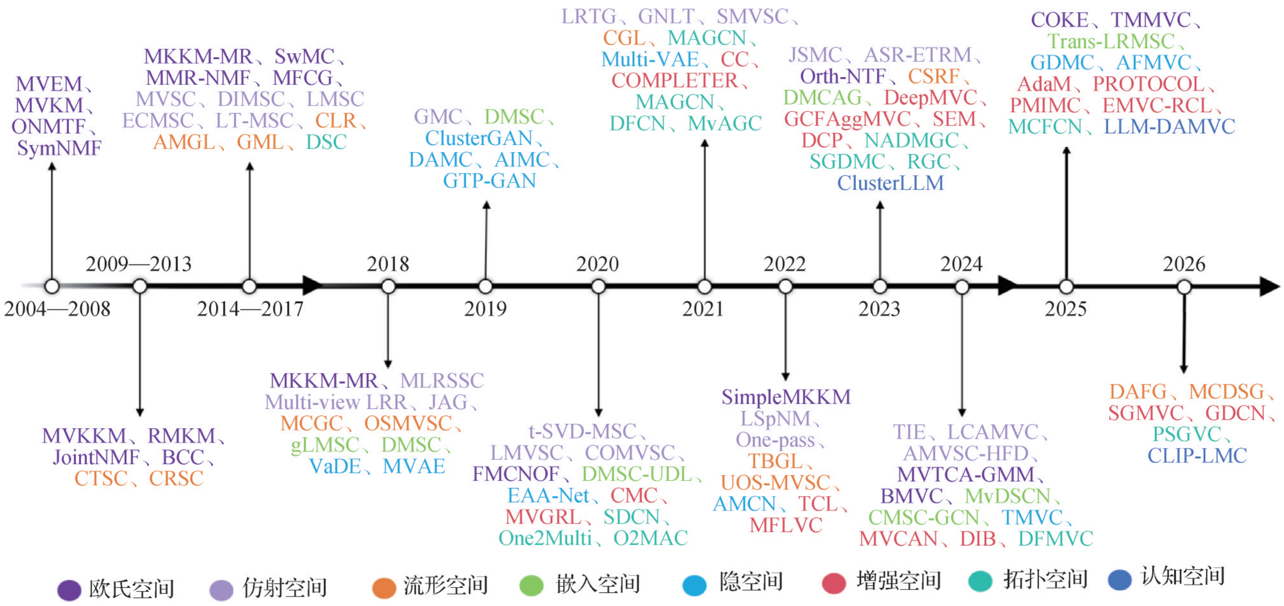
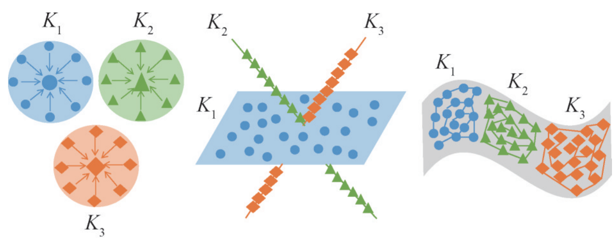


图2　多视图聚类发展范式趋势

Fig. 2　Evolution of development paradigms in multi-view clustering

1　几何结构先验的传统多视图聚类

传统多视图聚类主要建立在对原始特征空间中显式几何性质的统计假设之上，如图3所示。该类方法通过在欧氏、仿射及流形3种特征空间中构建显式的数学约束，旨在挖掘数据在浅层表征下的几何结构与物理关联。本节将从这3个空间维度对传统多视图聚类的代表性研究进行系统梳理。



(a) 欧氏空间：中心结构与原型聚类　(b) 仿射空间：线性结构与子空间聚类　(c) 流形空间：拓扑结构与图学习

图3　传统几何结构先验示意图

Fig. 3　Schematic：traditional geometric priors ((a)Euclidean space：centroid and proto-clusterings; (b) affine space：linear and subspace clusterings; (c) manifold space：topology and graph learning)

1.1　欧氏空间：中心结构与原型聚类

1.1.1　基于K-Means的迭代划分方法

基于K-Means的迭代划分方法是欧氏空间下确定原型的典型范式，其核心思想是通过最小化样本与类簇质心间的平方误差学习全局聚类结构。给定多视图数据集，该类方法通常致力于寻找一组共享的一致性聚类划分矩阵S以及各视图特定的质心矩阵$\boldsymbol{\mu}^v$，其基础数学模型可抽象为

$$\min_{S,\{\boldsymbol{\mu}^v\}} \sum_{v=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{k} s_{ij} \| \boldsymbol{x}_i^v - \boldsymbol{\mu}_j^v \|_2^2 \tag{1}$$

式中，$s_{i,j} \in \{0, 1\}$为一致性聚类指示变量（若样本i

属于簇j则为1,否则为0),$\boldsymbol{x}_i^v$为第v个视图下第i个样本的特征向量,$\boldsymbol{\mu}_j^v$表示第v个视图下第j个类簇的几何中心。

该领域早期研究深受协同训练思想影响,Bickel和Scheffer(2004)提出了多视图期望最大化(multi-view expectation-maximization, MVEM)算法和多视图K-Means(multi-view K-Means, MVKM)算法,通过在一个视图上学习聚类结构并利用其去约束另一个视图的学习,实现了视图间的一致性交互。然而,传统的K-Means方法受限于线性边界假设,难以处理非线性分布。针对这一问题,Tzortzis和Likas(2012)通过核方法提出了多视图核K-Means(multi-view kernel K-Means, MVKKM),利用核函数的凸组合将数据映射到高维再生核希尔伯特空间,从而在新的空间中寻找最优的簇划分。为了克服传统平方欧氏距离对异常值敏感的问题,Cai等人(2013)提出了鲁棒多视图K-Means(robust multi-view K-Means, RMKM),利用$\ell_{2,1}$范数抑制异常值噪声,并通过共识指示矩阵实现大规模多视图数据的一致性学习。为了减少基核间的冗余并提升判别力,Liu等人(2016)提出了基于矩阵诱导正则化的多视图核K-Means(multiple kernel K-Means clustering with matrix-induced regularization, MKKM-MR),通过引入矩阵诱导正则化降低组合核系数的相关性,强制模型学习出更加多样且正交的核表示,有效改善了传统方法中信息高度冗余的问题。进一步地,考虑到不同视图在聚类中的贡献度存在差异,Nie等人(2017)提出的带自加权机制的多视图聚类(self-weighted multi-view clustering, SwMC)框架,通过引入自加权机制初步实现了视图权重的自动分配,解决了手动调参的烦琐性。然而,这类启发式策略在理论上难以完全规避非凸目标函数带来的局部最优风险。为解决该问题,Liu等人(2023)提出简单多视图核K-Means(simple multi-view kernel K-Means, SimpleMKKM),该框架将优化重构为最小—最大(min-max)范式,通过最小化核对齐误差确保了全局最优性。针对大规模应用中的计算瓶颈,Liang等人(2025)提出了COKE(core kernel)框架,通过构建核心核实现了对核权重的高效近似,在保留SimpleMKKM全局最优特性的同时显著降低了复杂度,将最小—最大范式原型聚类扩展至大规模数据场景。

1.1.2 非负矩阵分解方法

另一类原型学习范式从非负矩阵分解(non-negative matrix factorization, NMF)视角出发,将观测数据解构为若干代表性基底的非负线性组合。其建模机理在于通过最小化低秩重构误差学习潜在原型,优化目标本质上锚定在欧氏空间下的距离度量,即通过最小化Frobenius范数实现紧凑的原型表征。在多视图场景下,典型的多视图矩阵分解模型可统一表示(Liu等,2013)为

$$\min \sum_{v=1}^{m} \| \boldsymbol{X}^v - \boldsymbol{V}^v \boldsymbol{U}^{v\mathrm{T}} \|_{\mathrm{F}}^2 + \sum_{v=1}^{m} \lambda_v \| \boldsymbol{V}^v - \boldsymbol{V}^* \|_{\mathrm{F}}^2$$
$$\text{s.t. } \| \boldsymbol{U}_{:,k}^v \|_1 = 1, \boldsymbol{U}^v, \boldsymbol{V}^v, \boldsymbol{V}^* \geqslant 0 \tag{2}$$

式中,$\boldsymbol{X}^v$表示第v个视图的原始数据矩阵。目标函数的第1项致力于在单一视图内将原始数据重构为基向量的线性加性组合,$\boldsymbol{U}^v$代表第v个视图的特征基向量矩阵,$\boldsymbol{V}^v$为特定视图的表示系数矩阵。式(2)的第2项为多视图协同约束,旨在融合各视图的表示矩阵,通过引入权重超参数λ_v,强制各视图的$\boldsymbol{V}^v$向一个全局一致的共识矩阵$\boldsymbol{V}^*$逼近。最终的聚类结果即可直接从该一致性表示矩阵$\boldsymbol{V}^*$中导出。此外,约束条件$\| \boldsymbol{U}_{:,k}^v \|_1 = 1$用于保证列基向量的归一化,而非负约束则赋予了模型基于部分构成整体的物理可解释性。

早期Ding等人(2006)的奠基性工作通过正交非负矩阵t分解(orthogonal nonnegative matrix t-factorizations, ONMTF)模型,在数学上严谨地建立了矩阵分解与K-Means聚类的理论纽带。Kuang等人(2012)的对称非负矩阵分解(symmetric non-negative matrix factorization, SymNMF)将此框架泛化至对称相似度矩阵,实现了核空间下的原型等价转换。在多视图融合层面,Liu等人(2013)提出的联合非负矩阵分解(joint non-negative matrix factorization, JointNMF)框架通过共享共识系数矩阵,实现了异构视图间共有原型结构的有效解耦。为了进一步弥补全局重构在捕捉局部拓扑方面的不足,Zong等人(2017)引入流形正则化约束,提出了多视图流形正则化非负矩阵分解(multi-manifold regularized non-negative matrix factorization, MMNMF),在原型提取中考虑了数据的局部几何结构。在表征维度的深度挖掘方面,Li等人(2023)与Zhang等人(2026)突破了传统矩阵表达的局限,分别利用正交非负张量分解(orthogonal non-negative tensor factorization, Orth-NTF)与张量化多维多视图聚类(tensorized multidimensional multi-view clustering，TMMVC)捕捉视图间的高阶关联,解决了异构特征维度的对齐难题。针对大规模数据场景,Yang等人(2021)提出的快速多视图共识非负与正交分解(fast multi-view clustering model via nonnegative and orthogonal factorization，FMCNOF)框架通过优化更新策略,显著降低了高维海量样本下的原型迭代成本。此外,Yang等人(2024b)提出了基于共识图矩阵分解(matrix factorization of consensus graph，MFCG)的大规模多视图聚类方法,该工作将多视图共识图的学习过程与矩阵分解框架无缝融合,在保证算法具备高扩展性的同时,显著提升了海量样本下跨视图潜在结构一致性的挖掘精度。

1.1.3　概率生成模型

与上述确定性分解方法不同,概率生成模型从统计推断角度解决原型学习问题。高斯混合模型(Gaussian mixture model，GMM)(McLachlan和Peel,2000)通过协方差矩阵建模簇的椭球状分布,克服了 K-Means 仅能处理球形簇的局限,并利用期望最大化(expectation-maximization，EM)算法实现视图间的软对齐。Lock和Dunson(2013)引入贝叶斯一致性聚类(Bayesian consensus clustering，BCC)框架,利用狄利克雷过程自适应推断原型数量。此外,Jiang等人(2024)将 GMM 引入多视图迁移聚类框架(multi-view transfer clustering based on Gaussian mixture，MT-GMM),通过概率原型的参数传递优化多视图任务。Shapiro和Battle(2022)则构建了更为复杂的贝叶斯多视图聚类(Bayesian multi-view clustering，BMVC)结构,以建模视图间非对称的依赖关系。尽管这类传统概率方法具备优良的统计解释性,但在处理高维非线性数据时仍存在瓶颈,关于结合深度学习的生成式方法将在第2.2小节详细阐述。

1.2　仿射空间:线性结构与子空间聚类

该类方法建立在仿射空间假设之上,即认为高维数据分布在若干低维线性子空间的并集之上。其核心机制是自表示性质,即每个样本可以由同一子空间内其他样本的线性组合重构(Vidal,2011)。子空间聚类的理论基础可追溯至经典的稀疏子空间聚类(sparse subspace clustering，SSC)(Elhamifar和Vidal,2013)与低秩表示(low-rank representation，LRR)(Vidal和Favaro,2014)。Gao等人(2015)提出了多视图子空间聚类(multi-view subspace clustering，MVSC),通过引入公共聚类结构约束,强制不同视图的表示系数矩阵在仿射子空间上实现对齐,其目标函数为

$$\min_{\boldsymbol{Z}^v,\boldsymbol{E}^v,\boldsymbol{F}}\sum_{v=1}^{m}\left\|\boldsymbol{X}^v-\boldsymbol{C}^v\boldsymbol{X}^v-\boldsymbol{E}^v\right\|_{\mathrm{F}}^{2}+$$
$$\lambda_1\sum_{v=1}^{m}\mathrm{tr}(\boldsymbol{H}^{*\mathrm{T}}(\boldsymbol{D}^v-\boldsymbol{W}^v)\boldsymbol{H}^*)+\lambda_2\sum_{v=1}^{m}\left\|\boldsymbol{E}^v\right\|_1$$
$$\mathrm{s.t.}\ \boldsymbol{C}^{v\mathrm{T}}\boldsymbol{1}=\boldsymbol{1},\boldsymbol{C}_{ii}^{v}=0,\boldsymbol{H}^{*\mathrm{T}}\boldsymbol{H}^*=\boldsymbol{I}\qquad(3)$$

式中,第1项表示与数据集自表示性质相关的重构误差,经过子空间学习后可得到第v个视图的子空间表示矩阵$\boldsymbol{C}^v$。由于真实数据点通常不会完全精确地分布在理想子空间内,因此引入$\boldsymbol{E}^v$表示偏离线性组合的离群项。第2项为多视图协同谱图模块,通过为所有视图引入一个公共划分矩阵$\boldsymbol{H}^*$(其中$\boldsymbol{E}^v$与$\boldsymbol{W}^v$分别代表图度矩阵与相似度矩阵),表明谱聚类过程是在所有视图上同步联合执行的,从而强制挖掘全局一致的拓扑结构。第3项是基于ℓ_1范数的正则化项,旨在保证离群项矩阵$\boldsymbol{E}^v$的稀疏性。此外,约束条件$\boldsymbol{C}_{ii}^v=0$避免了样本自表示的平凡解,而$\boldsymbol{H}^{*\mathrm{T}}\boldsymbol{H}^*=\boldsymbol{I}$则保证了共识表示的正交性。

在多视图场景下,Brbić和Kopriva(2018)提出多视图低秩稀疏子空间聚类(multi-view low-rank sparse subspace clustering，MLRSSC),利用ℓ_2范数和核范数的联合约束构建亲和矩阵,以实现在噪声干扰下对仿射结构的鲁棒恢复。为了平衡视图间的冗余与协同,Cao等人(2015)提出多样性诱导多视图子空间聚类(diversity-induced multi-view subspace clustering，DiMSC)框架,引入Hilbert-Schmidt独立性准则作为多样性约束,通过最小化视图间表示系数的相关性剔除冗余信息,从而迫使模型学习各视图特有的互补特征。Zhang等人(2017)提出了潜在多视图子空间聚类(latent multi-view subspace clustering，LMSC),该方法通过学习一个潜在的公共投影空间,在更纯净的特征空间重构数据的仿射几何结构,显著提升了算法的鲁棒性。在亲和图优化与复杂噪声抑制方面,Wang等人(2017)提出互斥——一致性正规化的多视图子空间聚类(exclusivity-consistency regularized multi-view subspace clustering，ECMSC),引入基于Hadamard积的互斥性正则项来约束表示系数,强制各视图捕捉样本间不同的线性组合关系,在提取共性信息的同时有效抑制了

重构分量的维度冗余。与此不同的是,Tang 等人(2019)侧重于图结构的全局特性,提出了联合亲和图学习(joint affinity graph learning, JAG)方法,直接在优化目标中对亲和矩阵施加秩约束,以诱导出具有明确 k 个连通分量的联合图结构,从而在亲和矩阵中显式地构造出理想的块对角属性。针对多模态数据中各模态噪声分布不一的难题,Abavisani 和 Patel(2018a)在多模态低秩稀疏子空间聚类(multi-modal low-rank sparse subspace clustering, MLRSSC)引入 $\mathcal{L}_{2,1}$ 范数与核范数,对重构误差进行分量化建模,利用结构化先验差异化地滤除噪声。考虑到单一的一致性建模难以应对复杂场景,Cai 等人(2023)联合平滑多视图聚类(jointly smoothing multi-view clustering, JSMC),将视图间的一致性与不一致性同时纳入统一的自表示学习框架,并结合联合平滑策略保留局部流形结构,进一步提升了模型对异质信息的表示能力。

随着张量理论的引入,多视图子空间聚类从单一的矩阵对齐转向了高阶张量约束。Zhang 等人(2015)提出低秩张量多视图子空间聚类(low-rank tensor multi-view subspace clustering, LT-MSC),将各视图的表示矩阵堆叠为三维张量,并利用张量核范数约束空间各维度间的低秩性,从而捕捉视图间的高阶一致性。考虑到奇异值在物理含义上的重要性差异,Gao 等人(2020b)引入基于张量奇异值分解(tensor singular value decomposition, t-SVD)的图学习方法,通过对张量 SVD 分解后的奇异值赋予差异化权重,增强了模型对光照波动与非结构化噪声的鲁棒性。针对张量核范数作为凸松弛算子难以获得最优低秩逼近的问题,Chen 等人(2022)提出低秩张量图学习(low-rank tensor graph learning, LRTG),在频域奇异值上施加非凸罚函数,从理论上证明了其在子空间恢复精度上优于传统凸优化方法。随后,Chen 等人(2021)进一步构建了广义非凸张量逼近框架(generalized non-convex tensor approximation, GNLTA),结合加权 Schatten-p 范数实现了对奇异值显著性的自适应度量,提升了复杂场景下的聚类稳定性。Guo 等人(2023)则引入张量对数 Schatten-p 范数最小化(tensor logarithmic Schatten-p norm minimization, TLSpNM),利用对数算子的平滑特性更准确地逼近张量秩,有效缓解了大规模奇异值对低秩结构的扰动。Gu 和 Feng(2024)将张量建模与字典学习相结合,提出了一种可扩展的三元信息增强(tri-information enhancement, TIE)框架,通过三元信息增强机制在降低计算复杂度的同时,提升了张量表示的判别力。

为解决全样本自表示矩阵带来的计算瓶颈,研究者通过引入锚点机制,将原始的 $n \times n$ 邻接图简化为 $n \times m$ ($m \leqslant n$) 的二部图,使算法复杂度降至线性级别。Kang 等人(2020)提出大规模多视图子空间聚类(large-scale multi-view subspace clustering, LMVSC)框架,通过引入锚点机制将全样本亲和图简化为二部图,使子空间聚类的时间复杂度从 $O(N^3)$ 降至线性量级。Sun 等人(2021)引入统一锚点学习机制,提出可扩展多视图子空间聚类(scalable multi-view subspace clustering, SMVSC)方法,通过联合优化锚点位置与表示矩阵,确立了大规模子空间聚类的基本范式。Ji 和 Feng(2023)则在此基础上融入张量理论,提出基于增强张量秩最小化的锚点结构正则化(anchor structure regularization with enhanced tensor rank minimization, ASR-ETR)方法,利用增强的张量秩最小化约束锚点结构,在保持计算效率的同时挖掘了视图间的高阶相关性。为了解决两步法(图构建与谱聚类分离)可能导致的次优解问题,Zhang 等人(2022)提出一致性一步法多视图子空间聚类(consistent one-step multi-view subspace clustering, COMVSC)框架,通过直接对子空间表示矩阵施加块对角秩约束,实现了表示学习与离散标签生成的同步优化。为将该理论拓展至大规模场景,Liu 等人(2022)结合锚点策略提出了单遍(one-pass)框架,利用低秩二部图的秩约束直接求解聚类标签,在维持数学可解释性的前提下提升了执行效率。针对跨视图锚点语义失配的问题,Zhang 等人(2024)提出簇感知锚点多视图聚类(cluster-wise anchor multi-view clustering, CAMVC)策略,利用共识聚类指示矩阵指导锚点生成,确保了不同视图在潜在语义空间的一致性。此外,Ou 等人(2024)针对多视图特征的异质性,提出基于锚点的层级特征递减多视图子空间聚类(anchor-based multi-view subspace clustering with hierarchical feature descent, MVSC-HFD),通过在锚点空间内进行多尺度特征融合,强化了模型对复杂大规模样本的判别能力。

1.3 流形空间:拓扑结构与图学习

基于流形空间的方法假设高维数据本质上分布

于低维的非线性流形结构中，其基本思想是在低维流形内最大程度地保留原始特征的局部拓扑关联。该类方法通常通过构建相似度图并结合谱图理论，将聚类任务转化为最优图切分问题。在多视图场景下，为了克服预定义静态图对噪声敏感的缺陷，当前的主流范式倾向于采用自适应图学习策略，通过联合优化特定视图的局部拓扑与全局共识图，其经典目标函数可表述（Wang等，2020）为

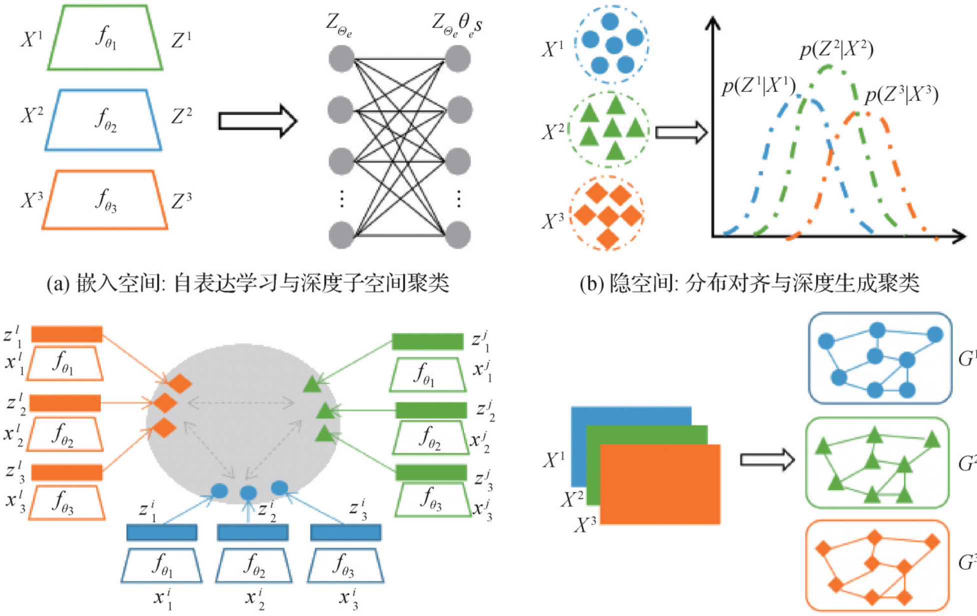
$$\min_{\{S^v\},U}\sum_{v=1}^{m}\sum_{i,j=1}^{n}\|\boldsymbol{x}_i^v-\boldsymbol{x}_j^v\|_2^2\boldsymbol{s}_{ij}^v+\beta\sum_{v=1}^{m}\sum_{i=1}^{n}\|\boldsymbol{s}_i^v\|_2^2+\sum_{v=1}^{m}\boldsymbol{w}^v\|\boldsymbol{U}-\boldsymbol{S}^v\|_F^2$$
$$\text{s.t. } \boldsymbol{s}_{ii}^v=0,\boldsymbol{s}_{ij}^v\geqslant 0,1^{\mathrm{T}}\boldsymbol{s}_i^v=1,\boldsymbol{u}_{ij}\geqslant 0,1^{\mathrm{T}}\boldsymbol{u}_i=1 \quad (4)$$

式中，$\boldsymbol{x}_i^v$与$\boldsymbol{x}_j^v$分别表示第v个视图下的第i个与第j个样本特征向量。$\boldsymbol{S}^v\in\mathbf{R}^{n\times n}$表示第$v$个视图动态学习到的局部相似度图矩阵，$\boldsymbol{U}\in\mathbf{R}^{n\times n}$是最终融合所有视图拓扑信息的全局共识图矩阵。

早期，Kumar 和 Daumé Ⅲ（2011a）提出了协同训练多视图谱聚类（co-training spectral clustering，CTSC）。在此基础上，Kumar等人（2011b）提出了协同正则化多视图谱聚类（co-regularized spectral clustering，CRSC），通过在潜在空间中强制不同视图的谱嵌入趋于一致，初步实现了对多视图互补信息的学习。针对预定义静态图对噪声敏感的问题，Nie等人（2016a）提出了约束拉普拉斯秩（constrained Laplacian rank，CLR）算法，通过显式约束拉普拉斯矩阵的零特征值重数，直接从数据中诱导出具有明确连通分量的理想拓扑结构。针对视图贡献差异，Nie等人（2016b）提出了参数无关的自动多图学习（auto-weighted multi-graph learning，AMGL）框架，通过重构残差自动调节视图权重，避免了手动调参的局限性。随后，图构建与聚类任务的深度耦合成为主流趋势。Zhan等人（2018）提出了多视图图学习（multi-view graph learning，GML）框架，通过在统一目标内交替更新相似度图与拉普拉斯秩约束，实现了图拓扑与聚类结构的同步优化。在其后续工作中，Zhan等人（2019）提出了多视图一致性图学习（multi-view consensus graph clustering，MCGC），通过最小化视图间差异并引入共识约束，进一步提升了模型对一致性拓扑信息的表达精度。Wang等人（2020）提出的图多视图聚类（graph-based multi-view clustering，GMC）框架联合优化了各视图的图学习与全局共识图的构建，实现了无需K-Means即可直接输出聚类结果的端到端学习。随着研究向高阶张量空间扩展，Li等人（2022b）提出一致性图学习（consensus graph learning，CGL）方法，利用加权张量核范数捕捉跨视图的高阶一致性。针对大规模数据的扩展性瓶颈，Xia等人（2023）构建了张量化二部图学习（tensorized bipartite graph learning，TBGL）框架，通过结合张量Schatten-p范数与二部图策略，在维持计算效率的同时实现了对高阶拓扑关系的判别性挖掘。此外，为解决谱嵌入与K-Means分步执行导致的次优解问题，端到端的一步法（one-step）成为关键路径。Zhu等人（2019）提出了一步多视图谱聚类（one-step multi-view spectral clustering，OMSC）方法，将谱表征求解与离散指示矩阵生成集成至单一优化过程。Tang等人（2023）设计了统一一步多视图谱聚类（unified one-step multi-view spectral clustering，UOMvSC）框架。Chen等人（2023）则提出了基于一致性谱旋转融合（consensus spectral rotation fusion，CSRF）的多视图聚类方法，利用正交变换直接逼近离散聚类标签。Wang等人（2026）提出了基于多图自适应双随机图（multi-graph adaptive doubly stochastic graph，MADSG）的多视图聚类方法，通过在图学习过程中施加双随机约束，确保了图结构的鲁棒性。Jiang等人（2026）提出了基于多样化锚点图融合（diverse anchor graph fusion，DAFG）的非对齐多视图聚类，针对视图间未严格对齐的挑战，实现了在大规模非对齐数据下的高效聚类。

2 语义协同先验的深度多视图聚类

不同于传统方法依赖原始空间的浅层几何先验，深度多视图聚类通过非线性映射将异构数据投射至高阶语义空间，实现了从显式几何约束向隐式语义关联的跨越。该类方法分别利用嵌入空间的特征重构、隐空间的分布建模、增强空间的对比判别以及拓扑空间的图关联学习，深度挖掘视图间的一致性，如图4所示。尽管各空间在建模手段上有所侧重，但其核心均指向对一致性语义的准确捕捉：嵌入空间经由重构提取一致性信息，隐空间借助分布映射类别属性，拓扑空间通过连接传递结构关联，增强空间则通过对比学习强化语义表达的判别性。



(a) 嵌入空间: 自表达学习与深度子空间聚类
(b) 隐空间: 分布对齐与深度生成聚类
(c) 增强空间: 互信息最大化与深度对比聚类
(d) 拓扑空间: 图结构学习与深度图聚类

图4　语义协同先验示意图

Fig. 4　Schematic: semantic collaborative priors ((a) embedding space: self-expression and deep subspace clustering; (b) latent space: distribution alignment and deep generative clustering; (c) augmented space: MI maximization and deep contrastive clustering; (d) topological space: graph structure and deep graph clustering)

2.1　嵌入空间:自表达学习与深度子空间聚类

深度子空间聚类(deep subspace clustering, DSC)利用深度神经网络将高维非线性数据映射到低维空间,并在此空间内重构数据的线性自表达关系。Ji等人(2017)提出的DSC网络是该领域的开山之作,通过在自编码器瓶颈层引入自表达层并施加$\boldsymbol{Z}=\boldsymbol{ZC}$约束,首次实现了端到端的非线性子空间分割。在多视图场景下,深度多视图子空间聚类进一步引入多视图协同机制,旨在学习能够同时捕捉视图间一致性与互补性的共识表示。为了实现高效的跨视图协同,Zhang等人(2020)提出了一种广义潜在多视图子空间聚类(generalized latent multi-view subspace clustering, gLMSC)框架,通过将所有视图映射至统一的潜在空间,在潜在表征上构建共享自表达约束,有效挖掘了视图间的一致性关联。Abavisani和Patel(2018b)将这一思想扩展至多模态领域,提出了深度多模态子空间聚类(deep multi-modal subspace clustering, DMSC)网络,通过分层融合策略强制不同模态共享同一系数矩阵,实现了语义级的深度协同。在此基础上,Wang等人(2021)强调了特征的判别性,提出了基于统一与判别学习的深度多视图子空间聚类(deep multi-view subspace clustering with unified and discriminative learning, DMSC-UDL),通过联合优化聚类与判别性损失,显著增强了多视图嵌入特征的可分性。而Zhu等人(2024)则提出了多视图深度子空间聚类网络(multi-view deep subspace clustering networks, MvDSCNs),通过引入多样性正则化,在追求共性的同时保留各视图的独特性,以构建包含丰富互补信息的关联图。

针对大规模数据导致系数矩阵计算复杂度呈平方级增长的瓶颈,Cui等人(2023)提出了基于锚点图的深度多视图子空间聚类(deep multi-view subspace clustering based on anchor graph, DMCAG),该方法在嵌入空间构建轻量级锚点图,将聚类复杂度降至线性量级。针对更复杂的非线性关联与全局依赖,Guan等人(2024)提出了基于对比学习与图卷积的多视图子空间聚类(contrastive multi-view subspace clustering with graph convolution, CMSCGC),将图卷积网络与对比学习相结合,在特征提取阶段最大化视图间的互信息(mutual information, MI),显著提升了高光谱等复杂数据的聚类效果。Lin等人(2025)进一步探索了Transformer架构的潜力,提出了基于自注意力的低秩多视图子空间聚类(transformer-based attentive low-rank multi-view subspace clustering，TALMSC），利用自注意力机制捕获全局上下文信息，提升了低秩多视图子空间聚类性能。

2.2 隐空间：分布对齐与深度生成聚类

作为深度生成聚类的代表性工作，Jiang 等人（2017）提出的变分深度嵌入（variational deep embedding，VaDE）将 GMM 作为变分自编码器（variational autoencoder，VAE）的先验分布，通过推导证据下界（evidence lower bound，ELBO）实现了特征重构与聚类分配的联合优化。针对多视图数据的融合瓶颈，Wu 和 Goodman（2018）提出的多视图变分自编码器（multi-view variational autoencoder，MVAE）引入专家乘积（product of experts，PoE）机制，通过计算各视图边缘后验分布的乘积来推断联合潜分布，为多模态数据的隐空间一致性建模提供了概率框架。为了获得更高质量的表示，Xu 等人（2021）提出的多视图变分自编码器（multi-view variational autoencoder，Multi-VAE）采用解耦学习策略，将特征拆分为视图共有因子与特有因子，通过仅对共有因子施加聚类约束，提升了模型对视图特定噪声的鲁棒性。

在对抗生成范式方面，Mukherjee 等人（2019）提出的聚类生成对抗网络（clustering generative adversarial network，ClusterGAN）探索了生成对抗网络（generative adversarial network，GAN）隐空间的结构化聚类，利用反向映射网络将数据投影回离散—连续混合的潜在流形中，为复杂数据的分布对齐提供了新思路。为了解决分布异构性与数据不完整挑战，Li 等人（2019b）提出的深度对抗多视图聚类（deep adversarial multi-view clustering，DAMC）框架通过在隐空间引入判别器，强制各视图潜分布向预设先验对齐，以此抑制视图特有噪声。针对缺失模态场景，Xu 等人（2019）提出的对抗不完整多视图聚类（adversarial incomplete multi-view clustering，AIMC）借助循环对抗训练，在隐空间内动态预测并补全缺失视图的潜在表示，确保了不完整样本在统一隐空间内的分布一致性。在提升融合质量方面，Zhou 和 Shen（2020）设计了端到端对抗多视图聚类（end-to-end adversarial multi-view clustering，EAMC）注意力网络，利用判别器对融合质量的置信度评估反馈调节注意力权重。Wang 等人（2023）则进一步提出了具有自适应融合的对抗多视图聚类（adversarial multi-view clustering，AMvC）网络，通过多判别器博弈迫使编码器在竞争中生成更具鲁棒性的共识表示。

此外，部分研究开始关注模型在复杂环境下的可信度、因果关联及社会属性。Yang 等人（2024a）提出可信多视图聚类框架（alternating generative adversarial representation learning，AGARL），通过交替优化的生成对抗表示学习增强了训练稳定性与结果可靠性。针对现实场景中跨视图样本非严格对齐的瓶颈，Yang 等人（2025）结合因果学习提出了因果多视图聚类（causal multi-view clustering，CauMVC）泛化深度学习方法，利用后门调整技术推断因果不变的隐变量，克服了部分对齐数据导致的分布偏移。同时，Jiang 等人（2025）关注算法的伦理约束，提出了对抗公平多视图聚类（adversarial fair multi-view clustering，AFMVC），利用对抗性去偏技术消除隐空间中与敏感属性相关的偏置信息，确保了聚类结果在性别、种族等维度上的公平性。这些工作从分布拟合出发，进一步拓展了深度生成聚类在可解释性、鲁棒性及公平对齐等维度的发展。

2.3 增强空间：互信息最大化与深度对比聚类

对比多视图聚类打破了传统方法对静态重构目标的依赖，转而在增强空间中挖掘视图间的不变性。其核心在于将同一样本的不同视图视为正样本对，利用信息噪声对比估计（information noise-contrastive estimation，InfoNCE）等对比损失最大化其在特征空间的一致性。Tian 等人（2020）提出的对比多视图编码（contrastive multi-view coding，CMC）确立了实例级跨视图对比范式，但 Trosten 等人（2021）指出盲目追求严格对齐可能抑制模型对高质量视图的特征提取，其后续研究深度多视图聚类（deep multi-view clustering，DeepMVC）（Trosten 等，2023）进一步从理论上揭示了过度强调实例对齐会损害簇结构的内在分布。为突破实例级对比忽视全局语义的局限，Li 等人（2021）提出对比聚类（contrastive clustering，CC），通过行列双重对比机制同时实现实例判别与聚类分配对齐，其后续工作进一步提出了双重对比网络（twin contrastive learning，TCL）（Li 等，2022a），将该范式扩展至在线聚类场景。在此基础上，Xu 等人（2022）设计了多层级特征多视图聚类（multi-level feature multi-view clustering，MFLVC），通过捕捉跨层级的语义关联增强表征的鲁棒性。此外，Xu 等人（2023a）提出了一个自加权多视图对比学习框架（self-weighted multi-view contrastive learning, SEM),以解决当不同视图间语义信息不一致时,学习到的特征往往会丧失判别力,进而导致表示退化的问题。Yan等人(2023)提出的GCFAggMVC则通过全局与跨视图特征聚合(global and cross-view feature aggregation, GCFAgg),构建了更具全局结构的样本对比对。在信息论层面,Yan等人(2024)提出的深度信息瓶颈(deep information bottleneck, DIB)框架摒弃了传统的变分近似,利用基于矩阵的熵估计直接优化多视图互信息(MI),为对比对齐提供了更严谨的解释。

在图结构数据方面,Hassani和Khas Ahmadi(2020)提出的多视图图表示学习(multi-view graph representation learning, MVGRL)利用图扩散机制捕捉高阶拓扑信息。Pan和Kang(2021)则通过图滤波技术消除噪声,实现了图对比与聚类目标的深度耦合。针对视图缺失挑战,Lin等人(2021a)提出COMPLETER框架,通过双向预测一致性学习和类别因子挖掘在不显式重构原始数据的前提下实现了缺失特征的补全与视图对齐。随后,其在双对比预测(dual contrastive prediction, DCP)(Lin等,2023)中进一步引入了双对比预测机制,利用一个视图的已知信息通过对比损失强制预测另一个视图的缺失表征,从而在潜在空间内建立起鲁棒的跨视图关联。随着研究的深入,对比多视图聚类进一步向结构感知与自适应方向发展。Fei等人(2025)提出了图结构感知的深度多视图对比聚类(deep multi-view contrastive clustering with graph structure awareness, DMvCGSA),在保持实例一致性的同时利用拓扑关联引导聚类划分。Yuan等人(2025)提出了基于原型匹配的不完整多视图聚类(prototype matching-based incomplete multi-view clustering, PMIMC)方法,通过引入原型匹配机制,在语义空间内对齐不同视图的类别中心,有效地解决了视图缺失导致的信息不对称问题。此外,Xue等人(2025)引入部分最优传输(partial optimal transport, POT)技术,提出了PROTOCOL以解决不平衡多视图场景下的对比偏移问题。Wang等人(2025b)提出了基于强化对比学习的有效多视图聚类(effective multi-view clustering with reinforcement contrastive learning, EMVC-RCL),引入强化学习动态筛选高质量对比对,实现了采样策略的自适应机制。Hu等人(2025)则引入了不确定性估计机制,利用置信度动态调整对比权重。Xue等人(2026)提出了基于结构增强对比学习的自适应多视图一致性聚类(adaptive multi-view contrastive learning, AdaM),通过结构增强的对比学习强化了视图间的一致性约束。Cui等人(2026)提出了结构引导的深度多视图聚类(structure-guided deep multi-view clustering, SGMVC)方法,通过引入结构约束来引导深度表示学习。Zhu等人(2026)提出了基于生成扩散对比网络(generative diffusion contrastive network, GDCN)的多视图聚类,将扩散模型的生成能力与对比学习相结合,提升了复杂场景下的聚类鲁棒性。

2.4 拓扑空间:图结构学习与深度图聚类

深度多视图图聚类利用图神经网络(graph neural network, GNN)将多视图数据映射到拓扑空间,旨在协同挖掘视图内的几何结构与视图间的互补语义。在早期探索中,Bo等人(2020)提出了结构化深度聚类网络(structural deep clustering network, SDCN)框架,通过传递算子将自编码器的深层表征融入图卷积网络(graph convolutional network, GCN),开启了结构化深度聚类的新范式。Fan等人(2020)提出的One2Multi模型利用图自编码器,通过从信息源视图重构多视图拓扑结构,实现了跨视图一致性特征的学习。为更精细地处理属性与拓扑的融合,Cheng等人(2020)提出了多视图属性图卷积网络(multi-view attribute graph convolution networks, MAGCN),设计了双通路编码器以显式对齐几何特征。Tu等人(2021)提出的深度融合聚类网络(deep fusion clustering network, DFCN)则引入结构与属性融合模块,利用互信息最大化动态整合多视图特征,有效缓解了深度GNN中的过平滑问题。为了突破传统GCN仅能捕捉局部邻域信息的局限,Lin等人(2021b)提出了多视图属性图聚类(multi-view attribute graph clustering, MAGC)方法,通过利用高阶图卷积算子捕捉视图间的高阶拓扑一致性,显著提升了模型对复杂流形结构的建模能力。

由于上述方法大多依赖预定义的静态图结构,难以应对原始数据中的噪声干扰,研究重心逐渐转向动态图构建与邻域感知优化。Xia等人(2022)提出了自监督图卷积多视图聚类(self-supervised graph convolutional multi-view clustering, SGCMC),引入自监督机制指导图卷积网络,利用数据自身的结构信号强化了特征的聚类友好性。Du等人

(2023)则针对特征空间与图空间的关联瓶颈，提出了邻域感知深度多视图图卷积聚类(neighborhood-aware deep multi-view graph convolutional clustering，NMvC-GCN)，通过协同优化特征邻域与图拓扑的一致性，增强了模型对局部结构扰动的鲁棒性。随后，Huang等人(2023)提出了自监督图深度多视图聚类(self-supervised graph deep multi-view clustering，SGDMC)，结合图注意力网络与自监督学习，通过注意力系数为不同视图自适应分配权重。针对伪连接问题，Wang等人(2024a)提出的SURER框架设计了结构自适应模块，在训练过程中动态修正拓扑逻辑。Ren等人(2024)提出了动态加权图融合深度多视图聚类(dynamic weighted graph fusion deep multi-view clustering，DFMVC)，通过动态加权融合机制生成高质量的一致性图结构。而Pei等人(2025)进一步构建了多视图共识图卷积网络(multi-view consensus fusion convolutional network，MCFCN)，确保了聚类分配的稳定性。此外，针对大规模图数据的扩展性瓶颈，Wang等人(2025a)提出了深度多视图锚点聚类(deep multi-view anchor clustering，DMAC)，通过在隐空间动态学习代表性样本点，实现了在线性时间复杂度下捕捉全局拓扑结构。Li等人(2026)提出了伪标签相似度图驱动的对比聚类(pseudo-label similarity graph-driven multi-view contrastive clustering，PSGVC)，通过在对比学习中引入伪标签构建相似度图，有效缓解了对比聚类中虚假负样本带来的偏差问题。

3 多模态大模型时代的多视图聚类

随着多模态大模型进入万亿参数与海量数据时代(Wang等，2024b)，多视图聚类的内涵发生了深刻变化。不同于传统方法致力于在孤立数据中挖掘几何结构，本节所述研究将LMM视为蕴藏人类通用常识与逻辑推理能力的知识基座。通过引入模型认知先验，聚类不仅成为大模型数据治理的核心手段，更深度嵌入到了模型的混合专家架构、检索增强及机械可解释性分析中。当前大模型实践中主要依赖以K-Means为代表的高效聚类算法。本节将系统梳理大模型语境下聚类技术的应用现状，并探讨多视图聚类在这一新兴领域中的潜在发展方向。

3.1 基础设施化：聚类在大模型全生命周期的实践

在现有的LMM工业链条中，聚类主要被视为一种大规模数据治理与动态路由的高效手段。在预训练阶段，Abbas等人(2023)提出的语义去重(semantic deduplication，SemDeDup)利用K-Means在嵌入空间构建语义桶，通过簇内去重策略缓解互联网规模(web-scale)数据的语义冗余。在指令微调中，Zhang等人(2025a)提出的D3与Ge等人(2024)提出的聚类与随机采样(clustering and random sampling，CaR)框架将聚类作为均衡器，通过在不同的语义簇中进行多样化采样，确保模型能够均匀覆盖各类任务视图。除了数据治理，聚类逻辑也深度嵌入模型架构中，如DeepSeekMoE(Dai等，2024)等混合专家模型(MoE)通过隐式的令牌(token)聚类实现专家特化，将输入流动态分发至特定的功能模块。在检索增强生成(RAG)领域，递归摘要树检索(recursive abstractive processing for tree-organized retrieval，RAPTOR)(Sarthi等，2024)与动态层级检索增强生成(dynamic hierarchical retrieval-augmented generation，DH-RAG)(Zhang等，2025b)则通过层级聚类对长文档或历史对话进行递归分割，构建起具备多尺度语义精度的检索拓扑。这些应用虽然在机理上触及了数据分布的一致性，但在算法执行层面仍倾向于K-Means。这种选择背后的逻辑在于，LMM产生的高维嵌入空间已具备极强的线性可分性，简单的欧氏距离聚类即可捕捉核心语义，且其O(N)的复杂度是处理海量互联网数据的唯一工业级选择。

3.2 协同赋能：多视图聚类与大模型融合展望

尽管目前K-Means占据主流，但在面对多模态协同、专家解耦及机械可解释性等高阶任务时，多视图聚类正展现出更强的表征建模能力，同时LMM的认知能力也为传统MVC注入了新的生命力。一方面，多视图聚类可显著增强大模型对异质信息的整合能力。例如，在多模态数据选择或MoE专家路由设计中，引入MVC的一致性与互补性约束，可比单一视图的K-Means更精准地解耦跨模态噪声，实现更高阶的专家特化。The geometry of concepts(Li等，2025)等机械可解释性研究也证明，LLM内部特征呈现出复杂的多尺度聚类结构，这正契合了MVC挖掘隐空间拓扑关系的长处，引入MVC的互补性约束可更准确地解析跨层级、跨模态的特征关联，从而优化MoE架构中的专家路由逻辑，减少视图间的维度

冗余。

另一方面，LMM正在重新定义MVC的先验引导模式。对比语言—图像预训练（contrastive language-image pre-training，CLIP）驱动的聚类工作展示了如何利用LLM的逻辑推理能力作为决策代理。ClusterLLM（Zhang等，2023）与大语言模型驱动的对抗多视图聚类（LLM-driven adversarial multi-view clustering，LLM-DAMVC）（Xu和Wang，2025）等工作展示了如何将人类的常识判断转化为聚类过程中的语义引导信号。Ouyang等人（2026）提出了CLIP驱动的终身多视图聚类（lifelong multi-view clustering，CLIP-LMC），利用大规模预训练模型CLIP提供的语义空间作为先验，解决了在持续流入的多视图数据流中进行知识积累与动态聚类的难题。这种融合范式解决了传统MVC在处理异构视图时相似度度量失效的挑战，将聚类从单纯的统计特征对齐提升至认知逻辑一致的新维度。此外，当前大模型多关注认知层面的学习而忽略了对感知的分析，而后者更有助于理解智能发展的底层逻辑。未来可借鉴格式塔心理学中关于涌现、补缺和连通等整体聚类能力的理论，结合多视图聚类的互补性获得与感知相关的、具有可解释性的特征表示。以上双向赋能不仅将聚类从几何找规律提升至语义找逻辑的高度，也为构建大规模自适应多视图聚类算法提供了全新的技术范式。

经过前3节详细探讨，本文系统还原了多视图聚类从底层几何先验、深层语义协同到大模型认知的技术发展过程。为了便于查阅与对比，表1

表1　多视图聚类核心范式与代表性算法总览

Table 1　Overview of core paradigms and representative algorithms in multi-view clustering

核心范式	空间与子分类	代表算法	核心机制与贡献
传统几何先验（数据驱动/浅层特征挖掘）	欧氏空间	MVKM (Bickel和Scheffer，2004)	早期多视图一致性交互的先驱工作，通过一个视图约束另一视图。
		RMKM (Cai等，2013)	利用共识指示矩阵集成，提升传统平方欧氏距离对异常值的鲁棒性。
		SimpleMKKM(Liu等，2023)	将优化重构为min-max范式，解决启发式交替迭代易陷入局部最优风险。
		COKE (Liang等，2025)	构建核心核实现对核权重的近似，复杂度降低，适用于超大规模数据。
		JointNMF (Liu等，2013)	共享共识系数矩阵，异构视图间共有原型结构有效解耦。
		TMMVC(Zhang等，2026)	张量化多维建模，突破传统矩阵表达局限，捕捉高阶关联。
		MFCG (Yang等，2024b)	实现高维度、海量样本下跨视图潜在结构一致性挖掘。
		MT-GMM(Jiang等，2024)	将GMM引入迁移聚类框架，传递概率原型参数，优化多视图迁移任务。
	仿射空间	MVSC (Gao等，2015)	引入公共聚类结构约束，是传统多视图子空间聚类基线模型。
		LT-MSC (Zhang等，2015)	将自表示矩阵对齐拓展至高阶张量空间。
		LMVSC/SMVSC (Kang等，2020; Sun等，2021)	引入统一锚点机制，突破全样本计算瓶颈，时间复杂度降至线性。
		CLR (Nie等，2016a)	从数据中直接诱导具有明确连通分量的理想拓扑。
	流形空间	GML (Zhan等，2018)	实现图拓扑与聚类结构的同步协同优化。
		GMC (Wang等，2020)	联合优化了各视图的图学习与全局共识图的构建，并施加拉普拉斯秩约束，实现了无需K-Means即可直接输出聚类结果的端到端学习。
		UOMvSC(Tang等，2023)	集成图学习、谱嵌入与标签预测于统一闭环，避免谱表征向离散解转化的精度下降（一步法）。

续表1　多视图聚类核心范式与代表性算法总览

Table 1　Overview of core paradigms and representative algorithms in multi-view clustering（continued）

核心范式	空间与子分类	代表算法	核心机制与贡献
深度语义协同先验 (非线性映射/高阶语义解耦)	嵌入空间	gLMSC (Zhang等,2020)	将所有视图映射至统一的潜在空间构建共享自表达约束,有效挖掘并强化了视图间的深层一致性关联。
		DMCAG (Cui等,2023)	在嵌入空间构建轻量级锚点图,结合自监督与对比学习,解决深度子空间特征系数矩阵平方级计算瓶颈。
		TALMSC (Lin等,2025)	利用Transformer自注意力机制捕获全局上下文,大幅提升低秩多视图子空间聚类的表示能力。
		MVAE (Wu和Goodman,2018)	引入专家乘积(PoE)机制推断联合潜分布,为多模态生成式隐空间的一致性建模提供概率框架。
	隐空间	ClusterGAN (Mukherjee等,2019)	利用反向映射网络将数据投影回离散—连续混合的潜在流形中。探索GAN隐空间的结构化聚类,为分布对齐提供新思路。
		AIMC (Xu等,2019)	借助循环对抗训练,动态预测补全隐空间缺失表示,应对极端开放环境下不完整视图(IMVC)难题。
		CauMVC (Yang等, 2025)	结合因果学习,利用后门调整推断因果不变隐变量,克服跨视图样本非严格对齐带来的分布偏移。
		CMC (Tian等,2020)	将同一样本不同视图视为正样本对,利用InfoNCE最大化一致性。
	增强空间	CC (Li等,2021)	提出实例与簇分配的行列双重对比机制,突破传统实例对齐忽视全局簇语义分布的局限。
		COMPLETER (Lin等,2021a)	双向预测一致性学习与类别因子挖掘,无需显式重构原始数据即可实现缺失特征逻辑补全。
		MFLVC (Xu等,2022)	捕捉跨层级的语义关联以进行多层级特征对比学习,显著增强了多视图特征表征的鲁棒性。
		GCFAggMVC (Yan等,2023)	通过全局与跨视图特征聚合构建样本对比对,提升了样本对比对的全局辨识度与聚类精度。
		SEM (Xu等,2023a)	引入自加权策略,有效缓解了多视图对比学习中存在的表示退化问题。
		PROTOCOL (Xue等,2025)	引入部分最优传输以优化对比损失,有效解决不平衡场景下的对比偏移。
	拓扑空间	SDCN (Bo等,2020)	通过传递算子将自编码器深层表征无缝融入图卷积网络,开启了属性特征与拓扑结构融合的深度图聚类新范式。
		OneZMulti (Fan等,2020)	利用图自编码器,从信息源视图重构多视图拓扑结构,实现了跨视图一致性嵌入的有效提取与重构。
		MAGCN (Cheng等,2020)	设计双通路属性图卷积编码器以显式对齐几何特征,更协同地处理了节点属性与网络拓扑的融合。
		DFCN (Tu等,2021)	引入结构与属性融合模块,利用互信息最大化整合特征,有效缓解了深度GNN中的过平滑问题。
		MAGC (Lin等,2021b)	利用高阶图卷积算子捕捉视图间的高阶拓扑一致性。显著提升了模型对复杂多流形拓扑结构的刻画能力。
大模型认知先验 (知识增强/常识逻辑对齐)	认知空间(底层基础设施)	SemDeDup (Abbas等,2023)	利用K-Means构建语义桶进行簇内去重,实现海量数据的冗余清洗。
		DeepSeekMoE (Dai等,2024)	通过隐式的token聚类机制实现专家特化解耦,驱动模型动态路由机制,为专家网络分配对应语义。
	认知空间(逻辑推理反哺)	ClusterLLM (Zhang等,2023)	将LLM的强推理能力作为决策代理,引入人类常识判断,将聚类从纯统计相似度对比升维至语义逻辑引导。
		LLM-DAMVC (Xu和Wang,2025)	利用LLM动态代理机制辅助多视图特征进行认知对齐。在极度复杂或异构的跨模态场景下实现高阶语义解耦。

汇总了三大核心范式下的空间划分、代表性算法及其核心机制。表1清晰勾勒出多视图聚类在理论基础与模型架构上的代际跨越,同时也为后续探讨其在复杂真实场景中的落地应用提供了方法论铺垫。

4 多视图聚类在真实场景中的应用

在前文探讨算法理论基础上,本节聚焦多视图聚类在现实场景中的应用。针对现实数据高噪声、强异构与标签稀缺等挑战,如图5所示,多视图聚类在生物医学信息、地理空间观测、社会感知与安全、工业网络风控4个领域展现出显著的应用价值。

4.1 智慧医疗与生物信息学:多组学融合

临床中单一维度的数据难以全面揭示复杂的疾病机理,而患者的基因组、转录组、表观遗传特征及医学影像等天然构成了多视图数据。 近年来,研究者逐渐引入多视图学习与无监督聚类框架,以有效挖掘不同组学间的互补信息。在具体的特征融合策略上,针对特定组学聚类结构易被忽略的问题,Miao等人(2025)构建了基于扩散增强相似性的聚类框架PartIES,在分区级别对多组学特征进行加权整合,挖掘出了与临床预后高度相关的癌症亚型。面对组学数据中普遍存在的噪声干扰与视图间信息冗余,Che等人(2025)提出了一种鲁棒的多视图表示学习模型,通过张量分离与流形正则化机制,增强了底层特征提取的抗噪能力。此外,为了协调不同组学和独立聚类算法带来的结果异质性,Brière等人(2021)开发了基于证据积累策略的一致性聚类工具ClustOmics,利用样本的共现频率实现了多源初步聚类结果的平滑与稳健整合。

4.2 遥感与空间感知:多源异构数据融合

广域地表覆盖物的自动化解析是城市规划和灾害监测的基础。单一光学遥感易受云层遮挡或异物同谱现象干扰,导致分类失效。多视图聚类通过空间—光谱融合机制实现了多源卫星数据的无监督解译。在实际工程中,系统常联合高光谱图像、激光雷达及合成孔径雷达输入。例如,Wang等人(2025c)利用多尺度锚点图技术实现了大规模多模态遥感地物的精准划分。Guan等人(2024)结合图卷积网络有效降低了高光谱数据的冗余维度。此外,Liu等人(2025)通过自适应权重融合海量轨迹与兴趣点数据,实现了对城市功能区的精细化识别。

4.3 公共安全与社会计算:异常挖掘与跨视角识别

在公共安全领域,多视图技术主要用于解决复杂场景下的模式挖掘与目标识别难题。一方面,在舆情监控与虚假信息检测中,多视图框架通过整合异构内容以识别潜在异常模式。Canh等人(2024)

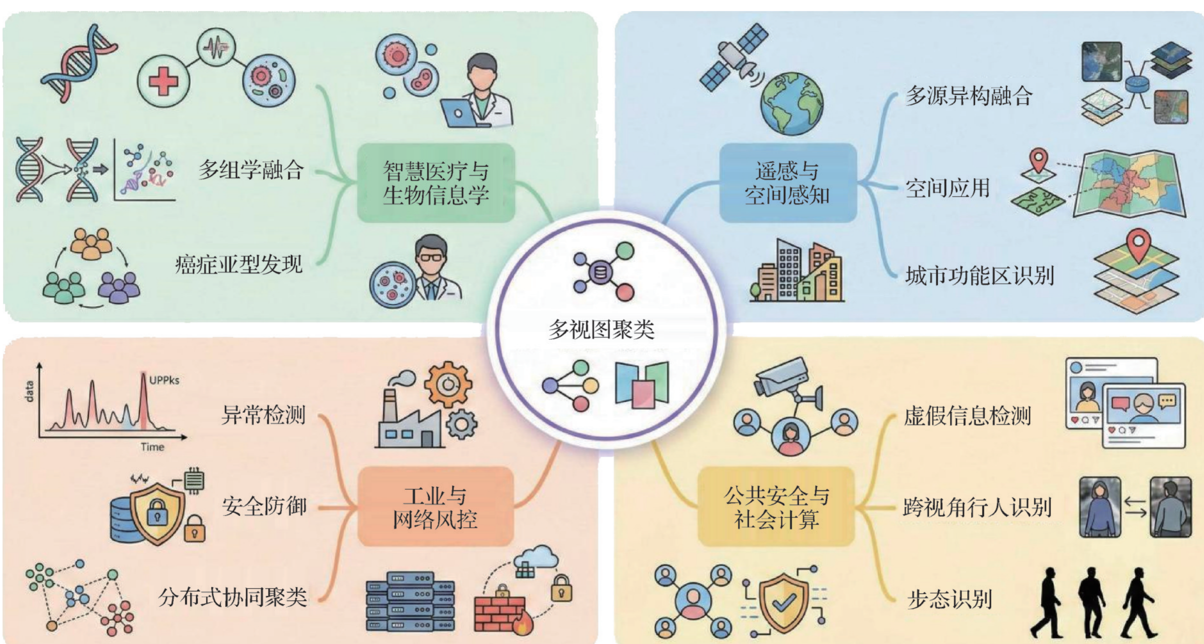


图5 多视图聚类的应用

Fig. 5 Applications of multi-view clustering

提取新闻文本与社交参与度特征，构建多视图模糊聚类模型以划分信息真伪。Cruickshank 与 Carley (2020)则联合分析用户、标签与链接网络，实现了对重大公共事件中话题社区及其演进趋势的有效刻画。另一方面，在智能视觉监控中，针对同一个体在不同摄像头视角下产生的表征偏移与高度异构数据难题(Xu 等，2023b)，多视图对齐技术通过挖掘步态序列间的语义一致性，在无标注条件下辅助完成鲁棒的行人特征重构与身份识别。

4.4 工业与网络风控：无监督异常检测

工业物联网与复杂网络中的异常事件往往高度隐蔽且缺乏标注，单一视图难以全面反映系统状态。因此，通过挖掘跨视图结构不一致性来发现异常信号的无监督多视图分析技术逐渐成为主流。早期浅层模型多依赖重构误差判定异常，Zhao 与 Fu (2015)利用多视图矩阵分解同步识别特征与结构离群点。Li 等人(2015)基于低秩分析，通过分离稀疏误差矩阵以量化样本异常程度。随着场景复杂度提升，研究重点转向分布式与深度架构。针对工业物联网的隐私保护需求，Wang 等人(2018)提出基于模糊 C-均值的协同聚类，在不共享原始数据的前提下实现分布式离群点检测。面对属性网络的复杂拓扑，Peng 等人(2022a)构建了深度框架，通过联合编码节点属性与网络结构视图，提升了对高隐蔽性异常节点的无监督挖掘能力。针对实际风控场景中存在的视图缺失问题，Wang 等人(2024c)引入正则化对比学习机制，在部分视图缺失情况下，通过拉近跨视图一致性表征以提升对微弱异常信号的捕捉精度。

5 未来挑战与展望

尽管多视图聚类在几何结构挖掘与深层语义建模上取得了长足进步，初步展现出与多模态大模型生态融合的潜力，但要实现从实验室基准向真实世界基础设施的跨越，该领域仍需攻克以下核心挑战：

1)高维海量数据的计算瓶颈与在线泛化。现代真实场景下的多视图数据往往具备海量样本与超高维特征的双重属性。然而，当前主流的 MVC 算法(尤其是基于自表示学习或谱聚类的方法)通常依赖于构建全局亲和度图或矩阵分解，其计算与存储复杂度高达 $O(N^2)$甚至 $O(N^3)$。这种复杂度在面对百万级样本(如大模型预训练数据的清洗、推荐系统中的用户行为切片)时会导致计算资源的直接崩溃。未来，如何设计具备近似线性复杂度、支持分布式计算且能够处理流式数据的增量/在线多视图聚类算法，是突破规模瓶颈的首要任务。

2)开放环境下的长尾失衡与极度不完整视图。近年来，多视图聚类在开放环境的鲁棒性建模取得了长足进展，特别是在解决视图缺失问题，学界已构建了丰富的解决方案。然而，现有的补全机制多局限于中低度缺失的理想化设定，当面临真实工业场景中极高比例的视图样本缺失或传感器完全失效等极端情况时，依赖局部图结构或浅层对齐的模型仍极易发生拓扑崩溃，这方面的研究亟待进一步深化。除了视图特征层面的极端不完整性，现实世界中的数据在类别层面还普遍呈现出长尾失衡。例如，在多组学罕见病诊断或金融多源异构反欺诈场景中，高价值的异常类或少数类样本极度稀缺，极易被多视图网络中占主导地位的多数类特征所淹没或过度平滑。面对这一极易导致模型表征同化的类别不平衡分布问题，现有研究鲜有涉足。因此，未来的研究不仅需突破极端模态缺失的特征恢复瓶颈，更需探索如何在极度不平衡的数据分布下，利用视图间的互补因果关联实现对少数类和异常样本的鲁棒发现，从而将优化的重心从单纯追求全局聚类准确率，向高价值长尾目标的精准挖掘转移。

3)评估体系重构。评估体系的滞后是当前多视图聚类领域迈向实际应用亟待解决的关键问题之一。尽管算法架构已从浅层模型发展至复杂的深度神经网络，甚至开始融入大模型对齐机制，但该领域的主流基准测试仍高度依赖于一些早期数据集，如表 2 所示。这些数据集通常通过手动提取 GIST (global image descriptor)、HOG (histogram of oriented gradients)、LBP(local binary pattern)、SIFT(scale-invariant feature transform)、PHOG(pyramid histogram of oriented gradients)等浅层特征构建而成，具备特征维度低、样本规模小且无噪声等理想化特征。由于评估数据与现实场景复杂度的脱节，当前部分研究在这些经典数据集上的性能提升已逐渐呈现边际效益递减的趋势，这在一定程度上限制了复杂算法在真实工业场景中的指导价值。

为进一步推动 MVC 领域的发展，未来的研究需逐步拓展现有的数据评估范式，使其更贴近真实世

表 2　当前多视图聚类领域常用数据集汇总

Table 2　Summary of current datasets commonly used in the multi-view clustering field

	数据集名称	样本数量	类别数	视图数	视图构成描述(多视图来源)
图像多特征提取	100Leaves (Van 和 Abbott,2003)	1 600	100	3	植物叶片的形状、纹理和边缘特征
	AR Face (Martínez 和 Benavente,1998)	数千	100+	3+	复杂遮挡人脸,像素、LBP、Gabor 等视图
	Caltech101-20 (Li 等,2004)	2 386	20	6	HOG, GIST, LBP, SIFT 等不同特征
	COIL-20 (Nene 等,1996)	1 440	20	3+	同一物体 360° 不同旋转角度的视角
	CUB-200 (Wah 等,2011)	11 788	200	2+	鸟类细粒度图像,CNN 特征与局部部位特征
	Flowers (Nilsback 和 Zisserman,2008)	8 189	102	2+	花卉图像,常提取颜色、形状及深度 CNN 特征
	Hdigit (Cai 等,2011)	10 000	10	2	手写数字的两种不同几何与纹理特征描述
	UCI (van Breukelen 等,1998)	2 000	10	6	手写数字的 6 种特征
	MSRC-v1 (Cai 等,2011)	210	7	5	颜色矩、纹理、HOG 等 5 种特征作为 5 个视图
	ORL / YALE / PIE Face (Georghiades 等,2001)	几百～数千	10～68	3～5	经典人脸数据集,由不同光照、姿态或表情构成
	Scene15 (Lazebnik 等, 2006)	4 485	15	3	空间包络、PHOG、LBP 特征组合
	YouTubeFace (Wolf 等,2011)	3 425	40	4+	视频帧提取的人脸多视图
图像深度增强与跨域视图	CIFAR-10/100 (Krizhevsky,2009)	5 万～6 万	10 / 100	2	对原始图像随机裁剪、颜色抖动等生成增强视图
	Fashion/Fashion-MNIST (Xiao 等,2017)	70 000	10	2	衣服/鞋包图像
	MNIST-USPS (Lecun 等,1998)	5 000	10	2	跨域多视图:MNIST 与 USPS
	Noisy MNIST (Wang 等,2015)	70 000	10	2	原始图像及注入高斯噪声
文本多重语义表示	3Sources (Greene 和 Cunningham,2009)	169	6	3	BBC、Reuters、Guardian 3 个新闻源的报道
	BBC/BBCSport (Greene 和 Cunningham, 2006)	2 225 / 737	5	2～4	将新闻文本划分为不同特征
	NGs (Lang,1995)	18 846	20	3	TF-IDF、LDA、Word2Vec 词向量 3 个视图
	Reuters-1200 (Amini 等,2009)	1 200	6	5	提取自路透社多语言文档的 5 种不同语言翻译
	WebKB (Blum 和 Mitchell,1998)	1 051	2	2	网页文本、网页指向的超链接锚文本
图结构与属性网络	ACM / DBLP / IMDB (Wang 等,2019)	数千	3～4	2	节点属性特征和学术/电影异质图拓扑结构连接
	Amazon Photos (Shchur 等,2018)	7 650	8	2	商品的用户评论和被共同购买的拓扑图连接结构
	Cora / Citeseer / PubMed (Sen 等,2008)	2 708～1.9 万	3~7	2	节点的属性特征和论文之间的引用拓扑图结构
生物信息与基因数据	BDGP (Cai 等,2012)	少量/特定	生物基因	2	果蝇胚胎的视觉图像特征与基因表达特征
	Prokaryotic (Brbić 等,2016)	551	4	3	原核生物的文本数据与基因组学特征多视图

注:“+” 表示视图数量根据提取策略可能增加,“～” 表示样本范围。

界的复杂需求。1)探索从严格对齐向弱对齐场景的泛化。真实互联网环境下的海量异构数据(如电商行为图谱、多模态社交流)往往难以满足严格的样本级对齐假设。因此,如何在仅存在局部对齐或语义级弱对齐的碎片化数据上有效诱导全局一致性,是未来算法设计的重要方向。2)引入更具挑战性的原生异构多模态场景作为测试基准,如表3所示。尽管表3中的部分数据发布较早,但其具备的图文语义关联、音视频时空同步等原生多模态属性,比单纯的图像特征堆叠更接近真实场景的需求。例如,在自动驾驶多源传感器融合场景中检验算法的鲁棒性,或直接将大模型提取的高维特征作为视图输入,检验算法在极高维特征空间中的解耦能力。3)构建具备现实噪声分布的大规模基准数据集。领域内亟待联合工业界,共同探索并推出类似于ImageNet规模的多视图聚类真实场景数据集。通过提供脱敏且包含复杂现实分布(如长尾失衡、高缺失率)的工业级评测基准,引导多视图聚类研究从理论机制的探讨,发展至复杂真实场景的落地应用。

6 结 语

多视图聚类是挖掘多源异构数据一致性与互补性信息的关键技术。本文跳出传统基于算法结构的分类框架,提出了一种先验驱动的分类体系,对该领域的发展脉络进行了系统性梳理。具体而言,本文总结了欧氏、仿射及流形空间下的几何结构先验挖掘机制,探讨了深度嵌入表示、生成、对比学习与深度图网络主导的语义协同先验建模方法,并分析了

表3 原生的异构多模态数据集

Table 3 Native heterogeneous multimodal datasets

类别	数据集名称	样本数量	类别数	模态1	模态2	在MVC领域的拓展潜力与挑战
图文跨模态对齐	AwA (Animals) (Lampert等,2009)	30 475	50	RGB图像特征	85维人工语义属性	探索视觉特征与连续/离散语义属性特征之间的异构对齐聚类
	IAPR TC-12 (Escalante等,2010)	20 000	255	旅游生活图像	多语言(英/德)描述	适合测试超大规模跨语言图文聚类
	MirFlickr-25K (Huiskes和Lew,2008)	25 000	24	社交媒体图像	用户标注的短文本	大规模图文匹配,文本模态极度稀疏且充满社交网络主观噪声
	MS-COCO (Lin等,2014)	超30万	80	复杂场景图像	人工高质量长文本描述	样本量极其庞大
	NUS-WIDE (Chua等,2009)	269 648	81	Web图像	Flickr用户关联标签	标签存在大量缺失
	Pascal Sentences (Farhadi等,2010)	1 000	20	自然场景图像	50句自然语言描述	经典图文跨模态基准
	Wikipedia (Wiki) (Rasiwasia等,2010)	2 866	10	维基百科图像	对应的维基百科长文章	长文本包含大量专业词汇,导致图像与长文本的特征对齐极具挑战
音视频物理跨模态	CCV (Jiang等,2011)	9 317	20	消费级视频流	对应的环境/事件音频	大规模复杂音视频事件聚类,非结构化特征极强
	CUAVE (Patterson等,2002)	36个视频	10	嘴唇运动视频	同步的语音音频信号	音视频(audio/visual)聚类基准,需解决声学与视觉的时序协同问题
深度与3D传感器跨模态	NYU-Depth V2 (Silberman等,2012)	1 449	26	RGB彩色图像	深度传感器图像	室内场景,典型的同源异构物理传感器数据,适合三维场景理解

在多模态大模型生态下,聚类技术在数据治理、混合专家架构及检索增强生成等前沿应用中的认知先验对齐作用。

展望未来,多视图聚类领域仍面临诸多挑战与发展机遇,其核心趋势可归纳为以下3个维度:1)在机制设计上,聚类过程正从依赖底层特征的统计相似度计算,向利用大模型推理与常识能力的逻辑先验引导过渡,以实现更精准的跨模态语义解耦;2)在复杂场景应用中,突破海量高维数据的计算瓶颈,并在极端长尾分布与极度样本视图缺失的非对称条件下维持模型的鲁棒性,仍是亟待解决的工程难点;3)在评估体系建设上,领域亟需拓展现有理想化的小规模基准,向包含真实环境噪声、弱对齐假设的工业级大规模数据集转移。总体而言,通过在理论机制与评估范式上的持续完善,多视图聚类将进一步深化其在无监督知识发现中的核心价值,并为现代人工智能基础设施提供稳固的底层结构化支撑。

参考文献(References)

Abavisani M and Patel V M. 2018a. Multimodal sparse and low-rank subspace clustering. Information Fusion, 39: 168-177 [DOI: 10.1016/j.inffus.2017.05.002]

Abavisani M and Patel V M. 2018b. Deep multimodal subspace clustering networks. IEEE Journal of Selected Topics in Signal Processing, 12(6): 1601-1614 [DOI: 10.1109/JSTSP.2018.2875385]

Abbas A, Tirumala K, Simig D, Ganguli S and Morcos A S. 2023. SemDeDup: Data-efficient learning at web-scale through semantic deduplication [EB/OL]. [2026-03-01]. https://arxiv.org/pdf/2303.09540.pdf

Amini M R, Usunier N and Goutte C. 2009. Learning from multiple partially observed views — an application to multilingual text categorization//Proceedings of the 23rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 28-36

Bickel S and Scheffer T. 2004. Multi-view clustering//Proceedings of the 4th IEEE International Conference on Data Mining. Brighton, UK: IEEE: 19-26 [DOI: 10.1109/ICDM.2004.10095]

Blum A and Mitchell T. 1998. Combining labeled and unlabeled data with co-training//Proceedings of the 11th Annual Conference on Computational Learning Theory. Madison, USA: ACM: 92-100 [DOI: 10.1145/279943.279962]

Bo D Y, Wang X, Shi C, Zhu M Q, Lu E M and Cui P. 2020. Structural deep clustering network//Proceedings of 2020 Web Conference. Taipei, China: ACM: 1400-1410 [DOI: 10.1145/3366423.3380214]

Brbić M and Kopriva I. 2018. Multi-view low-rank sparse subspace clustering. Pattern Recognition, 73: 247-258 [DOI: 10.1016/j.patcog.2017.08.024]

Brbić M, Piškorec M, Vidulin V, Kriško A, Šmuc T and Supek F. 2016. The landscape of microbial phenotypic traits and associated genes. Nucleic Acids Research, 44(21): 10074-10090 [DOI: 10.1093/nar/gkw964]

Breit A, Ott S, Agibetov A and Samwald M. 2020. OpenBioLink: a benchmarking framework for large-scale biomedical link prediction. Bioinformatics, 36(13): 4097-4098 [DOI: 10.1093/bioinformatics/btaa274]

Brière G, Darbo É, Thébault P and Uricaru R. 2021. Consensus clustering applied to multi-omics disease subtyping. BMC Bioinformatics, 22(1): #361 [DOI: 10.1186/s12859-021-04279-1]

Cai X, Nie F P and Huang H. 2013. Multi-view k-means clustering on big data//Proceedings of the 23rd International Joint Conference on Artificial Intelligence. Beijing, China: AAAI Press: 2598-2604

Cai X, Nie F P, Huang H and Kamangar F. 2011. Heterogeneous image feature integration via multi-modal spectral clustering//Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE: 1977-1984 [DOI: 10.1109/CVPR.2011.5995740]

Cai X, Wang H, Huang H and Ding C. 2012. Joint stage recognition and anatomical annotation of drosophila gene expression patterns. Bioinformatics, 28(12): i16-i24 [DOI: 10.1093/bioinformatics/bts220]

Cai X S, Huang D, Zhang G Y and Wang C D. 2023. Seeking commonness and inconsistencies: a jointly smoothed approach to multi-view subspace clustering. Information Fusion, 91: 364-375 [DOI: 10.1016/j.inffus.2022.10.020]

Canh H T, Thong P H, Giang L T and Hung P D. 2024. Fake news detection based on multi-view fuzzy clustering algorithm//Proceedings of the 14th International Conference on Ad Hoc Networks. Hanoi, Vietnam: Springer: 150-166 [DOI: 10.1007/978-3-031-55993-8_12]

Cao X C, Zhang C Q, Fu H Z, Liu S and Zhang H. 2015. Diversity-induced multi-view subspace clustering//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 586-594 [DOI: 10.1109/CVPR.2015.7298657]

Che H J, Guo W, Leung M F, Cao Y T and Liu C. 2025. Robust diverse multi-view learning for cancer subtyping. IEEE Transactions on Computational Biology and Bioinformatics, 2025, 22(6): 2685-2696 [DOI: 10.1109/TCBBIO.2025.3599217]

Chen J, Mao H, Peng D Z, Zhang C Q and Peng X. 2023. Multiview clustering by consensus spectral rotation fusion. IEEE Transactions on Image Processing, 32: 5153-5166 [DOI: 10.1109/TIP.2023.3310339]

Chen Y Y, Wang S Q, Peng C, Hua Z Y and Zhou Y C. 2021. Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering. IEEE Transactions on Image Processing, 30:

4022-4035［DOI：10.1109/TIP.2021.3068646］

Chen Y Y，Xiao X L，Peng C，Lu G M and Zhou Y C. 2022. Low-rank tensor graph learning for multi-view subspace clustering. IEEE Transactions on Circuits and Systems for Video Technology，32（1）：92-104［DOI：10.1109/TCSVT.2021.3055625］

Cheng J F，Wang Q Q，Tao Z Q，Xie D Y and Gao Q X. 2020. Multi-view attribute graph convolution networks for clustering//Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama，Japan：IJCAI：2973-2979［DOI：10.24963/ijcai.2020/411］

Chua T S，Tang J H，Hong R C，Li H J，Luo Z P and Zheng Y T. 2009. NUS-WIDE：a real-world web image database from National University of Singapore//Proceedings of 2009 ACM International Conference on Image and Video Retrieval. Santorini，Greece：ACM：#48［DOI：10.1145/1646396.1646452］

Cruickshank I J and Carley K M. 2020. Characterizing communities of hashtag usage on twitter during the 2020 COVID-19 pandemic by multi-view clustering. Applied Network Science，5(1)：#66［DOI：10.1007/s41109-020-00317-8］

Cui C H，Ren Y Z，Pu J Y，Pu X R and He L F. 2023. Deep multi-view subspace clustering with anchor graph//Proceedings of the 32nd International Joint Conference on Artificial Intelligence. Macao，China：IJCAI：3577-3585［DOI：10.24963/ijcai.2023/398］

Cui J R，Wu X H，Zhang H T，Dong C J and Wen J. 2026. Structure-guided deep multi-view clustering. Information Fusion，125：#103461［DOI：10.1016/j.inffus.2025.103461］

Dai D M，Deng C Q，Zhao C G，Xu R X，Gao H Z，Chen D L，et al. 2024. DeepSeekMoE：towards ultimate expert specialization in mixture-of-experts language models//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok，Thailand：Association for Computational Linguistics：1280-1297［DOI：10.18653/v1/2024.acl-long.70］

Ding C，Li T，Peng W and Park H. 2006. Orthogonal nonnegative matrix t-factorizations for clustering//Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Philadelphia，USA：ACM：126-135［DOI：10.1145/1150402.1150420］

Du G W，Zhou L H，Li Z X，Wang L Z and Lyu K. 2023. Neighbor-aware deep multi-view clustering via graph convolutional network. Information Fusion，93：330-343［DOI：10.1016/j. inffus. 2023. 01.001］

Elhamifar E and Vidal R. 2013. Sparse subspace clustering：algorithm，theory，and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence，35（11）：2765-2781［DOI：10.1109/TPAMI.2013.57］

Escalante H J，Hernandez C A，Gonzalez J A，Lopez-Lopez A，Montes M，Morales E F，et al. 2010. The segmented and annotated IAPR TC-12 benchmark. Computer Vision and Image Understanding，114(4)：419-428

Fan S H，Wang X，Shi C，Lu E M，Lin K and Wang B. 2020. One2Multi graph autoencoder for multi-view graph clustering//Proceedings of 2020 Web Conference. Taipei，China：ACM：3070-3076［DOI：10.1145/3366423.3380079］

Fang U，Li M，Li J X，Gao L X，Jia T and Zhang Y C. 2023. A comprehensive survey on multi-view clustering. IEEE Transactions on Knowledge and Data Engineering，35（12）：12350-12368［DOI：10.1109/TKDE.2023.3270311］

Farhadi A，Hejrati M，Sadeghi M A，Young P，Rashtchian C，Hockenmaier J，et al. 2010. Every picture tells a story：generating sentences from images//Proceedings of the 11th European Conference on Computer Vision. Heraklion，Greece：Springer：15-29［DOI：10.1007/978-3-642-15561-1_2］

Fei L K，He J L，Zhu Q，Zhao S P，Wen J and Xu Y. 2025. Deep multi-view contrastive clustering via graph structure awareness. IEEE Transactions on Image Processing，34：3805-3816［DOI：10.1109/TIP.2025.3573501］

Gao D H，Jin L B，Chen B，Qiu M H，Li P，Wei Y，et al. 2020a. FashionBERT：text and image matching with adaptive loss for cross-modal retrieval//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM：2251-2260［DOI：10.1145/3397271.3401430］

Gao H C，Nie F P，Li X L and Huang H. 2015. Multi-view subspace clustering//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago，Chile：IEEE：4238-4246［DOI：10.1109/ICCV.2015.482］

Gao Q X，Xia W，Wan Z Z，Xie D Y and Zhang P. 2020b. Tensor-SVD based graph learning for multi-view subspace clustering//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York，USA：AAAI Press：3930-3937［DOI：10.1609/aaai.v34i04.5807］

Ge Y，Liu Y L，Hu C，Meng W B，Tao S M，Zhao X F，et al. 2024. Clustering and ranking：diversity-preserved instruction selection through expert-aligned quality estimation//Proceedings of 2024 Conference on Empirical Methods in Natural Language Processing. Miami，USA：Association for Computational Linguistics：464-478［DOI：10.18653/v1/2024.emnlp-main.28］

Georghiades A S，Belhumeur P N and Kriegman D J. 2001. From few to many：illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence，23（6）：643-660［DOI：10.1109/34.927464］

Greene D and Cunningham P. 2006. Practical solutions to the problem of diagonal dominant ensembles//Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh，USA：ACM：377-384［DOI：10.1145/1143844.1143892］

Greene D and Cunningham P. 2009. A matrix factorization approach for integrating multiple data views//Machine Learning and Knowledge Discovery in Databases. Bled，Slovenia：Springer：423-438［DOI：

10.1007/978-3-642-04180-8_45]

Gu Z B and Feng S H. 2024. From dictionary to tensor: a scalable multi-view subspace clustering framework with triple information enhancement//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #3288

Guan R X, Li Z H, Tu W X, Wang J, Liu Y, Li X J, et al. 2024. Contrastive multiview subspace clustering of hyperspectral images based on graph convolutional networks. IEEE Transactions on Geoscience and Remote Sensing, 62: #5510514 [DOI: 10.1109/TGRS.2024.3370633]

Guo J P, Sun Y F, Gao J B, Hu Y L and Yin B C. 2023. Logarithmic Schatten-p norm minimization for tensorial multi-view subspace clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(3): 3396-3410 [DOI: 10.1109/TPAMI.2022.3179556]

Hassani K and Khas Ahmadi A H. 2020. Contrastive multi-view representation learning on graphs//Proceedings of the 37th International Conference on Machine Learning. Virtual, Austria (Vienna): PMLR: 4116-4126

Hu S Z, Tian B Y, Liu W B and Ye Y D. 2025. Self-supervised trusted contrastive multi-view clustering with uncertainty refined//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI Press: 17305-17313 [DOI: 10.1609/aaai.v39i16.33902]

Huang Z M, Ren Y Z, Pu X R, Huang S D, Xu Z L and He L F. 2023. Self-supervised graph attention networks for deep weighted multi-view clustering//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI Press: 7936-7943 [DOI: 10.1609/aaai.v37i7.25960]

Huiskes M J and Lew M S. 2008. The MIR flickr retrieval benchmark//Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval. Vancouver, Canada: ACM: 39-43 [DOI: 10.1145/1460096.1460104]

Ji J T and Feng S H. 2023. Anchor structure regularization induced multi-view subspace clustering via enhanced tensor rank minimization//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 19286-19295 [DOI: 10.1109/iccv51070.2023.01772]

Ji P, Zhang T, Li HD, Salzmann M and Reid I. 2017. Deep subspace clustering networks//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 23-32

Jiang H Y, Tao H, Jiang Z Q and Hou C P. 2026. Unaligned multi-view clustering via diversified anchor graph fusion. Pattern Recognition, 170: #111977 [DOI: 10.1016/j.patcog.2025.111977]

Jiang M D, Zhou J H, Hu L Y, Liu X Y, He Z Y and Chen Z K. 2025. Adversarial fair multi-view clustering [EB/OL]. [2026-03-01]. https://arxiv.org/pdf/2508.04071.pdf

Jiang Y, Li X C, Zhou J, Han S Y, Du T, Yang C, et al. 2024. Multi-view transfer clustering algorithm based on Gaussian mixture model//Proceedings of 2024 International Conference on Fuzzy Theory and Its Applications (iFUZZY). Kagawa, Japan: IEEE: 1-5 [DOI: 10.1109/iFUZZY63051.2024.10661371]

Jiang Y G, Ye G N, Chang S F, Ellis D and Loui A C. 2011. Consumer video understanding: a benchmark database and an evaluation of human and machine performance//Proceedings of the 1st ACM International Conference on Multimedia Retrieval. Trento, Italy: ACM: 29 [DOI: 10.1145/1991996.1992025]

Jiang Z X, Zheng Y, Tan H C, Tang B S and Zhou H N. 2017. Variational deep embedding: an unsupervised and generative approach to clustering//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: IJCAI: 1965-1972 [DOI: 10.24963/ijcai.2017/273]

Kang Z, Zhou W T, Zhao Z T, Shao J M, Han M and Xu Z L. 2020. Large-scale multi-view subspace clustering in linear time//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press: 4412-4419 [DOI: 10.1609/aaai.v34i04.5867]

Krizhevsky A. 2009. Learning multiple layers of features from tiny images [EB/OL]. [2026-03-01]. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf

Kuang D, Park H and Ding C H Q. 2012. Symmetric nonnegative matrix factorization for graph clustering//Proceedings of 2012 SIAM International Conference on Data Mining. Anaheim, USA: SIAM: 106-117 [DOI: 10.1137/1.9781611972825.10]

Kumar A and Daumé III H. 2011a. A co-training approach for multi-view spectral clustering//Proceedings of the 28th International Conference on Machine Learning (ICML-11). Bellevue, USA: Omnipress: 393-400

Kumar A, Rai P and Daume III H. 2011b. Co-regularized multi-view spectral clustering//Proceedings of the 25th International Conference on Neural Information Processing Systems. Granada, Spain: Curran Associates Inc.: 1413-1421

Lampert C H, Nickisch H and Harmeling S. 2009. Learning to detect unseen object classes by between-class attribute transfer//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 951-958 [DOI: 10.1109/CVPR.2009.5206594]

Lang K. 1995. NewsWeeder: learning to filter netnews//Proceedings of the 12th International Conference on Machine Learning. Tahoe City, USA: Morgan Kaufmann: 331-339

Lazebnik S, Schmid C and Ponce J. 2006. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories//Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE: 2169-2178 [DOI: 10.1109/CVPR.2006.68]

Lecun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324 [DOI: 10.1109/5.726791]

Li F F, Fergus R and Perona P. 2004. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories//Proceedings of 2004 Conference on Computer Vision and Pattern Recognition Workshop. Washington, USA: IEEE: #178 [DOI: 10.1109/CVPR.2004.383]

Li G J, Yu Z W, Yang K X, Lyu J M and Chen C L P. 2026. Pseudo-label similarity graph-driven multi-view contrastive clustering. IEEE Transactions on Multimedia [DOI: 10.1109/tmm.2026.3660170]

Li J, Gao Q X, Wang Q Q, Yang M and Xia W. 2023. Orthogonal non-negative tensor factorization based multi-view clustering//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #800

Li S, Shao M and Fu Y. 2015. Multi-view low-rank analysis for outlier detection//Proceedings of 2015 SIAM International Conference on Data Mining. Vancouver, Canada: SIAM: 748-756 [DOI: 10.1137/1.9781611974010.84]

Li Y F, Hu P, Liu Z T, Peng D Z, Zhou J T and Peng X. 2021. Contrastive clustering//Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI Press: 8547-8555 [DOI: 10.1609/aaai.v35i10.17037]

Li Y F, Yang M X, Peng D Z, Li T H, Huang J T and Peng X. 2022a. Twin contrastive learning for online clustering. International Journal of Computer Vision, 130(9): 2205-2221 [DOI: 10.1007/s11263-022-01639-z]

Li Y M, Yang M and Zhang Z F. 2019a. A survey of multi-view representation learning. IEEE Transactions on Knowledge and Data Engineering, 31(10): 1863-1883 [DOI: 10.1109/TKDE.2018.2872063]

Li Y X, Michaud E J, Baek D D, Engels J, Sun X Q and Tegmark M. 2025. The geometry of concepts: sparse autoencoder feature structure. Entropy, 27(4): #344 [DOI: 10.3390/e27040344]

Li Z L, Tang C, Liu X W, Zheng X, Zhang W and Zhu E. 2022b. Consensus graph learning for multi-view clustering. IEEE Transactions on Multimedia, 24: 2461-2472 [DOI: 10.1109/TMM. 2021.3081930]

Li Z Y, Wang Q Q, Tao Z Q, Gao Q X and Yang Z H. 2019b. Deep adversarial multi-view clustering network//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China: IJCAI: 2952-2958 [DOI: 10.24963/ijcai.2019/409]

Liang W X, Liu X W, Liang K, Liu J Y and Zhu E. 2025. COKE: core kernel for more efficient approximation of kernel weights in multiple kernel clustering//Proceedings of the 42nd International Conference on Machine Learning. Vancouver, Canada: PMLR: 37257-37280

Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]

Lin Y J, Gou Y B, Liu X T, Bai J F, Lyu J C and Peng X. 2023. Dual contrastive prediction for incomplete multi-view representation learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 4447-4461 [DOI: 10.1109/tpami.2022.3197238]

Lin Y J, Gou Y B, Liu Z T, Li B Y, Lyu J C and Peng X. 2021a. COMPLETER: incomplete multi-view clustering via contrastive prediction//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 11169-11178 [DOI: 10.1109/CVPR46437.2021.01102]

Lin Y X, Liu H, Yu X and Zhang C M. 2025. Leveraging Transformer-based autoencoders for low-rank multi-view subspace clustering. Pattern Recognition, 161: #111331 [DOI: 10.1016/j.patcog.2024.111331]

Lin Z P, Kang Z, Zhang L Z and Tian L. 2021b. Multi-view attributed graph clustering. IEEE Transactions on Knowledge and Data Engineering, 35(2): 1872-1880 [DOI: 10.1109/TKDE.2021.3101227]

Liu J L, Wang C, Gao J and Han J W. 2013. Multi-view clustering via joint nonnegative matrix factorization//Proceedings of 2013 SIAM International Conference on Data Mining. Austin, USA: SIAM: 252-260 [DOI: 10.1137/1.9781611972832.28]

Liu Q L, Lu Z X, Huan W H and Fan C. 2025. Adaptive weighted multi-view subspace clustering method for recognizing urban functions from multi-source social sensing data. Geo-Spatial Information Science, 28(3): 1255-1279 [DOI: 10.1080/10095020.2024.2356243]

Liu S Y, Wang S W, Zhang P, Xu K, Liu X W, Zhang C W, et al. 2022. Efficient one-pass multi-view subspace clustering with consensus anchors//Proceedings of the 36th AAAI Conference on Artificial Intelligence. AAAI Press: 7576-7584 [DOI: 10.1609/aaai.v36i7.20723]

Liu X W. 2023. SimpleMKKM: simple multiple kernel K-means. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 5174-5186 [DOI: 10.1109/TPAMI.2022.3198638]

Liu X W, Dou Y, Yin J P, Wang L and Zhu E. 2016. Multiple kernel k-means clustering with matrix-induced regularization//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA: AAAI Press: 1888-1894 [DOI: 10.1609/aaai.v30i1.10249]

Liu X W, Liang K, Wang J, Liu S Y, Wang X K and Wang H M. 2026. Two decades of multi-view clustering: taxonomy, application, and challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 48(3): 3744-3764 [DOI: 10.1109/TPAMI. 2025.3640109]

Lock E F and Dunson D B. 2013. Bayesian consensus clustering. Bioinformatics, 29(20): 2610-2616 [DOI: 10.1093/bioinformatics/btt425]

Martínez A M and Benavente R. 1998. The AR face database. CVC Technical Report No. 24, Bellatera, Spain: Computer Vision Center (CVC), Universitat Autonoma de Barcelona.

McLachlan G J and Peel D. 2000. Finite Mixture Models. New York, USA: Wiley [DOI: 10.1002/0471721182]

Miao Y Q, Xu H and Wang S. 2025. PartIES: a disease subtyping framework with partition-level integration using diffusion-enhanced similarities from multi-omics data. Briefings in Bioinformatics, 26(1): #bbae609 [DOI: 10.1093/bib/bbae609]

Mukherjee S, Asnani H, Lin E and Kannan S. 2019. ClusterGAN: latent space clustering in generative adversarial networks//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI Press: 4610-4617 [DOI: 10.1609/aaai.v33i01.33014610]

Nene S A, Nayar S K and Murase H. 1996. Columbia object image library (COIL-20) [EB/OL]. [2026-03-01]. https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php

Nie F P, Wang X Q, Jordan M I and Huang H. 2016a. The constrained laplacian rank algorithm for graph-based clustering//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA: AAAI Press: 1969-1976 [DOI: 10.1609/aaai.v30i1.10302]

Nie F P, Li J and Li X L. 2016b. Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA: AAAI: 1881-1887

Nie F P, Li J and Li X L. 2017. Self-weighted multiview clustering with multiple graphs//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: IJCAI: 2564-2570 [DOI: 10.24963/ijcai.2017/357]

Nilsback M E and Zisserman A. 2008. Automated flower classification over a large number of classes//Proceedings of the 6th Indian Conference on Compu Vision, Graphics and Image Processing. Bhubaneswar, India: IEEE: 722-729 [DOI: ter10.1109/ICVGIP.2008.47]

Ou Q Y, Wang S W, Zhang P, Zhou S H and Zhu E. 2024. Anchor-based multi-view subspace clustering with hierarchical feature descent. Information Fusion, 106: #102225 [DOI: 10.1016/j.inffus.2024.102225]

Ouyang S, Yuan H L, Wang D L, Sun Y, Ren Z W and Li X F. 2026. CLIP-driven lifelong multi-view clustering. Pattern Recognition, 178: #113425 [DOI: 10.1016/j.patcog.2026.113425]

Pan E L and Kang Z. 2021. Multi-view contrastive graph clustering//Proceedings of the 35th International Conference on Neural Information Processing Systems. Curran Associates Inc. 2148-2159

Patterson E K, Gurbuz S, Tufekci Z and Gowdy J N. 2002. CUAVE: a new audio-visual database for multimodal human-computer interface research//Proceedings of 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing. Orlando, USA: IEEE: 2017-2020 [DOI: 10.1109/ICASSP.2002.5745028]

Pei C P, Dornaika F and Bi J J. 2025. MCFCN: multi-view clustering via a fusion-consensus graph convolutional network [EB/OL]. [2026-03-01]. https://arxiv.org/pdf/2511.05554.pdf

Peng Z, Luo M N, Li J D, Xue L G and Zheng Q H. 2022a. A deep multi-view framework for anomaly detection on attributed networks. IEEE Transactions on Knowledge and Data Engineering, 34(6): 2539-2552 [DOI: 10.1109/TKDE.2020.3015098]

Rasiwasia N, Costa Pereira J, Coviello E, Doyle G, Lanckriet G R G, Levy R, et al. 2010. A new approach to cross-modal multimedia retrieval//Proceedings of the 18th ACM International Conference on Multimedia. Firenze, Italy: ACM: 251-260 [DOI: 10.1145/1873951.1873987]

Ren Y Z, Pu J Y, Cui C H, Zheng Y, Chen X Y, Pu X R, et al. 2024. Dynamic weighted graph fusion for deep multi-view clustering//Proceedings of the 33rd International Joint Conference on Artificial Intelligence. Jeju, Korea (South): IJCAI: 4842-4850 [DOI: 10.24963/ijcai.2024/535]

Sarthi P, Abdullah S, Tuli A, Khanna S, Goldie A and Manning C D. 2024. RAPTOR: recursive abstractive processing for tree-organized retrieval//Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: ICLR: 32628-32649

Sen P, Namata G, Bilgic M, Getoor L, Gallagher B and Eliassi-Rad T. 2008. Collective classification in network data. AI Magazine, 29(3): 93-106 [DOI: 10.1609/aimag.v29i3.2157]

Shapiro B D and Battle A. 2022. Bayesian multi-view clustering given complex inter-view structure. F1000Research, 11: #1460 [DOI: 10.12688/f1000research.126215.1]

Shchur O, Mumme M, Bojchevski A and Günnemann S. 2018. Pitfalls of graph neural network evaluation [EB/OL]. [2026-03-01]. https://arxiv.org/pdf/1811.05868.pdf

Silberman N, Hoiem D, Kohli P and Fergus R. 2012. Indoor segmentation and support inference from RGBD images//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer: 746-760 [DOI: 10.1007/978-3-642-33715-4_54]

Sun M J, Zhang P, Wang S W, Zhou S H, Tu W X, Liu X W, et al. 2021. Scalable multi-view subspace clustering with unified anchors//Proceedings of the 29th ACM International Conference on Multimedia. [s. l.]: ACM: 3528-3536 [DOI: 10.1145/3474085.3475516]

Tang C, Li Z L, Wang J, Liu X W, Zhang W and Zhu E. 2023. Unified one-step multi-view spectral clustering. IEEE Transactions on Knowledge and Data Engineering, 35(6): 6449-6460 [DOI: 10.1109/TKDE.2022.3172687]

Tang C, Zhu X Z, Liu X W, Li M M, Wang P C, Zhang C Q, et al. 2019. Learning a joint affinity graph for multiview subspace clustering. IEEE Transactions on Multimedia, 21(7): 1724-1736 [DOI: 10.1109/TMM.2018.2889560]

Tian Y L, Krishnan D and Isola P. 2020. Contrastive multiview coding. European Conference on Computer Vision (ECCV), Glasgow, UK: 776-794 [DOI: 10.1007/978-3-030-58621-8_45]

Trosten D J, Løkse S, Jenssen R and Kampffmeyer M. 2021. Reconsidering representation alignment for multi-view clustering//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 1255-1265 [DOI: 10.1109/CVPR46437.2021.00131]

Trosten D J, Løkse S, Jenssen R and Kampffmeyer M C. 2023. On the effects of self-supervision and contrastive alignment in deep multi-view clustering//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 23976-23985 [DOI: 10.1109/CVPR52729.2023.02296]

Tu W X, Zhou S H, Liu X W, Guo X F, Cai Z P, Zhu E, et al. 2021. Deep fusion clustering network//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Virtual, Online (Global): AAAI Press: 9978-9987 [DOI: 10.1609/aaai.v35i11.17198]

Tzortzis G and Likas A. 2012. Kernel-based weighted multi-view clustering//Proceedings of the 12th IEEE International Conference on Data Mining. Brussels, Belgium: IEEE: 675-684 [DOI: 10.1109/ICDM.2012.43]

van Breukelen M, Duin R P W, Tax D M J and Den Hartog J E. 1998. Handwritten digit recognition by combined classifiers. Kybernetika, 34(4): 381-386

Van G J and Abbott I. 2003. 100 leaf dataset [EB/OL]. [2026-03-01]. https://archive.ics.uci.edu/ml/datasets/One-hundred+plant+species+leaves+data+set

Vidal R. 2011. Subspace clustering. IEEE Signal Processing Magazine, 28(2): 52-68 [DOI: 10.1109/MSP.2010.939739]

Vidal R and Favaro P. 2014. Low rank subspace clustering (LRSC). Pattern Recognition Letters, 43: 47-61 [DOI: 10.1016/j.patrec.2013.08.006]

Wah C, Branson S, Welinder P, Perona P and Belongie S. 2011. The caltech-ucsd birds-200-2011 dataset [EB/OL]. [2026-03-01]. http://www.vision.caltech.edu/visipedia/CUB-200-2011.html

Wang B C, Zeng C S, Chen M L and Li X L. 2025a. Towards learnable anchor for deep multi-view clustering//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI Press: 21044-21052 [DOI: 10.1609/aaai.v39i20.35401]

Wang H, Yang Y and Liu B. 2020. GMC: graph-based multi-view clustering. IEEE Transactions on Knowledge and Data Engineering, 32(6): 1116-1129 [DOI: 10.1109/TKDE.2019.2903810]

Wang H M, Chen T H, Wang H T, Shao X Q and Su P. 2018. Fuzzy clustering based anomaly detection for distributed multi-view data//Proceedings of 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). Rio de Janeiro, Brazil: IEEE: 1-7 [DOI: 10.1109/fuzz-ieee.2018.8491464]

Wang J, Feng S H, Lyu G Y and Yuan J Z. 2024a. Surer: Structure-adaptive unified graph neural network for multi-view clustering//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 15520-15527 [DOI: 10.1609/aaai.v38i14.29478]

Wang M, Huang Z Z, He H G, Lu H C, Shan H M and Zhang J P. 2024b. Potential and prospects of segment anything model: a survey. Journal of Image and Graphics, 29(6): 1479-1509 (王淼，黄智忠，何晖光，卢湖川，单洪明，张军平. 2024. 分割一切模型SAM的潜力与展望：综述. 中国图象图形学报，29(6)：1479-1509) [DOI: 10.11834/jig.230792]

Wang N, Cui Z G, Li A H, Wang R and Nie F P. 2026. Multi-view clustering based on doubly stochastic graph. Signal Processing, 238: #110144 [DOI: 10.1016/j.sigpro.2025.110144]

Wang Q Q, Cheng J F, Gao Q X, Zhao G S and Jiao L C. 2021. Deep multi-view subspace clustering with unified and discriminative learning. IEEE Transactions on Multimedia, 23: 3483-3493 [DOI: 10.1109/TMM.2020.3025666]

Wang Q Q, Tao Z Q, Xia W, Gao Q X, Cao X C and Jiao L C. 2023. Adversarial multiview clustering networks with adaptive fusion. IEEE Transactions on Neural Networks and Learning Systems, 34(10): 7635-7647 [DOI: 10.1109/TNNLS.2022.3145048]

Wang Q Q, Xu H M, Zhang Z H, Tao Z Q and Gao Q X. 2025b. Efficient multi-view clustering via reinforcement contrastive learning//Proceedings of the 34th International Joint Conference on Artificial Intelligence. Montreal, Canada: IJCAI: 6361-6369 [DOI: 10.24963/ijcai.2025/708]

Wang W R, Arora R, Livescu K and Bilmes J A. 2015. On deep multi-view representation learning//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR: 1083-1092

Wang X, Ji H Y, Shi C, Wang B, Ye Y F, Cui P, et al. 2019. Heterogeneous graph attention network//Proceedings of 2019 Web Conference. San Francisco, USA: ACM: 2022-2032 [DOI: 10.1145/3308558.3313562]

Wang X B, Guo X J, Lei Z, Zhang C Q and Li S Z. 2017. Exclusivity-consistency regularized multi-view subspace clustering//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1-9 [DOI: 10.1109/CVPR.2017.8]

Wang X X, Zhang Y S and Zhou Y C. 2025c. Multimodal remote sensing image clustering with multi-scale spectral-spatial anchor graphs. IEEE Transactions on Geoscience and Remote Sensing, 63: #4405612 [DOI: 10.1109/TGRS.2025.3545460]

Wang Y J, Xu Q Q, Jiang Y B Y, Dai S R and Huang Q M. 2024c. Regularized contrastive partial multi-view outlier detection//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, Australia: Association for Computing Machinery: 8711-8720 [DOI: 10.1145/3664647.3681125]

Wolf L, Hassner T and Maoz I. 2011. Face recognition in unconstrained videos with matched background similarity//Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE: 529-534 [DOI: 10.1109/CVPR.2011.5995566]

Wu M K and Goodman N. 2018. Multimodal generative models for scalable weakly-supervised learning//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc.: 5580-5590

Xia W, Gao Q X, Wang Q Q, Gao X B, Ding C and Tao D C. 2023. Tensorized bipartite graph learning for multi-view clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 5187-5202 [DOI: 10.1109/TPAMI.2022.3187976]

Xia W, Wang Q Q, Gao Q X, Zhang X D and Gao X B. 2022. Self-supervised graph convolutional network for multi-view clustering. IEEE Transactions on Multimedia, 24: 3182-3192 [DOI: 10.1109/TMM.2021.3094296]

Xiao H, Rasul K and Vollgraf R. 2017. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms [EB/OL]. [2026-03-01]. https://arxiv.org/pdf/1708.07747.pdf

Xu C, Guan Z Y, Zhao W, Wu H C, Niu Y F and Ling B L. 2019. Adversarial incomplete multi-view clustering//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China: IJCAI: 3933-3939 [DOI: 10.24963/ijcai.2019/546]

Xu H M and Wang Q Q. 2025. LLM-DAMVC: a large language model assisted dynamic agent for multi-view clustering//Proceedings of the 39th Annual Conference on Neural Information Processing Systems. San Diego, USA: Curran Associates Inc.

Xu J, Chen S, Ren Y Z, Shi X S, Shen H T, Niu G, et al. 2023a. Self-weighted contrastive learning among multiple views for mitigating representation degeneration//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #54

Xu J, Ren Y Z, Tang H Y, Pu X R, Zhu X F, Zeng M, et al. 2021. Multi-VAE: learning disentangled view-common and view-peculiar visual representations for multi-view clustering//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9214-9223 [DOI: 10.1109/ICCV48922.2021.00910]

Xu J, Tang H Y, Ren Y Z, Peng L, Zhu X F and He L F. 2022. Multi-level feature learning for contrastive multi-view clustering//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 16030-16039 [DOI: 10.1109/CVPR52688.2022.01558]

Xu W Z, Huang T H, Ben X Y, Zeng Y and Zhang J P. 2023b. Cross-view gait recognition: a review. Journal of Image and Graphics, 28(5): 1265-1286 (许文正, 黄天欢, 贲晛烨, 曾翌, 张军平. 2023. 跨视角步态识别综述. 中国图象图形学报, 28(5): 1265-1286) [DOI: 10.11834/jig.220458]

Xue X Q, Cai Q, Zhang Z W, Lei Y M, Shan H M and Zhang J P. 2026. Adaptive multi-view consistency clustering via structure-enhanced contrastive learning. Pattern Recognition, 172: #112409 [DOI: 10.1016/j.patcog.2025.112409]

Xue X Q, Lei Y M, Cai Q, Shan H M and Zhang J P. 2025. PROTOCOL: partial optimal transport-enhanced contrastive learning for imbalanced multi-view clustering//Proceedings of the 42nd International Conference on Machine Learning. Vancouver, Canada: PMLR: 70105-70119

Yan W Q, Zhang Y Y, Lyu C L, Tang C, Yue G H, Liao L, et al. 2023. GCFAgg: global and cross-view feature aggregation for multi-view clustering//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 19863-19872 [DOI: 10.1109/CVPR52729.2023.01902]

Yan X Q, Jin Z X, Han F S and Ye Y D. 2024. Differentiable information bottleneck for deterministic multi-view clustering//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 27425-27434 [DOI: 10.1109/CVPR52733.2024.02590]

Yang B, Zhang X T, Nie F P, Wang F, Yu W Z and Wang R. 2021. Fast multi-view clustering via nonnegative and orthogonal factorization. IEEE Transactions on Image Processing, 30: 2575-2586 [DOI: 10.1109/TIP.2020.3045631]

Yang W Q, Wang M H, Tang C, Zheng X, Liu X W and He K L. 2024a. Trustworthy multi-view clustering via alternating generative adversarial representation learning and fusion. Information Fusion, 107: #102323 [DOI: 10.1016/j.inffus.2024.102323]

Yang X H, Wang S W, Jin J Q, Wang F D, Liu T R, Jin Y M, et al. 2025. Generalized deep multi-view clustering via causal learning with partially aligned cross-view correspondence [EB/OL]. [2026-03-01]. https:/arxiv.org/pdf/2509.16022.pdf

Yang Z B, Tan Y H and Yang T. 2024b. Large-scale multi-view clustering via matrix factorization of consensus graph. Pattern Recognition, 155: #110716 [DOI: 10.1016/j.patcog.2024.110716]

Yuan H L, Sun Y, Zhou F, Wen J, Yuan S H, You X J, et al. 2025. Prototype matching learning for incomplete multi-view clustering. IEEE Transactions on Image Processing, 34: 828-841 [DOI: 10.1109/TIP.2025.3529378]

Zhan K, Nie F P, Wang J and Yang Y. 2019. Multiview consensus graph clustering. IEEE Transactions on Image Processing, 28(3): 1261-1270 [DOI: 10.1109/TIP.2018.2877335]

Zhan K, Zhang C Q, Guan J P and Wang J S. 2018. Graph learning for multiview clustering. IEEE Transactions on Cybernetics, 48(10): 2887-2895 [DOI: 10.1109/TCYB.2017.2751646]

Zhang C, Jia X Y, Li Z C, Chen C L and Li H X. 2024. Learning cluster-wise anchors for multi-view clustering//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 16696-16704 [DOI: 10.1609/aaai.v38i15.29609]

Zhang C Q, Fu H Z, Hu Q H, Cao X C, Xie Y, Tao D C, et al. 2020. Generalized latent multi-view subspace clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(1): 86-99 [DOI: 10.1109/TPAMI.2018.2877660]

Zhang C Q, Fu H Z, Liu S, Liu G C and Cao X C. 2015. Low-rank tensor constrained multiview subspace clustering//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile：IEEE：1582-1590［DOI：10.1109/ICCV.2015.185］

Zhang C Q，Hu Q H，Fu H Z，Zhu P F and Cao X C. 2017. Latent multi-view subspace clustering//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu，USA：IEEE：4333-4341［DOI：10.1109/CVPR.2017.461］

Zhang J，Zhang C X，Liu Y，Jin Y X，Yang X W，Zheng B，et al. 2025a. D3：diversity，difficulty，and dependability-aware data selection for sample-efficient LLM instruction tuning//Proceedings of the 34th International Joint Conference on Artificial Intelligence. Montreal，Canada：IJCAI.org：8348-8356［DOI：10.24963/ijcai.2025/928］

Zhang F Y，Zhu D Z，Ming J，Jin Y L，Chai D，Yang L，et al. 2025b. DH-RAG：a dynamic historical context-powered retrieval-augmented generation method for multi-turn dialogue［EB/OL］.［2026-03-01］. https://arxiv.org/pdf/2502.13847.pdf

Zhang P，Liu X W，Xiong J，Zhou S H，Zhao W T，Zhu E，et al. 2022. Consensus one-step multi-view subspace clustering. IEEE Transactions on Knowledge and Data Engineering，34(10)：4676-4689［DOI：10.1109/TKDE.2020.3045770］

Zhang Y W，Wang Z H and Shang J B. 2023. ClusterLLM：large language models as a guide for text clustering//Proceedings of 2023 Conference on Empirical Methods in Natural Language Processing. Singapore，Singapore：Association for Computational Linguistics：13903-13920［DOI：10.18653/v1/2023.emnlp-main.858］

Zhang Y Z and Lu G F. 2026. Tensorized multi-dimensional multi-view clustering based on nonnegative matrix factorization. Neural Networks，197：#108507［DOI：10.1016/j.neunet.2025.108507］

Zhao H D and Fu Y. 2015. Dual-regularized multi-view outlier detection//Proceedings of the 24th International Joint Conference on Artificial Intelligence. Buenos Aires，Argentina：AAAI：4077-4083［DOI：10.24963/ijcai.2015/572］

Zhou L H，Du G W，Lyu K，Wang L Z and Du J W. 2024. A survey and an empirical evaluation of multi-view clustering approaches. ACM Computing Surveys，56(7)：1-38［DOI：10.1145/3645108］

Zhou R W and Shen Y D. 2020. End-to-end adversarial-attention network for multi-modal clustering//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：14607-14616［DOI：10.1109/CVPR42600.2020.01463］

Zhu J，Zou X，Wang X，Liu L，Tang C and Dai L R. 2026. Generative diffusion contrastive network for multi-view clustering. IEEE Signal Processing Letters，33：1316-1320［DOI：10.1109/LSP.2026.3668750］

Zhu P F，Yao X J，Wang Y，Hui B Y，Du D W and Hu Q H. 2024. Multiview deep subspace clustering networks. IEEE Transactions on Cybernetics，54(7)：4280-4293［DOI：10.1109/TCYB.2024.3372309］

Zhu X F，Zhang S C，He W，Hu R Y，Lei C and Zhu P F. 2019. One-step multi-view spectral clustering. IEEE Transactions on Knowledge and Data Engineering，31(10)：2022-2034［DOI：10.1109/TKDE.2018.2873378］

Zong L L，Zhang X C，Zhao L，Yu H and Zhao Q L. 2017. Multi-view clustering via multi-manifold regularized non-negative matrix factorization. Neural Networks，88：74-89［DOI：10.1016/j.neunet.2017.02.003］

作者简介

薛旭倩，女，博士研究生，主要研究方向为多视图聚类、深度学习和多模态学习。E-mail：xqxue22@m.fudan.edu.cn

张军平，通信作者，男，教授，主要研究方向为机器学习、图像处理、生物识别认证和智能交通系统。
E-mail：jpzhang@fudan.edu.cn

文杰，男，教授，主要研究方向为模式识别、多模态学习、机器学习和异常检测。E-mail：jiewen_pr@126.com

刘新旺，男，教授，主要研究方向为计算机视觉、图像处理、数据挖掘和模式识别。E-mail：xinwangliu@nudt.edu.cn