

DOI:10.16356/j.2097-6771.2026.02.022

基于双通道粒计算的深度多视图聚类方法

蔡超越¹, 马星如¹, 郭静¹, 胡鑫¹, 鞠恒荣^{1,2}, 丁卫平¹

(1. 南通大学人工智能与计算机学院, 南通 226019; 2. 南京大学计算机软件新技术国家重点实验室, 南京 210023)

摘要: 针对不同视图间存在质量差异、边界样本处理困难及局部语义结构不一致等问题, 本文提出了一种基于双通道粒计算的深度多视图聚类方法。通过双通道特征融合模块, 利用全局平均池化通道与全局最大池化通道分别获取视图的整体语义与显著判别特征, 并融合生成增强特征。同时引入双通道对比学习策略, 分别在样本级特征空间和局部模糊粒球结构进行对比学习, 模糊粒球级对比学习分为粒球内部模糊粒球对比学习和跨视图模糊粒球对比学习, 前者在优化聚类边界的同时使得粒球内部正样本更加靠近, 后者可以确保不同视图学习到一致的粒球结构。此外, 本文引入了视图自适应注意力权重分配机制, 提升高质量视图在聚类中的主导作用。在 8 个公开的多视图数据集上验证了本文方法的有效性。结果表明, 本方法和现有的 MFLVC, SCMVC 等多视图聚类方法相比, 提高了聚类的准确性。

关键词: 深度多视图聚类; 双通道对比学习; 跨视图模糊粒球; 视图自适应注意力权重分配; 双通道特征融合; 粒计算

中图分类号: TP183 文献标志码: A 文章编号: 1005-2615(2026)02-0457-14

Deep Multi-view Clustering with Dual-Channel Granular Computing

CAI Chaoyue¹, MA Xingru¹, GUO Jing¹, HU Xin¹, JU Hengrong^{1,2*}, DING Weiping¹

(1. School of Artificial Intelligence and Computer Science, Nantong University, Nantong 226019, China; 2. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China)

Abstract: In order to deal with the problems of quality differences among different views, ambiguous boundary samples and differences in semantic structures among different views, we propose deep multi-view clustering with dual-channel granular computing. A dual-channel feature fusion module is designed to strengthen key representations, where the global average pooling channel captures holistic semantics, and the global max pooling channel focuses on highly discriminative cues. Furthermore, a dual-channel contrast learning strategy is introduced for contrast learning at the sample and local fuzzy granular-ball structure level respectively. Fuzzy granular-ball level contrast learning is divided into intra-granular-ball and cross-view fuzzy granular-ball contrast learning. The former optimizes the clustering boundary by making positive samples inside the granular-ball closer. The latter ensures consistent granular-ball structures are learned across different views. Additionally, this paper introduces a view-adaptive attention weight assignment mechanism that enhances the leading role of high-quality views in clustering. We verify the effectiveness of our method on eight publicly available multi-view datasets. The results show that our method improves clustering accuracy compared to the existing multi-view clustering methods, such as MFLVC, SCMVC, etc.

基金项目: 国家自然科学基金(62006128); 南京大学计算机软件新技术国家重点实验室资助项目(KFKT2024B30); 南通市自然科学基金(JC2024044); 教育部产学研合作协同育人项目(2409233550)。

收稿日期: 2025-07-17; **修订日期:** 2025-10-29

通信作者: 鞠恒荣, 男, 副教授, E-mail: juhengrong@ntu.edu.cn。

引用格式: 蔡超越, 马星如, 郭静, 等. 基于双通道粒计算的深度多视图聚类方法[J]. 南京航空航天大学学报(自然科学版), 2026, 58(2): 457-470. CAI Chaoyue, MA Xingru, GUO Jing, et al. Deep multi-view clustering with dual-channel granular computing[J]. Journal of Nanjing University of Aeronautics & Astronautics (Natural Science Edition), 2026, 58(2): 457-470.

Key words: deep multi-view clustering; dual-channel contrastive learning; cross-view fuzzy granular-ball; view-adaptive attention weight assignment; dual-channel feature fusion; granular computing

随着信息技术的快速发展,大量数据从不同来源收集或者以不同的属性进行描述。通常,这些数据缺乏标签信息。例如,在视频记录的情况下,它可能从不同相机中捕获图像,包含不同声音音频和文本描述,在图像记录的情况下,可能获得不同角度的图像。在图像识别、社交网络、生物信息学等实际应用中普遍存在着多视图数据。为了以无监督的方式探索多个视图之间有用的一致和互补信息,多视图聚类^[1-2]旨在整合来自不同来源的数据,挖掘一致与互补信息,从而更全面地获取多视图特征,提高聚类性能。

传统多视图聚类方法按照融合策略的不同可以分为早期融合、中期融合和晚期融合 3 种方法。早期融合方法直接连接多视图特征,实现简单,但是容易被多视图间的异常值影响。中期融合方法是在特征提取或表示学习阶段进行信息整合,因此已逐渐成为主流研究方向。晚期融合则是每个视图先独立进行聚类,之后再对聚类结果加权整合,这种方法虽保留了各视图的特质,但难以充分提取跨视图一致性语义。

由于深度神经网络^[3]有强大的表示学习能力,深度聚类^[4-7]逐渐成为研究热点。目前很多研究方法采用自编码器从多视图数据里学习潜在表示来驱动聚类。对比学习^[8-9]作为自监督表示学习方法的重要分支,通过构造正负样本对来提高特征的分辨能力。对比学习目前在计算机视觉等领域已取得显著成果,所以将对比学习和聚类任务相结合可能进一步推动无监督学习性能提升。

尽管已有研究取得显著进展,但仍要面临以下挑战:(1)现有的深度多视图聚类方法多数仅关注样本级相似性,忽略局部结构信息,难以充分利用多层次语义。(2)现有方法普遍默认视图质量一致,低质量视图会干扰最终性能。(3)在多视图数据中,不同视图间语义结构可能存在差异,造成各视图局部结构不一致,阻碍高质量信息的融合和对齐。(4)边界样本是指在特征空间中位于不同簇群交界区域的数据点。这些样本由于其固有的模糊性和类别不确定性,导致了两个主要问题:(1)边界样本与多个簇的中心距离相近,使得模型难以做出明确的划分,从而降低了整体聚类准确性。(2)大量的边界样本会使得模型学习到的决策边界变得模糊,从而削弱模型的泛化能力和鲁棒性。通常使用的硬聚类方法将样本强制划分到某个集群,难以准确表达边界样本的归属情况。

针对上述问题,本文提出了一种基于双通道粒计算的深度多视图聚类方法。本方法把双通道对比学习作为核心内容,充分学习多视图数据中的多层次语义信息,还提出跨视图模糊粒球机制对齐不同视图中所对应的粒球结构。针对低质量视图影响聚类效果这一问题,本方法提出自适应视图注意力权重分配机制,赋予高质量视图更高的权重让其主导聚类过程。

1 相关工作

1.1 多视图聚类

基于子空间的多视图聚类方法^[10],通过将每个视图的高维数据映射到一个低维子空间,并在此基础上学习共享隐空间,得到更明确、更鲁棒的潜在子空间表示,最后完成聚类任务。基于图的多视图聚类方法^[11],利用多视图的数据构建保留融合图,并通过图切分完成聚类。基于矩阵分解的多视图聚类方法^[12],将每个视图分解为低秩矩阵,然后在低维空间对数据进行聚类。基于多核学习的多视图聚类方法^[13],使用不同的预定义内核来处理不同的视图,通过线性或非线性核融合提升表示能力与聚类性能。基于协同训练的多视图聚类方法^[14],通过使用先验知识引导不同视图的聚类,通过迭代执行策略,最大限度地使得所有视图的聚类结果趋于一致。然而这些传统的多视图聚类方法往往有特征表示能力不足、处理大规模高维数据效率低以及对异构视图的适应性较弱等缺陷。与传统方法相比,深度多视图聚类方法^[15]具有更强的特征学习能力和更高的灵活性,可以处理复杂的多视图数据。比如,Xie 等^[16]提出了深度嵌入式聚类方法,联合优化自编码器潜在嵌入与聚类分配。基于此,改进的 DEC^[17]引入聚类和重建目标之间的权衡,防止深度模型崩溃。DEMVC 方法^[18]通过协同训练多个神经网络,充分利用了多视图中的互补信息。Yu 等^[19]利用统一的编码器来处理每个视图的信息,并整合视图中不同的信息。Xu 等^[20]提出了一种用于深度多视图聚类的自监督判别特征学习,该方法利用自监督的全局引导有效挖掘多视图的互补信息,降低低质量视图的干扰。Tang 等^[21]提出的深度安全多视图聚类方法,首次系统定义并解决视图增加导致的性能退化问题,通过安全模块自适应选择有效特征,提升动态增视图场景下的鲁棒性。Zhu 等^[22]提出了一种可信 Mamba 融合网络,旨在缓解噪声干扰并解决多视图聚类中的表示

对齐问题。Gan等^[23]通过多阶段融合策略,有效整合了多视图数据中的互补信息,提升了聚类的准确性和鲁棒性。这些方法的一个显著局限在于其隐含假设视图质量均等。实际应用中,多视图数据集常存在显著的视图质量差异,某些视图可能存在噪声、不完整或信息量不足等问题。若缺乏适应性管理此类异质性的机制,这些方法易出现两大缺陷:(1)低质量视图可能引入了噪声,减弱了学习到的一致性特征表示的质量。(2)高质量视图的信息含量可能被稀释。为应对这一挑战,本文提出了一种视图自适应注意力权重策略。该策略通过衡量每个视图的一致性特征与全局特征表示之间的相似度,定量评估其重要性。这些相似度得分被转换为注意力权重,使模型能够自适应地突出高质量视图,并放大对最终聚类结果的贡献。

1.2 对比学习

对比学习的主要目标是学习具有良好判别性特征的特征空间,使正样本更接近、负样本更分离。在单视图聚类中,正对通常是通过增强同一个样本来构建,负对通常是来自同一队列的其他样本。这种思路自然扩展到多视图场景,将跨视图表示视作天然增强,形成多视图对比学习范式。在计算机视觉领域里代表性的对比学习方法有SimCLR^[24]、MoCo^[25]、BYOL^[26]等。这些方法一般通过数据增强生成同一图像的不同视图,接着利用对比损失函数让同一图像的不同视图在特征空间靠近,而不同图像的视角则要相互远离。SimCLR方法使用NT-Xent损失函数把同一图像的两个增强视图当作正样本对,并将批次里的其他视图作为负样本;MoCo方法引入动量编码器和队列机制扩大负样本的范围,提升训练效率与稳定性。把对比学习与聚类相结合也是近年来一个重要研究方向。PCL^[27]方法学习一组原型向量并通过最大化样本与其对应原型之间的互信息来进行对比学习;SwAV^[28]借助在线聚类分配实现样本与聚类分配之间的对比学习。这些方法不仅关注样本级的判别性,还考虑了簇群级的结构信息,为无监督表示学习提供新的思路。HCL^[29]方法则是引入了层次化的对比学习框架,通过在多层次上进行对比以捕获多尺度特征。多视图对比学习本质就是对比学习在多视图场景下的扩展,其目的是通过对比不同视图之间的特征来学习跨视图的一致性表示。MFLVC^[30]提出了一种多级特征多视图对比框架,这个模型以无融合的方式从原始特征中学习低级特征、高级特征和语义标签。在低级特征上进行重建,在高级特征上通过

对比学习探索共识。CVCL^[31]方法则是在聚类层面上设计对比学习,通过跨视图聚类分配对比,直接捕获语义一致性,有效学习视图的不变表示,在理论上也严格证明了损失下界与对齐可行性。然而,现有方法大多局限于样本级或聚类级对比学习,二者均存在关键局限性。样本级对比学习虽能有效建模精细相似性,却因假阴性问题导致语义相近样本被分离,从而削弱构建紧凑类内结构的能力。相反,聚类级对比学习专注于全局聚类分配,往往忽略关键抽象信息。这种粗粒度策略无法捕捉抽象数据的复杂关联,导致跨视图聚类结构的对齐效果欠佳。为解决这些局限性,本文提出一种双通道对比学习框架,其中引入名为粒球的中间结构。通过在样本级与粒球级同时进行对比学习,模型能够捕捉多粒度的表示。粒球作为抽象数据邻域的代表,实现了对更高层次结构关系的显式建模。此外,为缓解聚类边界的不确定性,模糊理论被融入粒球机制。该机制摒弃可能导致边界实例误判的硬性分配,转而采用软隶属度策略。该框架包含两个目标:视图内粒球旨在通过聚合高模糊共隶属度的样本增强邻域紧凑性;视图间目标则确保不同视图间模糊粒球的结构一致性。这种双通道方法使聚类性能更具鲁棒性与准确性。

2 本文方法

2.1 相关符号定义

给定多视图数据集 $\{X^v\}_{v=1}^M$,其包含 N 个样本,每个样本具有 M 个视图, $X^v = \{X_1^v, X_2^v, \dots, X_N^v\} \in \mathbb{R}^{N \times d_v}$,其中 d_v 表示原始多视图数据集中第 v 个视图的特征维度。多视图聚类旨在将 N 个样本划分为 k 个集群,为了提高清晰度和简洁性,在表1中列出了本文在研究中使用的主要符号。

表1 本文使用的主要符号的描述

Table 1 Descriptions of the main symbols used in this paper

符号	描述
X^v	第 v 个视图的原始特征
Z^v	第 v 个视图的嵌入特征
Z_{att}^v	第 v 个视图加强后的特征
G	融合后的全局特征
w_v	每个视图的注意力权重
u_{ik}	样本 i 对粒球 k 的隶属度
d_v	表示 X^v 的特征维度
d_f	潜在嵌入特征的特征维度
d_r	一致性特征的特征维度
d_g	全局特征的特征维度

2.2 算法整体框架

图 1 为本文所提出方法的整体框架。主要包含以下模块:双通道特征融合模块、视图自适应注意力权重分配机制、跨视图模糊粒球机制以及基于双通道的对比学习策略。首先,每个视图的原始数据都通过一个特定的编码器得到潜在特征 $\{Z^v\}_{v=1}^M$ 。接着潜在特征 $\{Z^v\}_{v=1}^M$ 通过一个双通道特征融合模块来获得增强后的特征 $\{Z_{\text{att}}^v\}_{v=1}^M$ 。此外,增强后的特征 $\{Z_{\text{att}}^v\}_{v=1}^M$ 通过 CIM 模块获得一致性特征 $\{R^v\}_{v=1}^M$ 。然后,通过 FFM 模块融合增强特征获得全局特征 G 。视图自适应注意力机制是用来衡量 $\{R^v\}_{v=1}^M$ 和全局特征 G 的一致性,赋予和全局特征相似性更高的视图更高的注意力分数,使得高质量视图能够在聚类中占据主导作用。最后通

过双通道对比学习,使得相似特征更紧凑,易于最后的聚类。本文针对传统聚类方法难以充分利用多视图数据中互补信息和一致性信息的问题,提出了基于双通道粒计算的深度多视图聚类方法。该方法首先利用独立编码器与双通道特征融合模块提取每个视图的关键特征,然后通过视图一致性模块和全局特征融合模块获得视图一致性特征和全局特征。基于此,设计了视图自适应注意力权重分配机制,赋予高质量视图更高权重,使得高质量视图能在聚类中占主导作用。本方法的核心是基于双通道的对比学习策略,充分利用多视图不同层次的特征信息。此外,设计了跨视图模糊粒球机制,提升边界样本建模能力并显式对齐视图间结构关系。

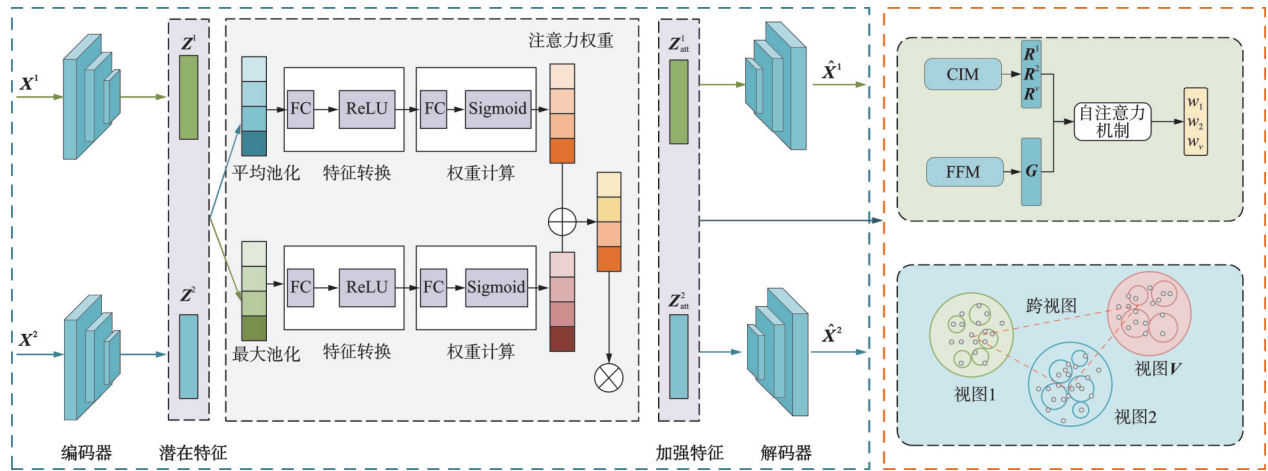


图 1 基于双通道粒计算的深度多视图聚类方法整体框架。

Fig.1 Overall framework of deep multi-view clustering with dual-channel granular computing

2.3 双通道特征融合

由于自编码器可以学习数据的低维表示,去除冗余信息,提取关键特征,对于聚类任务有着关键的作用。因此,本文对任一视图 v 都设计一个独立的编码器将原始特征映射到低维潜在特征空间中,有

$$Z^v = E_v(X^v; \varphi) \quad (1)$$

式中: $E_v(\cdot)$ 表示多视图数据中第 v 个视图的编码器,编码器由多个全连接层和 ReLU 激活函数组成; φ 表示编码器的神经网络参数; $Z^v \in \mathbb{R}^{N \times d_v}$ 表示第 v 个视图的潜在特征表示, d_v 表示潜在特征空间的维度。

为提升判别性特征表达,引入双通道特征融合模块,分别建模全局语义和显著细节。两个通道分别计算全局平均池化和全局最大池化的注意力权重,最大池化通道关注最显著的特征,它在特征图的一个区域内只选择值最大的那个像素最为输出,它旨在捕获区域内最突出、最具有判别力的特征,

能够有效地捕捉到图像的纹理和边缘等细节信息。平均池化通道关注整体性的特征,计算特征图一个区域内所有像素的平均值作为输出,捕获区域内的整体特征能够保留更多的背景信息,对微小的噪声和干扰有更强的鲁棒性。对于一些质量不佳的数据,单一的池化方法可能会失效。而双通道机制提供了一个更鲁棒的方式,当一个通道的信息不可靠时,另一个通道仍然可能提供有用的判别依据,使得模型更加可靠。同时,双通道的一个分支更专注于细节,另一个分支专注于全局,这种分工可以使网络的学习过程更加高效和有针对性,最终学习到的特征表示质量也更高。本文设计的双通道注意力权重机制参考了 SENet^[32] 的通道注意力思想以及 CBAM^[33] 的最大池化补充策略,在此基础上融合两种池化方式以增强特征表达。具体计算公式为

$$\begin{aligned} A_{\text{avg}}^v &= \sigma \text{MLP}(\text{AvgPool}(Z^v)) \\ A_{\text{max}}^v &= \sigma \text{MLP}(\text{MaxPool}(Z^v)) \end{aligned} \quad (2)$$

$$A^v = A_{\text{avg}}^v + A_{\text{max}}^v \quad (3)$$

式中: $A_{\text{avg}}^v \in \mathbf{R}^{N \times d_i}$ 表示经过平均池化得到的注意力权重,是经过 1 个多层感知器(Multilayer perceptron, MLP)和 Sigmoid 函数得到; $A_{\text{max}}^v \in \mathbf{R}^{N \times d_i}$ 表示经过最大池化得到的注意力权重; $A^v \in \mathbf{R}^{N \times d_i}$ 表示融合平均池化和最大池化之后的注意力权重; σ 表示一个 Sigmoid 函数; AvgPool 和 MaxPool 分别表示平均池化和最大池化。

最后将潜在特征与融合后的注意力权重相乘得到加权后的增强特征,具体计算公式为

$$Z_{\text{att}}^v = Z^v \odot A^v \quad (4)$$

式中: $Z_{\text{att}}^v \in \mathbf{R}^{N \times d_i}$ 表示经过加强后的特征, \odot 表示逐元素乘法。其逐元素乘法体现选择性增强,可以视作特征重加权映射,使得显著特征获得更高响应。

为了保留原始特征的信息,使用解码器来重构特征,对于每个视图都有一个独立的解码器,尝试从注意力增强的特征 Z_{att}^v 重建原始输入 X^v ,具体计算公式为

$$\hat{X}^v = D^v(Z_{\text{att}}^v; \phi) \quad (5)$$

式中: ϕ 为解码器的神经网络参数; $D^v(\cdot)$ 表示第 v 个视图的解码器; $\hat{X}^v \in \mathbf{R}^{N \times d_v}$ 表示为第 v 个视图的重构特征,解码器的结构与编码器的结构对称。

综上,衡量每个视图的原始输入 X^v 和重建结果 \hat{X}^v 之间的重构损失,可以采用均方误差(Mean squared error, MSE)^[16],计算公式为

$$L_{\text{recon}} = \frac{1}{V} \sum_{v=1}^V \|X^v - \hat{X}^v\|_F^2 \quad (6)$$

式中 $\|\cdot\|_F$ 表示 Frobenius 范数。

2.4 视图自适应注意力权重分配

众所周知,由于多视图数据普遍存在质量差异,若低质量视图被等权融合,将稀释高质量语义并降低聚类性能。因此,本文提出了一种自适应的视图注意力权重分配机制,给予高质量视图更高的权重,使得高质量视图能够更好地主导聚类效果,基于不同的注意力分数赋予不同视图权重。

为了学习到视图间的一致性表示,将每个视图的注意力增强特征 Z_{att}^v 通过 1 个共享的映射模块 CIM 映射到 1 个视图一致性特征空间,得到视图一致性特征,有

$$R^v = \text{normalize}(\text{CIM}(Z_{\text{att}}^v)) \quad (7)$$

式中: CIM 是一个简单的线性层; normalize 表示 L2 归一化操作,保证特征向量的长度为 1,得到 $R^v \in \mathbf{R}^{N \times d_i}$ 。

为了融合多视图的信息,将所有视图的注意力增强特征 $Z_{\text{att}}^v = [Z_{\text{att}}^1, Z_{\text{att}}^2, \dots, Z_{\text{att}}^v]$ 连接起来,通过 1 个特征融合模块 FFM 得到全局特征矩阵

$G \in \mathbf{R}^{N \times d_e}$, 有

$$G = \text{normalize}(\text{FFM}(\text{concat}(Z_{\text{att}}^1, Z_{\text{att}}^2, \dots, Z_{\text{att}}^v))) \quad (8)$$

式中: FFM 模块包含了多个线性层和 ReLU 激活函数; concat(\cdot) 表示特征连接操作; normalize 表示 L2 归一化。

基于点积的注意力评分公式参考 Transformer 通用框架^[34],为了计算注意力权重,将全局特征均值向量定义为查询向量,具体计算公式为

$$Q = \frac{1}{N} \sum_{i=1}^N G_i \quad (9)$$

式中: $G \in \mathbf{R}^{N \times d_e}$ 为融合后的特征矩阵; N 为样本数量; G_i 表示第 i 个样本的全局特征向量; Q 为所有样本的全局特征均值。

每个视图的一致性特征均值作为键向量,计算公式为

$$K_v = \frac{1}{N} \sum_{i=1}^N R_i^v \quad (10)$$

式中: K_v 为第 v 个视图的特征均值; R^v 为第 v 个视图的一致性特征矩阵。

计算每个视图的注意力分数。具体计算公式为

$$S_v = \frac{QK_v^T}{\sqrt{d}} \quad (11)$$

式中: S_v 为一个标量,表示第 v 个视图一致性特征与全局特征之间的注意力分数; d 表示每个向量的特征维度,这里作为分母开根号是为了保持梯度稳定,便于模型训练。

最后计算每个视图的注意力权重,具体计算公式为

$$w_v = \sigma(S_v) = \frac{1}{1 + e^{-S_v}} \quad (12)$$

式中: $\sigma(\cdot)$ 为 Sigmoid 激活函数,用于归一化注意力分数; w_v 为第 v 个视图的注意力权重; S_v 为第 v 个视图的注意力得分,这里的分母用来将注意力分数归一化。

2.5 跨视图模糊粒球

为了增强不同视图间粒球结构的一致性,本文引入跨视图模糊粒球机制。首先使用聚类算法初始化中心,得到 K 个中心 $C = \{c_k\}_{k=1}^K$ 。由于不同数据集的分布不同,确定的粒球中心数 K 也不同,本文针对不同的数据集手动设置不同的中心数 K ,以此来得到相对更适合该样本的中心数 K ,后续可以对此进行改进,进行自适应调整,自动获得最好的聚类中心数 K 。计算每个样本到各个中心的欧式距离,具体计算公式如式(13)所示,欧式距离函数基于一致性特征空间内度量,是一个通用定义。然后使用高斯模糊函数来计算模糊隶属度,隶属度函

数参考模糊聚类^[35]思想,采用高斯核更好刻画局部粒球密度连续性,同时允许边界样本拥有多粒球软归属。如图2所示,一个样本可以划分到多个不同的粒球,然后归一化隶属度,具体计算公式为

$$D_{ij}^v = \|\mathbf{r}_i^v - \mathbf{c}_j^v\|_2 \quad (13)$$

$$u_{ij}^v = \frac{e^{-\frac{(D_{ij}^v)^2}{2\rho_j^2}}}{\sum_{k=1}^K e^{-\frac{(D_{ik}^v)^2}{2\rho_k^2}}} \quad (14)$$

式中: \mathbf{r}_i^v 为第 v 个视图内第 i 个样本的一致性特征; \mathbf{c}_j^v 是第 v 个视图内第 j 个粒球中心; D_{ij}^v 为第 v 个视图内样本 i 到第 j 个中心的距离; ρ_j 为第 j 个粒球的半径。粒球半径是一个可学习参数,本文首先设置初始半径,后续会随着反向传播优化更新,自动调整模糊程度,适应不同粒球内部的密度结构。 u_{ij} 为第 i 个样本对第 j 个粒球的隶属度,越接近中心,隶属度越高。

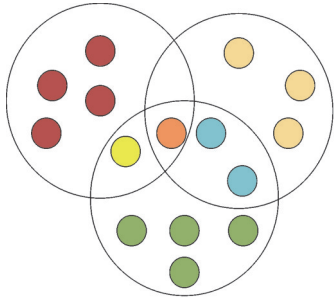


图2 模糊粒球划分

Fig.2 Fuzzy granular partitioning

对每个粒球 k 筛选隶属度高于阈值的作为粒球内的样本集合

$$S_k^v = \{\mathbf{r}_i^v | u_{ik}^v > \lambda\} \quad (15)$$

式中: λ 为隶属度阈值, S_k^v 为符合要求的粒球内样本集合。

计算粒球中心相似度矩阵,判断在不同视图下相似粒球是否有一致的表示,公式为

$$S_{pq}^{i,j} = \frac{\mathbf{c}_p^i \cdot (\mathbf{c}_q^j)^T}{\tau_g} \quad (16)$$

式中: \mathbf{c}_p^i 为第 i 个视图下的第 p 个粒球中心; \mathbf{c}_q^j 为第 j 个视图下的第 q 个粒球中心; $S^{i,j} \in \mathbf{R}^{K \times K}$ 为两个视图粒球中心之间的相似度。

基于上述结果,计算隶属度重叠矩阵。重叠越大说明这对粒球在不同视图中捕捉到了相同的样本结构,计算公式为

$$O_{pq}^{i,j} = \sum_{n=1}^N u_{np}^i \cdot u_{nq}^j \quad (17)$$

式中: N 表示样本数; u_{np}^i 表示第 n 个样本对视图 i 的粒球 p 的隶属度; O_{pq} 表示粒球 p,q 在不同视图中共享样本的程度。如果粒球对在不同视图中具有高

隶属度重叠,则表示其共同捕捉到局部一致性结构,可视作跨视图语义对齐证据。软重叠避免硬切分导致边界样本损伤。然后归一化为概率分布,有

$$\hat{O}_{pq}^{i,j} = \frac{O_{pq}^{i,j}}{\sum_{p,q} O_{pq}^{i,j}} \quad (18)$$

2.6 基于双通道的对比学习

现有的深度多视图聚类方法主要集中在样本级特征学习,难以捕捉到视图间局部的语义结构,为了获得多视图数据更丰富的语义信息,本文提出了基于双通道的对比学习策略,同时在样本级和粒球级局部结构进行对比学习。

首先计算样本级对比损失^[24],具体公式为

$$L_{\text{ins}} = \frac{1}{V} \sum_{v=1}^V \omega_v L_{\text{con}}(\mathbf{G}, \mathbf{R}^v) \quad (19)$$

式中 L_{con} 为InfoNCE对比损失,有

$$L_{\text{con}}(\mathbf{G}, \mathbf{R}^v) = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\text{sim}(\mathbf{G}_i, \mathbf{R}_i^v)/\tau_s}}{\sum_{j=1}^N \Pi[j \neq i] e^{\text{sim}(\mathbf{G}_i, \mathbf{R}_j^v)/\tau_s}} \quad (20)$$

式中: τ_s 为温度参数,控制相似度分布的平滑度; \mathbf{G}_i 表示为 \mathbf{G} 的第 i 个样本表示; \mathbf{R}_i^v 为 \mathbf{R}^v 的第 i 个样本表示;

$\text{sim}(\mathbf{G}_i, \mathbf{R}_j^v) = \frac{\langle \mathbf{G}_i, \mathbf{R}_j^v \rangle}{\|\mathbf{G}_i\| \|\mathbf{R}_j^v\|}$ 表示两个样本表示之间的余弦相似度; Π 为一个判断函数,如果满足条件就为1,否则为0。

计算粒球内部的对比损失,对于每个粒球 k ,计算其内部样本的加权对比损失,目标是拉近隶属于同一模糊粒球的样本表示,使得每一个粒球内部的结构更加紧凑,更利于之后的聚类,具体计算公式为

$$L_k^v = -\frac{\sum_{i \in S_k^v} u_{ik}^v \log \frac{e^{r_i^v \cdot (r_i^v)^T / \tau_g}}{\sum_{j \in S_k^v, j \neq i} e^{r_i^v \cdot (r_j^v)^T / \tau_g}}}{\sum_{i \in S_k^v} u_{ik}^v} \quad (21)$$

式中: \mathbf{r}_i^v 和 \mathbf{r}_j^v 为同一个粒球内的样本特征; τ_g 为粒球级对比学习的温度参数,用于控制相似度分布的平滑度。

粒球内整体损失表示为

$$L_{\text{granular}} = \frac{1}{V} \frac{1}{K} \sum_{v=1}^V \sum_{k=1}^K L_k^v \quad (22)$$

只对重叠度高的样本进行正样本训练,得到正样本的相似度的加权和,最后计算跨视图粒球之间的损失,具体计算公式为

$$L_{i,j} = -\log \frac{e^{\sum_{p,q} \hat{O}_{pq}^{i,j} \cdot \Pi(\hat{O}_{pq}^{i,j} > \tau)}}{\sum_{p,q} e^{S_{pq}^{i,j}}} \quad (23)$$

式中: r 为阈值; Π 为判断函数,判断重叠的隶属度是否大于阈值,如果大于阈值就为1,否则为0。当重叠隶属度大于阈值时,粒球对视为正对,否则视为负对。这项对比损失在局部结构上对齐视图间的局部语义,弥补样本级对比易产生假阴性的问题。

跨视图模糊粒球的最终损失为视图对之间的平均,有

$$L_{\text{cross-view}} = \frac{2}{V(V-1)} \sum_{i=1}^V \sum_{j=i+1}^V L_{i,j} \quad (24)$$

2.7 融合损失函数

在得到所有的损失函数之后,融合所有损失函数并最小化,即

$$L_{\text{overall}} = L_{\text{recon}} + L_{\text{ins}} + L_{\text{granular}} + L_{\text{cross-view}} \quad (25)$$

在最后的聚类阶段,本文对全局特征 G 使用 K -means 聚类算法进行聚类,获取到最终聚类结果 Y 。

2.8 算法流程

本文提出的基于双通道粒计算的深度多视图聚类方法的流程如算法1所示,首先进行预训练,获得自编码器的神经网络参数,然后进行基于双通道的对比学习训练,最后进行聚类。

算法1 基于双通道粒计算的深度聚类算法

输入:多视图数据集 $\{X^v\}_{v=1}^M$, 预训练轮次 Epochpre, 对比学习轮次 Epochcon, 批次大小 B , 样本级对比学习温度参数 τ_s , 粒球级对比学习温度参数 τ_g ;

输出:聚类结果 Y 。

- (1) for epoch=1 to Epochpre do
- (2) for 每个批次 do
- (3) 对每个视图
- (4) 根据式(1)计算潜在特征表示;
- (5) 根据式(4)应用双通道特征融合,计算经过注意力增强之后的特征;
- (6) 根据式(5)计算重构输入;
- (7) 根据式(6)计算重构损失;
- (8) 更新神经网络参数和优化器参数;
- (9) end for;
- (10) end for;
- (11) for epoch=1 to Epochcon do:
- (12) for 每个批次 do:
- (13) 对每个视图:
- (14) 重复自编码预训练的步骤;
- (15) 根据式(7)获得一致性语义特征;
- (16) 根据式(8)融合获得全局特征 G ;
- (17) 根据式(12)计算每个视图的注意力权重;
- (18) 计算双通道对比损失;

(19) 根据式(19)计算样本级对比损失;

(20) 根据式(22)计算粒球内部对比损失;

(21) 根据式(24)计算跨视图的粒球对比损失;

(22) 根据式(25)计算总损失;

(23) end for;

(24) end for;

(25) 用 K -means 计算聚类结果 Y 。

在本节中,对本方法的时间复杂度进行了全面的分析。主要分为了两阶段训练:第一阶段主要是针对自编码器的预训练。其中双通道特征融合中的平均池化和全局池化时间复杂度都为 $O(Nd_f)$, 每个视图计算重构损失的时间复杂度为 $O(Nd_v d_f)$, 预训练的时间复杂度可以计算为 $O(VNd_v d_f)$ 。第二阶段主要是基于双通道的对比学习训练,首先重复自编码器预训练,时间复杂度和第一阶段一样,都为 $O(VNd_v d_f)$ 。计算一致性语义特征的时间复杂度为 $O(VNd_v d_r)$, 计算全局特征的时间复杂度和一致性语义特征的复杂度相同,样本级对比损失中需要计算自适应注意力权重,时间复杂度为 $O(VNd_v)$, 同时需要计算视图内所有样本对的相似度,时间复杂度为 $O(VN^2 d_r)$ 。计算粒球级内部对比学习时间复杂度为 $O(KN^2 d_r)$, 跨视图粒球级对比学习中需要计算粒球中心相似度矩阵,时间复杂度为 $O(V^2 K^2 d_r)$, 计算隶属度重叠矩阵的时间复杂度为 $O(V^2 K^2 N)$, 考虑到上述分析,算法1的时间复杂度为 $O(VNd_v d_f) + O(VNd_v d_r) + O(VNd_v) + O(VN^2 d_r) + O(KN^2 d_r) + O(V^2 K^2 (N + d_r))$, 在 N 足够大时,时间复杂度可近似为 $O((V+K)N^2 d_r)$ 。

3 实验分析

3.1 实验设置

为了评估本文提出模型的性能,如表2所示,在8个公开的多视图数据集上进行实验。包括了MNIST-USPS和BDGP两个小规模数据集,其包

表2 本文所使用的所有数据集描述

数据集	样本	视图	类别
Caltech-2V	1 400	2	7
Caltech-3V	1 400	3	7
Caltech-4V	1 400	4	7
Caltech-5V	1 400	5	7
BDGP	2 500	2	5
MNIST-USPS	5 000	2	10
Cifar10	50 000	3	10
Fashion	10 000	3	10

含了两个视图。也使用了 Cifar10 和 Fashion 这两个大规模数据集,以及 Caltech-XV 系列数据集,用来验证随着视图的增加本模型实验的稳定性。

本文选取了 11 种聚类方法进行了对比。包含 3 种传统聚类方法和 8 种流行的多视图聚类方法。3 种传统的聚类方法:K-means(K 均值聚类方法),GMM(高斯混合模型聚类),HC(层次聚类方法)^[36]。8 种流行的深度多视图聚类方法包括:MFLVC(对比多视图聚类的多级特征学习)^[30],DealMVC(多视图聚类的双对比标准)^[37],SCMVC(自加权对比融合的深度多视图聚类方法)^[38],CMVC(多分支空间共享特征的对比多视图聚类方法)^[39],DSMVC(深度安全多视图聚类)^[21],DEMVC(基于协同训练的深度嵌入式多视图聚类)^[18],SDMVC(用于深度多视图聚类的自监督判别特征学习)^[20],CVCL(通过对比聚类分配进

行深度多视图聚类)^[31]。

此外,本文采用了 3 种广泛使用的指标来评估所有对比算法和本文提出方法的聚类性能,包括了聚类精度 ACC,归一化互信息 NMI,纯度 Pur。对于这些聚类指标,其取值越接近 1,说明聚类性能越好。

3.2 实验结果与分析

本文方法和对比算法在 8 个公共数据集上的性能结果如表 3 和表 4 所示,粗体表示最佳结果,下划线表示次优结果。本文提出的方法在多个数据集上都表现出了最佳的结果,在 Caltech-2V 数据集上的准确率 ACC 比次佳结果高出了 8.7%,归一化互信息 NMI 比次佳结果高出了 6.9%,在 Caltech-3V 数据集上 ACC 比次佳结果高出了 6.1%,在 Caltech-4V 数据集上 ACC 比次佳结果高出了 3.5%,NMI 比次佳结果高 0.9%。同时可以注

表 3 4 个公共数据集上的性能指标

Table 3 Results on four public datasets

Method	MNIST-USPS			BDGP			Fashion			Cifar10		
	ACC	NMI	Pur	ACC	NMI	Pur	ACC	NMI	Pur	ACC	NMI	Pur
K-means	0.768	0.724	0.768	0.639	0.517	0.639	0.771	0.747	0.775	0.992	0.981	0.992
HC	0.830	0.824	0.836	0.576	0.464	0.576	0.770	0.781	0.771	0.990	0.975	0.990
GMM	0.797	0.745	0.797	0.502	0.434	0.526	0.608	0.699	0.672	0.846	0.936	0.894
MFLVC	0.995	0.988	0.995	0.976	0.941	0.976	0.993	0.982	0.993	0.992	0.978	0.992
CVCL	0.995	0.987	0.995	<u>0.990</u>	<u>0.967</u>	<u>0.990</u>	<u>0.993</u>	<u>0.981</u>	<u>0.993</u>	<u>0.994</u>	<u>0.984</u>	<u>0.994</u>
DealMVC	0.927	0.926	0.927	0.980	0.945	0.980	0.895	0.955	0.895	0.987	0.968	0.987
DSMVC	0.830	0.810	0.830	0.611	0.399	0.611	0.793	0.776	0.804	0.790	0.773	0.790
DEMVC	0.996	0.989	0.996	0.751	0.750	0.751	0.827	0.856	0.827	0.435	0.366	0.435
SDMVC	0.912	0.939	0.912	0.934	0.869	0.934	0.891	0.933	0.891	0.868	0.832	0.868
SCMVC	<u>0.997</u>	<u>0.990</u>	<u>0.997</u>	0.993	0.974	0.993	0.994	0.983	0.994	0.995	0.986	0.995
CMVC	0.991	0.976	0.991	0.983	0.949	0.983	0.993	0.980	0.993	0.987	0.965	0.987
Ours	0.997	0.992	0.997	0.986	0.959	0.986	0.992	0.979	0.992	0.995	0.986	0.995

表 4 不同视图数量的 Caltech 数据集上性能指标

Table 4 Results on the Caltech dataset with different numbers of views

Method	Caltech-2V			Caltech-3V			Caltech-4V			Caltech-5V		
	ACC	NMI	Pur	ACC	NMI	Pur	ACC	NMI	Pur	ACC	NMI	Pur
K-means	0.500	0.421	0.561	0.658	0.541	0.664	0.672	0.593	0.719	0.849	0.756	0.849
HC	0.486	0.440	0.534	0.665	0.581	0.672	0.708	0.674	0.719	0.866	0.787	0.866
GMM	0.561	0.394	0.574	0.650	0.542	0.654	0.664	0.582	0.712	0.812	0.726	0.812
MFLVC	0.611	0.526	0.622	0.601	0.584	0.612	0.739	0.653	0.739	0.816	0.715	0.816
CVCL	0.524	0.458	0.524	0.689	0.594	0.689	<u>0.836</u>	0.716	<u>0.836</u>	0.745	0.648	0.745
DealMVC	0.631	0.521	0.631	0.636	0.535	0.655	0.792	0.692	0.792	0.864	0.777	0.864
DSMVC	0.646	0.532	0.646	<u>0.763</u>	<u>0.684</u>	<u>0.764</u>	0.801	<u>0.750</u>	0.803	0.910	<u>0.832</u>	0.910
DEMVC	0.466	0.364	0.465	0.554	0.429	0.561	0.524	0.439	0.560	0.554	0.453	0.554
SDMVC	0.464	0.427	0.521	0.634	0.553	0.684	0.727	0.683	0.762	0.899	0.805	0.899
SCMVC	<u>0.654</u>	<u>0.544</u>	<u>0.654</u>	0.748	0.637	0.748	0.824	0.719	0.824	<u>0.906</u>	0.836	<u>0.906</u>
CMVC	0.589	0.467	0.598	0.644	0.575	0.672	0.826	0.721	0.826	0.887	0.796	0.887
Ours	0.741	0.613	0.741	0.824	0.691	0.824	0.871	0.759	0.871	<u>0.906</u>	0.826	<u>0.906</u>

意到在 MNIST-USPS 和 Cifar10 这两个公共数据集上也了一定的提升,在 BDGP 数据集和 Fashion 数据集上虽然不是最佳的结果,但是本文的方法依旧很接近最佳效果。

3.3 收敛性分析

如图 3 所示,可以观察到 MNIST-USPS 数据集和 Fashion 数据集的损失值(Loss)在训练过程中呈单调下降的趋势,并且随着轮次增加,趋于收敛。而 ACC 和 NMI 这些关键聚类指标在训练过

程中性能呈现出逐渐增加的趋势,到达顶峰之后,然后在狭窄的范围内略微波动,这些结果证明了本方法的收敛性。

图 4 为 MNIST-USPS 和 Fashion 数据集的潜在特征和视图一致性特征训练过程的性能变化,分别对多视图数据集提取到的潜在特征和一致性特征直接聚类,明显看出视图一致性特征在训练过程中更稳定,且训练性能更好,也证实了在对潜在特征进行双通道特征融合有效。

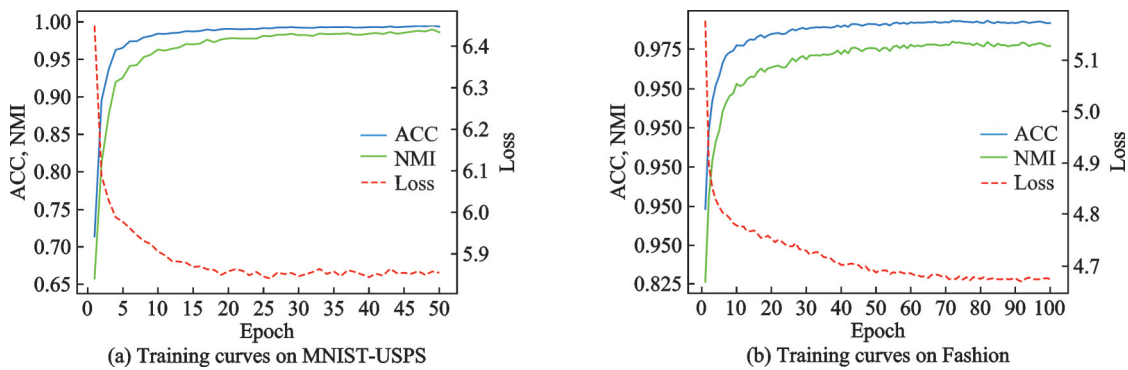


图 3 MNIST-USPS 和 Fashion 数据集训练过程中的损失和性能变化

Fig.3 Loss and performance variations during the training process of the MNIST-USPS and Fashion datasets

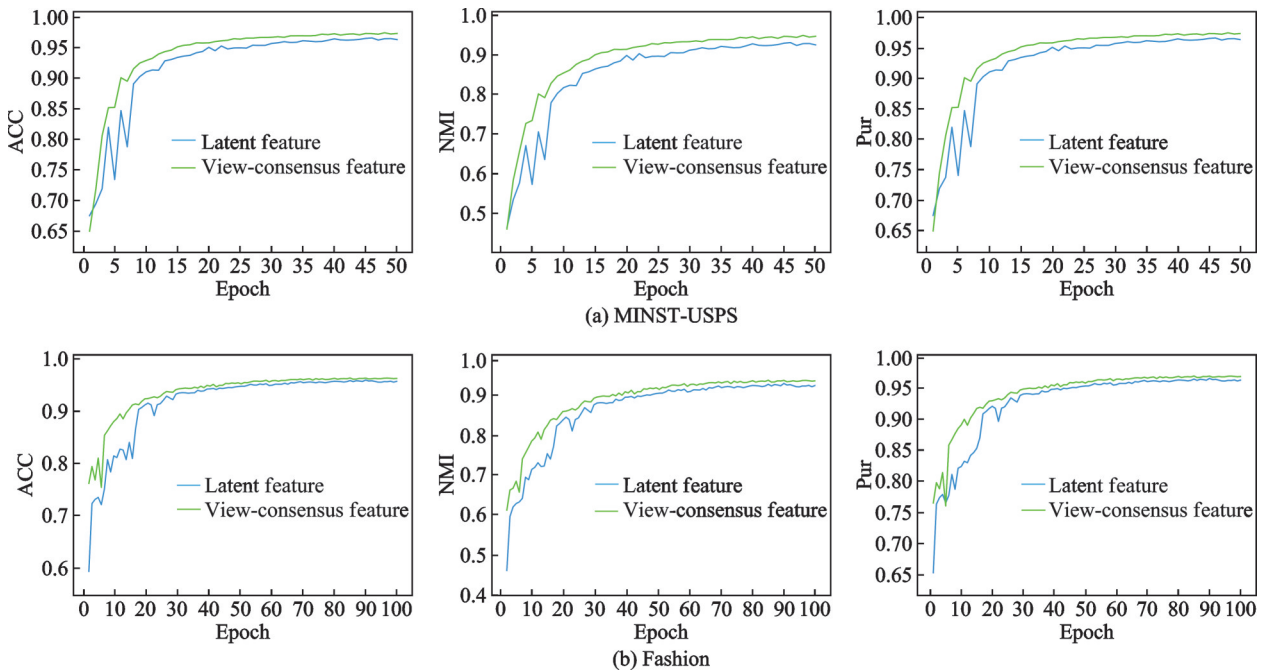


图 4 MNIST-USPS 和 Fashion 数据集的潜在特征和视图一致性特征训练过程的性能变化

Fig.4 Performance variations of the training process of potential features and view consistency features on the MNIST-USPS and Fashion datasets

3.4 混淆矩阵

如图 5 的混淆矩阵所示,在 MNIST-USPS 和 Fashion 数据集上主对角线极其清晰,每个真实类基本都集中在一个预测簇群上,对角线之外的非零值很少,表示误差聚类很少,聚类精度很高。而在 Caltech-4V 和 Caltech-5V 数据集上主对角线比较

清晰,但也有比较明显的非对角干扰,尤其是类 0 和类 6,说明有些类之间混淆程度比较高,表现为一类样本分配到多个聚类簇,或者多个类被分配到同一个簇。类 0 和类 6 明显有较大的错误分类,可能是这两个类别在视觉上有相似特征,比如背景、结构轮廓等。

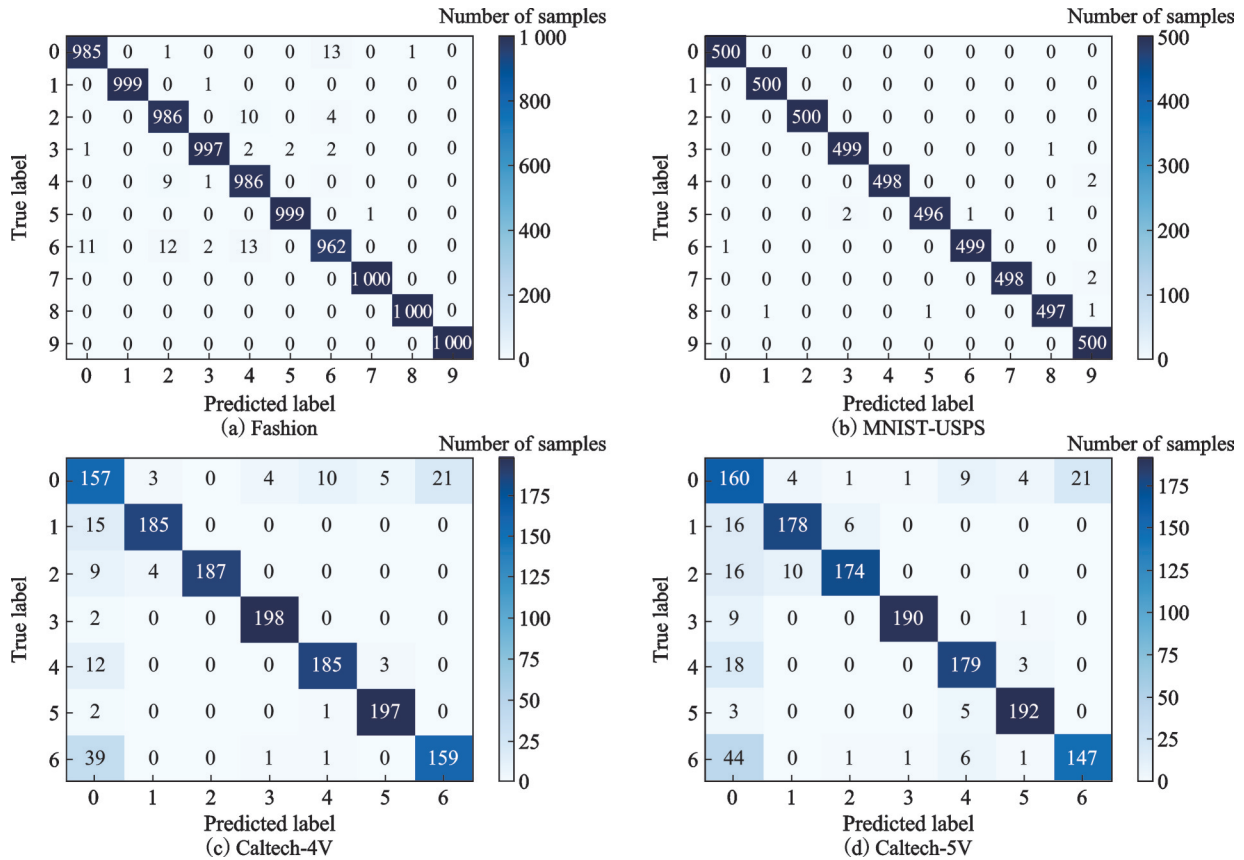


图5 Fashion、MNIST-USPS、Caltech-4V 和 Caltech-5V 数据集的混淆矩阵

Fig.5 Confusion matrices of the Fashion, MNIST-USPS, Caltech-4V and Caltech-5V datasets

3.5 T-SNE 可视化分析

为了直观地研究本文所提出的双通道对比学习策略,本文利用 T-SNE 方法将对训练过程可视化,如图 6 所示,分别为 MNIST-USPS、Caltech-5V 和 Fashion 数据集的 T-SNE 可视化,将这 3 个数据集未进行对比训练时的簇群结构,第一次对比学习训练后的簇群结构,中期训练后的簇群结构和最后训练结束之后的簇群结构分别进行可视化。从这些可视化图中可以看到当第一阶段的自动编码器的预训练完成,但还未进行多粒度对比学习训练时,多视图特征的判别程度还很低,聚类结构仍然很散乱,聚类不能反映真实的聚类结构,在接下来的多粒度对比学习过程中,可以看到嵌入特征的聚类结构变得越清晰,直到最后能将各个样本有效划分到相应的集群中,这也证实了本文方法的有效性。

3.6 参数敏感性分析

在本节中,主要讨论样本级对比学习温度参数 τ_s 和粒球级对比学习温度参数 τ_g 的最佳设置。如图 7 所示,展示了在 MNIST-USPS 数据集和 Caltech-4V 数据集中样本级对比学习温度参数 τ_s 在指定的区间 $\{0.3, 0.5, 0.7, 1.0, 1.1, 1.3, 1.5\}$,粒球级对比学习温度参数 τ_g 在区间 $\{0.3, 0.5, 0.7,$

$1.0, 1.1, 1.3, 1.5\}$ 内调整时,本方法的 ACC、NMI、PUR 的变化。当 τ_g 较小时,本文所提出的方法聚类精度会降低,这可能是因为低温下,粒球级相似度分布相对尖锐,一些本质上属于同一类别但不完全重合的样本可能被认为当作负样本,一些模糊簇的边界区域也难以被挖掘。当 τ_s 较小时,聚类精度也有所下降,这可能是因为正样本之间的相似度被过度放缩,特征空间收缩严重,容易类内过拟合,失去语义区分能力。当 τ_s 和 τ_g 值逐渐恢复时,聚类精度逐渐恢复。根据经验,本文为 MNIST-USPS 数据集设置 $\tau_s = 0.7$ 和 $\tau_g = 1.3$,为 Caltech-4V 数据集设置 $\tau_s = 0.5$ 和 $\tau_g = 1.1$ 。

3.7 消融实验

为了展示本文所提出方法的有效性,单独删除每个模块组件来观察性能的变化,如表 5 所示,表中 DCF 表示为双通道特征融合模块,Ins 表示为样本级对比学习,GB 表示为粒球级对比学习,WA 表示为视图自适应注意力权重机制,w/o 表示模块消融。在缺少了双通道特征融合 DCF 模块的情况下,数据集的聚类性能都有一定幅度的下降,这是因为双通道特征融合加强了关键特征的权重,使得重要的特征被更加关注,减少了噪声冗余信息;在缺少了样

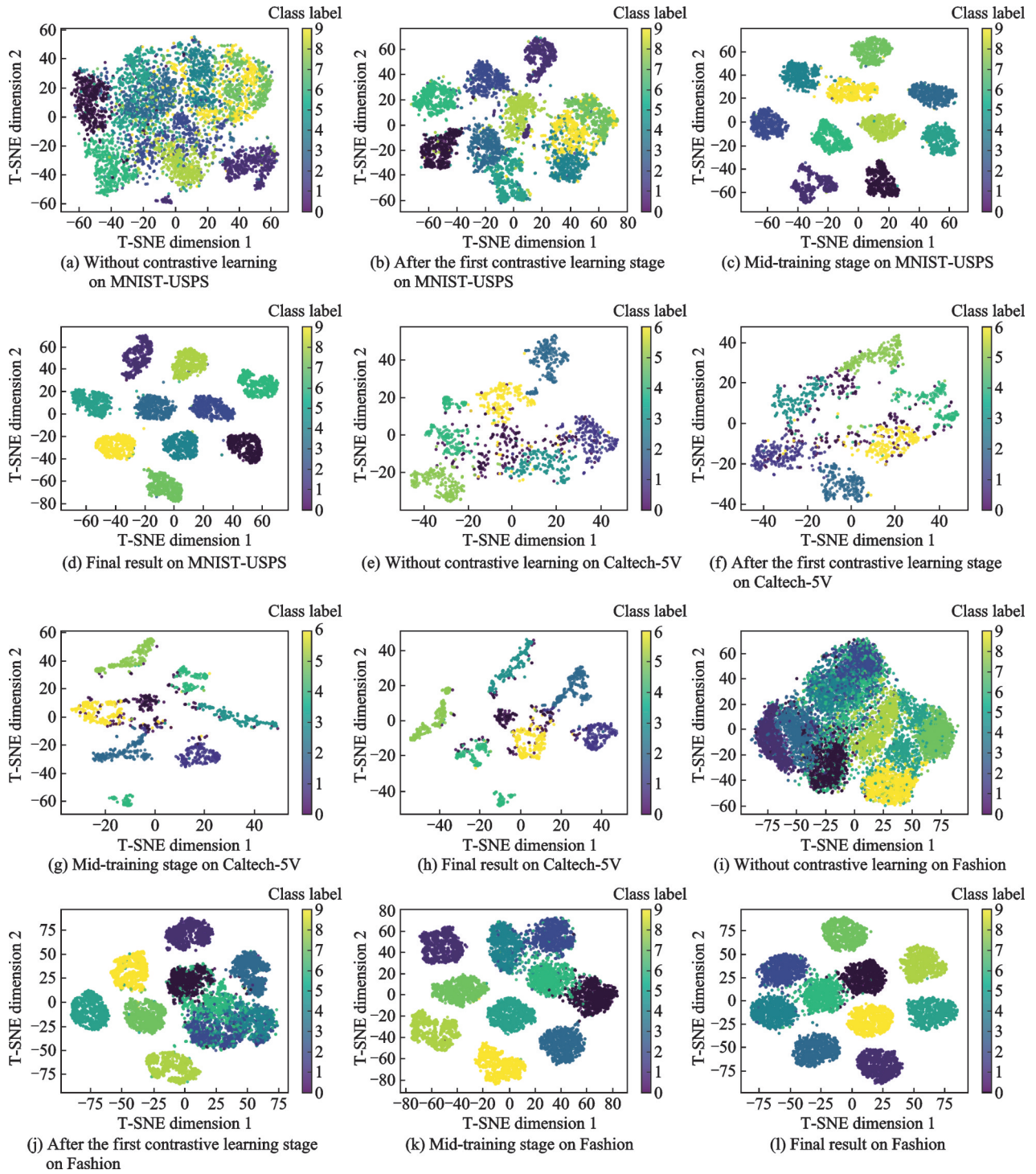


图 6 在 MNIST-USPS, Caltech-5V 和 Fashion 数据集上不同训练阶段的 TSNE 可视化结果

Fig.6 TSNE visualizations at different training stages on MNIST-USPS, Caltech-5V and Fashion datasets

本级对比学习 Ins 的情况下,聚类性能指标下降最多,这是因为样本级对比学习能够学习到最基本的细粒度特征;在缺少了粒球级对比学习 GB 的情况下,数据集的效果也都有了一定的下降,这是因为粒球级对比学习能够学习到数据样本的局部结构,同时对齐不同视图的聚类分布;缺少了自适应权重 WA 的情况下,也有所下降,表明不同质量的视图对聚类效果也有一定的影响。通过这些结果表明,本文所提出的组件模块都是有效的。

3.8 对比算法时间复杂度分析

在 CVCL 中,让 m 和 s 分别表示所提网络架构中的小批量大小和隐藏层中神经元的最大数量。在微调阶段前馈网络的复杂度为 $O(n_v m d_v s^{(r+1)} + n_v m d_v K)$ 。重构损失,跨视图对比损失和跨视图一致性损失分别为 $O(n_v m d_v)$, $O(m^2 K n_v ((n_v - 1) + n_v (K - 1)))$, $O(K n_v)$ 。因此, CVCL 的整体复杂度为 $t(n_v m d_v s^{(r+1)} + n_v m d_v K + n^2 m^2 K^2)$, 其中 t 为与训练和微调的最大迭代数。在 SDMVC 中, K, V, M 分

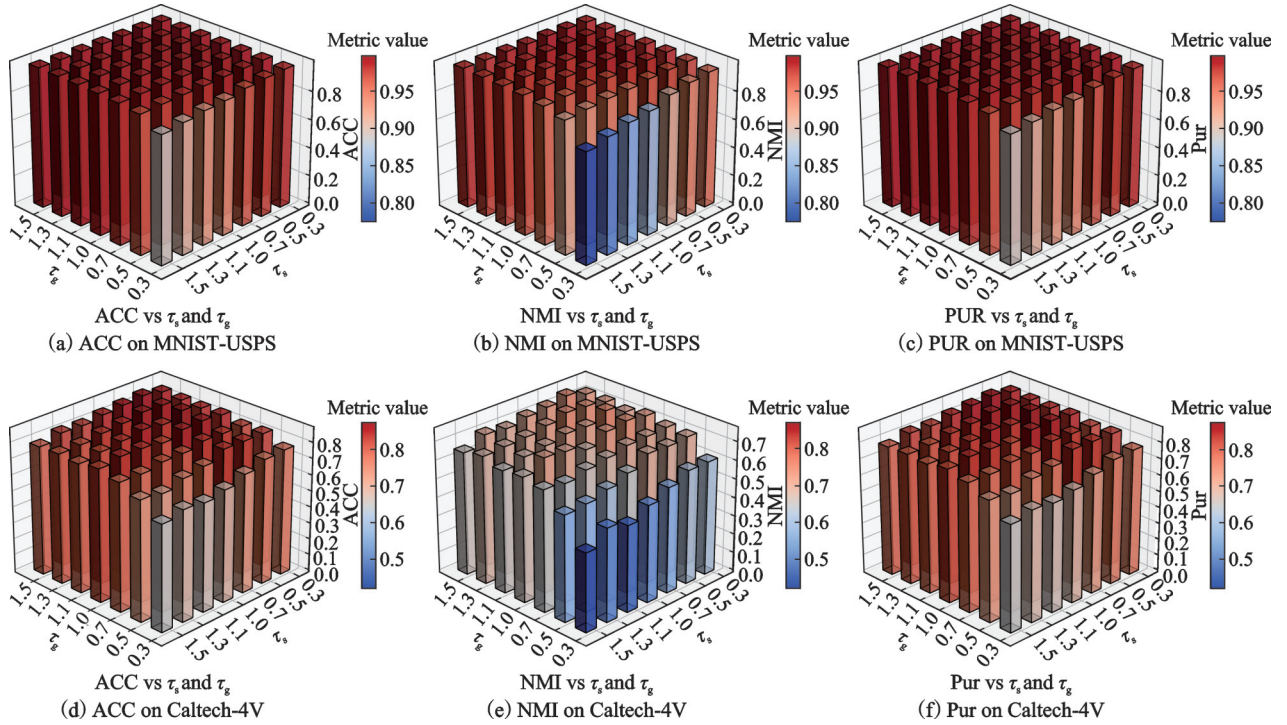


图7 在MNIST-USPS和Caltech-4V数据集上不同温度参数下的聚类性能变化

Fig.7 Clustering performance under different temperature parameters on MNIST-USPS and Caltech-4V

表5 损失模块的消融实验

Table 5 Ablation studies on loss components

Method	Caltech-2V			Caltech-3V			Caltech-4V		
	ACC	NMI	Pur	ACC	NMI	Pur	ACC	NMI	Pur
w/o DCF	0.674	0.562	0.674	0.752	0.615	0.752	0.836	0.722	0.836
w/o Ins	0.537	0.362	0.555	0.648	0.458	0.648	0.671	0.474	0.671
w/o GB	0.690	0.553	0.690	0.677	0.574	0.677	0.850	0.755	0.850
w/o WA	0.611	0.490	0.624	0.689	0.508	0.716	0.862	0.754	0.862
Ours	0.741	0.613	0.741	0.824	0.691	0.824	0.871	0.759	0.871

别为聚类数,视图数和样本数。 M 表示自编码器中隐藏层神经元的最大数量, Z 表示嵌入特征的最大维度。其中 $V, K, Z \ll M$ 。 K -means和目标分布的复杂度都为 $O(NZK)$ 。 V 个自编码器的复杂度为 $O(VNM^2)$ 。整个算法的复杂度与样本数 N 线性相关。在MFLVC中重构损失的复杂度为 $O(Nd_vL)$,特征MLP提取的复杂度为 $O(NLHV)$,标签MLP的复杂度为 $O(NLKV)$,特征对比损失时间复杂度为 $O(V^2N^2H)$,整体的时间复杂度由特征对比损失主导, V 为视图数, H 为高级特征。在DEMVC中,时间复杂度的主导项为自编码器的前向与反向传播,表示为 $O(Nd_vLVE)$,其中 E 表示为训练轮数, L 为网络参数量,整个算法的时间复杂度也与样本数 N 线性相关。在DSMVC中,每个特征提取器的时间复杂度为 $O(Nd_vL)$,安全模块的时间复杂度为 $O(Nd_vV)$,损失函数的总复杂度为 $O(N^2K+N^2d)$ 。在CMVC中,视图重构损失复杂度为 $O(Nd_vL)$ 多分支空间共享特征聚合的时间

复杂度为 $O(IN(Vd_v)^2)$,其中 I 为分支数量,对比学习的损失复杂度为 $O(N(K+d))$,整体时间复杂度为 $O(E[IN(Vd_v)^2+N^2(K+d)])$, E 为总训练迭代次数。在SCMVC中,时间复杂度主要由自编码器的预训练与重构,视图共识特征提取,全局特征融合对比损失计算以及使用MMD距离的权重计算组成,其中总时间复杂度由对比损失和MMD计算主导,为 $O(VN^2d_v)$ 。本文方法由于引入了更多的结构建模,在计算上相比于一些多视图聚类方法略高,但换来了更好的聚类性能。同时,本文方法在时间复杂度上具有良好的可扩展性和计算效率,适用于中大规模的多视图聚类任务,在保持高性能的同时,有效控制了计算复杂度。

4 结 论

本文提出一种基于双通道粒计算的深度多视图聚类方法,通过运用双通道特征融合模块提取并凸显关键特征。为了充分利用多视图数据不同层

次的语义信息,提出了基于双通道的对比学习策略,同时在样本级和模糊粒球局部结构进行对比学习,在模糊粒球级对比学习中采用模糊粒球内部对比学习和跨视图模糊粒球对比学习,使得模型能更好处理边界样本,同时保持跨视图粒球结构的一致性。此外,为了让高质量视图能够在聚类中占据主导地位,设计了视图自适应注意力权重分配机制。本文在8个公开数据集上进行了验证,表明了本文所提出方法的有效性。在未来的工作中,将进一步探索本方法在实际场景中的应用。

参考文献:

- [1] CHAO G, SUN S, BI J. A survey on multiview clustering [J]. *IEEE Transactions on Artificial Intelligence*, 2021, 2(2): 146-168.
- [2] WANG Y, CHANG D, FU Z, et al. Consistent multiple graph embedding for multi-view clustering [J]. *IEEE Transactions on Multimedia*, 2021, 25: 1008-1018.
- [3] 周天奕, 丁卫平, 黄嘉爽, 等. 模糊逻辑引导的多粒度深度神经网络[J]. *模式识别与人工智能*, 2023, 36(9): 778-792.
ZHOU Tianyi, DING Weiping, HUANG Jiashuang, et al. Fuzzy logic guided deep neural network with multi-granularity [J]. *Pattern Recognition and Artificial Intelligence*, 2023, 36(9): 778-792.
- [4] XU C, GUAN Z, ZHAO W, et al. Deep multi-view concept learning [C]//*Proceedings of IJCAI*. [S.l.]: International Joint Conferences on Artificial Intelligence Organization, 2018: 2898-2904.
- [5] JU H, GUO J, DING W, et al. D3WC: Deep three-way clustering with granular evidence fusion [J]. *Information Fusion*, 2025, 114: 102699.
- [6] JU H, LU Y, DING W, et al. Three-way evidence theory-based density peak clustering with the principle of justifiable granularity [J]. *Applied Soft Computing*, 2024, 152: 111217.
- [7] 陈俊芬, 赵佳成, 翟俊海, 等. 基于无监督学习视觉特征的深度聚类方法 [J]. *南京航空航天大学学报*, 2021, 53(5): 718-725.
CHEN Junfen, ZHAO Jiacheng, ZHAI Junhai, et al. Deep clustering method based on unsupervised visual features learning [J]. *Journal of Nanjing University of Aeronautics & Astronautics*, 2021, 53(5): 718-725.
- [8] 胡深, 钱宇华, 王婕婷, 等. 基于对比学习的超多类深度图像聚类模型 [J]. *计算机科学*, 2023, 50(9): 192-201.
HU Shen, QIAN Yuhua, WANG Jieting, et al. Super multi-class deep image clustering model based on contrastive learning [J]. *Computer Science*, 2023 50(9): 192-201.
- [9] 冯天婵. 基于对比学习的深度多视图聚类方法研究 [D]. 太原: 山西大学, 2024.
FENG Tianchan. Research on deep multi-view clustering method with contrastive learning [D]. Taiyuan: Shanxi University, 2024.
- [10] WANG Q, CHENG J, GAO Q, et al. Deep multi-view subspace clustering with unified and discriminative learning [J]. *IEEE Transactions on Multimedia*, 2020, 23: 3483-3493.
- [11] LIN Z, KANG Z. Graph Filter-based multi-view attributed graph clustering [C]//*Proceedings of IJCAI*. [S.l.]: International Joint Conferences on Artificial Intelligence Organization, 2021: 2723-2729.
- [12] ZHANG C, WANG S, LIU J, et al. Multi-view clustering via deep matrix factorization and partition alignment [C]//*Proceedings of the 29th ACM International Conference on Multimedia*. [S.l.]: ACM, 2021: 4156-4164.
- [13] LIU J, LIU X, YANG Y, et al. Contrastive multi-view kernel learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(8): 9552-9566.
- [14] ZHAO X, EVANS N, DUGELAY J L. A subspace co-training framework for multi-view clustering [J]. *Pattern Recognition Letters*, 2014, 41: 73-82.
- [15] JU H, LU Y, DING W, et al. Multigranularity information fused contrastive learning with multiview clustering [J/OL]. *IEEE Transactions on Neural Networks and Learning Systems*, 2025. DOI: 10.1109/TNNLS.2025.3574885
- [16] XIE J, GIRSHICK R, FARHADI A. Unsupervised deep embedding for clustering analysis [C]//*Proceedings of International Conference on Machine Learning*. [S.l.]: PMLR, 2016: 478-487.
- [17] GUO X, GAO L, LIU X, et al. Improved deep embedded clustering with local structure preservation [C]//*Proceedings of IJCAI*. New York: Association for Computing Machinery, 2017, 17: 1753-1759.
- [18] XU J, REN Y, LI G, et al. Deep embedded multi-view clustering with collaborative training [J]. *Information Sciences*, 2021, 573: 279-290.
- [19] YU T, XU Y, YAN N, et al. Robust and fast subspace representation learning for multi-view subspace clustering [J]. *Applied Soft Computing*, 2025, 175: 113050.
- [20] XU J, REN Y, TANG H, et al. Self-supervised discriminative feature learning for deep multi-view clustering [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 35(7): 7470-7482.

- [21] TANG H, LIU Y. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2022: 202-211.
- [22] ZHU J, ZOU X, LIU L, et al. Trusted mamba contrastive network for multi-view clustering[C]//Proceedings of ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2025: 1-5.
- [23] GAN Y, YOU Y, HUANG J, et al. Multi-view clustering via multi-stage fusion[J]. IEEE Transactions on Multimedia, 2025, 27: 4571-4583.
- [24] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2020: 1597-1607.
- [25] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 9729-9738.
- [26] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent-A new approach to self-supervised learning[J]. Advances in Neural Information Processing Systems, 2020, 33: 21271-21284.
- [27] LI J, ZHOU P, XIONG C, et al. Prototypical contrastive learning of unsupervised representations[EB/OL]. (2020-5-11). <https://doi.org/10.48550/arXiv.2005.04966>.
- [28] CARON M, MISRA I, MAIRAL J, et al. Unsupervised learning of visual features by contrasting cluster assignments[J]. Advances in Neural Information Processing Systems, 2020, 33: 9912-9924.
- [29] FENG X, XU Y, LU G, et al. Hierarchical contrastive learning for pattern-generalizable image corruption detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2023: 12076-12085.
- [30] XU J, TANG H, REN Y, et al. Multi-level feature learning for contrastive multi-view clustering[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern recognition. [S.l.]: IEEE, 2022: 16051-16060.
- [31] CHEN J, MAO H, WOO W L, et al. Deep multi-view clustering by contrasting cluster assignments[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2023: 16752-16761.
- [32] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 7132-7141.
- [33] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 3-19.
- [34] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [35] BEZDEK J C, EHRLICH R, FULL W. FCM: The fuzzy c-means clustering algorithm[J]. Computers & Geosciences, 1984, 10(2/3): 191-203.
- [36] DASGUPTA S, LONG P M. Performance guarantees for hierarchical clustering[J]. Journal of Computer and System Sciences, 2005, 70(4): 555-569.
- [37] YANG X, JIAQI J, WANG S, et al. Dealmvc: Dual contrastive calibration for multi-view clustering[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: Association for Computing Machinery, 2023: 337-346.
- [38] WU S, ZHENG Y, REN Y, et al. Self-weighted contrastive fusion for deep multi-view clustering[J]. IEEE Transactions on Multimedia, 2024, 26: 9150-9162.
- [39] ZHANG Y, YAN W, TANG C, et al. Multi-branch Space Sharing Feature Aggregation for contrastive multi-view clustering[J]. Pattern Recognition, 2025, 167: 111704.

(编辑:刘彦东)