



# 毕 业 论 文

题 目 基于多尺度对齐和扩展边界  
的增强多视图分类模型研究

姓 名 王逸远

学 号 21074219

指导教师 吕庚育

日 期 2025 年 5 月 28 日

北京工业大学

毕业设计（论文）任务书

题目	基于多尺度对齐和扩展边界的增强多视图分类模型研究				
专业	计算机科学与技术	学号	21074219	姓名	王逸远

【主要内容】

本课题拟提出基于对比学习和模糊集合理论的多视图分类模型。该模型结合对比学习在特征空间中的应用和模糊集合理论对视图分类边界进行清晰拟合，从而提高预测的准确性。

【基本要求】

知识与理解：掌握机器学习和深度学习基本理论，了解领域前沿。

技术技能：熟练掌握 PyTorch 框架及 Linux 。

个人素质：具备良好心态，能坚持应对学术挑战。

文献能力：具备中英文学术论文的阅读与写作能力。

【主要参考资料】

[1]Jing Wang. "Triple-granularity contrastive learning for deep multi-view subspace clustering". In: Proceedings of the 31st ACM International Conference on Multimedia, pp. 2994–3002, 2023.

[2]Ting Chen. "A simple framework for contrastive learning of visual representations". In: International Conference on Machine Learning, pp. 1597–1607 (2020).

完成期限：2025-05-24

指导教师签章：

专业负责人签字：2024 年 12 月 18 日

# 毕业设计(论文)诚信声明书

本人郑重声明：在毕业设计（论文）工作中严格遵守学校有关规定，恪守学术规范；所提交的论文是我个人在导师指导下独立研究、撰写的成果，毕业设计（论文）中所引用他人的文字、研究成果，均已在毕业设计（论文）中加以说明；在本人的毕业生设计（论文）中未剽窃、抄袭他人的学术观点、思想和成果，未篡改实验数据。

本毕业设计（论文）和资料若有不实之处，本人愿承担一切相关责任。

学生签名：\_\_\_\_\_ 日期： 2025 年 5 月 28 日

## 关于论文使用授权的说明

本人完全了解北京工业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

（保密的论文在解密后应遵守此规定）

签名：\_\_\_\_\_ 导师签名：\_\_\_\_\_ 日期： 2025 年 5 月 28 日

## 摘 要

多视图数据广泛存在于现实世界的各类应用中，是数据挖掘领域的重要研究对象。在多视图分析的下游任务中，多视图分类被认为是最具挑战性和价值的任务之一。通过融合多个视图的数据，模型能够获得更全面的特征表示。然而，特征异质性和信息冗余是多视图分类任务面临的两大主要难题。特征异质性指不同视图的数据来源于不一致的分布，缺乏统一的格式，使得视图特征融合极为困难；信息冗余则表现为数据中包含无关信息和模糊样本，导致模型无法学习清晰的决策边界，同时进一步影响模型泛化能力和决策可靠性。因此，需要设计一种能够同时缓解特征异质性和信息冗余的高效方法。

为了解决这些问题，本文提出了一种基于多尺度对齐和扩展边界的增强多视图分类模型研究方法，命名为 MAMC。该框架包括两个关键模块：多尺度对齐模块和自适应扩展边界模块。首先，多尺度对齐模块从视图间的实例尺度和类别间的原型尺度两个层次挖掘特征的共性与差异，有效缓解特征异质性问题。具体来说，实例尺度对齐模块对每个样本的多视图特征表示进行对齐，以挖掘视图之间的公共性语义特征，从而提升多视图数据一致性；原型尺度对齐模块从类别角度对齐样本原型，强化类间特征差异，帮助模型构建更清晰的决策边界。其次，自适应扩展边界模块基于数学模糊集理论，在表示空间中对原始决策边界进行动态扩展。通过将扩展边界模块集成到原型尺度对齐模块中，MAMC 模型能够缩紧类内表示并拉大类间距离，从而更好地处理模糊边界并缓解信息冗余问题。此外，与原始的决策边界相比较，扩展后的决策边界为潜在的未见样本保留更多分类空间，保证了模型的适应性和泛化能力。

MAMC 在多个广泛使用的多视图公开数据集上的结果表明，其分类性能显著优于多视图领域现有最先进的模型。同时，消融实验验证了多尺度对齐模块和扩展边界模块的有效性，超参数敏感性分析和噪声分析等实验进一步表明模型具有较强的鲁棒性和稳定性。此外，本文从理论上详细阐释了模型设计的合理性，并通过定性与定量实验证明，MAMC 能有效解决传统方法中的异构特征融合困难和决策边界模糊问题。凭借其先进的理论创新和优异性能表现，MAMC 为多视图数据的分类任务提供了一种创新且高效的解决方案，为多视图学习相关领域的研究提供了新的思路和方向。

**关键词：**多视图分类；对比学习；模糊集合



## Abstract

Multi-view data are widely present in various real-world applications and serve as a critical research focus in the field of data mining. Among the downstream tasks of multi-view analysis, multi-view classification is considered one of the most challenging and valuable tasks. By integrating data from multiple views, models can achieve more comprehensive feature representations. However, feature heterogeneity and information redundancy are two major challenges faced by multi-view classification tasks. Feature heterogeneity refers to the fact that data from different views often originate from inconsistent distributions and lack a unified format, making the fusion of view-specific features extremely difficult. Information redundancy manifests as the presence of irrelevant features and ambiguous samples in the data, which can blur the classification boundaries and reduce the model's generalization capability and decision reliability. Therefore, an efficient method that can simultaneously alleviate feature heterogeneity and information redundancy is urgently needed.

To address these challenges, this thesis proposes a novel multi-view classification framework based on multi-scale alignment and expanded boundaries, named MAMC. The framework comprises two key modules: a multi-scale alignment module and an adaptive expanded boundary module. First, the multi-scale alignment module mitigates feature heterogeneity by mining the commonalities and differences of features at both inter-view and inter-class levels. Specifically, the instance-scale alignment module aligns the multi-view features of each instance to extract common features among views, thereby improving the consistency of multi-view data. Meanwhile, the prototype-scale alignment module aligns the prototypes of samples across different classes to enhance inter-class feature differences, enabling the construction of clearer decision boundaries. Second, the adaptive expanded boundary module, grounded in mathematical fuzzy set theory, expands the decision boundaries in the representation space. By integrating the expanded boundary module into the prototype-scale alignment module, the MAMC model enhances intra-class compactness and enlarges inter-class separation, improving its ability to handle ambiguous boundaries and reduce redundancy. Compared to the original decision boundary, the expanded boundary reserves more classification space for unseen samples, boosting adaptability and generalization.

Extensive experiments on multiple widely-used publicly available datasets demonstrate that MAMC achieves significantly superior classification performance compared to existing state-of-the-art models. Ablation studies validate the effectiveness of the multi-scale alignment module and the extended boundary module, while hyperparameter sensitivity analysis and noise experiments further indicate the robustness and stability of the model. This thesis explains the theory behind the model design and shows through both qualitative and quantitative experiments that MAMC overcomes feature fusion challenges and ambiguous decision boundaries found in traditional methods. With its theoretical innovations and strong performance, MAMC provides an efficient solution for multi-view classification, offering new insights for multi-view learning.

**Keywords:** Multi-View Classification ; Contrastive Learning ; Fuzzy Set

## 目 录

摘 要 .....	I
Abstract .....	II
1. 绪论 .....	1
1.1 课题研究背景及意义 .....	1
1.2 基于深度学习的多视图模型研究现状 .....	2
1.2.1 多视图学习 .....	2
1.2.2 多视图分类 .....	4
1.2.3 研究现状总结 .....	5
1.3 本文研究内容及创新点 .....	6
1.4 论文组织框架 .....	7
2. 本文涉及的相关理论方法 .....	9
2.1 自编码器介绍和相关应用 .....	9
2.1.1 自编码器 .....	9
2.1.2 图自编码器 .....	9
2.1.3 自编码器在多视图领域的应用 .....	10
2.2 对比学习介绍和相关应用 .....	12
2.2.1 对比学习理论基础 .....	12
2.2.2 对比学习的应用 .....	13
2.3 模糊集理论 .....	15
2.4 交叉熵损失函数 .....	15
2.5 本章小结 .....	16
3. 基于多尺度对齐和扩展边界的增强多视图分类模型研究 .....	17
3.1 模型整体结构 .....	17
3.2 基于视图特定的自编码器特征提取 .....	20
3.3 基于实例层面的对齐模块 .....	21
3.4 基于原型层面的对齐模块 .....	23
3.4.1 扩展边界 .....	23
3.4.2 损失函数 .....	25

3.4.3 组合操作符的确定 .....	27
3.5 分类预测 .....	28
3.6 数学证明 .....	29
3.6.1 证明: $Q_{\text{pro}}$ 是 $Q$ 的上界 .....	29
3.6.2 证明: 选取 Min Pooling 为 Comb 操作符时类别方差最大化 .....	30
3.7 本章小结 .....	31
4. 实验结果与分析 .....	32
4.1 数据集介绍 .....	32
4.2 对比模型介绍 .....	33
4.3 评价指标 .....	34
4.3.1 分类任务指标 .....	34
4.3.2 复杂性分析指标 .....	35
4.4 实验设置 .....	35
4.5 分类结果 .....	36
4.6 消融实验与分析 .....	38
4.7 超参数分析 .....	38
4.8 噪声实验 .....	39
4.9 可视化分析 .....	41
4.10 收敛性分析 .....	42
4.11 复杂性分析 .....	43
4.12 本章小结 .....	44
总结与展望 .....	45
参考文献 .....	46

## 1. 绪论

本章首先在1.1节介绍多视图领域的课题研究背景和意义。其次在1.2节分别介绍利用多视图数据的多视图学习和多视图分类，同时对现有模型的不足之处进行分析并对领域研究现状进行总结。然后在1.3节阐述本文的研究内容和实现方案，以及分析所提出的模型如何解决领域中存在的问题。最后在1.4节介绍本文总体的组织架构。

### 1.1 课题研究背景及意义

多视图数据（Multi-View Data）是指通过不同视图、模态和异构特征集合对同一物体进行描述的数据形式，在现实世界中得到了广泛的应用<sup>[1][2]</sup>。例如，在目标检测分类任务中同一个物体可以由视频、图片和文本三个不同渠道的信息进行表示。又或者，通过结合同一个图像的不同特征（如颜色、纹理、形状等），可以将图像聚类成不同的类别。此外，在个性化推荐系统中通过结合用户的历史行为数据、社交网络以及商品的描述等不同角度的信息，能够为用户提供更精确的推荐结果。多视图分类（Multi-View Classification, MVC）提供了一个有效的框架，用于学习所需的多视图融合分类器，以连接这些异构特征，并对未见实例进行合理预测<sup>[3]</sup>。然而，多视图数据中隐藏的特征异质性和信息冗余问题已经成为多视图学习的主要障碍<sup>[4]</sup>。其中，特征异质性是指多视图数据来源于不同的数据分布，缺乏统一的数据格式，因此导致视图特征之间的融合极为困难<sup>[5]</sup>。而信息冗余则表现为数据中包含不必要的重复特征信息或模糊信息，这种冗余阻碍了对决策边界的合理划分，进而导致模型分类结果的不确定性。

现有的多视图学习主要通过寻找信息丰富的公共子空间<sup>[6][7][8]</sup>或理想的视图权重<sup>[9][10]</sup>来缓解特征异质性问题。其中，前者致力于挖掘多视图数据之间的公共性语义信息和用于互补的视图特异性语义信息<sup>[3][11]</sup>，但忽略了对于构建可靠的决策边界至关重要的类别差异；而后者假设各视图对样本信息的贡献度是稳定的<sup>[12]</sup>，并试图为不同视图分配合理的权重<sup>[13][14]</sup>。然而，由于不同视图的质量在实际应用中难以匹配多样化的样本<sup>[15]</sup>，这一假设难以成立，最终导致了不可靠的决策边界。例如，老虎和斑马具有相似的条纹但颜色不同，这使得模型会对颜色视图赋予更高的权重；然而，斑马和熊猫的颜色相同但条纹不同，这又促使模型对条纹视图给予更高的权重。这种情况下，视图的贡献度是动态变化的，而不是静态的。因此，传统方法中假设视图权重是固定不变的，就无法适应这种动态变化，导致模型构建出的决策边界不可靠。

为了解决这些问题，本文提出了一种基于多尺度对齐和扩展边界的多视图分类模型，命名为 MAMC。在模型设计方面，MAMC 包括两个关键模块：多尺度对齐模块和自适应扩展边界模块。首先，多尺度对齐模块通过实例尺度和原型尺度的双重对齐，分别从视图间和类别间挖掘样本特征共性与差异，从而有效缓解特征异质性问题；其次，

自适应扩展边界模块基于数学模糊集理论，专注于在表示空间中将原始边界动态扩展为更合理的决策边界。此外，原型尺度对齐模块在扩展边界模块的辅助下，能够帮助模型有效缩紧类内表示并拉大类间距离，缓解信息冗余问题，同时增强模型的泛化能力。

## 1.2 基于深度学习的多视图模型研究现状

多视图学习是一种机器学习方法，专注于处理具有异构特征的样本实例，尤其是那些包含不同模态数据的情况。其主要挑战之一是如何有效整合这些多样化的特征表示，以提取更全面有效的模态信息。本节将简要回顾关于多视图学习（Multi-View Learning, MVL）和多视图分类（Multi-View Classification, MVC）的一些相关工作和最新成果。

### 1.2.1 多视图学习

聚类（Clustering）<sup>[16] [17]</sup>是一种数据分析方法，旨在根据样本之间的相似性或差异性，将一组样本划分为若干子群体。聚类技术广泛应用于多个领域，包括统计学、计算机科学和机器学习，因其在模式识别、图像处理<sup>[18]</sup>、市场分析和生物信息学<sup>[19]</sup>中的有效性而备受关注。传统的聚类方法通常基于单一视图，即利用样本的单一特征集或数据来源进行分组。然而，在许多实际问题中，样本通常包含来自多个视图的数据。例如，在生物医学中，样本可能具有多种视图信息如基因表达、代谢组学数据和临床指标等；在社交网络中，用户的文本、图像和行为数据均可提供有价值的信息。这种单一视图的方法往往难以充分挖掘数据中潜在的复杂模式，因为它忽略了不同视图之间的互补性和相关性。多视图学习旨在融合来自多个视图的特征数据，充分利用各视图间的协同关系，有效克服了上述局限性。过去，多视图学习在现实世界中得到了广泛的应用，例如分类<sup>[20] [21]</sup>、聚类<sup>[22] [23]</sup>和检索<sup>[24]</sup>等，同时这些任务可大致分为五类：例如协同训练<sup>[25] [26]</sup>利用一个视图的先验知识来引导另一个视图的聚类以强调视图间的互助性、多核学习<sup>[27] [28]</sup>通过为每个视图预定义一个核函数并组合它们来捕捉不同视图数据中的非线性关系、多任务多视图聚类<sup>[29]</sup>通过设计一个多任务学习框架来捕捉任务间的相关性以实现知识转移、以及多视图图聚类<sup>[30] [31]</sup>和多视图子空间聚类<sup>[32]</sup>。其中，多视图图聚类和多视图子空间聚类是两类比较典型且具有代表性的任务。

在多视图图聚类领域，方法进一步分为基于谱的、基于图的和基于网络的三类聚类方法。在基于谱的研究中，又可以进一步细分为协同训练多视图谱聚类和协同正则化多视图谱聚类。其中前者的核心思想是通过视图间的协同学习，利用一个视图中的聚类信息来指导和优化另一个视图中的聚类过程，例如 CCT<sup>[33]</sup>在仅有少量标注样本的情况下提出了一种基于概率近似框架的共训练方法应用于网页分类问题；后者则专注于将不同视图对应的预测函数之间的差异最小化，例如 CMVSC<sup>[34]</sup>提出了一种共正则化谱聚类框架，通过进行正则化聚类来确保不同视角中的样本数据趋向一致的聚类结果。在基于图的方法中，如图 1-1 所示 MFLVC<sup>[35]</sup>提出了一种新颖的多层次特征学习框架，有效平衡

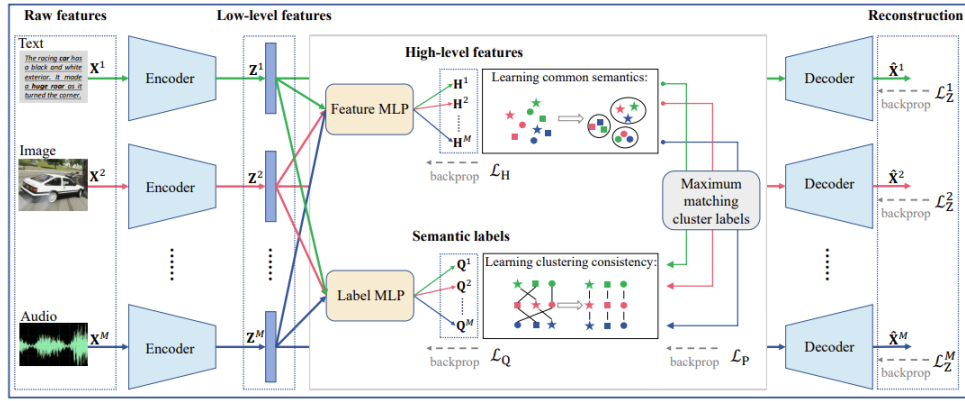


图 1-1 多视图图聚类方法 MFLVC 框架示意图

了多视图聚类中视图一致性与特异性提取冲突的问题。具体来说，MFLVC 首先通过自编码器对原始数据进行特征重建，并将这些特征应用于对比学习范式和标签信息提取，避免了传统方法中视图自身信息带来的冗余干扰，提高了在多个公开数据集上多视图图聚类的准确性。此外，LMF<sup>[36]</sup> 是一种基于链接矩阵分解的多视图融合方法，通过将多个视图自身的特异性特征与彼此之间的公共特征相结合有效整合多视图数据。同时，SWMCM<sup>[37]</sup> 通过引入拉普拉斯秩约束（Laplacian Rank Constraint）来自动学习每个视角的权重，提出了一种新颖的自加权多视图聚类方法。在基于网络的研究中，NoNClus<sup>[38]</sup> 通过非负矩阵分解在多个网络中进行聚类，该方法允许不同网络拥有独立的聚类结构，并利用网络相似性进行正则化。此外，MCA<sup>[39]</sup> 通过联合跨域聚类对齐，提出了一种鲁棒的多网络聚类算法，在多个真实数据集上表现出色。

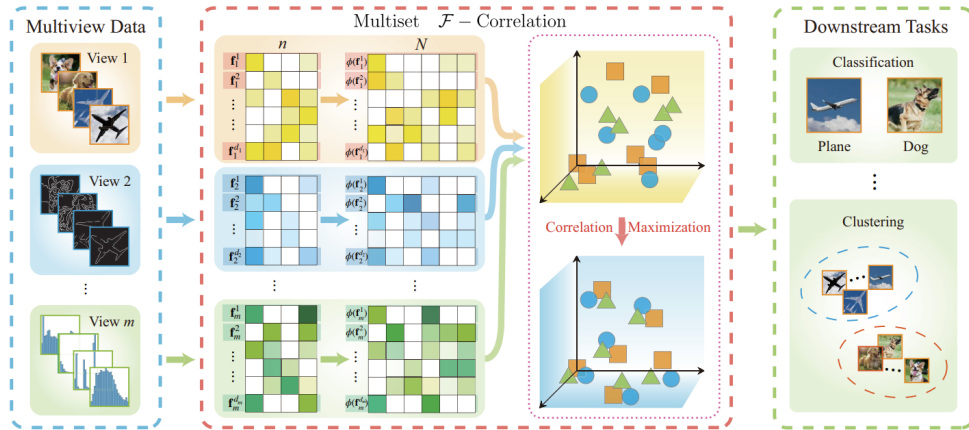


图 1-2 多视图子空间聚类方法 CFMR 框架示意图

对于多视图子空间聚类，这一类任务主要包括基于子空间学习方法<sup>[23] [40]</sup> 和非负矩阵分解方法。基于子空间学习方法的主要研究方向是寻找一个公共子空间，将不同视图的原始数据投影变换到该子空间中以生成一致的特征表示。例如，基于典型相关分析（Canonical Correlation Analysis, CCA）的方法<sup>[12]</sup> 通过最大化不同视图间的相关性，将

它们投影到一个共享的子空间中,并通过目标函数找到最终的视图联合表示。其次,如图 1-2 所示,CFMR<sup>[41]</sup>提出了一种基于标准  $\mathcal{F}$ -相关 ( $\mathcal{F}$ -correlation) 的框架,将每个特征通过非线性映射投影到特征子空间中。在该框架中,每个视图的数据通过数学中的非线性函数映射到一个新的空间,然后 CFMR 通过最大化各视图之间的相关性来减少样本中包含的冗余信息,以确保提取出来的视图特征保持高度相关性。最终,得到的特征表示将被用于多视图领域的下游任务,如分类和聚类。在基于非负矩阵分解的工作中,MVMF<sup>[42]</sup>提出了一种基于非负矩阵分解的多视图聚类算法,通过联合矩阵分解并引入归一化策略有效整合多视图信息。同时,DMvNMF<sup>[43]</sup>提出了一种基于正则化非负矩阵分解的多视图算法,通过在特征空间中保持几何结构并学习特定视图亲和关系,实现了有效的多视图数据聚类。此外,基于二值矩阵分解和视图权重的方法也受到关注。例如 AW-BMVC<sup>[13]</sup>提出了一种动态学习策略,通过给予视图特征合理的权重,以自适应方式有效利用视图特定信息,并通过损失融合策略实现性能提升。

### 1.2.2 多视图分类

在多视图分类任务领域,多视图数据本身编码了丰富的语义信息。通过整合来自不同视图的信息,可以增强多视图分类模型的理解和预测能力。现有的多视图分类方法通常可分为两类:基于特征融合的方法和基于决策的方法。基于特征融合的方法主要关注如何有效地融合来自各个视图的特征<sup>[44]</sup>。在过去,lrMMC<sup>[45]</sup>提出了一个基于矩阵分解的框架,为多个视图寻找一个共享表示,并在特征子空间上通过矩阵补全进行分类;其次,LSA-MML<sup>[46]</sup>利用矩阵分解捕捉不同视图之间的互补信息,并引入希尔伯特-施密特独立性准则 (Hilbert-Schmidt Independence Criterion) 来学习共享的子空间表示。同时,L-VSM<sup>[3]</sup>提出了一种新颖的多视图分类方法,通过标签驱动的特征图构建和特征融合策

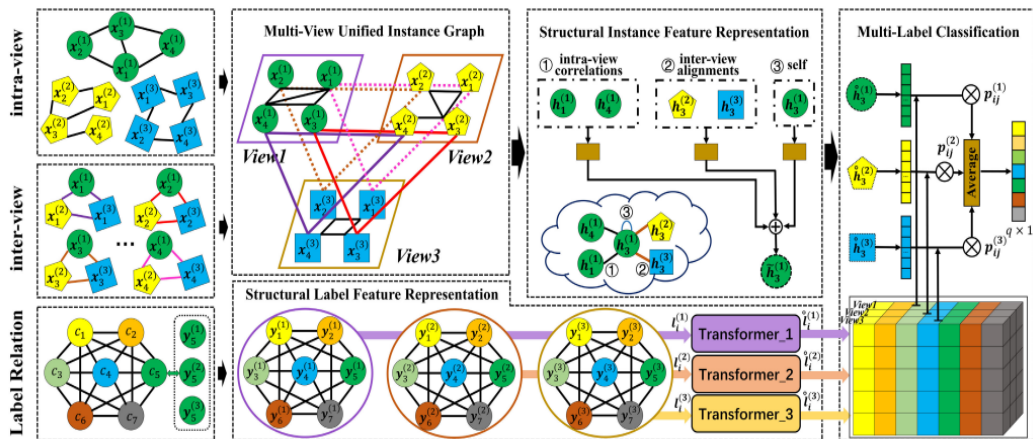


图 1-3 基于特征融合的 L-VSM 框架示意图

略,结合图注意力机制和 Transformer 架构,有效整合多视图特征与标签相关性并取得良好的性能表现。具体来说,如1-3所示 L-VSM 首先构建统一的多视图实例图以捕捉视



图内相关性和视图间对齐性，然后生成结构化实例特征表示和标签特征表示并将它们结合在一起通过亲和力得分来去优化模型的分类性能。近年来，MV-HFMD<sup>[47]</sup>通过引入一种创新的融合方案和互助蒸馏机制，使得神经网络能够更好地适应多视图分类任务。此外，mmdynamics<sup>[48]</sup>通过动态评估不同样本中视图特征表示的信息量，实现了对多个视图的异构特征可信融合。与基于特征融合的方法不同，基于决策的方法则更侧重于在决策层面做出合理的判断。在过去，DEMVC<sup>[49]</sup>提出了一种利用视图自身所包含的特定信息并自适应加权来生成深度交互特征的方法，最终通过多视图损失融合策略来实现联合决策。现如今，如图1-4所示 ETMC<sup>[15]</sup>通过在证据层面动态整合不同视图信息，提出了一个新的多视图学习范式，从而提升了分类结果的可靠性。其次，RCML<sup>[50]</sup>关注不同视

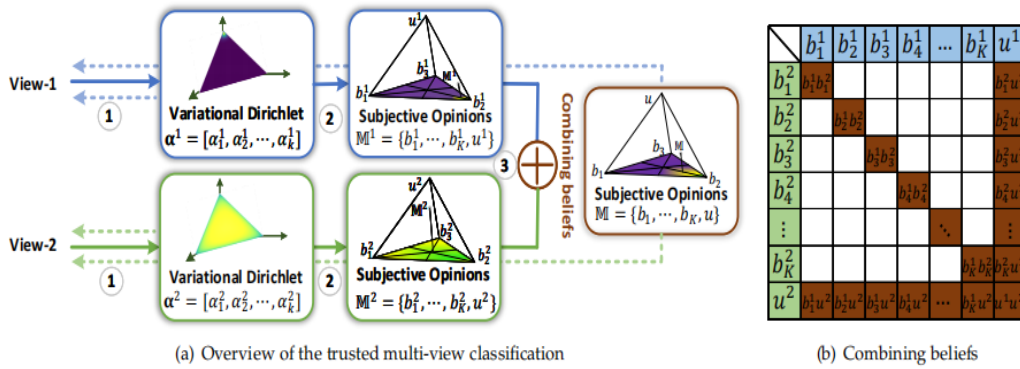


图 1-4 基于决策层面的 ETMC 方法框架示意图

图中潜在的冲突信息，并提出了一种新颖的特征聚合策略，有效建模了不同视图中的共性语义与它们自身可靠的特异性信息之间的关系。同时 MVA-KNN<sup>[51]</sup>引入邓普斯特-谢弗理论（Dempster-Shafer Theory）来整合各视图的信息，成功地将自适应的 KNN 算法扩展应用到多视图情景中。此外，IPMVSC<sup>[52]</sup>构建了一个基于模糊规则的分类模型，具有良好的可解释性，同时多视图领域的公开数据集上取得了良好的性能表现。

尽管现有的基于特征融合的方法能够将多视图领域不同视图中存在的异构特征有效地融合在一起，但是这类方法往往没有彻底地解决各视图信息中存在的异质性问题，从而导致特征融合没有达到令人满意的效果；而基于决策的方法虽然能够有效地在决策阶段整合不同视图的信息，但它们通常仅限于在最终决策阶段与视图进行交互，未能充分融合所有视图的信息。相比之下，本文的模型通过在表示学习阶段对视图特征进行对齐，并在决策阶段进行有效结合，实现了跨视图信息的高效互动和整合。

### 1.2.3 研究现状总结

本章节主要介绍了基于深度学习的多视图模型研究现状，首先在1.2.1小节中对多视图学习进行了介绍，重点介绍了多视图图聚类和多视图子空间聚类这两类比较具有代表性的任务；然后在1.2.2小节中介绍了多视图分类，重点是从基于特征融合的方法和基于决策的方法这两个角度介绍了目前研究领域的现状和代表性工作。通过对多视图学习和



多视图分类研究领域的系统性梳理，可以发现现有研究在融合多视图数据异构特征、进行可信合理决策方面取得了显著进展，但仍无法解决特征异质性和信息冗余问题。

例如，在特征融合层面，现有的方法通常基于子空间投影来去捕获不同视图之间的公共性语义信息，但过度强调一致性可能削弱模型对判别性特征的提取能力，同时这些方法往往忽视类别间的差异，难以界定清晰的决策边界，也就无法缓解多视图数据中存在的特征异质性问题；而在决策层面，许多方法希望通过为不同的视图分配合理的权重来提升多视图领域下游任务的性能，但这些方法通常假设视图权重是固定不变的，无法动态适应不同的样本实例，也就无法解决多视图信息冗余的问题，特别是在模糊样本接近决策边界时。因此，本文的研究重点是在特征表示空间中挖掘视图间的共性语义信息，同时帮助模型学习清晰的决策边界，从而有效缓解特征异质性和信息冗余的问题。

针对上述挑战，本文引入数学模糊集理论，提出 MAMC 模型以系统性优化多视图分类任务。相较于现有研究，本工作的创新定位体现在：第一，通过多尺度对齐模块实现实例级特征交互与原型级语义校准的双重优化，在保留视图特异性的同时增强跨视图语义一致性；第二，基于模糊隶属度函数构建自适应扩展边界模块以动态扩展原始决策边界，通过将扩展边界模块集成到原型尺度对齐模块中，模型能够有效调节类内紧凑性与类间分离性，从而缓解模糊样本和信息冗余问题。这些设计使得模型能够更精准地刻画多视图数据的内在结构，帮助模型形成清晰的决策分类边界。

### 1.3 本文研究内容及创新点

近年来，多视图学习的研究致力于通过估计多视图数据中的不确定性信息来寻求有效整合多视图特征，从而实现可信的决策<sup>[53][50]</sup>。实际上，尽管这些研究通过调整数据分布以实现更好的分类决策<sup>[54]</sup>，但仍然无法解决上述提到在多视图研究中存在的特征异质和信息冗余问题。为了直观地说明这些问题以及本文的研究动机，本文在第四章节的图4-3中提供了 t-SNE 的可视化结果，其中包括本文的 MAMC 模型（基于多尺度对齐和扩展边界的增强多视图分类模型）以及四个先进的多视图模型：mmdynamics<sup>[48]</sup>、ETMC<sup>[15]</sup>、PDMF<sup>[53]</sup>、RCML<sup>[50]</sup>和原始数据（Raw）的可视化结果。所有的可视化结果是基于 Scene15 数据集<sup>[55]</sup>来生成的。t-SNE 图像呈现的结果反映了当前多视图分类模型存在的两个问题：首先是这些模型无法充分挖掘不同视图之间的共性以收紧类内特征表示；其次是它们忽略了类别差异对清晰化模糊决策边界的贡献。

因此，本文设计的 MAMC 模型专注于挖掘不同视图之间的共性并在表示空间中学习清晰的决策边界，以解决特征异质性和信息冗余的问题。MAMC 模型的核心组成在于多尺度对齐模块和辅助原型尺度对齐模块的扩展边界模块。具体而言，多尺度对齐模块分别基于实例尺度和类别尺度进行对齐，其中实例级对齐模块通过对齐每个样本实例的视图表示来挖掘视图间的共性，原型级对齐模块则利用具有相同标签的实例之间的类别差异，来寻找清晰的决策边界。扩展边界的设计基于数学模糊集理论<sup>[56]</sup>，其扩展过

程涉及模糊表示和清晰表示。其中模糊表示定义在模糊表示空间，而清晰表示定义在常规表示空间。扩展边界的核心假设是清晰表示是模糊表示的一种特殊类型，这使得本文能够在模糊表示空间中找到清晰表示的可学习正样本，随后这些样本被用于构造扩展边界的元素。最终，扩展的决策边界被集成到原型级对齐过程中，以收紧类内特征表示并放大类间距离。因此，本文的主要贡献可以归纳为：

(1) 提出了创新的多视图分类模型 MAMC，旨在解决多视图研究领域由特征异质性和信息冗余引发的问题。通过引入多尺度对齐模块，MAMC 分别从视图间和类别间挖掘视图特征共性语义和差异信息，从而有效缓解特征异质性问题。

(2) 提出了一种新颖的自适应扩展边界，用于辅助原型尺度对齐模块解决信息冗余和模糊边界问题。通过将原始边界进行动态扩展，为潜在属于该类别的样本保留了额外的空间，帮助模型形成清晰的分类边界。

(3) 从理论上解释了模型设计的合理性，并在广泛的多视图公开数据集上设计了一系列相关实验证明了所提模型表现出的显著优越性。

## 1.4 论文组织框架

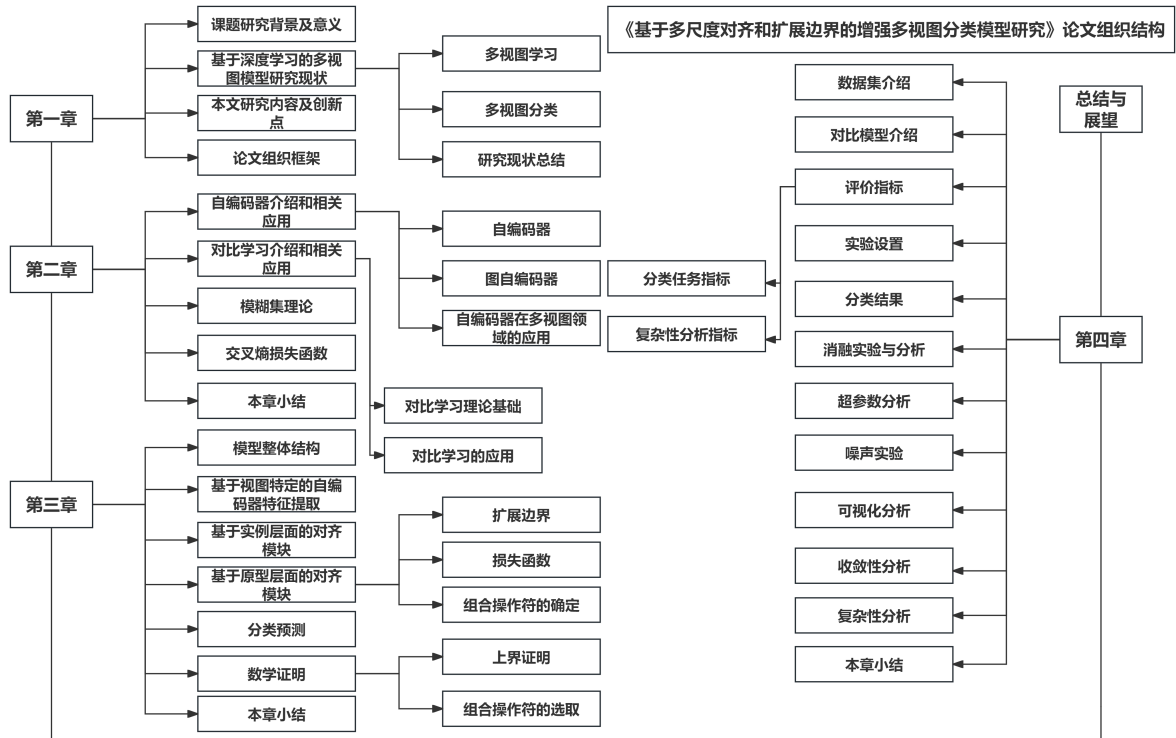


图 1-5 全文组织结构示意图

如图1-5所示，本文共计四个章节，组织结构如下：

**第一章 绪论** 第一章介绍本文的研究背景及意义，以及基于深度学习的多视图模型研究现状，同时总结本文的研究动机、主要创新点并说明论文的组织结构。

**第二章 本文涉及的相关理论方法** 第二章首先介绍了自编码器和对比学习的相关知识和代表性工作，然后介绍模糊集理论和交叉熵损失函数，为构建模型提供前置知识。

**第三章 基于多尺度对齐和扩展边界的增强多视图分类模型研究** 第三章详细介绍本文提出的基于多尺度对齐和扩展边界的增强多视图分类模型研究的原理和设计思路，其中主要结构分别有用于特征提取的自编码器模块、基于对比学习构建的多尺度对齐模块、辅助原型尺度对齐模块的扩展边界模块和联合分类器。

**第四章 实验结果和分析** 第四章介绍了本文在多视图分类任务中使用的公开数据集和评价指标，以及用于比较本文提出模型性能的对比基线模型。然后介绍了实验的详细设置，并进行了各种实验以全面地对模型有效性进行评估。

最后的**总结与展望**部分归纳了本文的全部工作内容，分析了本研究提出模型的缺点和可以改进的方向，并对未来的进一步发展和应用进行了展望。

## 2. 本文涉及的相关理论方法

本章节主要介绍本文所涉及到的相关理论方法，包括自编码器、对比学习、模糊集理论和交叉熵损失函数，以及它们的相关应用。首先，本文在2.1.1小节中介绍自编码器相关理论基础，随后在2.1.2小节中介绍图自编码器，以及在2.1.3小节中介绍自编码器应用在多视图领域中的相关工作。此外，分别在2.2.1小节和2.2.2小节中介绍对比学习的理论基础和相关工作，然后在2.3节介绍模糊集理论的相关概念和定义，最后在2.4节中介绍交叉熵损失函数以及在2.5节中进行本章总结。

### 2.1 自编码器介绍和相关应用

#### 2.1.1 自编码器

自编码器 (Autoencoder)<sup>[57] [58]</sup> 是一种无监督学习算法，广泛用于数据降维、特征学习和数据生成等任务。它的核心思想是通过神经网络将输入数据映射到一个低维的表示 (编码过程)，然后再通过神经网络将该低维表示重建回原始数据 (解码过程)。自编码器的训练目标是使得输入数据和重建数据之间的差异尽可能小。自编码器的基本架构包括两个主要部分：编码器 (Encoder) 和解码器 (Decoder)。编码器的任务是将输入数据  $\mathcal{X}$  通过编码函数映射到低维潜在表示  $\mathcal{Z}$ ：

$$\mathcal{Z} = \mathcal{F}_\theta(\mathcal{X}) \quad (2-1)$$

其中， $\mathcal{F}$  为编码函数， $\theta$  为编码器参数。解码器将潜在表示  $\mathcal{Z}$  重新映射回原始空间：

$$\hat{\mathcal{X}} = \mathcal{G}_\phi(\mathcal{Z}) \quad (2-2)$$

其中， $\mathcal{G}$  为解码函数， $\phi$  为解码器参数。自编码器的目标是最小化输入数据  $\mathcal{X}$  和重建数据  $\hat{\mathcal{X}}$  之间的重构误差，常用的损失函数为均方误差 (Mean Squared Error, MSE)：

$$Loss(\mathcal{X}, \hat{\mathcal{X}}) = \frac{1}{N} \sum_{i=1}^N (\mathcal{X}_i - \hat{\mathcal{X}}_i)^2 \quad (2-3)$$

其中， $Loss(\mathcal{X}, \hat{\mathcal{X}})$  是损失函数，用于表示输入数据  $\mathcal{X}$  和重建数据  $\hat{\mathcal{X}}$  之间的均方误差。 $N$  是样本数量，表示一批数据中包含的样本总数。 $\mathcal{X}_i$  是第  $i$  个样本的真实值， $\hat{\mathcal{X}}_i$  是第  $i$  个样本的重建值， $(\mathcal{X}_i - \hat{\mathcal{X}}_i)^2$  度量了它们之间的差异。

#### 2.1.2 图自编码器

图自编码器 (Graph Autoencoder, GAE) 是一种用于处理图结构数据的自编码器，旨在学习图中节点的低维特征表示，并重建图的邻接矩阵或节点特征。图数据 (如社交网

络、分子图等）不同于图像或文本等欧几里得数据，具有非规则的拓扑结构。GAE 通常与图卷积网络  $\mathcal{GCN}$  结合使用，能够捕捉图的拓扑结构和节点特征，适用于图表示学习、链接预测等任务。在这里，使用  $\mathcal{GCN}$  将图中的节点映射到低维嵌入空间：

$$\mathcal{Z} = \mathcal{F}_\theta(\mathcal{A}, \mathcal{X}) = \mathcal{GCN}_\theta(\mathcal{A}, \mathcal{X}) \quad (2-4)$$

其中， $\mathcal{A}$  是图的邻接矩阵， $\mathcal{X}$  是原始的节点特征矩阵， $\mathcal{Z}$  是节点的嵌入矩阵， $\mathcal{GCN}_\theta$  是图卷积网络， $\theta$  是编码器的参数。GAE 的解码器从节点嵌入  $\mathcal{Z}$  重建图结构。在这里，最常用的解码器是内积解码器，用于重建邻接矩阵：

$$\hat{\mathcal{A}} = g_\phi(\mathcal{Z}) = \sigma(\mathcal{Z}\mathcal{Z}^T) \quad (2-5)$$

其中， $\sigma$  表示 sigmoid 函数，通常用来将输出限制在 0 到 1 之间，确保邻接矩阵的元素值符合图的连接概率（边的存在概率）。 $\hat{\mathcal{A}}$  表示重建的邻接矩阵，即从节点低维表示  $\mathcal{Z}$  恢复出来的邻接矩阵。同时，GAE 通过交叉熵损失来最小化原始邻接矩阵  $\mathcal{A}$  和重构邻接矩阵  $\hat{\mathcal{A}}$  之间的差异，在这里定义为：

$$Loss(\mathcal{A}, \hat{\mathcal{A}}) = -\frac{1}{|\mathcal{E}|} \sum_{(i,j) \in \mathcal{E}} [\mathcal{A}_{ij} \log \hat{\mathcal{A}}_{ij} + (1 - \mathcal{A}_{ij}) \log(1 - \hat{\mathcal{A}}_{ij})] \quad (2-6)$$

其中， $\mathcal{A}_{ij}=1$  表示原图中的节点  $i$  和  $j$  之间有边， $\mathcal{A}_{ij}=0$  表示它们之间无边， $\hat{\mathcal{A}}_{ij}$  表示模型预测出的该边存在的概率。此外， $\mathcal{E}$  表示在训练中所考虑的节点对（通常包含所有真实边及相应的负采样）， $|\mathcal{E}|$  是边的总数，用于归一化损失。

### 2.1.3 自编码器在多视图领域的应用

自编码器在多视图领域得到了广泛的使用并呈现出令人满意的效果，其中涌现了许多具有代表性的工作，在这里本文主要介绍 AE2-Nets<sup>[59]</sup>、TRUST<sup>[1]</sup> 和 O2MAC<sup>[60]</sup> 这三个比较具有代表性的工作。AE2-Nets 是一个新颖的无监督多视图表示学习框架，旨在从

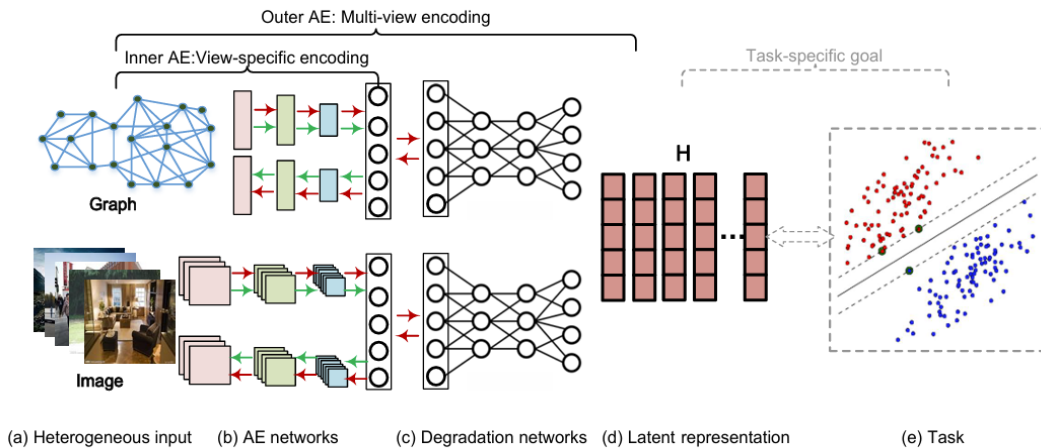


图 2-1 AE2-Nets 框架示意图

多个异构视图学习中学习一个综合低维表示，其创新在于嵌套的自动编码器网络设计，通过内层和外层两个部分的协同工作实现视图特定表示学习与多视图信息编码的联合优化。具体而言，如图2-1所示内层自动编码器独立地针对每个视图提取特征并重构原始数据，而外层自动编码器则将这些视图特定表示整合为一个联合的低维表示，来捕捉不同视图中的不一致信息。这种设计结构避免了传统的多视图方法通过子空间投影再组合所导致的信息丢失和冗余问题，同时实验结果表明其在分类和聚类任务中优于现有方法。

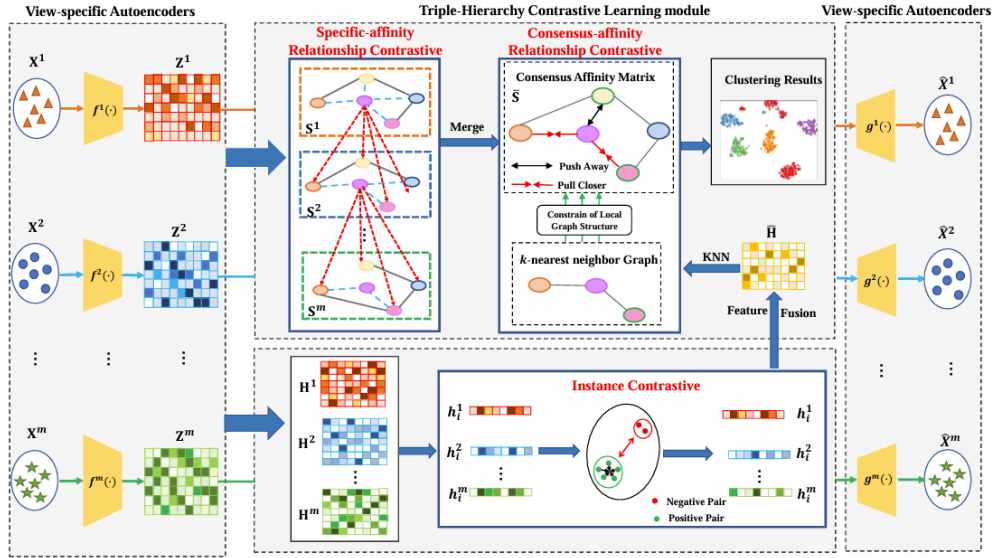


图 2-2 TRUST 框架示意图

TRUST 是一个基于深度学习的多视图子空间聚类框架。如图2-2所示整体结构分为三个部分：左边是基于视图特定的自编码器，负责将不同视图的原始数据（如  $X^1, X^2, \dots, X^m$ ）通过编码器（如  $f^1(\cdot), f^2(\cdot), \dots, f^m(\cdot)$ ）提取潜在特征嵌入向量（如  $Z^1, Z^2, \dots, Z^m$ ），并将它们变换成更高层次的特征  $S$ （如  $S^1, S^2, \dots, S^m$ ）和  $H$ （如  $H^1, H^2, \dots, H^m$ ）供对比学习模块使用；中间部分通过三个层次分别是实例层次、特定亲和关系层次和一致性亲和关系层次来构建对比学习范式，充分利用多视图数据中的互补信息，在特征空间中拉近正样本对的距离和推远负样本对的距离以形成清晰的聚类边界；右边部分则是对应的自编码器输出，通过解码器（如  $g^1(\cdot), g^2(\cdot), \dots, g^m(\cdot)$ ）重构原始视图数据（ $\hat{X}^1, \hat{X}^2, \dots, \hat{X}^m$ ）。TRUST 通过在不同层次使用对比学习联合优化视图特征表示，实现了多视图数据在一致性与结构关系上的深度融合，从而显著提升了聚类性能和鲁棒性。

O2MAC 是一个专为多视图图聚类设计的深度学习框架。多视图图聚类是一项近年来备受关注的研究任务，其目标是从具有多个视图的图数据中寻找一个合理的分区。O2MAC 的创新在于，它首次将图深度学习技术应用于具有属性信息的多视图图聚类任务，通过处理图结构和节点内容，来学习高质量的节点表示并优化聚类结果。如图2-3所示 O2MAC 框架主要由两个核心部分组成：分别是单视图到多视图的图自动编码器和自训练图聚类模块。图自编码器模块通过编码器从选定的信息量最丰富的视图中提取共



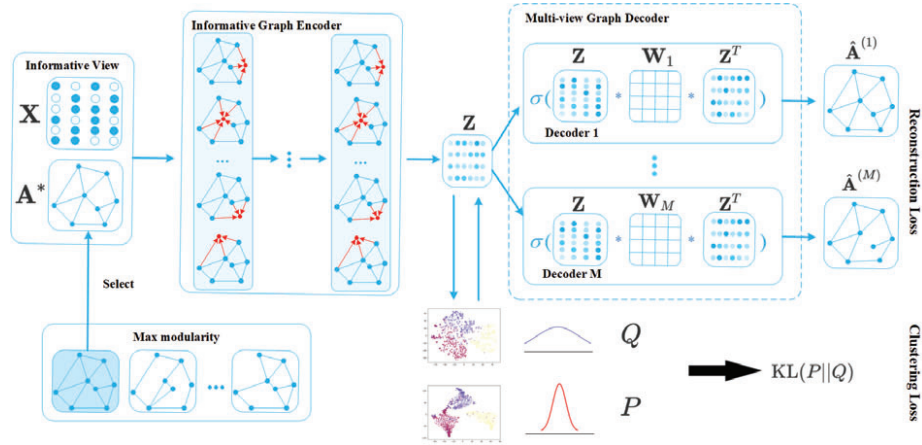


图 2-3 O2MAC 框架示意图

享的节点表示，并利用解码器重建多个视图的图结构和内容，以捕捉多视图的共享特征和视图特异性信息；自训练图聚类模块则通过生成软标签并结合 KL 散度损失，迭代优化聚类结果和节点表示，实现了对多视图的有效聚类。

## 2.2 对比学习介绍和相关应用

### 2.2.1 对比学习理论基础

对比学习（Contrastive Learning）<sup>[61]</sup> 是一种强大的自监督学习范式，模型通过区分相似的正样本和不相似的负样本数据来学习特征表示。其核心思想是拉近正样本对在特征空间中的距离，同时将负样本对距离推远，从而获得判别性特征。InfoNCE（Info Noise Contrastive Estimation）对比损失函数<sup>[62]</sup> 是一种在深度学习中广泛使用的损失函数，特别是在自监督学习和对比学习领域，旨在衡量特征嵌入空间中正样本对和负样本对的相似性。该损失函数通过最大化正样本对的相似度，同时最小化负样本对的相似度，以帮助模型学习良好的决策边界。给定两个随机变量  $X$  和  $Y$ ，互信息（Mutual Information, MI）<sup>[63]</sup> 度量的是它们之间的相关程度或信息共享量：

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (2-7)$$

其中， $p(x,y)$  表示联合分布， $p(x)$  和  $p(y)$  分别表示边缘分布。当  $X$  与  $Y$  完全独立时， $p(x,y) = p(x)p(y)$ ，此时  $X$  和  $Y$  的互信息为 0；当  $X$  完全由  $Y$  决定（或二者有确定的函数映射关系）时，互信息趋向于一个较大的值。同时互信息的熵形式定义为：

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \quad (2-8)$$

其中， $H(X), H(Y)$  分别是  $X$  和  $Y$  的熵，描述每个随机变量自身的不确定性； $H(X,Y)$  是  $X$  与  $Y$  的联合熵，描述二者共同的不确定性。当二者越相关，知道其中一个能更多地减

少对另一个的不确定性，于是表达式  $(H(X)+H(Y)-H(X,Y))$  的值就越大。在引入了互信息的概念之后，将 InfoNCE 对比损失函数定义为：

$$\mathcal{L} = -\log \frac{\exp(\sin(q, k^+)/\tau)}{\exp(\sin(q, k^+)/\tau) + \sum_{k^- \in \mathcal{K}^-} \exp(\sin(q, k^-)/\tau)} \quad (2-9)$$

其中，查询样本  $q$  与正样本  $k^+$  构成相似对，通常源于同类样本或数据增强视图；负样本集合  $\mathcal{K}^-$  则包含与  $q$  不相关的样本特征。相似度函数  $s(\cdot, \cdot)$  多采用余弦相似度或点积计算，温度系数  $\tau$  用于调整相似性分数的尺度。该损失函数通过 softmax 形式构建分类任务，最大化  $q$  与  $k^+$  的归一化相似度同时抑制  $q$  与  $\mathcal{K}^-$  中样本的关联强度，促使模型在特征空间中构建具有判别性的表示结构，使相似样本聚集而非相似样本分离。

此外，从信息论的角度，InfoNCE 损失可以被视为互信息的一种可计算的下界。具体来说，InfoNCE 损失的期望与互信息  $I(X;Y)$  之间存在如下不等式：

$$I(X;Y) \geq \log(N) - \mathcal{L} \quad (2-10)$$

在这里，当本文最小化 InfoNCE 损失  $\mathcal{L}$  时，就等价于最大化  $(\log(N) - \mathcal{L})$ 。当 InfoNCE 损失作为互信息的下界时，优化该损失等同于间接提高  $X$  和  $Y$  之间的互信息。因此，如果希望通过神经网络提取的特征嵌入表示尽量保留原始数据中的有用信息，那么就可以通过最小化 InfoNCE 损失以最大化互信息来实现。

### 2.2.2 对比学习的应用

近年来，几种对比学习框架由于其在学习强大表示方面的有效性而在深度学习中获得了突出的地位。例如，SimCLR<sup>[64]</sup> 是一种基于对比学习的无监督学习框架，它使用数据增强生成多个视图，通过最大化样本和其增强视图的相似度并最小化不同样本之间的相似度来学习图像的表达。具体来说如算法 2-1 所示，首先输入包括批次大小  $N$ 、常量  $\tau$ 、表示函数  $f$ 、投影函数  $g$  以及数据增强策略  $T$ 。对于每个样本  $\{x_k\}_{k=1}^N$ ，首先应用两种不同的数据增强方法  $t \sim T$  和  $t' \sim T$ ，生成增强后的图像  $\tilde{x}_{2k-1} = t(x_k)$  和  $\tilde{x}_{2k} = t'(x_k)$ ，并通过神经网络  $f$  提取出特征表示  $h_{2k-1} = f(\tilde{x}_{2k-1})$  和  $h_{2k} = f(\tilde{x}_{2k})$ 。接着，投影网络  $g$  将表示映射到投影空间，得到新的特征表示  $z_{2k-1} = g(h_{2k-1})$  和  $z_{2k} = g(h_{2k})$ 。然后，使用余弦相似度  $s_{i,j} = \frac{z_i^T z_j}{\|z_i\| \|z_j\|}$  来衡量表示  $z_i$  和  $z_j$  之间的相似度，以进一步计算对比损失  $\ell(i,j)$ ，其中正样本对和负样本对的相似度分别通过  $\ell(i,j) = -\log \frac{\exp(s_{i,j}/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp(s_{i,k}/\tau)}$  进行度量。损失函数  $\mathcal{L}$  被定义为所有样本对的双向平均对比损失： $\mathcal{L} = \frac{1}{2N} \sum_{k=1}^N [\ell(2k-1, 2k) + \ell(2k, 2k-1)]$ 。通过最小化此损失，网络参数  $f$  和  $g$  得到优化，从而提高特征表示的质量。最后，返回训练得到的表示编码器  $f(\cdot)$ ，并丢弃投影网络  $g(\cdot)$ 。如图 2-4 所示 MoCo<sup>[65]</sup> 是一种基于对比学习的无监督学习框架，其将对比学习看作是一个字典查询任务，通过维护一个动态更新的动量队列来存储负样本，从而减少计算负担。该图展示了三种不同的对比损失机制。第一种方法通过反向传播实时更新编码器从而令字典中的键（key）具有很高的



算法 2-1: SimCLR's main learning algorithm.

输入: batch size  $N$ , constant  $\tau$ , structure of  $f, g, T$ .

输出: return encoder network  $f(\cdot)$ , and throw away  $g(\cdot)$

```

1: for sampled minibatch  $\{x_k\}_{k=1}^N$  do
2:   for all  $k \in \{1, \dots, N\}$  do
3:     draw two augmentation functions  $t \sim T, t' \sim T$ 
4:      $\tilde{x}_{2k-1} = t(x_k)$  { first augmentation}
5:      $h_{2k-1} = f(\tilde{x}_{2k-1})$  { representation}
6:      $z_{2k-1} = g(h_{2k-1})$  { projection}
7:      $\tilde{x}_{2k} = t'(x_k)$  { second augmentation}
8:      $h_{2k} = f(\tilde{x}_{2k})$  { representation}
9:      $z_{2k} = g(h_{2k})$  { projection}
10:   end for
11:   for all  $i \in \{1, \dots, 2N\}$  and  $j \in \{1, \dots, 2N\}$  do
12:      $s_{i,j} = z_i^T z_j / (\|z_i\| \|z_j\|)$  { pairwise similarity}
13:   end for
14:   define  $\ell(i, j)$  as  $\ell(i, j) = -\log \frac{\exp(s_{i,j}/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp(s_{i,k}/\tau)}$ 
15:    $\mathcal{L} = \frac{1}{2N} \sum_{k=1}^N [\ell(2k-1, 2k) + \ell(2k, 2k-1)]$ 
16:   update networks  $f$  and  $g$  to minimize  $\mathcal{L}$ 
17: end for
    
```

特征一致性，但字典大小依赖于 Mini-batch 的大小，受到 GPU 内存的限制。第二种方法将键表示存储在内存库中，优化时从中抽取键，尽管内存库的内容不会通过反向传播更新，可以支持更大的字典大小，但是内存库里面的键可能导致过时问题。第三种方法则采用动量编码器，通过动量更新的方式生成键表示，同时维护一个队列作为动态字典，结合了内存库和端到端机制的优势。MoCo 通过这种方式提升了在大规模数据集上的效率，并有效地学习视觉表示，缩小了无监督和有监督表示学习之间的差距。此外，BYOL<sup>[66]</sup> 是一种自监督学习方法，设计了一种不需要负样本的对比学习框架。它通过创建目标网络 and 在线网络结合数据增强生成不同的视图来训练模型。BYOL 的核心思想是，通过最小化两个网络之间的表示距离来学习有用的特征，而不依赖负样本的存在，突破了传统对比学习中需要使用负样本的限制。在这里，本文使用对比学习范式来去设计了创新的多尺度对齐模块来去应用在多视图领域的下游任务上，分别是实例尺度对齐模块和原型尺度对齐模块。其中，相比于传统的对比学习范式，实例尺度对齐模块通过对齐每个样本实例的不同视图表示来挖掘视图之间潜藏的共性；而原型尺度对齐模块则在扩展边界模块的辅助下，通过利用同一标签实例之间的类别差异来寻找更加明确的决策边界，具体来说样本实例在特征空间中应该更加靠近其对应的类别中心同时远离其他类别的中心。通过应用对比学习范式，MAMC 取得了极具竞争力的效果。

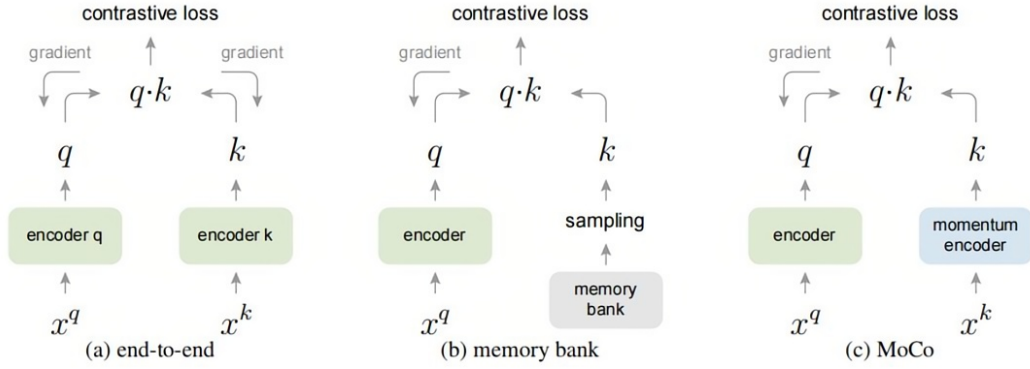


图 2-4 MoCo 对比学习框架示意图

## 2.3 模糊集理论

模糊集理论（Fuzzy Set Theory）是在数学经典集合理论的基础上，引入成员隶属度（Membership Degrees）的概念，用于对某一元素是否属于集合进行更加细粒度的定义<sup>[67]</sup>。宇宙集（Universe） $\mathcal{U}$  上的一个模糊集合（Fuzzy Set） $\mathcal{A}$  通过成员隶属函数  $f_{\mathcal{A}}(x)$  来表征，该函数为  $\mathcal{U}$  中的每个元素分配一个介于  $[0, 1]$  的实数，其中  $\mathcal{U}$  被定义为包含当前讨论或问题中所有考虑元素的集合，而  $f_{\mathcal{A}}(x)$  表示元素  $x$  在集合  $\mathcal{A}$  中的隶属度<sup>[68]</sup>。模糊集理论为不确定性集合的推理提供了强大的框架，但成员隶属函数的具体形式很大程度上取决于相关领域知识。因此在这里具体定义如下：一个函数  $f_{\mathcal{A}} : \mathcal{U} \rightarrow \mathcal{R}$  被称为成员隶属函数，这里  $\mathcal{R}$  为实数集（The Set Of Real Numbers）。同时一个包含宇宙集  $\mathcal{U}$  和成员隶属函数  $f_{\mathcal{A}}$  的元组  $(\mathcal{U}, f_{\mathcal{A}})$  被称为模糊集。基于上述定义，当  $L = [0, 1]$  时， $\mathcal{A}$  是模糊集。当成员隶属函数仅取 0 或 1 的值，即  $L = \{0, 1\}$  时， $\mathcal{A}$  退化为一个经典集合。

## 2.4 交叉熵损失函数

交叉熵损失函数（Cross-Entropy Loss）常被用于分类任务中，能够很好地对模型预测输出与真实分布之间的差异进行度量。在信息论中，熵（Entropy）最初用于度量一个离散随机变量所包含的不确定性。对于一个离散分布  $P$ ，其熵可定义为：

$$H(P) = - \sum_{x \in \mathcal{X}} P(x) \log P(x) \quad (2-11)$$

其中， $\mathcal{X}$  是所有可能取值的集合， $\log$  常默认为自然对数。当分布越均匀时，不确定性越大，熵值也越高。相对熵（或称 KL 散度）用于衡量两个分布  $P$  与  $Q$  之间的差异，当  $P$  和  $Q$  完全相同时， $D_{KL}(P \parallel Q) = 0$ 。该值越大，说明  $Q$  与  $P$  的差异越大，定义如下：

$$D_{KL}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} \quad (2-12)$$

在信息论中，交叉熵  $H(P, Q)$  的定义以及和 KL 散度之间的关系可以写为：

$$H(P, Q) = - \sum_{x \in \mathcal{X}} P(x) \log Q(x) \quad (2-13)$$

$$H(P, Q) = H(P) + D_{\text{KL}}(P \parallel Q) \quad (2-14)$$

如果将  $P$  当作“真实分布”，把  $Q$  当作“模型预测分布”，那么最小化交叉熵即是尽量让  $Q$  贴近  $P$ ，从而间接最小化 KL 散度，同时最小化交叉熵本质上等价于最大化对数似然（Log-Likelihood）。因此，对于二分类问题，交叉熵损失函数定义通常写作：

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (2-15)$$

其中， $N$  表示样本总数， $y_i \in \{0, 1\}$  表示第  $i$  个样本的真实标签， $\hat{y}_i \in (0, 1)$  表示模型对第  $i$  个样本属于正类的预测概率。在这里，当  $y_i = 1$ （样本为正类），损失项主要是  $-\log(\hat{y}_i)$ ，如果模型对正类的预测概率  $\hat{y}_i$  越大，则损失越小；当  $y_i = 0$ （样本为负类），损失项主要是  $-\log(1 - \hat{y}_i)$ ，如果模型对负类的预测概率  $1 - \hat{y}_i$  越大，则损失越小。

对于多分类任务，假设有  $C$  个类别，模型输出为一个概率分布  $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_C)$ ，其中  $\hat{y}_j$  表示预测为第  $j$  个类别的概率。真实标签可以用独热编码（One-Hot Encoding）表示为  $\mathbf{y} = (y_1, y_2, \dots, y_C)$ ，其中只有真实类别对应的向量为 1，其余为 0。

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{i,j} \log(\hat{y}_{i,j}) \quad (2-16)$$

其中， $N$  表示样本总数， $\hat{y}_{i,j}$  表示第  $i$  个样本预测为类别  $j$  的概率， $y_{i,j} \in \{0, 1\}$  表示第  $i$  个样本的真实类别。最后，在训练神经网络模型时，如果把交叉熵用作损失函数，模型会通过反向传播（Back-propagation, BP）来最小化该损失。具体来说，最小化交叉熵等价于最大化对数似然，从而使得预测值更贴近真实的概率分布。

## 2.5 本章小结

本章节主要涉及 MAMC 设计过程使用到的相关理论方法，首先在 2.1 节介绍了自编码器理论（2.1.1 小节和 2.1.2 小节）和其应用（2.1.3 小节）；其次在 2.2 节介绍了对比学习理论（2.2.1 小节）和其相关工作（2.2.2 小节）；然后在 2.3 节介绍了模糊集理论，最后在 2.4 节介绍了交叉熵损失函数。在 MAMC 中，本文利用自编码器来提取特征表示，同时使用对比学习范式来去设计了多尺度对齐模块，来解决特征异质性和信息冗余问题。

### 3. 基于多尺度对齐和扩展边界的增强多视图分类模型研究

本章主要介绍基于多尺度对齐和扩展边界的增强多视图分类模型框架 MAMC，该模型设计旨在解决多视图分类中的两个主要问题：特征异质性和信息冗余。MAMC 模型由两个关键模块：多尺度对齐模块和扩展边界模块组成。首先，多尺度对齐模块从实例和原型两个层次进行对齐，挖掘视图特征中存在的共性与差异，有效缓解特征异质性。其次，借助自适应扩展边界模块的辅助，原型尺度对齐模块能够动态扩展原始类别边界缩紧类内样本间距并拉大类间距离来对模糊样本进行处理，从而提升模型的泛化能力和性能表现。本章首先在3.1节介绍模型的整体结构和各个组成模块所实现的功能，主要包括特征提取器、多尺度对齐模块和联合分类器。然后在第3.2节详细介绍组成特征提取器的自编码器，接着在第3.3节和第3.4节分别说明基于实例尺度和基于原型尺度的对齐模块。此外，在第3.5节讲解模型分类预测策略以及在第3.6节中对构建模型过程中所涉及到的数学理论进行推导证明。最后在第3.7节中对本章内容进行了总结。

#### 3.1 模型整体结构

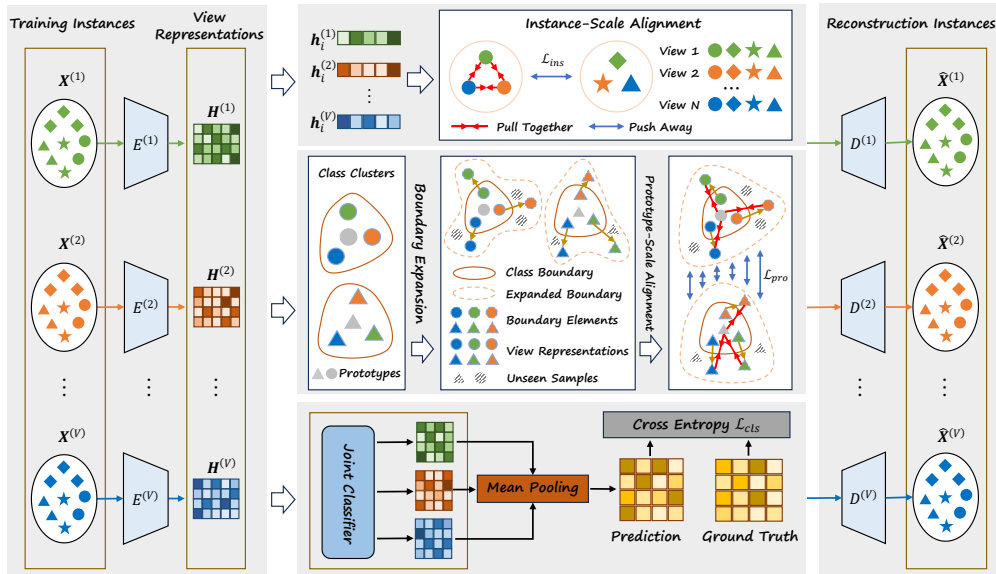


图 3-1 MVMC 结构示意图，主要包括特征提取器、多尺度对齐模块和联合分类器

如图 3-1 所示 MVMC 模型主要由三个部分组成：（1）基于视图特定的特征提取器（Representation Extractors）：多个基于视图特定的自编码器从原始的多视图数据  $\{X^{(v)}\}_{v=1}^V$  (Training Instances) 中抽取出特征表示  $\{H^{(v)}\}_{v=1}^V$ ；（2）多尺度对齐模块（Multi-Scale Alignment）：实例尺度对齐模块（Instance-Scale Alignment）对每个实例  $x_i$  的视图表示  $\{h_i^{(v)}\}_{v=1}^V$  进行对齐，其中图标的各种形状对应不同的实例，不同颜色的图标对应来自同一实例不同视图的特征表示；原型尺度对齐模块（Prototype-Scale Alignment）结

合自适应扩展边界模块 (Expanded-Boundary) 进行动态边界扩展, 并对具有相同标签的视图表示进行对齐; (3) 联合分类器 (Joint classifier): 所有视图表示被送入联合分类器以获得融合后的预测结果。整体上看, MAMC 模型最后基于原始输入数据  $\{X^{(v)}\}_{v=1}^V$  所获得的输出预测值为  $\hat{p}_i$ , 以下为 MAMC 的具体运行流程:

$$\{H^{(v)}\}_{v=1}^V = \text{Representation Extractors}(\{X^{(v)}\}_{v=1}^V), 1 \leq v \leq V \quad (3-1)$$

首先, 本文对特征表示进行提取, 公式3-1从原始数据  $\{X^{(v)}\}_{v=1}^V$  中抽取出特征表示  $\{H^{(v)}\}_{v=1}^V$ 。其中, 多视图数据集  $\mathcal{S} = \{(\{x_i^{(v)}\}_{v=1}^V, y_i) | 1 \leq i \leq N\}$  被定义为具有  $N$  个独立同分布的样本, 同时包括样本对应的  $V$  个视图以及标签。此外, 使用实值向量  $x_i^{(v)} \in \mathbb{R}^{D_v} (1 \leq v \leq V, i = 1, \dots, N)$  表示第  $i$  个样本在第  $v$  个视图上的特征向量, 其中  $D_v$  是第  $v$  个视图的特征维度。对于标签, 使用自然数  $y_i \in \mathbb{N}$  表示第  $i$  个样本的真实标签, 同时  $\hat{y}_i \in \{0, 1\}^K$  表示对应的独热编码 (One-hot),  $K$  代表了标签类别的数量。

$$\{\hat{H}^{(v)}\}_{v=1}^V = \text{Multi-Scale Alignment}(\{H^{(v)}\}_{v=1}^V) \quad (3-2)$$

公式3-2表示将特征表示  $\{H^{(v)}\}_{v=1}^V$  作为多尺度对齐模块的输入并对其进行更新。

$$\{\hat{H}_1^{(v)}\}_{v=1}^V = \text{Instance-Scale Alignment}(\{H^{(v)}\}_{v=1}^V, \tau_{ins}) \quad (3-3)$$

$$\{Z^{(v)}\}_{v=1}^V = \text{Expanded-Boundary}(\{H^{(v)}\}_{v=1}^V) \quad (3-4)$$

$$\{\hat{H}_2^{(v)}\}_{v=1}^V = \text{Prototype-Scale Alignment}(\{Z^{(v)}\}_{v=1}^V, \tau_{pro}) \quad (3-5)$$

$$\{\hat{H}^{(v)}\}_{v=1}^V = \text{Joint Alignment Optimization}(\{\hat{H}_1^{(v)}\}_{v=1}^V, \{\hat{H}_2^{(v)}\}_{v=1}^V) \quad (3-6)$$

在这里, 多尺度对齐模块可以进一步分解为实例尺度对齐模块 (公式3-3) 和原型尺度对齐模块 (公式3-5)。其中实例尺度对齐部分的输入是  $\{H^{(v)}\}_{v=1}^V$ , 在进行原型尺度对齐之前需要通过辅助的扩展边界模块 (公式3-4) 获得更新后的表示  $\{Z^{(v)}\}_{v=1}^V$  将其作为原型尺度对齐模块的输入,  $\tau_{ins}$  和  $\tau_{pro}$  分别为两个对齐模块的超参数, 最后不同尺度对齐模块的输出分别为  $\{\hat{H}_1^{(v)}\}_{v=1}^V$  和  $\{\hat{H}_2^{(v)}\}_{v=1}^V$ 。公式3-6展示了实例对齐尺度和原型对齐尺度模块对特征表示  $\{H^{(v)}\}_{v=1}^V$  进行的联合优化过程, 本质上是一种损失融合策略。

$$\hat{p}_i = \text{Joint classifier}(\{\hat{H}^{(v)}\}_{v=1}^V) \quad (3-7)$$

最后, 公式3-7展示了将多尺度对齐模块更新后获得的表示  $\{\hat{H}^{(v)}\}_{v=1}^V$  送入联合分类器中获得最后的模型预测值  $\hat{p}_i$ 。MAMC 的具体框架如算法 3-1 所示。具体实现方面, MAMC 的第一步是利用自编码器模块从原始输入数据  $\mathcal{S} = \{(\{x_i^{(v)}\}_{v=1}^V, y_i) | 1 \leq i \leq N\}$  中提取特征表示  $\{h_i^{(v)}\}_{v=1}^V$ , 该操作具体是通过逐层的 MLP 网络 (包括线性层、ReLU 激活函数和

## 算法 3-1: MAMC 整体流程

输入：多视图数据集： $\mathcal{S} = \{(\{\mathbf{x}_i^{(v)}\}_{v=1}^V, \mathbf{y}_i) | 1 \leq i \leq N\}$   $\{\{\mathbf{x}_i^{(v)}\}_{v=1}^V \in \mathbb{R}^{B \times D}, \mathbf{y}_i \in \mathbb{R}^{B \times k}\}$ ，其中  $B$  为每一批次的样本数， $D$  为样本的原始特征维度， $K$  为样本类别个数；MLP 线性层的设定维度  $\mathcal{H}$ ；超参数  $\tau_{ins}, \tau_{pro}, \alpha, \beta$ ；训练轮数  $T$

输出：模型预测值  $\hat{\mathbf{p}}_i$   $\{\hat{\mathbf{p}}_i \in \mathbb{R}^{B \times K}\}$

- 1: 原始的多视图数据  $\{\mathbf{x}_i^{(v)}\}_{v=1}^V$  首先进入自编码器模块来获得特征表示矩阵  $\{\mathbf{h}_i^{(v)}\}_{v=1}^V$
- 2: 对于编码器来说，每一层由 MLP 线性层、RELU 激活函数和 Dropout 层组成
- 3: 首先： $\{\mathbf{x}_i^{(v)'}\}_{v=1}^V = \text{Dropout}\left(\text{ReLU}\left(\mathbf{W}^{(1)}\{\mathbf{x}_i^{(v)}\}_{v=1}^V + \mathbf{b}^{(1)}\right), 0.15\right) \{\{\mathbf{x}_i^{(v)'}\}_{v=1}^V \in \mathbb{R}^{B \times 1.4 \times H}\}$
- 4: 其次： $\{\mathbf{x}_i^{(v)''}\}_{v=1}^V = \text{Dropout}\left(\text{ReLU}\left(\mathbf{W}^{(2)}\{\mathbf{x}_i^{(v)'}\}_{v=1}^V + \mathbf{b}^{(2)}\right), 0.15\right) \{\{\mathbf{x}_i^{(v)''}\}_{v=1}^V \in \mathbb{R}^{B \times 1.2 \times H}\}$
- 5: 最后： $\{\mathbf{h}_i^{(v)}\}_{v=1}^V = \text{Dropout}\left(\text{ReLU}\left(\mathbf{W}^{(3)}\{\mathbf{x}_i^{(v)''}\}_{v=1}^V + \mathbf{b}^{(3)}\right), 0.15\right) \{\{\mathbf{h}_i^{(v)}\}_{v=1}^V \in \mathbb{R}^{B \times H}\}$
- 6: 特征提取器总体过程表示为： $\{\mathbf{h}_i^{(v)}\}_{v=1}^V = \text{Representation Extractors}(\{\mathbf{x}_i^{(v)}\}_{v=1}^V)$
- 7: 对于解码器来说，同样每一层由 MLP 线性层、RELU 激活函数和 Dropout 层组成
- 8: 首先： $\{\hat{\mathbf{x}}_i^{(v)''}\}_{v=1}^V = \text{Dropout}\left(\text{ReLU}\left(\mathbf{W}^{(4)}\{\mathbf{h}_i^{(v)}\}_{v=1}^V + \mathbf{b}^{(4)}\right), 0.1\right) \{\{\hat{\mathbf{x}}_i^{(v)''}\}_{v=1}^V \in \mathbb{R}^{B \times 0.6 \times D}\}$
- 9: 其次： $\{\hat{\mathbf{x}}_i^{(v)'}\}_{v=1}^V = \text{Dropout}\left(\text{ReLU}\left(\mathbf{W}^{(5)}\{\hat{\mathbf{x}}_i^{(v)''}\}_{v=1}^V + \mathbf{b}^{(5)}\right), 0.1\right) \{\{\hat{\mathbf{x}}_i^{(v)'}\}_{v=1}^V \in \mathbb{R}^{B \times 0.8 \times D}\}$
- 10: 最后： $\{\hat{\mathbf{x}}_i^{(v)}\}_{v=1}^V = \text{Dropout}\left(\text{ReLU}\left(\mathbf{W}^{(6)}\{\hat{\mathbf{x}}_i^{(v)'}\}_{v=1}^V + \mathbf{b}^{(6)}\right), 0.1\right) \{\{\hat{\mathbf{x}}_i^{(v)}\}_{v=1}^V \in \mathbb{R}^{B \times D}\}$
- 11: **for**  $epoch = 1$  到  $T$  **do**
- 12: 通过  $\{\hat{\mathbf{x}}_i^{(v)}\}_{v=1}^V$  和  $\{\mathbf{x}_i^{(v)}\}_{v=1}^V$  计算自编码器重构损失  $\mathcal{L}_{rec}$
- 13: 实例尺度对齐模块挖掘视图特征共性，缓解特征异质性
- 14: 通过  $\{\mathbf{h}_i^{(v)}\}_{v=1}^V$  和超参数  $\tau_{ins}$  计算实例尺度损失  $\mathcal{L}_{ins}$
- 15: 扩展边界模块进行动态边界扩展以实现清晰的分类边界
- 16:  $\{\mathbf{z}_i^{(v)}\}_{v=1}^V = \text{Expanded-Boundary}(\{\mathbf{h}_i^{(v)}\}_{v=1}^V) \{\{\mathbf{z}_i^{(v)}\}_{v=1}^V \in \mathbb{R}^{B \times H}\}$
- 17: 原型尺度对齐模块进一步挖掘视图特征共性，从而更好地缓解特征异质性
- 18: 通过  $\{\mathbf{z}_i^{(v)}\}_{v=1}^V$  和超参数  $\tau_{pro}$  计算原型尺度损失  $\mathcal{L}_{pro}$
- 19: 将  $\{\mathbf{h}_i^{(v)}\}_{v=1}^V$  输入到联合分类器中获得模型预测值  $\hat{\mathbf{p}}_i$   $\{\hat{\mathbf{p}}_i \in \mathbb{R}^{B \times K}\}$
- 20: 通过预测值  $\hat{\mathbf{p}}_i$  和训练集真实标签  $\mathbf{y}_i$  计算交叉熵损失  $\mathcal{L}_{cls}$
- 21: 最后，联合优化总体损失  $\mathcal{L}_{all} = \mathcal{L}_{rec} + \mathcal{L}_{cls} + \alpha \mathcal{L}_{ins} + \beta \mathcal{L}_{pro}$
- 22: **end for**

Dropout 层) 完成的。自编码器由两个主要部分组成：编码器 (Encoder) 和解码器 (Decoder)。它们分别实现了从原始数据到隐藏表示的映射 (编码)，以及从隐藏表示恢复到原始数据的映射 (解码)。在这里提取到的特征表示  $\{\mathbf{h}_i^{(v)}\}_{v=1}^V$  是通过编码器产生的，而解码器的目的是将特征表示  $\{\mathbf{h}_i^{(v)}\}_{v=1}^V$  映射回到原始数据空间  $\{\mathbf{x}_i^{(v)}\}_{v=1}^V$ 。自编码器的训练目标是 minimized 重建误差，因此通过重建后的特征表示  $\{\hat{\mathbf{x}}_i^{(v)}\}_{v=1}^V$  和原始特征表示  $\{\mathbf{x}_i^{(v)}\}_{v=1}^V$  可以计算需要优化的重建损失  $\mathcal{L}_{rec}$ 。实例尺度对齐模块将同一实例的不同视图表示视作正样本对，不同实例的特征表示视作负样本对，通过应用对比学习范式在特征空间中拉近正样本对距离和拉远负样本对距离捕获多视图数据中的特征共性，在这

里将  $\{\mathbf{h}_i^{(v)}\}_{v=1}^V$  和超参数  $\tau_{ins}$  作为实例尺度对齐模块的输入来计算实例尺度损失  $\mathcal{L}_{ins}$ 。扩展边界模块利用特征表示  $\{\mathbf{h}_i^{(v)}\}_{v=1}^V$  进行动态边界扩展以形成新的样本类别中心缓解信息冗余问题，通过将扩展边界模块集成到原型尺度对齐模块中，后者进一步将样本实例表示和其所对应的类别中心表示视作正样本对，样本实例表示和其不属于的类别中心视作负样本对，通过对比学习进一步挖掘视图特征共性，在这里通过扩展边界模块的输出  $\{\mathbf{z}_i^{(v)}\}_{v=1}^V$  和超参数  $\tau_{pro}$  计算原型尺度损失  $\mathcal{L}_{pro}$ 。最后，将  $\{\mathbf{h}_i^{(v)}\}_{v=1}^V$  输入到参数共享的联合分类器中获得模型预测值  $\hat{\mathbf{p}}_i$ ，通过  $\hat{\mathbf{p}}_i$  和真实标签  $\mathbf{y}_i$  计算分类任务交叉熵损失  $\mathcal{L}_{cls}$ 。在这里，基于本文应用的损失叠加融合策略，对总体损失  $\mathcal{L}_{all} = \mathcal{L}_{rec} + \mathcal{L}_{cls} + \alpha\mathcal{L}_{ins} + \beta\mathcal{L}_{pro}$  ( $\alpha, \beta$  为权重超参数) 进行联合优化，通过反向传播更新模型参数完成训练过程。

本章的后续内容将详细介绍构建 MAMC 模型的模块组件，包括特征提取器、多尺度对齐模块和扩展边界模块，并探究它们如何有效地利用多视图数据来完成分类任务。此外，MAMC 的设计动机一方面是希望解决多视图数据特征异质性问题，研究如何将不同的视图表示有效地融合在一起贡献于最后的下游任务，成为该领域最新且效果最好的多视图 SOTA 分类模型，另一方面则是致力于通过结合模糊集理论进行动态扩展边界来缓解多视图特征中存在的信息冗余问题。正是解决了多视图领域之前受到重点关注的问题，MAMC 才可以在多个领域中的公开数据集上取得优异的性能表现，并在和同领域其他先进模型的比较中获得领先地位，成为该领域内的目前提出的最先进的模型之一。因此本章除了介绍各模块组件的具体实现细节之外，还会详细分析 MAMC 是如何解决多视图领域存在的问题，从而在多视图分类任务上取得良好性能。

### 3.2 基于视图特定的自编码器特征提取

多视图数据中的异构性隐藏了许多潜在问题，例如视图特征维度的不一致和存在噪声等。为了解决这些问题，通过采用基于多个视图特定的自编码器，可以在相同的维度内生成对应的视图特征表示以便进行进一步的融合和处理操作<sup>[69]</sup>。自编码器的核心思想是通过编码（Encoding）和解码（Decoding）过程将输入数据映射到一个低维的隐空间，再从该隐空间重建出尽量接近原始输入的数据。编码器的结构通常由一个或多个神经网络层组成，其任务是提取输入数据中的重要特征，生成一个低维的表示（通常称为潜在变量或隐向量）。首先，编码器将  $\mathcal{X}$  作为输入数据， $\mathcal{F}$  作为映射函数，并输出潜在表示  $\mathcal{Z} = \mathcal{F}_{\text{Encoding}}(\mathcal{X})$ 。其次，解码器将隐空间中的潜在表示  $\mathcal{Z}$  映射回原始数据空间  $\hat{\mathcal{X}} = \mathcal{F}_{\text{Decoding}}(\mathcal{Z})$ 。在这里，本文将原始的多视图数据输入编码器中以获得它们对应的特征嵌入向量。其中，编码器和解码器由多层感知机（MLPs）网络构成，每一层的网络结构是由线性层（Linear Layer）、ReLU 激活层（ReLU Activation Layer）和 Dropout 层（Dropout Layer）来去堆叠而成的。线性层，也叫全连接层（Fully Connected Layer, FC），是一个基础的神经网络层，作用是将输入的特征向量通过一个线性变换映射到一个新的空间；ReLU（Rectified Linear Unit，修正线性单元）是最常用的激活函数之一，它的作

## 算法 3-2: 自编码器整体流程

输入： 原始特征  $\mathcal{X}$ ；每一批次的样本数  $\mathcal{B}$ ；原始特征维度  $\mathcal{D}$ ；线性层的设定维度  $\mathcal{H}$ 。

输出： 编码器：隐层表示  $\mathcal{Z}$ ；解码器：重建原始特征  $\hat{\mathcal{X}}$

- 1: 编码器：每一层由 MLP 线性层、RELU 激活函数和 Dropout 层组成
- 2: 首先：  $\mathcal{X}' = \text{Dropout}(\text{ReLU}(\mathbf{W}^{(1)}\mathcal{X} + \mathbf{b}^{(1)}), 0.15) \{ \mathcal{X}' \in \mathbb{R}^{B \times (1.4 \times H)} \}$
- 3: 其次：  $\mathcal{X}'' = \text{Dropout}(\text{ReLU}(\mathbf{W}^{(2)}\mathcal{X}' + \mathbf{b}^{(2)}), 0.15) \{ \mathcal{X}'' \in \mathbb{R}^{B \times (1.2 \times H)} \}$
- 4: 最后：  $\mathcal{Z} = \text{Dropout}(\text{ReLU}(\mathbf{W}^{(3)}\mathcal{X}'' + \mathbf{b}^{(3)}), 0.15) \{ \mathcal{Z} \in \mathbb{R}^{B \times H} \}$
- 5: 解码器：每一层由 MLP 线性层、RELU 激活函数和 Dropout 层组成
- 6: 首先：  $\hat{\mathcal{X}}'' = \text{Dropout}(\text{ReLU}(\mathbf{W}^{(4)}\mathcal{Z} + \mathbf{b}^{(4)}), 0.1) \{ \hat{\mathcal{X}}'' \in \mathbb{R}^{B \times (0.6 \times D)} \}$
- 7: 其次：  $\hat{\mathcal{X}}' = \text{Dropout}(\text{ReLU}(\mathbf{W}^{(5)}\hat{\mathcal{X}}'' + \mathbf{b}^{(5)}), 0.1) \{ \hat{\mathcal{X}}' \in \mathbb{R}^{B \times (0.8 \times D)} \}$
- 8: 最后：  $\hat{\mathcal{X}} = \text{Dropout}(\text{ReLU}(\mathbf{W}^{(6)}\hat{\mathcal{X}}' + \mathbf{b}^{(6)}), 0.1) \{ \hat{\mathcal{X}} \in \mathbb{R}^{B \times D} \}$

用是引入非线性，使得神经网络能够学习到更复杂的特征；Dropout 是一种正则化技术，目的是通过冻结一定比例和数量的神经元减少神经网络的过拟合问题，达到正则化效果，以提高模型的泛化能力。此外，如算法 3-2 所示编码器和解码器的维度分别设置为  $\{\mathcal{D}, 1.4 \times \mathcal{H}, 1.2 \times \mathcal{H}, \mathcal{H}\}$  和  $\{\mathcal{H}, 0.6 \times \mathcal{D}, 0.8 \times \mathcal{D}, \mathcal{D}\}$  ( $\mathcal{D}$  为输入的原始特征维度， $\mathcal{H}$  为线性层所设置的维度)。对于第  $v$  个视图数据  $X^{(v)}$ ，本文应用一个视图特定的编码器  $E^{(v)}$  来生成视图特征表示矩阵  $H^{(v)}$ ，并使用解码器  $D^{(v)}$  来恢复原始的视图特征。因此，自编码器的重构过程由以下公式进行定义：

$$\mathcal{L}_{rec} = \frac{1}{V} \sum_{v=1}^V \left\| X^{(v)} - D^{(v)} \left( E^{(v)}(X^{(v)}) \right) \right\|_F^2 \quad (3-8)$$

自编码器通过将不同视图的数据映射到共享的隐空间，可以帮助模型学习到每个视图的共同特征。这使得模型可以捕获不同视图之间的公共性语义信息，同时避免单个视图中的噪声影响。同时，自编码器的设计本质上具有去噪和特征压缩的作用，在多视图任务中，多个视角的数据可能包含冗余信息，自编码器通过学习低维隐空间，能够在一定程度上缓解信息冗余问题并保留有用的信息。多视图数据通常包含来自不同来源（例如图像、文本和视频等）的信息，而每种视角可能提供不同的上下文信息。通过为每个视图分配对应的特定编码器，在保留不同视图独立性的同时可以有效提取各个视图的特征，去除冗余的噪声并将它们放缩到同一个特征维度，帮助模型在隐空间中有效融合各个视角的信息，从而形成更加丰富和完整的样本特征表示。

### 3.3 基于实例层面的对齐模块

解决多视图数据特征融合困难的关键之一在于挖掘视图间的共性。这是因为不同视图中的特征通常在表达相同对象时会存在差异，而这些差异会导致特征融合的困难。



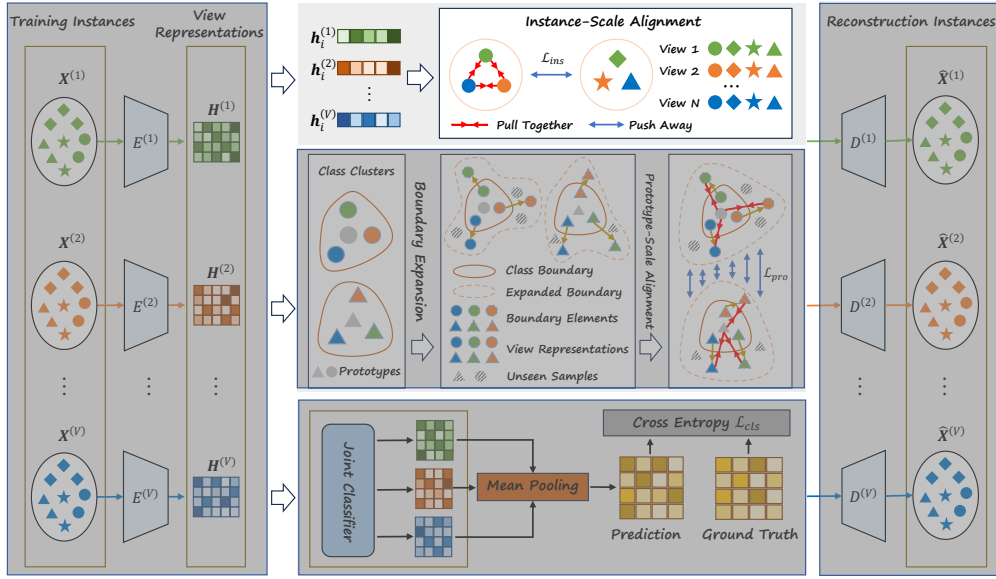


图 3-2 实例尺度对齐模块示意图

对比学习通过在特征空间中将正样本对拉近、负样本对推远来利用数据之间潜藏的共性<sup>[70]</sup>。实例级对齐模块结合对比学习，将每个实例和这些实例自身所对应的所有视图表示视为正样本对，其他实例对应的视图表示视为负样本对。具体来说，从视图特定编码器中抽取的同一样本对应的不同视图特征表示  $\{h_i^{(v)}\}_{v=1}^V$  被用于构建正样本对，而其他实例对应视图表示则用于构建负样本对。因此，实例尺度损失函数定义为：

$$\mathcal{L}_{ins} = -\frac{1}{NV} \sum_{v=1}^V \sum_{u \neq v} \sum_{i=1}^N \log \frac{e^{-Dis(h_i^{(v)}, h_i^{(u)}) \cdot \tau_{ins}}}{e^{-Dis(h_i^{(v)}, h_i^{(u)}) \cdot \tau_{ins}} + \sum_{r=u, v} \sum_{j \neq i} e^{-Dis(h_i^{(v)}, h_j^{(r)}) \cdot \tau_{ins}}} \quad (3-9)$$

其中， $\mathcal{L}_{ins}$  用于衡量预测值与实际值之间的差异，训练时模型将会最小化该损失。 $N$  代表样本的总数量， $V$  表示不同视图类别的总数。同时  $h_i^{(v)}$  和  $h_i^{(u)}$  分别为第  $i$  个样本在不同视图下的特征表示，用于构建正样本对； $h_i^{(v)}$  和  $h_j^{(r)}$  分别为不同样本实例对应的视图特征表示，用于构建负样本对。 $Dis(\cdot, \cdot)$  是一个距离度量函数，用于衡量两个样本之间的相似度。 $\tau_{ins}$  是一个温度系数，用于调节距离的权重，从而控制模型的学习过程。此外，本文通过对数概率来去表示损失项，其中分子代表正样本对的相似度，分母包含正样本对和所有负样本对的相似度之和，优化的目标是通过最小化损失值，使得正样本对的相似度占比最大化。例如，当分子增大（正样本对相似度高）、分母减小（负样本对相似度低）时，分数趋近于 1，损失值降低；反之若分子减小（正样本对相似度低）、分母增大（负样本对相似度高），分数趋近于 0，损失值显著升高。在训练过程中，损失函数的值会随着模型的优化而变化，并最终收敛到一个稳定的最小值。通过对正样本视图表示进行对齐，MAMC 可以提取相似样本中的关键公共语义信息，缓解特征异质性问题；通过判别负样本视图表示，不同样本间的差异逐渐显现，从而获得清晰的决策边界。

### 3.4 基于原型层面对齐模块

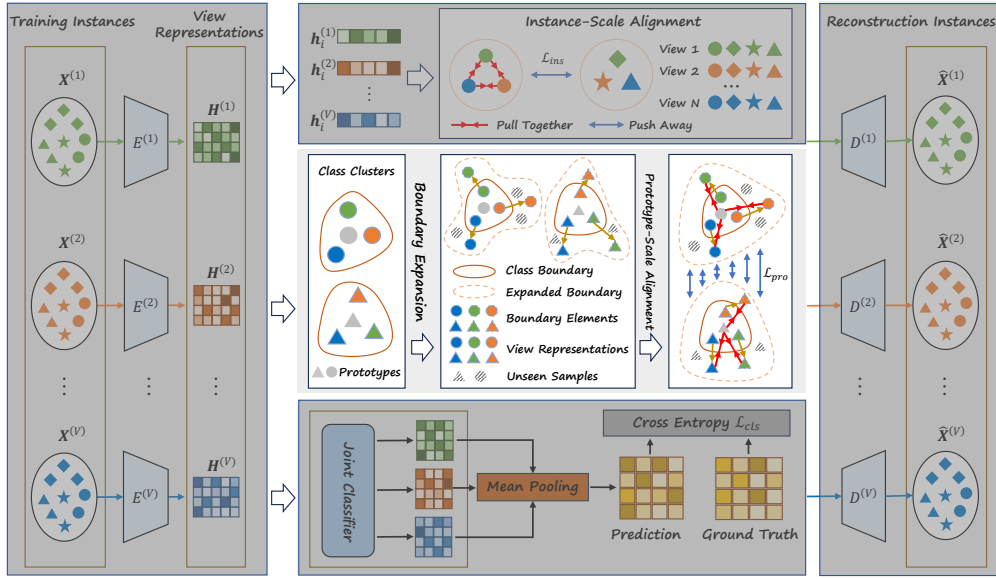


图 3-3 原型尺度对齐模块示意图

与实例尺度对齐模块相比，原型尺度对齐模块的目标是帮助模型利用类间的差异来划分明确的决策边界，实现这一目标的核心是构建扩展的边界。在这一章节中，首先本文会在3.4.1小节详细解释 MAMC 是如何动态进行扩展边界的，同时探讨扩展边界模块是如何缓解多视图数据中存在的信息冗余问题；其次在3.4.2小节，将会介绍原型尺度对齐模块的详细内容：包括新类原型的出现和原型尺度对比损失函数；最后在3.4.3小节，将讨论如何选取3.4.1小节中提到的组合操作符算子并对其进行理论分析。

#### 3.4.1 扩展边界

传统的原型学习主要对属于同一类别下的所有样本特征进行均值处理，以得到该类别的一个原型，同时原型所代表的平均向量表示也被称为类别中心。但这种方法由于采取了平均化操作导致数据被进行了过度的平滑，使得当面对类别内部信息多样化或者类别间差异明显情况时模型无法有效利用类别之间的差异信息。为了解决多视图数据中的信息冗余问题以实现清晰的分类边界，本文提出了一种自适应扩展边界来辅助原型尺度对齐模块。在这里，本文汇总了具有相同标签的所有视图表示来构造类别原型并进一步中心化视图表示。对于第  $k$  类训练数据，原型构造和中心化过程定义为：

$$\mathbf{c}_k = \frac{1}{V|Y_k|} \sum_{v=1}^V \sum_{y_i=k} \mathbf{h}_i^{(v)} \quad (3-10)$$

在这里，对所有标签为  $k$  的样本实例，将它们对应的特征表示  $\mathbf{h}_i^{(v)}$  进行求和并取均值，得到该类别的中心向量  $\mathbf{c}_k$ 。其中， $Y_k$  是包含标签为  $k$  的训练数据的集合， $|\cdot|$  表示集合

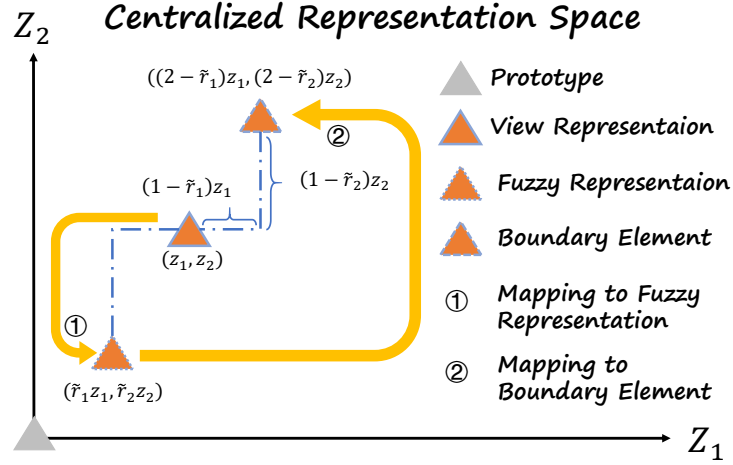


图 3-4 扩展边界过程示意图

的基数， $V$  为视图数，该平均操作可以有效滤除原始数据中存在的高频噪声。

$$\mathbf{z}_i^{(v)} = \mathbf{h}_i^{(v)} - \mathbf{c}_k, \quad v = 1, \dots, V \quad (3-11)$$

此外， $\mathbf{z}_i^{(v)}$  是样本  $i$  第  $v$  个视图中的中心化表示向量，原始特征表示  $\mathbf{h}_i^{(v)}$  通过减去类别中心  $\mathbf{c}_k$  去掉了自身所包含的全局特征信息，使得中心化处理后得到的特征表示  $\mathbf{z}_i^{(v)}$  更加关注样本与类别中心的偏差，而这些偏差信息往往是决定样本分类的重要因素。例如，假设特征表示  $\mathbf{h}_i^{(v)}$  在原始嵌入空间中靠近类别中心  $\mathbf{c}_k$ ，但此时  $\mathbf{h}_i^{(v)}$  所处的位置也可能是  $\mathbf{c}_k$  所对应类别与其他类别的重叠区域，则  $\mathbf{h}_i^{(v)}$  有可能靠近多个不同的类别中心，从而处于交界模糊地带无法清晰决策。去中心化操作相当于将样本移动到以  $\mathbf{c}_k$  为原点的新坐标系中，使得模型更关注样本与类别中心的内部差异，而不是绝对位置。扩展边界的关键是构造具有隶属度函数的模糊表示空间，在这个过程中通常使用高斯隶属函数<sup>[71]</sup>。较大的隶属函数数量可以产生更复杂、灵活的决策边界，但存在过拟合的风险；较小的隶属函数数量会使得决策边界更加平滑，但可能无法准确捕捉复杂的类别分布，因此合理数量的隶属函数设置将赋予扩展边界更强的自适应能力。为此，当采用  $L$  个高斯隶属函数，第  $l$  个高斯隶属函数的隶属度将在这里被定义为：

$$\mathbf{r}_{i,j,l}^{(v)} = \exp \left\{ -\frac{\left( \mathbf{z}_{i,j}^{(v)} - \mathbf{m}_{j,l} \right)^2}{2\delta_{j,l}^2} \right\}, \quad l = 1, 2, \dots, L, \quad j = 1, 2, \dots, d \quad (3-12)$$

其中  $d$  是视图表示的维度， $\mathbf{z}_{i,j}^{(v)}$  是第  $v$  个视图中第  $i$  个样本中心化表示向量的第  $j$  个元素， $\mathbf{m}_{j,l}$  和  $\delta_{j,l}$  分别是元素对应的第  $l$  个高斯隶属函数的可训练均值和标准差。这些高斯函数定义了模糊边界的分布和性质，使得模型可以适应复杂的特征分布从而在特征空间中通过动态扩展获得合理的决策分类边界。

$$\tilde{r}_{i,j}^{(v)} = \text{Comb} \left\{ r_{i,j,l}^{(v)} \right\}, l = 1, 2, \dots, L \quad (3-13)$$

Comb 是融合隶属度的组合操作（如选取最大值、最小值或均值等），其具体定义和选取将在3.4.3小节中进行讨论，同时  $\tilde{r}_{i,j}^{(v)}$  是组成隶属度矩阵  $\tilde{R}^{(v)}$  的元素。

$$\tilde{Z}^{(v)} = \tilde{R}^{(v)} \odot Z^{(v)} \quad (3-14)$$

其中， $\odot$  是 Hadamard 乘积（逐元素相乘），公式3-14展示了通过 Hadamard 乘积将隶属度矩阵  $\tilde{R}^{(v)}$  和去中心化特征表示  $Z^{(v)}$  相结合生成模糊化表示矩阵  $\tilde{Z}^{(v)}$ 。该操作为每个样本特征分配对应的模糊权重，使得模糊化特征表示  $\tilde{Z}^{(v)}$  能够增强模型对模糊边界的感知能力，从而提升对复杂类别分布的建模能力。根据隶属函数的性质，隶属度矩阵  $\tilde{R}^{(v)}$  的每个值均限制在  $[0, 1]$  范围内，这意味着清晰表示可以被视为模糊表示的一种特殊情况。基于此，本文将清晰表示和模糊表示均视为正样本，使得模型能够在训练过程中有效地利用所有类型的样本以增强其对复杂类别分布的感知能力。此外，本文假设坐标差异反映的偏差是可容忍的，因为正样本之间的差异通常不会显著违反公共性语义信息。也就是说，即使在某些情况下，样本的特征稍微有些偏差，只要这些偏差不会破坏类别的整体特点，它们就不会对分类决策造成显著影响。受此启发，如图 3-4 对原始类别边界进行动态扩展来获得新的扩展边界元素：

$$\bar{Z}^{(v)} = \tilde{Z}^{(v)} + 2 \left( 1 - \tilde{R}^{(v)} \right) \odot Z^{(v)} = \left( 2 - \tilde{R}^{(v)} \right) \odot Z^{(v)} \quad (3-15)$$

从数学角度理解，扩展边界的元素通过模糊表示的坐标变换生成。其中， $\bar{Z}^{(v)}$  代表了经过公式3-15生成的扩展边界元素，本质上是进行扩展边界后新的模糊表示，公式中的 1 和 2 分别是两个实值矩阵，它们的每个元素均为 1 和 2。其中，新生成的扩展边界元素  $\bar{Z}^{(v)}$  继承了原始模糊表示  $\tilde{Z}^{(v)}$  的两个期望属性：语义正性和自适应性。语义正性指的是获得的边界元素应该被视为原始清晰表示的正样本集合，这将有助于识别远离类别中心的模糊样本。这种做法增强了模型对类别重叠和模糊边界的适应能力，从而使模型能在复杂的分类任务中表现得更加精确和鲁棒。同时自适应性指的是扩展边界基于可学习的隶属函数，能够自适应地调整以将模糊样本纳入训练过程。这两个特性旨在捕获每个边界内的几乎所有样本，以充分利用类间的差异帮助模型学习以获得更好的分类性能。

### 3.4.2 损失函数

原始类别边界的扩展导致由新生成的扩展边界元素  $\bar{Z}^{(v)}$  组成的类原型  $\tilde{C}$  出现，这引发了原始特征表示组成的  $C$  和  $\tilde{C}$  之间的相互作用以达到平衡状态。本文通过线性组合对两者进行混合，以获得融合后的原型  $\bar{C}$ ，该过程定义如下：

$$\bar{C} = \eta C + (1 - \eta) \tilde{C} \quad (3-16)$$

其中，每个原型向量  $C$  通过公式3-10计算，在此之后通过将每个类别的原始类原型  $C$  添加到更新后的扩展边界表示矩阵  $\bar{Z}$  中，将  $\bar{Z}$  恢复为模糊表示矩阵  $\bar{H}$ ， $\tilde{C}$  通过对每个类别所有的模糊表示矩阵  $\bar{H}$  进行平均计算得到。在这里，原始类原型  $C$  的添加可以帮助新生成的扩展边界类原型  $\tilde{C}$  保留原始类别的核心信息，避免过度偏离真实类别中心。 $\eta$  是用于协调  $C$  和  $\tilde{C}$  之间的平衡系数，以生成最后的融合原型  $\bar{C}$ 。

同时，在这里本文希望最大化每个类别的特征值分布范围，这等价于最大化每个类别内的方差，这样可以使得每个类别内部存在更多的差异性从而增强模型对未知数据的泛化能力。同时，本文期望在类别之间实现更好的分离性，并在每个类别内保持紧凑性，这旨在最大化不同类别之间的距离，同时最小化每个类别内的方差。考虑到这些因素，本文将损失函数形式定义为一个极小极大博弈：

$$\mathcal{L}_{pro} = \min_{\mathcal{P}+\mathcal{V}} \max_{\mathcal{V}} (\mathcal{P} + \mathcal{V}) \quad (3-17)$$

其中， $\mathcal{P}$  被定义为用于度量不同类别之间的距离，而  $\mathcal{V}$  被定义为在泛化未见数据时用于缩紧类别内样本的度量。这里的  $\min_{\mathcal{P}+\mathcal{V}}$  表示最小化  $\mathcal{P} + \mathcal{V}$  的整体目标， $\max_{\mathcal{V}}$  是针对  $\mathcal{V}$  的操作。在极小极大博弈中  $\max_{\mathcal{V}}$  表示本文希望通过调整  $\mathcal{V}$  来增大类别内部的方差，从而使每个类别在特征空间内拥有更广泛的特征分布以提高模型对未知数据的适应能力，然后在这种情况下最小化  $\mathcal{P} + \mathcal{V}$ 。通过这种极大极小博弈的方式，模型能够在训练过程中自动寻找一个平衡点，使得类别间的分离度和类别内的紧凑性达到最优状态。这样，模型不仅能够很好地区分不同类别，还能在每个类别内部保持样本的紧凑性，从而增强对未见数据的泛化能力。 $\mathcal{P}$  和  $\mathcal{V}$  的具体定义如下：

$$\begin{cases} \mathcal{P} = -\frac{1}{NKV} \sum_{v=1}^V \sum_{k=1}^K \sum_{n=1}^N \left\| \mathbf{h}_n^{(v)} - \bar{\mathbf{c}}_k \right\|_2^2 \\ \mathcal{V} = \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k \right\|_2^2 \end{cases} \quad (3-18)$$

其中， $N$  为样本数量， $V$  为视图数量， $K$  为类别数量。同时上面的公式定义了两个重要的概念  $\mathcal{P}$  和  $\mathcal{V}$ ，分别用于度量类别之间的分离度和类别内的紧凑性。具体来说， $\mathcal{P}$  衡量了不同类别之间的距离，通过计算每个样本与其对应类别原型的欧几里得距离来度量类别间的分离度。另一方面， $\mathcal{V}$  衡量了类别内部的紧凑性，计算每个类别内所有样本与该类别原型之间的欧几里得距离，从而度量类别内样本的聚集性。在这里  $\bar{\mathbf{c}}_k$  表示第  $k$  类的融合原型向量， $\mathbf{h}_n^{(v)}$  和  $\bar{\mathbf{h}}_i^{(v)}$  分别表示第  $v$  视图中的第  $n$  个和第  $i$  个样本恢复后的扩展边界元素特征表示。通过最小化  $\mathcal{P} + \mathcal{V}$ ，模型在训练过程中能够实现类别之间的良好分离，同时保持每个类别内的样本紧凑，从而增强模型的泛化能力。为了方便起见，本文

令  $Q = P + V$ ，并在模型中提出了  $Q$  的一个上界，期望通过优化其上界来进一步优化  $Q$ 。在这里，上界  $Q_{pro}$  的定义如公式3-19所示。其中， $\bar{c}_k$  是第  $k$  类的融合原型向量， $\bar{c}_j$  表示所有的类别原型， $\bar{h}_i^{(v)}$  是第  $v$  个视图中第  $i$  个样本恢复的扩展边界元素， $\tau_{pro}$  是温度系数，用于控制类别的紧凑性。通常，较大的  $\tau_{pro}$  会使类别更加紧凑，能够减少类别内部

$$Q_{pro} = -\frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \log \frac{e^{-\tau_{pro} \|\bar{h}_i^{(v)} - \bar{c}_k\|_2^2}}{\sum_j e^{-\tau_{pro} \|\bar{h}_i^{(v)} - \bar{c}_j\|_2^2}} \quad (3-19)$$

的散度。该损失函数通过最大化样本与其真实类别原型的相似性，同时最小化样本与其他类别原型的相似性，优化类别边界的区分性。最后，本文在3.6.1小节中提供了  $Q_{pro}$  是  $Q$  的上界的理论证明，并对其作用进行分析。

在原型尺度对齐损失中，本文整合了所有视图表示作为类别样本，以全面充分地提取各个视图中存在的关键信息。通过这种方式，所有视图信息被聚合在一起，共同形成了类别间的差异信息，从而限定决策边界。此外，对比学习评估了原始特征  $\mathbf{x}_i$  与恢复的扩展边界元素  $\bar{h}_i^{(v)}$  之间的互信息<sup>[62]</sup>，其定义为：

$$I\left(\mathbf{x}_i, \left\{\bar{h}_i^{(v)}\right\}_{v=1}^V\right) \geq \log(\mathcal{N}) - Q_{pro} \quad (3-20)$$

其中， $I(\cdot)$  用于计算互信息，旨在评估特征表示之间的共享信息量，在这里较大的互信息表示恢复的扩展边界元素能够很好地保留原始数据的信息。 $\mathcal{N}$  是对比学习中构成负样本对的样本数量。该不等式表明，更多的负样本数量以及对损失函数  $Q_{pro}$  进行优化，能够提高不等式的下界，保留更多从原始数据中提取的完整信息，从而实现更好的对齐性能。通常，整合所有视图表示进行模型训练可以灵活地倍增负样本的数量。

### 3.4.3 组合操作符的确定

在公式3-12中的组合操作符 **Comb** 对于扩展边界至关重要。一方面，**Comb** 操作符会影响扩展边界的自适应能力；另一方面，**Comb** 操作符可以控制类别方差，以便对未见数据进行泛化。前者通过可学习的隶属函数实现，后者则受到 **Comb** 操作符自身的确定性所影响。当采用最小池化（Min Pooling）作为 **Comb** 操作符时，每个类别的整体方差达到最大化（证明见3.6.2小节）。因此，当理想的 **Comb** 操作符被定义为隶属函数的最小值时，每个类别的方差被最大化，这意味着扩展边界可以泛化更多的潜在样本。最终，原型级对齐损失  $\mathcal{L}_{pro}$ （见公式3-17）被简化为公式3-21形式。

$$\mathcal{L}_{pro} = \min -\frac{1}{NV} \sum_{v=1}^V \sum_{i=1}^N \log \frac{e^{-\tau_{pro} \|\bar{h}_i^{(v)} - \bar{c}_i\|_2^2}}{\sum_j e^{-\tau_{pro} \|\bar{h}_i^{(v)} - \bar{c}_j\|_2^2}} \quad (3-21)$$

在这里， $\mathcal{L}_{pro}$  是原型尺度对齐损失函数，用于优化类内紧致性和类间分离性。 $\bar{\mathbf{h}}_i^{(v)}$  为恢复的扩展边界元素对应的特征表示， $\bar{\mathbf{c}}_i$  为样本实例对应的正确类别中心， $\bar{\mathbf{c}}_j$  表示所有的类别中心（包括正确类别和其他类别中心），优化目标是使得  $\mathcal{L}_{pro}$  的损失尽可能降低。样本实例特征表示应当靠近其对应的正确类别中心，同时远离其他的类别中心。

### 3.5 分类预测

在分类过程中，本文采用了一种参数共享策略的联合分类器，该分类器具有以下优点：首先共享参数能够使得模型同时接收来自所有视图的信息，有助于为多视图数据构建一个稳健的分类器；其次多尺度对齐模块挖掘了视图间的共性和类别间的差异，从而构建清晰的决策边界，减少了对拥有大量参数的复杂分类器需求。模型的预测通过对联合分类器的输出进行平均得到，具体定义如下：

$$\hat{\mathbf{p}}_i = \frac{1}{V} \sum_{v=1}^V \mathbf{p}_i^{(v)} = \frac{1}{V} \sum_{v=1}^V g(\mathbf{h}_i^{(v)}) \quad (3-22)$$

其中， $g(\cdot)$  是一个逻辑（Logistic Regression）分类器， $\mathbf{p}_i^{(v)}$  是第  $v$  个视图中  $g(\cdot)$  的输出， $\hat{\mathbf{p}}_i$  是最终的预测值。在训练过程中，本文使用交叉熵损失函数（Cross-Entropy Loss Function）来帮助模型计算分类损失，相关公式定义如下：

$$\mathcal{L}_{cls} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K \hat{\mathbf{y}}_{ij} \log \hat{\mathbf{p}}_{ij} \quad (3-23)$$

此外，本文综合考虑了多个损失函数，提出了一个总损失函数  $\mathcal{L}_{all}$ 。该函数将重构损失、分类损失、实例级对齐损失和原型级对齐损失结合在一起，以实现多方面的优化目标。重构损失用于确保模型能够还原输入数据的原始结构；分类损失则是通过最大化正确分类的概率来引导模型学习准确的决策边界；实例级对齐损失帮助对齐不同视图中的实例特征，以提取视图中的公共语义信息；原型级对齐损失则关注于从类别差异角度利用对比学习范式，确保类别间的分离度和类别内部的紧凑性。通过这些损失的联合优化，模型能够有效地学习到不同视图间的共性并消除决策边界附近的模糊区域，从而成功缓解多视图领域特征异质性和信息冗余所带来的问题。

$$\mathcal{L}_{all} = \mathcal{L}_{rec} + \mathcal{L}_{cls} + \alpha \mathcal{L}_{ins} + \beta \mathcal{L}_{pro} \quad (3-24)$$

其中， $\alpha$  和  $\beta$  是惩罚系数，在算法3-1中提供了 MAMC 算法的整体联合优化训练流程。综上所述，MAMC 模型通过多尺度对齐模块和自适应扩展边界模块的协同作用，结合多目标联合优化策略，形成了一个完整的多视图分类解决方案。该框架不仅在理论上具有创新性，还在实践中展现出优异的性能，能够有效处理多视图数据中的关键挑战，为多视图学习领域提供了一种系统性且高效的技术路径。



### 3.6 数学证明

#### 3.6.1 证明： $Q_{pro}$ 是 $Q$ 的上界

首先，在这里为了增强泛化性，设  $\tau_{pro} = 1$ ，对公式3-19进行化简变换得到：

$$\begin{aligned}
 Q_{pro} &= -\frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \log \frac{e^{-\tau_{pro} \|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k\|_2^2}}{\sum_j e^{-\tau_{pro} \|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2}} \\
 &= \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \log \frac{\sum_j e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2}}{e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k\|_2^2}} \\
 &= \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left( \log \sum_j e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2} - \log e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k\|_2^2} \right) \\
 &= \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left( \log \sum_j e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2} + \|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k\|_2^2 \right)
 \end{aligned} \tag{3-25}$$

然后，引入常数  $K$  和期望表示  $\mathbb{E}[\cdot]$  将原始表达式简化为一个更易于计算的形式：

$$Q_{pro} \approx \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left( \log K \mathbb{E} \left[ e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2} \right] + \|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k\|_2^2 \right) \tag{3-26}$$

在原始表达式中，内部求和运算  $\sum_j$  将遍历所有类别  $j$ ，同时在这里总类别数为  $K$ 。在近似过程中，本文假设各类别对目标函数的贡献度是均匀的。基于这一假设，原式中的求和项  $\sum_j e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2}$  可近似等效为  $K$  个独立项期望值的线性组合。具体而言，式中  $\mathbb{E} \left[ e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2} \right]$  表示对  $e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2}$  的期望值，该期望值本质上反映了所有类别平均贡献的统计特性，并可以简化复杂式子的计算。基于上述近似，进一步化简可得：

$$Q_{pro} = \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left( \log \mathbb{E} \left[ e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2} \right] + \|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k\|_2^2 + \log K \right) \tag{3-27}$$

将常数  $K$  忽略以得到新的损失函数  $Q'_{pro}$ ：

$$Q'_{pro} = \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left( \log \mathbb{E} \left[ e^{-\|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j\|_2^2} \right] + \|\bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k\|_2^2 \right) \tag{3-28}$$

Jensen 不等式 (Jensen's inequality)<sup>[72]</sup> 是数学中一个关于凸函数的重要不等式，其核心是：对于一个凸函数，函数值的平均值总是大于或等于函数在平均值点的函数值。具体来说，假设  $f(x)$  是一个凸函数，且  $X$  是一个随机变量，则 Jensen 不等式可以表示为：

$$f(E[X]) \leq E[f(X)] \tag{3-29}$$

其中  $E[X]$  表示随机变量  $X$  的期望值， $E[f(X)]$  表示函数  $f(X)$  的期望值。



因此，根据 Jensen 不等式本文可以推出：

$$\exp \left\{ \mathbb{E} \left[ -\left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j \right\|_2^2 \right] \right\} \leq \mathbb{E} \left[ \exp \left\{ -\left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j \right\|_2^2 \right\} \right] \quad (3-30)$$

由于对数函数是单调递增的，联立公式3-28和3-29最终得到：

$$\mathcal{Q}'_{pro} \geq \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left( \mathbb{E} \left[ -\left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j \right\|_2^2 \right] + \left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k \right\|_2^2 \right) \approx \mathcal{Q} \quad (3-31)$$

### 3.6.2 证明：选取 Min Pooling 为 Comb 操作符时类别方差最大化

首先，为了增强泛化性，同样地设  $\tau_{pro} = 1$ ，对公式 3-19 进行化简变换得到：

$$\begin{aligned} \mathcal{Q}_{pro} &= -\frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \log \frac{e^{-\left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k \right\|_2^2}}{\sum_j e^{-\left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j \right\|_2^2}} \\ &= \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left( \log \sum_j e^{-\left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j \right\|_2^2} + \left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k \right\|_2^2 \right) \\ &= \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left( \log \sum_j e^{-\left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j \right\|_2^2} \right) + \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left( \left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k \right\|_2^2 \right) \end{aligned} \quad (3-32)$$

设  $Var(\bar{\mathbf{c}}_k)$  表示第  $k$  类簇的方差。因此，进一步变换得到：

$$\mathcal{V} = \frac{1}{NV} \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_k \right\|_2^2 = \sum_{k=1}^K Var(\bar{\mathbf{c}}_k) \quad (3-33)$$

$$\mathcal{Q}_{pro} \approx \frac{1}{NV} \sum_{v=1}^V \sum_{i=1}^N \left( \log \sum_j e^{-\left\| \bar{\mathbf{h}}_i^{(v)} - \bar{\mathbf{c}}_j \right\|_2^2} \right) + \sum_{k=1}^K Var(\bar{\mathbf{c}}_k) \quad (3-34)$$

关于3.4.2小节中应用的公式3-16 ( $\bar{\mathbf{C}} = \eta \mathbf{C} + (1 - \eta) \tilde{\mathbf{C}}$ )，在实际使用中通常会更加权衡清晰的原型表示  $\mathbf{C}$ ，于是将  $\eta$  设置为近似于 1，然后得到：

$$\lim_{\eta \rightarrow 1} \mathcal{V} = \lim_{\eta \rightarrow 1} \sum_{k=1}^K Var(\bar{\mathbf{c}}_k) = \sum_{k=1}^K Var(\mathbf{c}_k) \quad (3-35)$$

因此，当应用中心化表示构造扩展边界时，每个类簇内的最大总体方差形式化为：

$$\begin{aligned} \max_{\eta \rightarrow 1} \lim_{\eta \rightarrow 1} \mathcal{V} &= \max \sum_{k=1}^K Var(\mathbf{c}_k) \\ &= \max \sum_{v=1}^V \sum_{k=1}^K \left| \sum_{y_i=k} \left\| \bar{\mathbf{h}}_i^{(v)} - \mathbf{c}_k \right\|_2^2 \right| \end{aligned} \quad (3-36)$$

继续将公式3-11和公式3-15代入公式3-36可以得到：

$$\begin{aligned}
 \max_{\eta \rightarrow 1} \lim \mathcal{V} &= \max \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left\| \bar{\mathbf{z}}_i^{(v)} \right\|_2^2 \\
 &= \max \sum_{v=1}^V \sum_{k=1}^K \sum_{y_i=k} \left\| \left( 2 - \tilde{\mathbf{r}}_i^{(v)} \right) \odot \mathbf{z}_i^{(v)} \right\|_2^2 \\
 &= \max \sum_{v=1}^V \sum_{i=1}^N \left\| \left( 2 - \tilde{\mathbf{r}}_i^{(v)} \right) \odot \mathbf{z}_i^{(v)} \right\|_2^2
 \end{aligned} \tag{3-37}$$

显然，对于每个表示，当  $\tilde{\mathbf{r}}_i^{(v)}$  的每个元素减小时， $\mathcal{V}$  增大，这与下面的式子相符合：

$$\left\| \left( 2 - \left\{ \tilde{\mathbf{r}}_i^{(v)} \right\}_{min} \right) \odot \mathbf{z}_i^{(v)} \right\|_2^2 \geq \left\| \left( 2 - \left\{ \tilde{\mathbf{r}}_i^{(v)} \right\}_{other} \right) \odot \mathbf{z}_i^{(v)} \right\|_2^2 \tag{3-38}$$

其中  $\left\{ \tilde{\mathbf{r}}_i^{(v)} \right\}_{min}$  和  $\left\{ \tilde{\mathbf{r}}_i^{(v)} \right\}_{other}$  是隶属度向量，同时是当使用 Min Pooling 和其他 Comb 操作符（例如 Mean Pooling 或者 Max Pooling）时公式3-13的输出，最后本文获得：

$$\sum_{v=1}^V \sum_{i=1}^N \left\| \left( 2 - \left\{ \tilde{\mathbf{r}}_i^{(v)} \right\}_{min} \right) \odot \mathbf{z}_i^{(v)} \right\|_2^2 \geq \sum_{v=1}^V \sum_{i=1}^N \left\| \left( 2 - \left\{ \tilde{\mathbf{r}}_i^{(v)} \right\}_{other} \right) \odot \mathbf{z}_i^{(v)} \right\|_2^2 \tag{3-39}$$

因此，最后获得的结论是使用 Min Pooling 可以最大限度地提高每个类的总体方差。

### 3.7 本章小结

本章介绍了 MAMC 模型的整体结构以及各个组成模块的基本原理和相关知识。首先在3.1节从整体上对 MAMC 模型进行了概述，介绍了特征提取器、多尺度对齐模块和联合分类器。然后在3.2节详细介绍了自编码器框架，并对其如何通过编码器和解码器将输入的原始多视图数据转换为可以在神经网络中流通的特征嵌入向量进行了介绍。同时在3.3节和3.4节中分别详细说明了如何基于实例尺度和原型尺度来去构建对比学习范式以实现清晰的分类边界。其中对于3.4节（原型对齐模块），又进一步细分成了3.4.1小节（扩展边界理论介绍），3.4.2小节（损失函数的构建）以及3.4.3小节（组合操作符的确定）。在3.5节中，继续介绍了模型所使用的联合分类器，同时将前面所介绍的模块全部串联了起来以获得最后需要优化的完整损失函数。最后在3.6节中补充了构建 MAMC 模块相关的理论分析和数学证明，以增强模型可解释性。

## 4. 实验结果与分析

本章主要介绍 MAMC 实验设置与实验结果分析。首先在4.1节介绍了多视图领域广泛使用的八个公开数据集，并4.2节说明了本文实验中用于对比的七个多视图领域先进模型。然后在4.3节展示了实验过程使用的评价指标，包括分类任务指标和复杂性分析指标。其次分别在4.4节和4.5节展示了实验环境和汇总的实验分类结果表格，并对它们进行分析以说明 MAMC 在分类任务上所取得的优越性能表现。同时在4.6节进行了消融实验与分析以说明多尺度对齐模块的效果和操作符选取对实验结果的影响。最后，从4.7节至4.11节依次进行了超参分析实验、噪声实验、可视化分析实验、收敛性分析实验和复杂度分析实验，并在4.12节中对整个实验部分进行总结。

### 4.1 数据集介绍

表 4-1 数据集统计信息，包括实例数量、类别数量、视图数量和视图维度

数据集	实例数量	类别数量	视图数量	视图维度
HandWritten	2,000	10	6	240/76/216/47/64/6
Scene15	4,485	15	3	20/59/40
PIE	680	68	3	484/256/279
CCV	6,773	20	3	20/20/20
Animal	11,673	20	4	2,689/2,000/2,001/2,000
100Leaves	1,600	100	3	60/60/60
Hdigit	10,000	10	2	784/256
YouTubeFace	101,499	31	5	64/64/512/647/838

在多视图学习领域，多个数据集为本文提供了不同视图下的图像或视频数据，帮助训练和评估模型在多个角度和特征上的表现。实验中，本文在八个公开的多视图数据集上对模型的性能进行了评估，包括 HandWritten<sup>[73]</sup>、Scene15<sup>[55]</sup>、PIE<sup>[74]</sup>、CCV<sup>[75]</sup>、Animal<sup>[76]</sup>、100Leaves<sup>[73]</sup>、Hdigit<sup>[77]</sup>和 YouTubeFace<sup>[77]</sup>。以下是各个数据集的介绍：

**HandWritten 数据集<sup>1</sup>：**HandWritten 包含 2,000 个手写数字样本，涵盖了数字 0 到 9，并提供了颜色直方图和 SIFT 特征等多种公共特征，适用于手写数字识别任务。

**Scene15 数据集<sup>2</sup>：**Scene15 包含 15 类自然场景下的图像，如高速公路、城市、高楼大厦等，每个类别有 241 到 374 张图像，用于测试和训练。

**PIE 数据集<sup>3</sup>：**PIE 是面部识别的经典数据集，包含超过 750,000 张图像，涵盖 337

<sup>1</sup><http://archive.ics.uci.edu/ml/datasets/Multiple+Features>

<sup>2</sup>[http://www-cvr.ai.uiuc.edu/ponce\\_grp/data/scene15/](http://www-cvr.ai.uiuc.edu/ponce_grp/data/scene15/)

<sup>3</sup><https://www.cs.cmu.edu/afs/cs/project/PIE/>

个不同个体，提供 15 个视图和 19 种光照条件，在这里使用了其中的子集。

**CCV 数据集<sup>4</sup>**：CCV 包含 9,317 个来自网络的视频，涉及 20 个语义类别，包括棒球、游行等事件，海滩场景和猫等物体。

**Animal 数据集<sup>5</sup>**：Animal 包括 20 个动物类别的图像，帮助训练图像分类模型。

**100Leaves 数据集<sup>6</sup>**：100Leaves 包含 1,600 个样本和 100 种植物，以纹理直方图、细尺度边缘和形状描述符三种特征表示每个植物样本，适合用于植物识别和分类研究。

**Hdigit 数据集<sup>7</sup>**：Hdigit 是结合了 MNIST 和 USPS 手写数字的多视图数据。

**YoutubeFace 数据集<sup>8</sup>**：YoutubeFace 是一个包含来自 YouTube 视频库中 5 个不同视图的面部图像 RGB 数据集，适用于面部识别和与面部相关的计算机视觉任务。

所有这些数据集为多视图学习提供了丰富的训练数据，促进了该领域的发展和应用。表 4-1 展示了本文所使用的数据集统计信息，包括实例数量、类别数量、视图数量和视图维度。对于每个数据集，本文将 80% 的样本用于训练，其余样本用于测试。同时为了获得可靠的结果，在划分数据集时进行了乱序处理。

## 4.2 对比模型介绍

为了展示所提出模型的有效性，本文选择了七种最新的先进多视图学习方法进行比较。选择这些方法的原因在于它们代表了当前在多视图学习领域中的领先水平，并且涵盖了不同的技术路线和解决方案。这些模型分别是 mmdynamics<sup>[48]</sup>、ETMC<sup>[15]</sup>、UMDL<sup>[78]</sup>、PDMF<sup>[53]</sup>、IPMVSC<sup>[52]</sup>、MV-HFMD<sup>[47]</sup>和 RCML<sup>[50]</sup>。相关模型的介绍如下：

**mmdynamics**：mmdynamics 是首个针对不同样本同时建模特征和模态变化，以提供可信融合的多模态分类方法。具体来说，该模型引入了稀疏门控机制以捕捉每个模态内特征的信息变化，并采用真实类别概率来评估每个模态的分类信度，在医学分类数据集上和其他模型相比取得了优越的性能。

**ETMC**：ETMC 提出了一种新型的多视图分类算法，称为可信多视图分类，通过动态集成不同视图中的特征信息，为多视图学习提供了一种新的范式。在面对噪声、损坏或分布外数据时，模型在准确性、鲁棒性和可信度上有着优越的表现。

**UMDL**：UMDL 提出了一种稀疏多视图融合范式，以灵活地显式捕捉多视图数据的互补性，有效解决视图特征维度不平衡问题。此外，该模型通过平衡相似性图来构建正负样本对，并采用对比学习以自监督的方式训练。在实验结果上，UMDL 优于基准方法，可以很好地应用于下游分类和分割任务。

**PDMF**：PDMF 通过引入渐进式学习策略克服了现有方法在处理缺失视图数据时的缺陷，以充分利用多视图数据的互补性。通过在一个合成的玩具数据集和 4 个真实世界

<sup>4</sup><http://www.ee.columbia.edu/ln/dvmm/CCV/>

<sup>5</sup><https://www.cs.ucf.edu/~xmzhang/datasets/>

<sup>6</sup><https://archive.ics.uci.edu/ml/datasets/One-hundred+plant+species+leaves+data+set>

<sup>7</sup><https://archive.ics.uci.edu/ml/index.php>

<sup>8</sup><https://www.cs.tau.ac.il/wolf/ytfaces/>

数据集上的实验结果表明，PDMF 优于现有的最先进方法。

**IPMVSC** : IPMVSC 提出了一种新型的多视图 Takagi–Sugeno–Kang (TSK) 模糊系统框架，以处理多视图数据的分类问题。该模型通过区分性的锚点和子空间图聚类策略分别为每个视图建模潜在数据分布，捕捉多视图数据结构。实验表明，与不同模型相比该方法在多个公开基准数据集上性能表现良好。

**MV-HFMD** : MV-HFMD 将结合了卷积神经网络 CNN 和 Transformer 模型的整体架构用于多视图分类，它引入了一种新的融合方案和互信息蒸馏方法，同时仅增加了少部分的额外参数。实验表明，该方法在多视图分类任务中表现优越，具备良好的泛化能力。

**RCML** : RCML 提出了一个可信多视图学习问题，解决了现实应用中多视图数据可能包含低质量样本实例（同一实例在其不同的视图特征中呈现出冲突的信息）问题。该模型从实验和理论上证明提出的策略可以准确地建模多视图共性和视图可靠性关系，在 6 个数据集上呈现了良好的实验结果。

通过与这些先进模型的比较，本文能够全面评估所提出模型的优势和改进之处，从而进一步验证其在实际应用中的有效性和可靠性。

### 4.3 评价指标

本小节介绍 MAMC 模型所使用的评价指标，包括分类任务指标和复杂性分析指标。

#### 4.3.1 分类任务指标

分类评价指标包括准确率 (Accuracy, Acc)、精确率 (Precision)、召回率 (Recall) 和宏 F1 值 (Macro-F1)，其定义和计算公式如下：

**准确率**：准确率是分类模型中最直观的评估指标之一，它表示模型预测正确的样本占有所有样本的比例。准确率可以表示为：

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3-40)$$

**精确率**：精确率用于衡量模型预测为正类的样本中有多少是正确的。它关注的是在所有被预测为正类的样本中，真正为正类的比例。精确率定义为：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3-41)$$

**召回率**：召回率衡量的是所有真实为正类的样本中，有多少被模型成功预测为正类。它反映了模型识别正类样本的能力。召回率定义为：

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3-42)$$

其中， $TP$  (True Positive) 表示被正确分类为正类的样本数， $TN$  (True Negative) 表示被正确分类为负类的样本数， $FP$  (False Positive) 表示被错误分类为正类的负样本数， $FN$

(False Negative) 表示被错误分类为负类的正样本数。

**宏 F1 值:** F1 值是精确率和召回率的调和平均数, 旨在综合评估模型的精确性与全面性。宏 F1 值是对所有类别 F1 值的算术平均, 适用于多类分类任务, 尤其是类别不平衡的情况。该评价指标对每个类别给予同等的权重, 因此能够避免模型对某一类偏向的影响。在这里, F1 值和宏 F1 值的定义分别如下:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3-43)$$

$$\text{Macro-F1} = \frac{1}{C} \sum_{i=1}^C F1_i \quad (3-44)$$

其中,  $C$  是类别数,  $F1_i$  是类别  $i$  的 F1 值。宏 F1 值通过对各类别的 F1 值取均值, 使得每个类别都能均衡地影响最终的评估结果。

#### 4.3.2 复杂性分析指标

**MACs:** Multiply-Accumulate Operations (乘加运算) 是衡量模型计算复杂度的一个重要指标, 指的是模型在进行前向传播或推理时所执行的乘加操作的总次数。在神经网络中, 尤其是在卷积神经网络 (CNN) 中, 每次卷积操作都会进行大量的乘加运算, 因此 MACs 对于评估神经网络的计算量非常重要。具体而言, MACs 衡量的是每个层 (例如卷积层、全连接层等) 中所有乘加操作的总和。

**Running Time:** Running Time (运行时间) 是指模型在进行推理或训练时所消耗的时间。它是衡量模型效率的关键指标, 通常用来评估模型在实际应用中的响应速度和实时性。对于深度学习模型, 尤其是在嵌入式设备、移动设备或其他计算资源受限的环境中, 运行时间是一个非常重要的考虑因素。

**Parameters:** Parameters (参数量) 指的是模型中需要训练的参数的总数。在神经网络中, 参数主要指的是权重和偏置。对于卷积神经网络 (CNN), 每个卷积核的权重和偏置都需要作为参数进行训练; 对于全连接层, 每个连接的权重也是模型的参数。参数量是衡量模型存储需求和训练复杂度的一个重要指标。更大的参数量通常意味着模型拥有更强的表达能力, 同时也需要更多的存储空间。

#### 4.4 实验设置

PyTorch 是一个由 Facebook 公司开源的深度学习框架, 广泛应用于机器学习研究和工业应用中。随机梯度下降 (SGD, Stochastic Gradient Descent) 是一种优化算法, 广泛应用于训练机器学习模型, 尤其是深度学习模型。MAMC 模型使用 PyTorch 实现, 并采用 SGD 优化器进行训练。自编码器模块中的编码器和解码器的维度分别设置为  $\{\mathcal{D}, 1.4 \times 512, 1.2 \times 512, 512\}$  和  $\{512, 0.6 \times \mathcal{D}, 0.8 \times \mathcal{D}, \mathcal{D}\}$  ( $\mathcal{D}$  为样本的原始特征维

项目	详情
GPU	NVIDIA GeForce RTX 3090 (24GB)
CPU	Intel(R) Xeon(R) Gold 6148 CPU @ 2.40GHz
操作系统	Ubuntu 24.04 LTS
编程语言	Python 3.8
CUDA 版本	12.2
推理框架	Pytorch
torch 版本	2.4.1+cu121
torchaudio 版本	2.4.1+cu121
torchvision 版本	0.19.1+cu121
管理环境	Anaconda

表 4-2 实验环境配置（硬件层和软件层）

度），以提取良好的特征嵌入表示。隶属函数的数量  $L$  设置为 5，赋予扩展边界模块强大的适应能力。学习率从组合  $\{1e^{-3}, 3e^{-3}, 5e^{-3}, 1e^{-2}\}$  中选择，系数  $\alpha$  和  $\beta$  从组合  $\{0.1, 0.25, 0.5, 0.75, 1\}$  中选择。温度系数  $\tau_{ins}$  和  $\tau_{pro}$  分别从组合  $\{0.5, 0.75, 1.0, 1.5, 2.0\}$  和组合  $\{0.5, 0.75, 1.0, 1.2, 1.5\}$  中选择，更加详细的超参数选取需要通过实验和交叉验证来找到最佳的组合。表 4-2 展示了实验硬件与软件环境信息，其中硬件部分包括 NVIDIA GeForce RTX 3090 显卡（24GB）和 Intel Xeon Gold 6148 CPU（2.40GHz）。操作系统为 Ubuntu 24.04 LTS，使用的编程语言为 Python 3.8，CUDA 版本为 12.2。深度学习推理框架为 PyTorch，Torch 版本为 2.4.1+cu121，torchaudio 版本为 2.4.1+cu121，torchvision 版本为 0.19.1+cu121。整个环境通过 Anaconda 管理工具进行管理。

## 4.5 分类结果

表4-3展示了本文的模型与其他基线模型的分类结果，并采用准确率（**Accuracy, Acc**）、精确率（**Precision**）、召回率（**Recall**）和宏 F1 值（**Macro-F1**）作为评价指标。其中，最佳结果以粗体显示，“OOM”表示模型出现内存溢出错误。观察图表数据得到，MAMC 模型在和其他七个先进模型在八个多视图领域公开数据集和四个评价指标上比较取得了全面的领先。同时，可以得出以下重要结论：

(1) 本文提出的 MAMC 模型在所有比较方法中展现出非凡的优越性，特别是在 Scene15 数据集上。与现有方法相比，本文的模型能够理想地对齐所有视图信息，同时学习令人满意的决策边界，这对于分类任务至关重要。

(2) 对于所有模型都能取得较好结果的数据集（例如 HandWritten、100Leaves 和 Hdigit 数据集），良好的决策边界较容易确定。然而，本文所提出的模型进一步提升了性能，这表明决策边界附近的样本模糊性得到了进一步缓解。

表 4-3 实验分类结果汇总 (%)；最佳结果以粗体显示；“OOM”表示模型出现内存溢出错误

Datasets	Metrics	mmdynamics	ETMC	UMDL	PDMF	IPMVSC	MV-HFMD	RCML	Ours
HandWritten	Acc	98.5±0.5	96.8±0.7	97.7±0.5	98.2±0.7	98.7±0.4	98.2±0.6	97.1±0.9	<b>98.9±0.4</b>
	Precision	98.5±0.5	96.8±0.7	97.8±0.5	98.3±0.7	98.7±0.4	98.2±0.5	97.1±0.9	<b>98.9±0.4</b>
	Recall	98.5±0.5	96.9±0.7	97.7±0.5	98.3±0.7	98.7±0.4	98.2±0.6	97.2±0.9	<b>98.9±0.4</b>
	Macro F1	98.5±0.5	96.8±0.7	97.7±0.5	98.2±0.7	98.7±0.4	98.1±0.5	97.1±0.9	<b>98.9±0.4</b>
Scene15	Acc	62.0±2.2	66.5±1.8	63.3±0.4	67.8±1.1	71.8±2.1	80.7±1.2	70.0±1.0	<b>81.5±0.7</b>
	Precision	57.3±2.8	66.6±2.5	61.2±0.7	65.8±2.8	73.8±0.7	80.6±1.6	70.6±1.3	<b>81.0±0.6</b>
	Recall	60.1±2.2	65.3±1.6	62.7±0.5	65.9±1.1	70.5±0.8	80.3±1.3	68.9±1.3	<b>80.8±0.7</b>
	Macro F1	57.6±2.5	62.8±1.8	61.1±0.7	63.0±1.4	69.1±0.9	80.0±1.4	67.1±1.1	<b>80.5±0.7</b>
PIE	Acc	72.5±3.7	90.5±2.3	71.1±3.3	88.7±3.0	91.9±2.0	86.8±2.4	91.8±2.9	<b>93.1±1.7</b>
	Precision	70.6±4.0	88.0±3.8	74.6±3.9	87.6±3.8	81.7±2.1	87.7±2.7	91.1±3.4	<b>92.8±2.1</b>
	Recall	70.6±3.7	89.6±3.6	71.1±3.3	87.3±4.9	82.1±1.8	88.9±2.4	91.0±3.5	<b>93.6±1.8</b>
	Macro F1	67.4±3.8	87.4±3.8	69.3±3.5	85.7±4.5	80.5±2.5	83.8±6.1	89.8±3.9	<b>92.1±2.0</b>
CCV	Acc	29.6±1.1	42.6±1.4	36.2±1.5	50.5±1.4	42.6±1.8	53.4±0.9	42.4±1.7	<b>54.0±1.1</b>
	Precision	22.8±2.8	41.4±2.6	35.5±1.3	48.5±1.3	48.3±2.0	51.1±1.1	42.3±2.2	<b>53.0±1.3</b>
	Recall	23.2±1.0	37.1±1.4	33.7±1.4	45.4±1.3	35.5±1.3	50.5±0.8	37.0±1.4	<b>50.6±1.0</b>
	Macro F1	20.1±1.1	36.0±1.6	33.1±1.4	45.1±1.5	34.3±1.5	50.4±0.9	35.9±1.5	<b>51.0±0.9</b>
Animal	Acc	56.7±1.4	56.6±0.8	34.0±1.1	57.0±0.7	39.5±1.9	59.6±0.7	56.8±1.2	<b>60.2±1.1</b>
	Precision	52.4±2.2	53.3±3.2	33.7±0.9	49.5±0.6	37.3±3.9	56.5±0.6	54.5±2.5	<b>56.7±1.3</b>
	Recall	49.9±1.3	49.7±1.0	30.4±0.8	47.2±0.7	33.3±2.5	53.7±0.4	49.5±1.2	<b>53.8±0.8</b>
	Macro F1	50.4±1.4	49.7±1.1	30.4±0.7	46.7±0.6	33.3±2.7	54.0±0.5	49.6±1.3	<b>54.1±0.8</b>
100Leaves	Acc	93.5±1.5	90.8±2.1	98.4±0.8	97.7±0.7	66.1±4.1	98.3±0.2	88.6±1.5	<b>98.5±1.1</b>
	Precision	93.6±1.2	90.5±2.5	<b>98.7±0.8</b>	97.7±0.9	70.2±3.4	98.4±0.0	88.7±2.0	98.5±1.0
	Recall	93.9±1.1	91.3±2.4	98.4±0.9	98.0±1.0	69.0±2.0	98.9±0.2	89.4±1.4	<b>99.0±0.7</b>
	Macro F1	92.7±1.3	89.3±2.7	98.3±0.9	97.5±1.0	64.0±2.7	98.4±0.1	87.0±1.7	<b>98.5±1.0</b>
Hdigit	Acc	99.6±0.1	90.8±2.1	98.0±0.2	99.4±0.2	97.8±0.3	84.1±2.0	98.3±0.3	<b>99.8±0.1</b>
	Precision	99.6±0.1	98.4±0.2	98.0±0.2	99.4±0.2	97.8±0.3	88.9±1.7	98.3±0.3	<b>99.8±0.1</b>
	Recall	99.6±0.1	98.4±0.2	98.0±0.2	99.4±0.2	97.9±0.3	87.5±2.0	98.3±0.3	<b>99.8±0.1</b>
	Macro F1	99.6±0.1	98.4±0.2	98.0±0.2	99.4±0.2	97.8±0.3	86.5±2.3	98.3±0.3	<b>99.8±0.1</b>
YoutubeFace	Acc	56.3±0.3	71.9±2.1	OOM	85.6±0.3	28.2±2.3	83.4±0.8	52.6±1.0	<b>87.1±0.2</b>
	Precision	74.8±0.6	83.2±0.8	OOM	89.2±0.4	29.9±4.0	85.6±1.0	83.1±1.0	<b>89.5±0.4</b>
	Recall	46.0±0.3	67.7±3.1	OOM	84.3±0.4	5.4±0.3	82.8±0.8	38.7±1.3	<b>86.2±0.3</b>
	Macro F1	53.5±0.5	73.0±2.8	OOM	86.6±0.3	5.2±0.4	83.9±0.8	46.6±1.4	<b>87.7±0.1</b>

(3) MV-HFMD 展现出非常优秀的性能，这意味着在多视图学习中有效的特征融合至关重要。然而，在特征异质性问题上，该模型并没有对其进行有效解决，而特征异质性在促进特征融合中发挥了重要作用，因此与 MAMC 相比，其性能在某种程度上受限。

(4) 与使用模糊规则的多视图学习模型 IPMVSC 相比，IPMVSC 专注于实现可解释的分类器，本文的模型则关注特征级别的问题，并在表示空间中应用模糊集理论。此外，从结果中本文发现，IPMVSC 在大规模数据集上表现不佳，而本文提出的 MAMC 模型在小规模和大规模的公开数据集上均表现良好。

综上所述，本文所提出的 MAMC 模型成功缓解了视图特征异质性问题，并通过自适应扩展边界模块减轻了信息冗余的影响，使得模型具有良好的性能表现。



#### 4.6 消融实验与分析

机器学习中的消融实验（Ablation Study）是指在模型或算法的不同部分进行有针对性地去除或修改，以分析每一模块的贡献和对整体性能的影响。本文在 HandWritten、Animal、PIE 和 CCV 这几个数据集上进行了消融实验，主要目的是评估多尺度对齐模块的有效性以及组合算子（Comb）选择的合理性。首先，本文通过去除和保留实例尺度对齐模块与原型尺度对齐模块来分析多尺度对齐模块的有效性。其次，本文讨论了组合算子的选取对实验结果的影响，在这里分别选取了（Min，Max，Mean）这三种算子进行比较。实验结果如表4-4所示，并得出以下结论：

表 4-4 消融实验结果（%），关注多尺度对齐模块和 Comb 操作符

$\mathcal{L}_{cls}$	$\mathcal{L}_{rec}$	$\mathcal{L}_{ins}$	$\mathcal{L}_{pro}$	Comb	HandWritten	Animal	PIE	CCV
✓	✓				96.08±0.47	55.78±0.04	84.93±0.36	42.48±0.58
✓	✓	✓			97.44±0.54	56.95±0.42	89.71±1.20	49.34±3.33
✓	✓		✓		96.76±0.68	56.95±0.65	90.81±0.82	51.01±2.39
✓	✓	✓	✓		97.46±1.10	57.80±0.15	91.54±0.82	51.69±2.18
✓	✓		✓	Min	98.15±0.41	59.20±0.75	92.40±0.92	52.34±0.34
✓	✓	✓	✓	Min	<b>98.85±0.42</b>	<b>60.16±1.03</b>	<b>93.11±1.72</b>	<b>53.97±1.14</b>
✓	✓	✓	✓	Max	97.67±0.42	58.26±0.31	91.92±0.60	52.18±2.68
✓	✓	✓	✓	Mean	98.69±0.30	59.74±0.81	92.46±1.09	53.11±0.40

(1) 结果表明，实例级对齐模块和原型级对齐模块均能有效挖掘视图间的共性和类间的差异性，这对于解决由特征异质性和信息冗余引起的问题至关重要。

(2) 与实例级对齐模块相比，原型级对齐模块显示出了更大的性能提升，其原因在于后者利用了有助于划定清晰决策边界的类间差异。

(3) 在这里，本文选取了不同的组合算子进行实验，其中 *Min Pooling* 展现了最佳性能，证明了其在分类未知数据时保留更多边界间距上的优越性。

#### 4.7 超参数分析

在机器学习中，超参数（Hyperparameters）是指在学习过程中，需要手动设置并且不会通过算法进行学习的参数。它们是训练模型之前设定的变量，可以显著影响模型的表现和训练效率。本小节，本文针对两个惩罚系数  $\alpha$  和  $\beta$ ，以及两个温度系数  $\tau_{ins}$  和  $\tau_{pro}$  进行了超参数分析。对于  $\alpha$  和  $\beta$ ，本文将其取值范围设置为  $\{0.1, 0.25, 0.5, 0.75, 1\}$ ；对于温度系数  $\tau_{ins}$  和  $\tau_{pro}$ ，本文分别设置其取值范围为  $\{0.5, 0.75, 1.0, 1.5, 2.0\}$  和  $\{0.5, 0.75, 1.0, 1.2, 1.5\}$ 。基于 HandWritten 数据集和 Hdigit 数据集的超参实验结果如图4-1所示。其中，(a) 和 (b) 探讨了惩罚系数  $\alpha$  和  $\beta$  的影响，(c) 和 (d) 探讨了温度系数  $\tau_{ins}$  和  $\tau_{pro}$  的影响。通过观

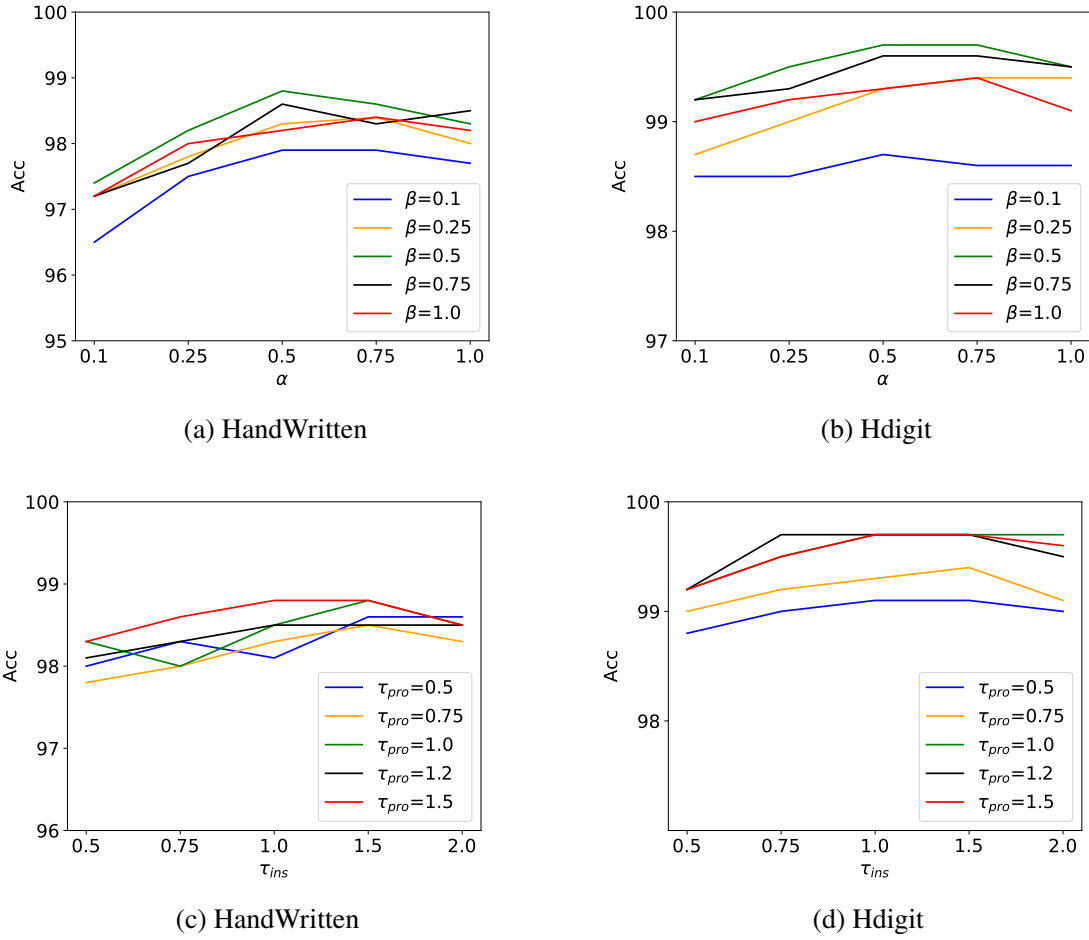


图 4-1 基于 HandWritten 和 Hdigit 数据集的超参数分析

察图像的变化趋势可以得到：与温度系数相比，惩罚系数对结果的影响更加显著，这说明了多尺度对齐模块的有效性。此外，最佳结果倾向于较大的温度系数，这表明在模型训练过程中对样本表示赋予更高权重具有更好的效果。

#### 4.8 噪声实验

在机器学习中，噪声实验通常是指在数据集中人为添加噪声，并观察模型在这种条件下的表现。噪声通常是指那些在原始数据中蕴含的随机干扰性信息，会使模型的学习变得更加困难。在本文的实验中，本文向一半的视图添加了不同水平的高斯噪声，并提供了图4-2中的对比结果。本文发现，当噪声增加时，所有模型的性能都会迅速下降，因为较大的噪声会导致更多模糊的边界。然而，本文模型的性能下降速度较其他模型慢，尤其是在噪声较小的情况下。例如，在 HandWritten 和 CCV 这两个数据集上，MAMC 直到高斯噪声增加的量级达到  $10^4$  左右性能出现了较大幅度的下降，和其他对比模型相比取得了显著的提升；同时在另外四个数据集（PIE, Animal, Scene15, Hdigit）上的表现，也证实了 MAMC 模型比其他对比模型更加稳定和具有鲁棒性。实际上，真实多视

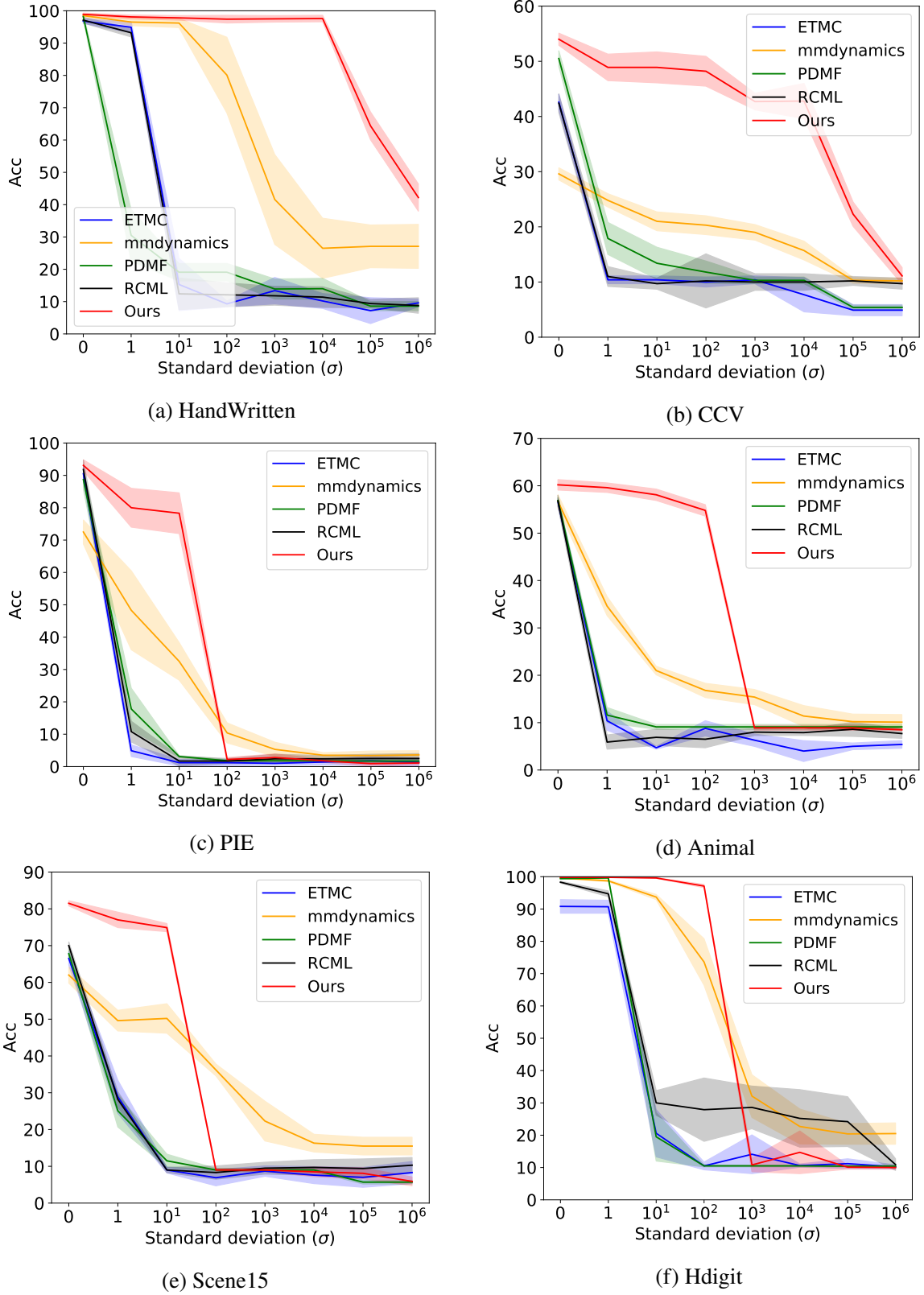


图 4-2 不同数据集上的高斯噪声实验

图数据中的噪声并不会达到一个非常大的量级。在这里，MAMC 的表现说明其能够有效利用所有视图信息中蕴含的特异性信息，以清除模糊的决策边界。

## 4.9 可视化分析

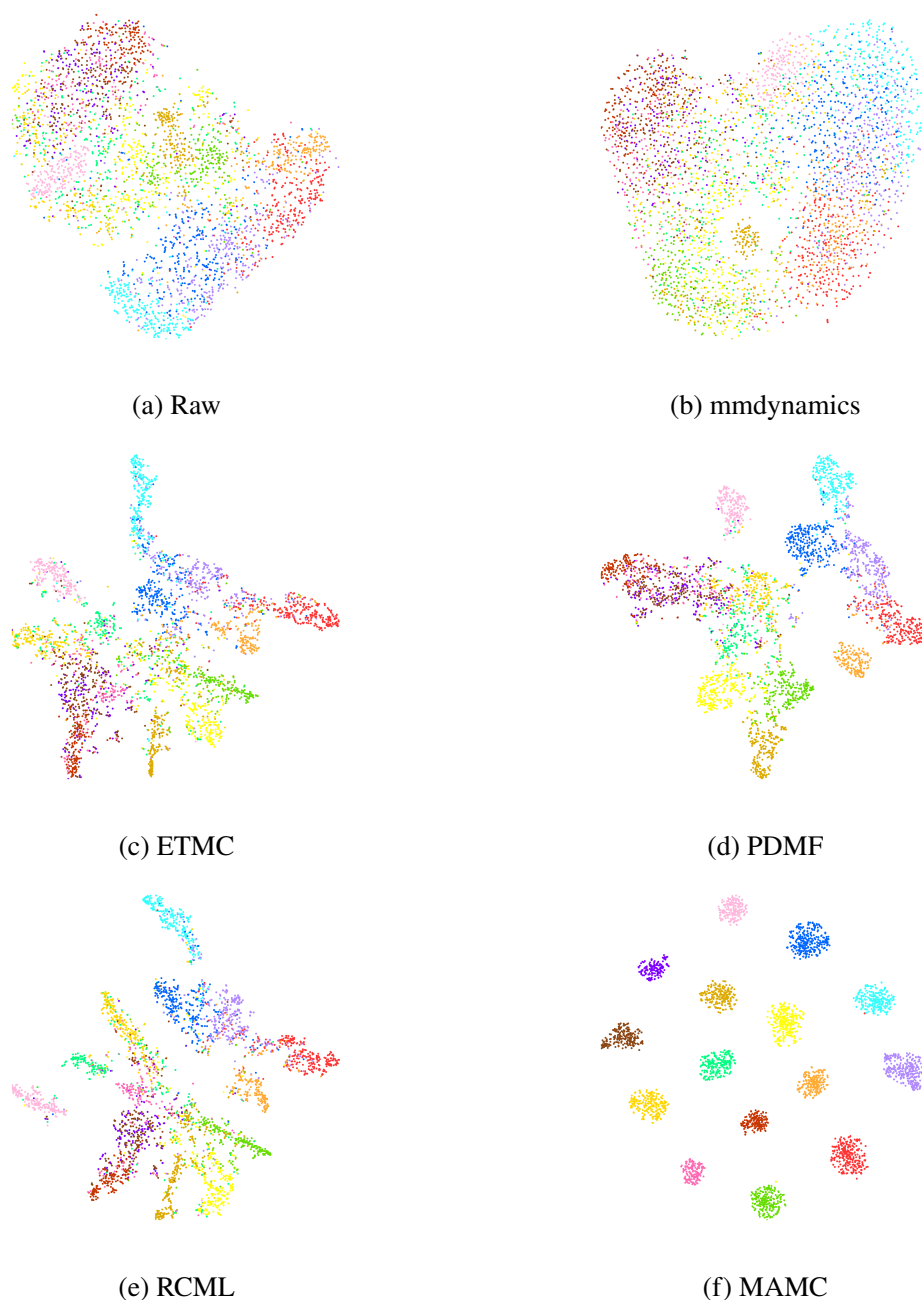
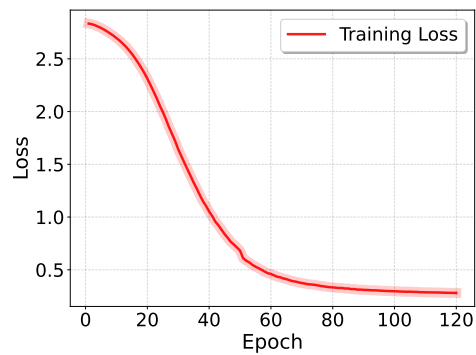


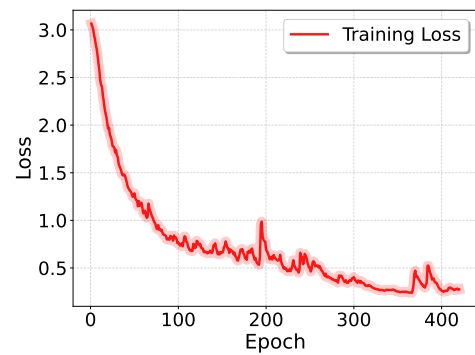
图 4-3 MVMC 和对比模型的 t-SNE 可视化分析

在这个实验中，本文提供了原始数据和不同多视图模型的性能对应的 t-SNE 可视化结果<sup>[79]</sup>，如图4-3所示。在这里，所有的视图特征表示将被拼接成一个最终的表示向量作为 t-SNE 算法的输入，以获得可视化结果。从结果中本文可以发现，尽管原始数据（Raw）存在模糊性以及呈现出较为杂乱的数据分布，大多数对比模型在一定程度上能够分辨出不同类别。然而，本文的模型 MAMC 显示出更好的结果，因为它挖掘了特征的共性，将同一类别的样本集中，并利用类间差异来区分不同类别。

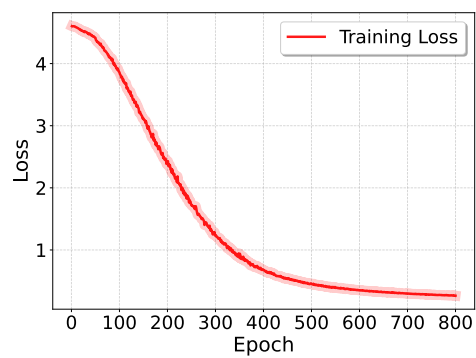
## 4.10 收敛性分析



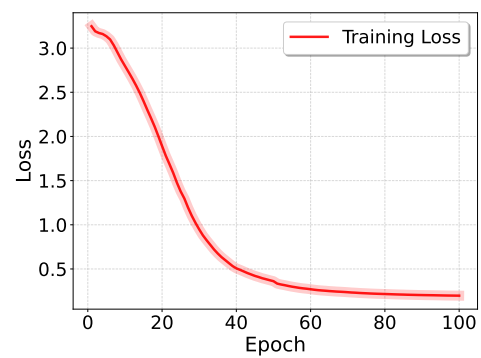
(a) HandWritten



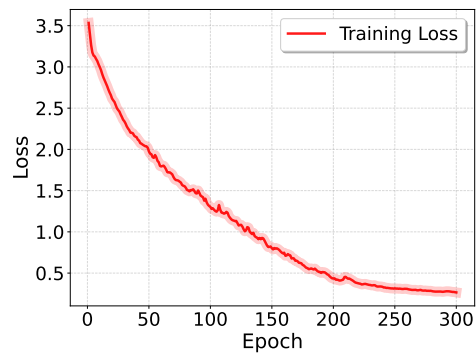
(b) Scene15



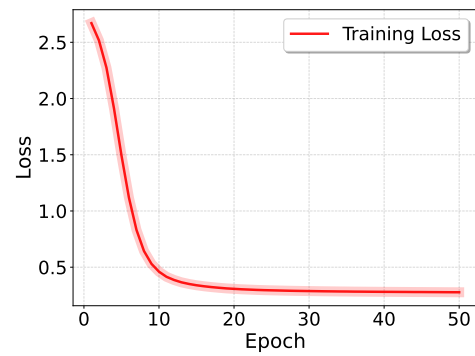
(c) PIE



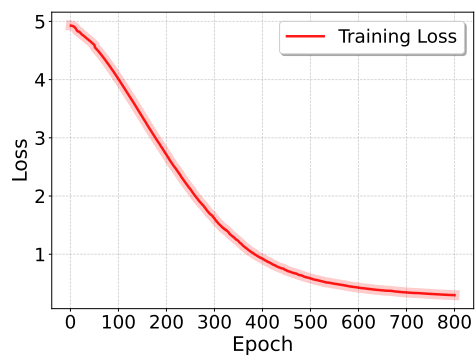
(d) Animal



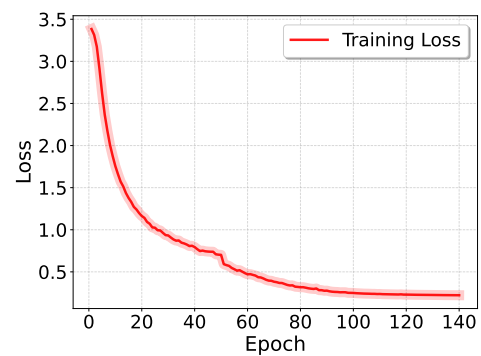
(e) CCV



(f) Hdigit



(g) 100Leaves



(h) YoutubeFace

图 4-4 不同数据集上的收敛性分析实验

在机器学习和优化算法中，模型的收敛性分析（Convergence Analysis）是用来研究模型在训练过程中如何逐步接近最优解（或近似最优解）的理论分析方法，旨在评估模型训练算法是否能够在有限的时间或迭代次数内达到稳定的解。本文在所有的数据集上对所提出的方法进行了收敛性分析，实验结果如图4-4所示。据观察，在所有的基准数据集中，模型训练过程产生的损失稳步减少，并在 50 到 800 个训练 epoch 之间收敛。最终，该实验结果证实了 MAMC 模型的收敛性。

#### 4.11 复杂性分析

表 4-5 复杂性实验结果汇总；MAMC 结果以粗体显示；“-”表示模型出现内存溢出错误

Datasets	Metrics	mmdynamics	ETMC	UMDL	PDMF	RCML	Ours
HandWritten	Running Time	21.911s	35.739s	10057.537s	21.972s	17.064s	<b>23.302s</b>
	MACs	267.871M	3.323M	1273.548M	189.989M	1.661M	<b>1373.655M</b>
	Parameters	1.0543M	0.0066M	5.2389M	1.722M	0.0066M	<b>5.353M</b>
Scene15	Running Time	28.996s	42.596s	31938.726s	45.678s	22.922s	<b>41.651s</b>
	MACs	66.604M	0.457M	132.890M	68.387M	0.457M	<b>617.869M</b>
	Parameters	0.264M	0.002M	20.632M	1.098M	0.002M	<b>3.197M</b>
PIE	Running Time	11.525s	11.233s	2721.368s	9.214s	7.775s	<b>17.499s</b>
	MACs	536.394M	17.739M	248.090M	188.088M	2.217M	<b>996.040M</b>
	Parameters	2.099M	0.070M	1.429M	1.540M	0.070M	<b>4.622M</b>
CCV	Running Time	40.678s	61.986s	29783.373s	69.529s	37.905s	<b>65.490s</b>
	MACs	56.521M	0.307M	125.338M	9.978M	0.307M	<b>263.633M</b>
	Parameters	0.225M	0.001M	46.361M	1.331M	0.001M	<b>3.323M</b>
Animal	Running Time	650.847s	181.911s	52103.473s	221.021s	119.074s	<b>177.946s</b>
	MACs	7643.710M	44.493M	1269.197M	576.355M	44.493M	<b>9360.186M</b>
	Parameters	29.873M	0.174M	141.214M	10.937M	0.174M	<b>37.353M</b>
100Leaves	Running Time	12.697s	16.173s	7386.190s	19.825s	10.078s	<b>44.393s</b>
	MACs	247.370M	4.915M	142.234M	80.740M	4.915M	<b>672.518M</b>
	Parameters	0.971M	0.020M	3.113M	0.814M	0.020M	<b>3.323M</b>
Hdigit	Running Time	57.370s	58.236s	44589.832s	85.163s	41.918s	<b>59.256s</b>
	MACs	506.520M	0.666M	211.558M	42.877M	2.662M	<b>883.593M</b>
	Parameters	1.982M	0.010M	100.825M	2.614M	0.010M	<b>4.245M</b>
YoutubeFace	Running Time	1660.614s	598.449s	-	513.350s	523.801s	<b>702.112s</b>
	MACs	4393.923M	67.456M	-	1466.827M	67.456M	<b>7988.087M</b>
	Parameters	4.328M	0.066M	-	15.840M	0.066M	<b>8.544M</b>

本文在表4-5和表4-6中展示了不同模型的复杂度对比。表4-5提供了详细的计算开销分析，使用的指标包括 MACs（模型乘加运算的总次数）、Running Time（运行时间）和 Parameters（参数量），表4-6则列出了所有模型的复杂度。为了公平比较，本文为所

有模型设置了相同的批量大小和训练轮数。从表4-5所有的统计数据来看，本文提出的 MVMC 模型并不繁琐，单个模型在一张 NVIDIA GeForce RTX 3090 (24GB) 显卡上便可

表 4-6 MAMC 和其他对比模型的复杂度

Model	mmdynamics	ETMC	UMDL	PDMF	RCML	Ours
Complexity	$O(N)$	$O(N)$	$O(N^3)$	$O(N^2)$	$O(N)$	$O(N^2)$

以完成训练和推理部署，并且在和一些先进模型的比较中中具有竞争力。表4-6展示了对比模型与本文所提出的 MVMC 模型之间的复杂度比较，通过观察发现 MVMC 模型与同一领域的先进模型在复杂度上基本处于相同水平。

#### 4.12 本章小结

本章介绍了关于 MAMC 模型实验部分的所有细节和设置，首先在4.1节介绍了多视图领域常用的公开数据集，通过这些数据集可以全面地评估模型的性能表现。4.2节对实验过程中使用的对比模型进行了介绍，通过和这些模型进行比较展现了 MVMC 在多视图领域中具有强大的竞争力和优越表现。4.3节对分类任务指标和复杂性分析指标进行了介绍，通过这些指标可以更加清晰地对 MAMC 进行性能和开销水平测试。4.4节展示了实验环境和相关配置，以帮助学术界和工业界的研究人员可以复现本文的研究，推动多视图领域的发展。4.5节展示了所有模型在八个公开数据集上的四个评价指标上的分类性能，直观地展示了 MAMC 在实验上相比其他模型具有更好的表现。4.6节针对不同尺度对齐模块和组合操作符的选取设计了消融实验，来去评估单个模块和多个模块在不同操作符下的性能表现。4.7节进行了超参分析实验，帮助本文选取最优的组合系数。4.8节展示了模型在不同水平下的高斯噪声实验，验证了 MAMC 的鲁棒性。4.9节进行了 t-sne 可视化分析，清晰地展现了不同模型对数据的分类效果。4.10节进行了收敛性分析实验，在经过一定轮次的训练之后，MAMC 在所有的数据集上均成功收敛。最后，在4.11节进行了复杂度分析实验，以评估模型的计算开销和参数量水平。

## 总结与展望

多视图数据因其能够从不同视角描述同一对象，在现实世界的各类应用中广泛存在，成为数据挖掘领域的重要研究对象。在多视图领域的下游任务中，多视图分类任务因其挑战性和价值而备受关注。通过融合多个视图的数据，多视图分类模型能够获得更全面的特征表示。然而，特征异质性和信息冗余是多视图分类任务面临的两大核心难题。特征异质性问题的产生来源于不同视图数据分布的不一致，从而导致视图特征难以有效融合；信息冗余则表现为原始的多视图数据中包含无关或模糊信息，使得分类边界模糊，进而降低了模型的泛化能力和决策可靠性。

为应对这些挑战，本文提出了一种增强多视图分类模型，命名为 MAMC。该模型通过创新的多尺度对齐模块和扩展边界理论，有效缓解了上述问题。首先，多尺度对齐模块从实例尺度和原型尺度两个层次挖掘视图间的共性语义和类别间存在的差异信息。具体来说：实例尺度对齐模块通过对齐每个实例的多视图特征，增强视图间的公共特征提取，提升多视图数据的一致性；原型尺度对齐模块通过类别间原型对齐，强化类间特征差异，助力构建更清晰的决策边界。其次，扩展边界模块基于数学模糊集理论，该模块在表示空间中扩展决策边界，通过辅助原型尺度对齐模块缩紧类内表示并拉大类间距离，有效处理模糊样本并缓解信息冗余。同时，该模块为潜在未见样本保留分类空间，进一步提升模型的适应性和泛化能力。MAMC 在多个公开数据集上的实验结果表明，其分类性能显著优于现有先进模型。消融实验验证了多尺度对齐模块和组合操作符选取的有效性，超参数敏感性分析和噪声分析等实验则展示了模型的鲁棒性和稳定性。此外，本文从理论上阐述了模型设计的合理性，并通过定性与定量实验证明了 MAMC 在解决特征融合不足和决策边界模糊问题上的优势。这些成果不仅为多视图分类任务提供了创新且高效的解决方案，也为多视图学习相关领域的研究开辟了新方向。

尽管 MAMC 模型在多视图分类任务中表现出色，但其仍存在改进和拓展的空间。当前 MAMC 主要聚焦于分类任务，其在其他多视图分析子任务（如聚类、缺失视图等场景下）中的潜力尚未充分挖掘。未来可进行探索将 MAMC 应用于这些任务上，以拓宽其应用场景和范围。同时 MAMC 的计算复杂度方面相比较于其他的对比模型，并没有取得显著的优势，训练和推理效率有待改进。未来可探索模型轻量化技术，以支持 MAMC 在资源受限环境下的部署和减少开销。此外，虽然 MAMC 在当前的公开数据集上表现出色，未来工作将进一步探索其在更大规模的真实世界数据集上的应用，使其能够真正成为一项可以从学术界落地到工业界投入使用的研究。



## 参考文献

- [1] Wang J, Feng S, Lyu G, et al. Triple-granularity contrastive learning for deep multi-view subspace clustering[C]. Proceedings of the ACM International Conference on Multimedia. 2023: 2994-3002.
- [2] Zhang C, Liang X, Zhou P, et al. Scalable multi-view unsupervised feature selection with structure learning and fusion[C]. Proceedings of the ACM International Conference on Multimedia. 2024: 5479-5488.
- [3] Lyu G, Yang Z, Deng X, et al. L-vsm: Label-driven view-specific fusion for multiview multilabel classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(12):6569-6583.
- [4] Tan X, Zhao C, Liu C, et al. A two-stage information extraction network for incomplete multi-view multi-label classification[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2024: 15249-15257.
- [5] Wang B, Li X, Li C, et al. Escaping the neutralization effect of modality features fusion in multimodal fake news detection[J]. Information Fusion, 2024, 111:102500.
- [6] Zhang C, Lou Z, Zhou Q, et al. Multi-view clustering via triplex information maximization[J]. IEEE Transactions on Image Processing, 2023, 32:4299-4313.
- [7] 张悦辰, 葛洪伟, 李婷. 基于特征级联的联合平滑多视图子空间聚类算法[J]. 南京大学学报(自然科学), 2024, 60(04):586-599.
- [8] 徐亚维. 关于加权多视图子空间聚类算法的研究与探索[D]. 华东师范大学, 2024.
- [9] Hu S, Geng R, Cheng Z, et al. A parameter-free multi-view information bottleneck clustering method by cross-view weighting[C]. Proceedings of the ACM International Conference on Multimedia. 2022: 3792-3800.
- [10] 刘溯源, 王思为, 唐厂, 等. 基于加权锚点的多视图聚类算法[J]. 自动化学报, 2024, 50(06):1160-1170.
- [11] Wang J, Feng S, Lyu G, et al. Surer: Structure-adaptive unified graph neural network for multi-view clustering[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2024: 15520-15527.
- [12] Kumar D, Maji P. Discriminative deep canonical correlation analysis for multi-view data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(10):14288-14300.
- [13] Houfar K, Samai D, Dornaika F, et al. Automatically weighted binary multi-view clustering via deep initialization[J]. Pattern Recognition, 2023, 137:109281.
- [14] Liu C, Wen J, Luo X, et al. Incomplete multi-view multi-label learning via label-guided masked view-and category-aware transformers[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2023: 8816-8824.
- [15] Han Z, Zhang C, Fu H, et al. Trusted multi-view classification with dynamic evidential fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2):2551-2566.
- [16] 章永来, 周耀鉴. 聚类算法综述[J]. 计算机应用, 2019, 39(07):1869-1882.
- [17] Xu R, Wunsch D. Survey of clustering algorithms[J]. IEEE Transactions on Neural Networks, 2005, 16(3):645-678.
- [18] 陈容珊. 两类深度聚类算法研究及在图像处理中的应用[D]. 西安电子科技大学, 2023.
- [19] 刘翔宇. 基因表达分析中双聚类算法研究[D]. 山东大学, 2021.

- [20] Fu K, Du C, Wang S, et al. Multi-view multi-label fine-grained emotion decoding from human brain activity[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(7):9026-9040.
- [21] Jiang Z, Luo T, Liang X. Deep incomplete multi-view learning network with insufficient label information[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2024: 12919-12927.
- [22] Liu S, Liu X, Wang S, et al. Fast incomplete multi-view clustering with view-independent anchors[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(6):7740-7751.
- [23] Shen Q, Chen Y, Zhang C, et al. Pick-and-place transform learning for fast multi-view clustering[J]. IEEE Transactions on Image Processing, 2024, 33:1272-1284.
- [24] Yan C, Gong B, Wei Y, et al. Deep multi-view enhancement hashing for image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(4):1445-1451.
- [25] Kumar A, Daumé H. A co-training approach for multi-view spectral clustering[C]. Proceedings of the International Conference on Machine Learning. 2011: 393-400.
- [26] Xu J, Ren Y, Li G, et al. Deep embedded multi-view clustering with collaborative training[J]. Information Sciences, 2021, 573:279-290.
- [27] Liu X, Dou Y, Yin J, et al. Multiple kernel  $k$ -means clustering with matrix-induced regularization[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2016, 30(1):1888-1894.
- [28] Zhao Y, Dou Y, Liu X, et al. A novel multi-view clustering method via low-rank and matrix-induced regularization[J]. Neurocomputing, 2016, 216:342-350.
- [29] Zhang X, Zhang X, Liu H, et al. Multi-task clustering through instances transfer[J]. Neurocomputing, 2017, 251:145-155.
- [30] 徐杰. 多视图聚类的特征表示研究[D]. 电子科技大学, 2024.
- [31] 周余琳, 王长鹏. 基于一致性和差异性的低秩张量多视图聚类算法[J]. 吉林大学学报 (理学版), 2025, 63(02):537-550.
- [32] Li R, Zhang C, Fu H, et al. Reciprocal multi-layer subspace learning for multi-view clustering[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2019: 8172-8180.
- [33] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training[C]. Proceedings of the Eleventh Annual Conference on Computational Learning Theory. 1998: 92-100.
- [34] Kumar A, Rai P, Daume H. Co-regularized multi-view spectral clustering[C]. Advances in Neural Information Processing Systems: volume 24. 2011.
- [35] 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. 2022: 16051-16060.
- [36] Tang W, Lu Z, Dhillon I S. Clustering with multiple graphs[C]. 2009 Ninth IEEE International Conference on Data Mining. 2009: 1016-1021.
- [37] Nie F, Li J, Li X. Self-weighted multiview clustering with multiple graphs[C]. Proceedings of the International Joint Conference on Artificial Intelligence. 2017: 2564-2570.
- [38] Ni J, Tong H, Fan W, et al. Flexible and robust multi-network clustering[C]. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015: 835-844.
- [39] Liu R, Cheng W, Tong H, et al. Robust multi-network clustering via joint cross-domain cluster alignment [C]. IEEE International Conference on Data Mining. 2015: 291-300.
- [40] Shen Q, Xu T, Liang Y, et al. Robust tensor recovery for incomplete multi-view clustering[J]. IEEE Transactions on Multimedia, 2024, 26:3856-3870.

- [41] Yuan Y H, Li J, Li Y, et al. Learning canonical f-correlation projection for compact multiview representation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 19260-19269.
- [42] Liu J, Wang C, Gao J, et al. Multi-view clustering via joint nonnegative matrix factorization[C]. Proceedings of the SIAM international conference on data mining. 2013: 252-260.
- [43] Luo P, Peng J, Guan Z, et al. Dual regularized multi-view non-negative matrix factorization for clustering [J]. Neurocomputing, 2018, 294:1-11.
- [44] Meng H, Deng Y, Zhong Q, et al. Federated multi-view multi-label classification[J]. IEEE Transactions on Big Data, 2024:1-13.
- [45] Liu M, Luo Y, Tao D, et al. Low-rank multi-view learning in matrix completion for multi-label image classification[C]. Proceedings of the AAAI Conference on Artificial Intelligence: volume 29. 2015: 2778-2784.
- [46] Zhang C, Yu Z, Hu Q, et al. Latent semantic aware multi-view multi-label classification[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2018: 4414-4421.
- [47] Black S, Souvenir R. Multi-view classification using hybrid fusion and mutual distillation[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 270-280.
- [48] Han Z, Yang F, Huang J, et al. Multimodal dynamics: Dynamical fusion for trustworthy multimodal classification[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 20707-20717.
- [49] Xu J, Li W, Liu X, et al. Deep embedded complementary and interactive information for multi-view classification[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 6494-6501.
- [50] Xu C, Si J, Guan Z, et al. Reliable conflictive multi-view learning[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2024: 16129-16137.
- [51] Fan Z, Huang Y, Xi C, et al. Multiview adaptive k-nearest neighbor classification[J]. IEEE Transactions on Artificial Intelligence, 2023, 5(3):1221-1234.
- [52] Hu X, Liu X, Pedrycz W, et al. Multi-view fuzzy classification with subspace clustering and information granules[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(11):11642-11655.
- [53] Xu C, Zhao W, Zhao J, et al. Progressive deep multi-view comprehensive representation learning[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2023: 10557-10565.
- [54] Sensoy M, Kaplan L, Kandemir M. Evidential deep learning to quantify classification uncertainty[C]. Advances in Neural Information Processing Systems. 2018: 1-11.
- [55] Fei-Fei L, Perona P. A bayesian hierarchical model for learning natural scene categories[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005: 524-531.
- [56] Zhang W, Deng Z, Zhang T, et al. One-step multiview fuzzy clustering with collaborative learning between common and specific hidden space information[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(10):14031-14044.
- [57] Michelucci U. An introduction to autoencoders[J]. arXiv:2201.03898, 2022.
- [58] 来杰, 王晓丹, 向前, 等. 自编码器及其应用综述[J]. 通信学报, 2021, 42(09):218-230.
- [59] Zhang C, Liu Y, Fu H. Ae<sup>2</sup>-nets:autoencoder in autoencoder networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2577-2585.
- [60] Fan S, Wang X, Shi C, et al. One2multi graph autoencoder for multi-view graph clustering[C]. Proceedings of The Web Conference. 2020: 3070-3076.

- [61] 张重生, 陈杰, 李岐龙, 等. 深度对比学习综述[J]. 自动化学报, 2023, 49(01):15-39.
- [62] van den Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv:1807.03748, 2018.
- [63] Poole B, Ozair S, Van Den Oord A, et al. On variational bounds of mutual information[C]. International Conference on Machine Learning. 2019: 5171-5180.
- [64] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]. International Conference on Machine Learning. 2020: 1597-1607.
- [65] 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. 2020: 9729-9738.
- [66] 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.
- [67] Zadeh L A. Fuzzy sets[J]. Information and Control, 1965, 8(3):338-353.
- [68] Zhelezniak V, Savkov A, Shen A, et al. Don't settle for average, go for the max: Fuzzy sets and max-pooled word vectors[J]. arXiv:1904.13264, 2019.
- [69] Zhong Q, Lyu G, Yang Z. Align while fusion: A generalized nonaligned multiview multilabel classification method[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(4): 7627-7636.
- [70] Wang J, Qu J, Wang K, et al. Improving the robustness of knowledge-grounded dialogue via contrastive learning[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2024: 19135-19143.
- [71] Li A, Deng Z, Zhang W, et al. Multiview transfer representation learning with tsk fuzzy system for eeg epilepsy detection[J]. IEEE Transactions on Fuzzy Systems, 2023, 32(1):38-52.
- [72] Jensen J L W V. Sur les fonctions convexes et les inégalités entre les valeurs moyennes[J]. Acta Mathematica, 1906, 30(1):175-193.
- [73] Tan Y, Liu Y, Huang S, et al. Sample-level multi-view graph clustering[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 23966-23975.
- [74] Gross R, Matthews I, Cohn J, et al. Multi-pie[C]. Proceedings of the IEEE International Conference on Automatic Face Gesture Recognition. 2010: 807-813.
- [75] Jiang Y G, Ye G, Chang S F, et al. Consumer video understanding: A benchmark database and an evaluation of human and machine performance[C]. Proceedings of the ACM International Conference on Multimedia Retrieval. 2011: 1-8.
- [76] Li Y, Shi X, Du C, et al. Manifold regularized multi-view feature selection for social image annotation [J]. Neurocomputing, 2016, 204:135-141.
- [77] Liu S, Yuan H, Yuan Z, et al. Deep contrastive multi-view clustering under semantic feature guidance [C]. International Conference on Advanced Data Mining and Applications. Springer Nature Singapore, 2024: 417-431.
- [78] Xu C, Li Z, Guan Z, et al. Unbalanced multi-view deep learning[C]. Proceedings of the ACM International Conference on Multimedia. 2023: 3051-3059.
- [79] Van der Maaten L, Hinton G. Visualizing data using t-sne[J]. Journal of Machine Learning Research, 2008, 9(11):2579-2605.