

基于多元关系建模的少样本分类算法研究



重庆大学硕士学位论文

(学术学位)

学生姓名：XXX

指导教师：XXX 教授

学科门类：工 学

学科名称：软件工程

研究方向：计算机视觉

答辩委员会主席：X X 教授

授位时间：2024 年 6 月

摘 要

待补全

关键词：少样本分类；关系建模；对比学习；语义信息表示

Key words:

目 录

摘 要	I
图目录	IV
表目录	V
1 问题重述与分析	1
1.1 问题背景	1
1.2 问题分析	1
1.2.1 问题一	1
1.2.2 问题二	1
1.2.3 问题三	1
1.2.4 问题四	2
2 模型的假设	3
2.1 少样本分类	3
2.2 对比学习	4
2.2.1 无监督对比学习	4
2.2.2 有监督对比学习	4
2.3 语义信息表示	5
2.3.1 Word2Vec	5
2.3.2 GloVe	6
2.3.3 BERT	7
2.3.4 CLIP	8
2.4 数据集及评价指标	9
2.4.1 数据集	9
2.4.2 评价指标	10
2.5 本章小结	10
3 基于多粒度样本关系建模的少样本分类研究	11
3.1 引言	11
3.1.1 研究动机	11
3.1.2 方法概述	12
3.2 基于多粒度样本关系对比学习的少样本特征学习算法	12
3.2.1 符号定义	12

3.2.2 整体框架	12
3.2.3 基础特征学习网络	13
3.2.4 多粒度样本关系对比学习算法.....	14
3.2.5 模型优化	16
3.2.6 模型推理	16
3.3 实验设置及结果分析.....	16
3.4 本章小结	16
4 基于语义-视觉多空间关系建模的少样本分类研究	18
4.1 引言	18
4.2 基于语义-视觉多空间关系建模的少样本特征适配算法	18
4.3 实验设置及结果分析.....	18
4.4 本章小结	18
5 总结与未来展望	19
5.1 总结	19
5.2 未来展望	20
参考文献.....	21
附 录.....	23
A. 作者在攻读硕士学位期间的论文目录	23
B. 作者在攻读硕士学位期间参与的科研项目	23
C. 学位论文数据集.....	24

图目录

图 2.1 少样本分类测试任务示意图 3

图 2.2 无监督对比学习与有监督对比学习..... 5

图 2.3 CBOW 模型与 Skip-Gram 模型示意图..... 6

图 2.4 BERT 的整体预训练和微调过程..... 7

图 2.5 CLIP 模型预训练示意图..... 8

图 3.1 样本关系示意图.....11

图 3.2 多粒度样本关系对比学习模型示意图13

图 3.3 MGSRL 模型推理过程示意图17

表目录

表 2.1 miniImageNet、CIFAR-FS 和 CUB 的数据集划分	9
表 2.2 tieredImageNet 的数据集划分	9

1 问题重述与分析

1.1 问题背景

STR, 也即短串联重复序列是法医 DNA 分析中最常用到的遗传标记。单一 STR 检测可以进行有效的辨别和区分; 但混合 STR 分型却难以区分。国内目前依靠法医进行人工分析, 精度与效率存在问题。//引用通过数学建模与计算方法实现混合 STR 智能化分析具有重要现实意义。

1.2 问题分析

1.2.1 问题一

确定混合样本中 DNA 贡献者的人数 (Number of Contributors, NOC) 是混合谱分析的第一步, 也是后续步骤的前提。当下研究得出检测样本中的 DNA 模板量越大, 则检测样本生成的 STR 图谱上的等位基因的峰面积越大/// (基于全局最小残差法快速分析混合 STR 图谱)。因此我们可以通过峰面积推断可能的比例。但由于可能出现的等位基因重叠和漏检导致实际的峰数量和面积少于理论值。

因此, 我们提出基于峰高分布的高斯混合模型并结合信息准则的方法估计贡献者人数。我们假设每个贡献者的 DNA 含量不同并且峰高总值近似为高斯分布则 N 个人的集合可以建模为 N 个高斯分布的混合。我们对候选的不同人数假设 $N = 1, 2, 3, \dots$ 进行 GMM 拟合, 用 EM 算法估计混合模型参数 (各高斯分布的均值、方差、权重等)。然后计算每个 N 假设下模型的赤池信息准则 (AIC) 或贝叶斯信息准则 (BIC)。信息准则综合考虑模型拟合优度和复杂度, 值越低表示平衡后效果越好。选择使 AIC/BIC 最小的 N 作为估计的贡献者人数。

1.2.2 问题二

确定了混合样本中的贡献者人数后, 我们需要估计每位贡献者的 DNA 相对含量, 即混合比例向量 $\mathbf{r} = (r_1, \dots, r_N)$ 。然而当比例向量相对接近时, 等位基因峰高将高度重叠, 难以正确区分等位基因。因此我们需要建立模型将等位基因峰高表达为多个贡献体的叠加并求解比例。

我们采用////

1.2.3 问题三

第三个子问题要求我们将混合的 STR 图谱拆解还原出每个贡献者各自的基因型, 并评估准确性。这实际上是对混合 DNA 进行个体分型的过程。如果有一个已知的嫌疑人/个体基因型数据库 (附件 3 提供), 则还需要将分离得到的基因型与数据库中的个体进行匹配, 从而识别出混合样本中都有哪些人。此问题是整个混合谱分析的核心与难点: 我们需要在不知道哪条峰属于哪个人的情况下, 将彼此

交织的等位基因峰正确归属到各个人的两个等位基因序列中去。对于一个 N 人混合、每人基因座数为 L 的情况，可能的基因型组合空间非常巨大（每个位点需要在观测峰中为 N 个人各选择 2 个等位基因，组合数爆炸式增长）。因此必须设计高效算法进行合理搜索和判断。

我们提出////

1.2.4 问题四

混合 STR 图谱中经常伴随各种实验噪声和假信号，如果不加以处理，可能干扰前述分析步骤的准确性。常见的噪声包括：电泳基线噪声导致的微小峰等。此外，还有信号本身的随机波动。若直接将所有观测峰都视为真实等位基因，将导致贡献者人数高估或错误的基因型分配。因此，需要在混合图谱分析中引入降噪机制，提高信噪比。

对此我们采用////

2 模型的假设

本章内容共分为五节，第一节详细介绍少样本分类任务；第二节介绍与第三章方法相关的对比学习工作；第三节介绍第四章方法应用到的语义信息表示及一些用来获得语义信息的自然语言处理模型和多模态模型；第四节对本文所使用的数据集和评价指标进行介绍，第五节对本章进行小结。

2.1 少样本分类

少样本分类，旨在模拟人类识别新类别的过程，希望模型在拥有大规模标注数据的类别上进行训练之后，能够总结并迁移所学知识和新的类别，以实现在新类别上仅用少量标注数据进行训练便能够达到良好效果的目的。与常规分类任务将数据集划分为训练集与测试集不同，少样本分类数据集被划分为基类数据和新类数据，两者类别互不相交。其中，基类数据与普通分类任务的训练集一致，所有数据均可以被用来训练模型，无论是以元学习还是以普通分类任务的训练方式。而新类数据则是用来测试模型性能，在少样本分类的测试过程中，会在新类数据集上随机采样大量分类任务，每个任务的数据又被划分为支持集与查询集，如图2.1所示。其中，支持集数据为带有标注的样本，可用来微调整个模型和重新训练分类器，而查询集作用则是类似普通分类任务中的测试集，用来评估模型准确率。根据采样任务中类别数目 N 和样本数目 K 的多少，其又可被称为 N -way K -shot 任务， N 通常取 5， K 通常取 1 或 5。最终，通过对大量采样任务分别进行评估，并计算这些任务的平均准确率作为模型性能的最终评价指标。



图 2.1 少样本分类测试任务示意图。

Fig. 2.1 Illustration of few-shot classification testing tasks.

2.2 对比学习

在计算机视觉领域，特征学习的方法越来越多样化。其中，对比学习以其独特的学习机制，即通过比较样本之间的相似性和差异性来提取鲜明且有区分度的特征表征，近年来受到了广泛的关注和研究。在图像处理任务中，对比学习已经证明了其在提高模型泛化能力和识别精度方面的显著效果，并被广泛应用到少样本分类问题中。根据对比学习是否使用数据集标签信息，可以将其分为无监督对比学习和有监督对比学习，以下将分别进行介绍。

2.2.1 无监督对比学习

无监督对比学习不依赖于标注数据，它通常采用正负样本对的形式来构建训练任务。正样本对通常来自于同一实例的不同视角（例如，同一图像的不同数据增强版本），而负样本对则来自于不同实例。模型的目标是使得正样本对在表示空间中彼此接近，而负样本对彼此远离。该过程一般通过最小化 InfoNCE 损失函数实现，该损失函数如下式所示，

$$\mathcal{L} = -\log \frac{\exp(\cos(f(x), f(x^+))/\tau)}{\exp(\cos(f(x), f(x^+))) + \sum_{j=1}^N \exp(\cos(f(x), f(x^-)))}. \quad (2.1)$$

其中，图像 x 经由网络 $f(\cdot)$ 后映射到特征空间， x^+ ， x^- 分别代表 x 的正样本以及负样本， N 为负样本数量。 $\cos(\cdot)$ 是余弦相似度， $\exp(\cdot)$ 为以 e 为底的指数函数。

Chen 等人^[1]提出了一个简单有效的无监督对比学习框架-SimCLR，旨在通过比较不同视角下图像的特征表示来学习强大的特征提取网络。SimCLR 的核心思想是利用数据增强来产生正样本对，即从同一张图像中通过随机的数据增强操作（如裁剪、颜色变换等）生成两个视角，然后使来自同一图像的特征相互靠近，同时使得来自不同图像的特征尽可能地远离。尽管 SimCLR 在无监督特征学习方面取得了显著的成果，但其有一个明显缺点，即 SimCLR 的效果很大程度上依赖于对比损失函数中大量不同的负样本对，为了达到最佳性能，需要批次大小很大，这对计算资源的要求较高。He 等人^[2]提出的 MoCo 算法通过引入一个动态字典来存储样本特征表示解决了此问题。这个字典是一个队列，新的样本特征进入队列时，旧的样本特征被移除，以保持队列的固定大小。MoCo 可通过此字典高效地采样大量负样本，因此不再需要使用很大的批次便可达到最佳效果。这些无监督对比学习方法特别适合于数据量大但未标注的场景，能够有效地利用大量未标注数据来学习有意义的特征表示。

2.2.2 有监督对比学习

虽然无监督对比学习为使用大量无标注数据训练一个好的预训练模型提供了有效途径，但因为其在样本建模过程中将样本 x 与其负样本距离推远，而负样本中可能包含 x 的同类样本，这可能会学习到错误的样本关系。因此，Khosla 等人^[3]

提出了有监督对比学习（Supervised Contrastive Learning，简称 SupCon）对这个问题进行解决。SupCon 是对比学习的一种变体，它结合了监督信号来进一步提升学习效率和特征表示的质量。与无监督对比学习相比，有监督对比学习在构造正负样本对时利用了标签信息，以确保模型不仅学会区分不同的样本，而且能够区分不同的类别，如图2.2所示（此图来源于 SupCon^[3]）。SupCon 不仅保留了无监督对比学习中正样本对的概念，更进一步地，将属于同一类别的不同样本也视为正样本对，负样本对则是来自不同类别的样本，以此强化模型对不同类别间差异的识别能力。这种方法有效地缩小了同类样本间的表征距离，同时增强了不同类别间表征的区分度，有助于提升模型在复杂视觉任务中的表现。

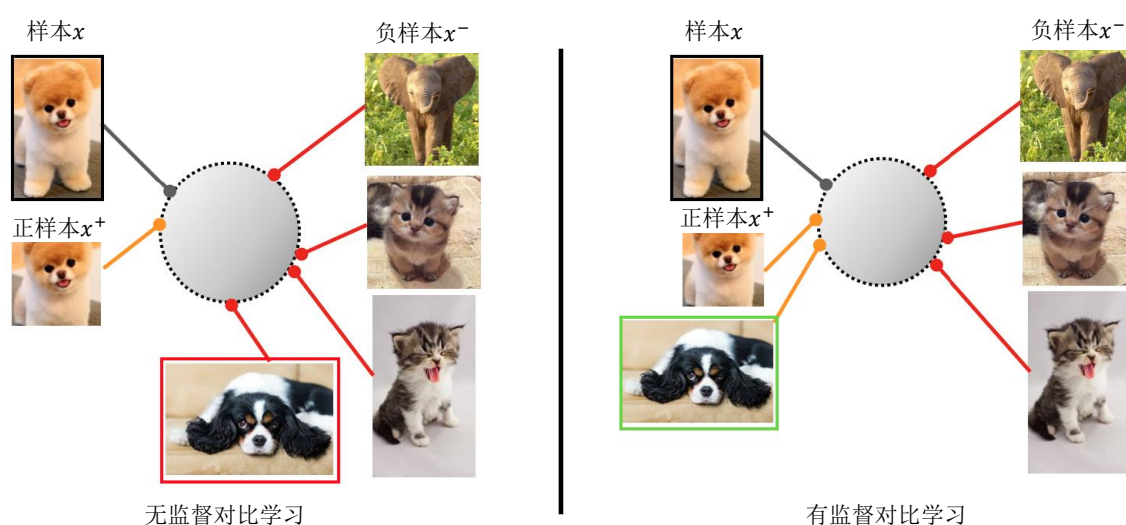


图 2.2 无监督对比学习与有监督对比学习。

Fig. 2.2 Unsupervised Contrastive Learning VS Supervised Contrastive Learning.

2.3 语义信息表示

目前，在少样本分类问题中，很多工作开始使用语义信息以对视觉信息进行补充，使用的语义信息一般为用自然语言处理（Natural Language Processing，简称 NLP）模型或多模态模型的文本编码器提取的语义特征。提取语义特征时，会将类别名称或提示文本与类别名称进行拼接之后的文本输入文本编码器，然后得到编码器输出的语义向量作为语义特征。以下对少样本分类中经常使用的语义特征提取模型进行介绍。

2.3.1 Word2Vec

Mikolov 等人^[4,5]提出的 Word2Vec 是一种广泛使用的自然语言处理技术，它从大量文本语料中以无监督的方式学习语义知识，旨在将词汇映射到稠密向量空间中，其中语义相似的词汇会在向量空间中彼此接近。Word2Vec 包含两种训练模型：连续词袋（Continuous Bag-of-Words，简称 CBOW）模型和跳跃（Continuous

Skip-gram, 简称 Skip-Gram) 模型, 如图2.3所示 (此图来源于 Word2Vec^[5])。

CBOW 模型: CBOW 模型通过上下文 (周围的词汇) 来预测当前词, 如图2.3 (左) 所示。具体来说, 它将上下文中的多个词汇作为输入, 并尝试预测在这些上下文词汇中间的目标词汇。这个模型特别适合处理较小的数据集。

Skip-Gram 模型: 与 CBOW 相反, Skip-Gram 模型使用一个词来预测其周围的上下文, 如图2.3 (右) 所示。给定一个特定的词, 目标是预测在一个特定范围内的前后词汇。Skip-Gram 模型在处理大数据集时表现更好, 尤其是对罕见词汇的表示更为有效。

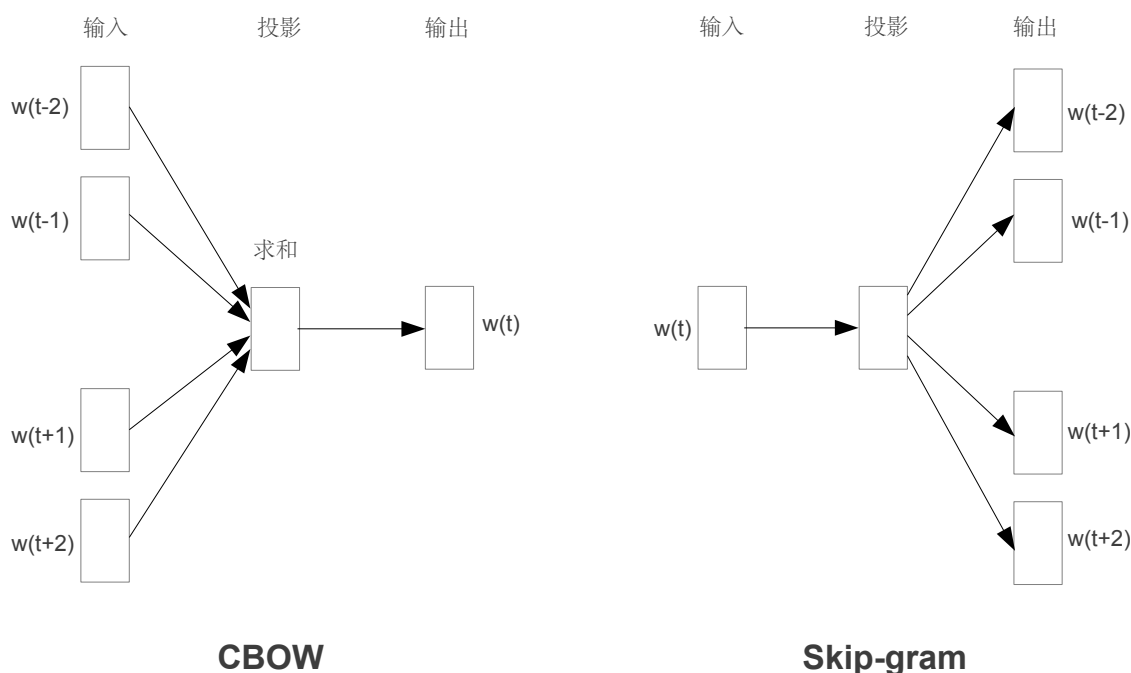


图 2.3 CBOW 模型与 Skip-Gram 模型示意图。

Fig. 2.3 Illustration of CBOW model and Skip-Gram model.

Word2Vec 的核心优势在于它能够捕捉到词汇之间的细微语义关系, 并通过向量运算来揭示词汇之间的语义相似性。这使得 Word2Vec 在诸多自然语言处理任务中得到广泛应用, 包括文本相似性度量、情感分析、机器翻译以及作为深度学习模型的预训练层等, 另外, 很多计算机视觉任务也使用 Word2Vec 来提取语义特征以对视觉特征进行修正或补充。

2.3.2 GloVe

Pennington 等人^[6]提出的 GloVe (Global Vectors for Word Representation) 也是一种用于词嵌入的无监督学习算法。该模型旨在将单词映射到一个向量空间中, 使得这些向量能够捕捉到词与词之间的共现关系, 从而反映出词义的复杂模式和结构。GloVe 模型的关键创新在于它结合了两种主流的词表示方法的优点: 基于

全局矩阵分解（Global Matrix Factorization）的方法和基于局部上下文窗口（Local Context Window）的方法。

GloVe 的核心思想是首先构建一个全局词共现矩阵，记录整个语料库中各个词之间的共现次数，然后通过优化一个目标函数来学习词向量。这个目标函数旨在让共现次数的对数值与相应词向量的点积尽可能接近，同时引入偏置项来进一步提升模型的灵活性和准确性。具体来说，GloVe 构建一个大型的词-词共现矩阵，矩阵中的每个元素代表了两个词在一定窗口大小内共同出现的次数。这一步捕获了全局的共现统计信息。然后其定义了一个特殊的损失函数，该损失函数不仅关注词对之间的共现概率，而且关注共现概率的比例，这有助于捕获词义之间更细微的差别。这个损失函数同时考虑到了共现次数的稀疏性和不均匀性。通过最小化损失函数，模型学习到的词向量能够反映出词与词之间的共现概率，这意味着词向量空间中的距离可以表示词义之间的相似度。这一步既利用了局部信息（通过具体的共现频率），也综合了全局信息（通过整个语料库的统计数据）。

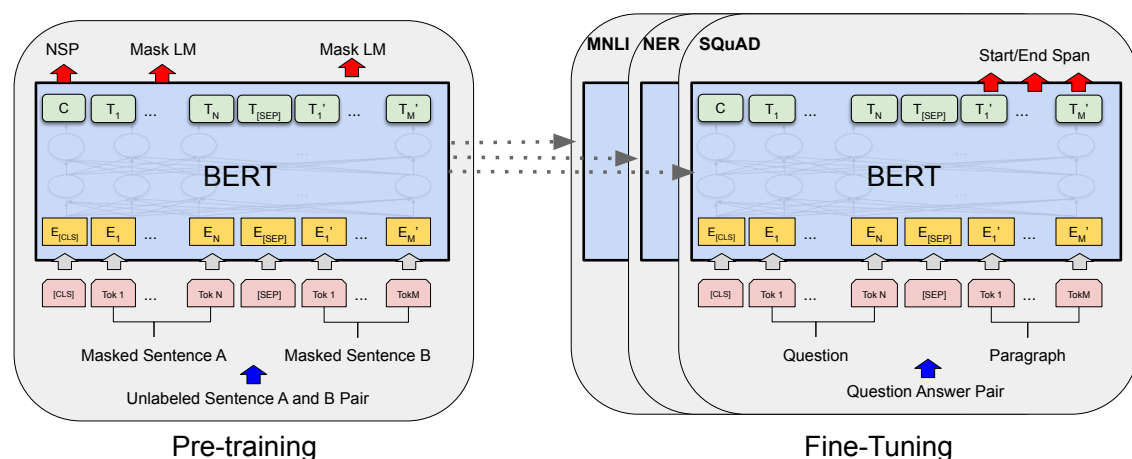


图 2.4 BERT 的整体预训练和微调过程。

Fig. 2.4 Overall pre-training and fine-tuning procedures for BERT.

2.3.3 BERT

Devlin 等人^[7]提出的 BERT（Bidirectional Encoder Representations from Transformers）模型是一种革命性的自然语言处理模型。该模型利用了 Transformer 架构的双向编码器，能够理解语言的深层语义和上下文关系。BERT 的创新之处在于其基于 Transformer 模型的编码器，使得它能够同时考虑词语左侧和右侧的上下文信息，这与以往的单向模型或浅层双向模型不同，使其能够更准确地理解词义。如图 2.4 所示（此图来源于 BERT^[7]），BERT 模型首先在大规模的文本语料库上进行预训练，学习通用的语义表示，然后针对具体的 NLP 任务进行微调，这一过程极大提升了模型在特定任务上的性能。由于其良好的性能与开创性，后续又出现了诸如 SBERT^[8]、RoBERTa^[9]、ALBERT^[10] 等改进工作。BERT 模型通过两种类型

的预训练任务学习语义表示：

- (1) **掩码语言模型 (Masked Language Model, 简称 MLM)**: 在训练过程中, BERT 会随机遮蔽模型输入句子中的一部分词语 (使用 [MASK] token 代替原有输入), 然后让模型预测这些遮蔽的词语, 这可以迫使模型学习到词语的双向上下文关系。另外, 为了解决模型微调期间从未看到 [MASK] token 的问题, BERT 模型不总是直接用 [MASK] token 代替所选单词, 而是将所选单词 80% 的概率替换为 [MASK] token, 10% 的概率用一个随即单词替换所选单词, 剩下 10% 的概率则是保持其不变。
- (2) **下一句预测 (Next Sentence Prediction, 简称 NSP)**: 由于很多 NLP 下游任务都是基于理解两个句子之间的关系, 如问答和自然语言推断, 因此 BERT 设计了一个下一句预测的任务。给定两个句子 A 和 B, 模型需要预测 B 是否是 A 的下一句, 这可以帮助模型理解句子间的关系。

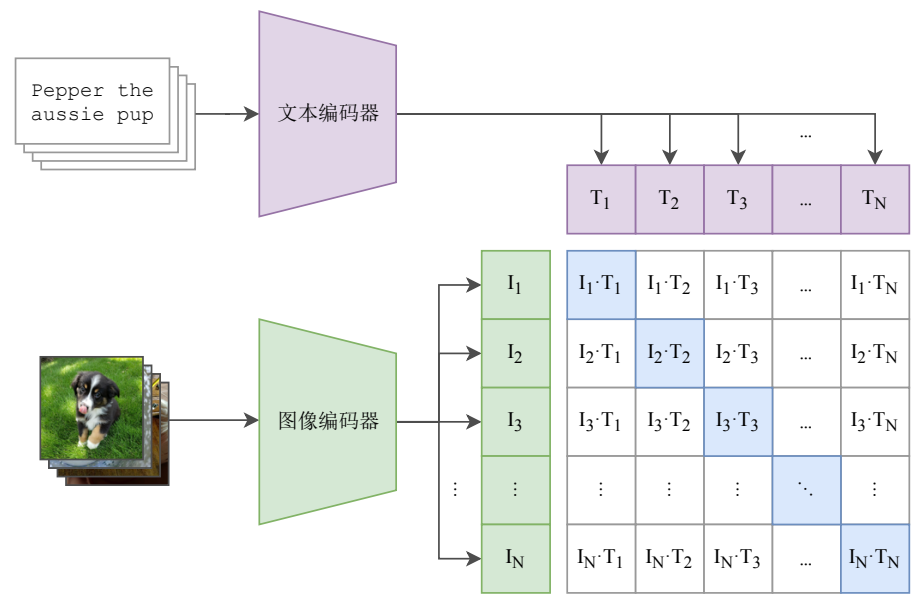


图 2.5 CLIP 模型预训练示意图。
Fig. 2.5 Illustration of pre-training CLIP model.

2.3.4 CLIP

近年来, Radford 等人^[11]提出的 CLIP (Contrastive Language-Image Pre-training) 模型受到了很多研究者的关注, 并促进了多模态大模型和一些其他任务的发展。CLIP 旨在通过大规模的图文对比学习来同时理解图像和文本, 并建立它们之间的联系。CLIP 模型的创新之处在于其跨模态能力, 它不仅能理解图片内容, 也能理解与图片内容相对应的文本描述, 从而在多种视觉任务上展示出了卓越的性能和强大的泛化能力, 并提供了一个充分建模文本关系的文本编码器。

如图2.5所示 (此图来源于 CLIP^[11]), CLIP 由两部分组成: 一个图像编码器

和一个文本编码器。图像编码器负责提取图像的视觉特征，而文本编码器则提取文本的语义特征。这两个编码器可以是任何形式的神经网络。在原始 CLIP 模型中，图像编码器基于 Vision Transformer (ViT) 或 ResNet 架构，而文本编码器基于 Transformer 架构。CLIP 的训练过程涉及大量图像和文本对的对比学习。具体来说，模型训练的目标是最大化相匹配的图像和文本对之间的相似度，同时最小化不匹配对的相似度。这种训练方式使得 CLIP 学习到的特征表示能够跨越视觉和语义的界限，理解两种模态之间的对应关系。

2.4 数据集及评价指标

本文使用四个少样本分类基准数据集对模型性能进行评估，包括三个普通少样本分类数据集：miniImageNet^[12]、tieredImageNet^[13]、CIFAR-FS^[14]，以及一个细粒度数据集：CUB-200-2011 (CUB)^[15]，以下对其进行分别介绍，并对少样本分类评价指标进行描述。

表 2.1 miniImageNet、CIFAR-FS 和 CUB 的数据集划分。

Table 2.1 Dataset partition of miniImageNet, CIFAR-FS and CUB.

数据集	类别数目			
	训练集	验证集	测试集	总数
miniImageNet	64	16	20	100
CIFAR-FS	64	16	20	100
CUB	100	50	50	200

表 2.2 tieredImageNet 的数据集划分。

Table 2.2 Dataset partition of tieredImageNet.

类别层级	类别数目			
	训练集	验证集	测试集	总数
超类	20	6	8	34
子类	351	97	160	608

2.4.1 数据集

miniImageNet 数据集^[12]和 tieredImageNet 数据集^[13]均为 ImageNet^[16]的子集。其中，miniImageNet 数据集包含 100 个类别，每个类别有 600 张图像。本文遵循 Ravi 等人^[17]提出的划分准则，训练集、验证集和测试集分别包含 64、16 和 20 个类别。tieredImageNet 数据集则包含 34 个超类（608 个子类），分为 20 个训练类别（351 个子类）、6 个验证类别（97 个子类）和 8 个测试类别（160 个子类）。CIFAR-FS 数据集^[14]源自 CIFAR-100 数据集，该数据集包含 64 个训练类别、16 个验证

类别和 20 个测试类别，每个类别同样有 600 张图像。Caltech-UCSD Birds(CUB)-200-2011（简称 CUB）数据集^[15] 则是一个包含不同种类的鸟类细粒度图像数据集，包含 11788 个图像样本，分为 200 个类别。根据 Triantafillou 等人^[18] 的划分准则，该数据集包含 100 个训练类别、50 个验证类别和 50 个测试类别。各数据集划分如表2.1和2.2所示。

2.4.2 评价指标

对于所有数据集，本文评估 5-way 1-shot 以及 5-way 5-shot 少样本分类任务性能。在一次模型评估中，本文方法采样 2000 个少样本分类任务，并计算了 95% 置信区间的平均分类准确率作为模型的评价指标。在一个少样本分类任务中，每个类别的支持集样本数目为 1 或 5（根据任务决定），查询集样本数目为 15，与其他方法^[19,20] 保持一致。

2.5 本章小结

本章首先详细介绍了少样本分类任务的定义及其训练测试过程。然后对后续研究工作所涉及到的相关技术进行了介绍，其中包括第三章所使用到的对比学习技术，根据是否使用数据集标签信息将其分为无监督对比学习和有监督对比学习进行了详细阐述；以及第四章所使用到的语义信息表示，介绍了如何提取语义信息表示和少样本分类中常用的语义特征提取模型。最后，介绍了本文方法所使用到的少样本分类数据集和评价指标。

3 基于多粒度样本关系建模的少样本分类研究

本章研究基于多粒度样本关系建模的少样本特征学习算法，通过挖掘多种粒度的样本关系并对其进行建模从而增强模型的特征提取能力，进而提升少样本分类任务的准确率。本章内容共分为四节，第一节介绍研究动机和方法概述；第二节介绍本章提出的基于多粒度样本关系对比学习的少样本特征学习算法；第三节给出实验设置和结果分析；第四节对本章进行小结。

3.1 引言

3.1.1 研究动机

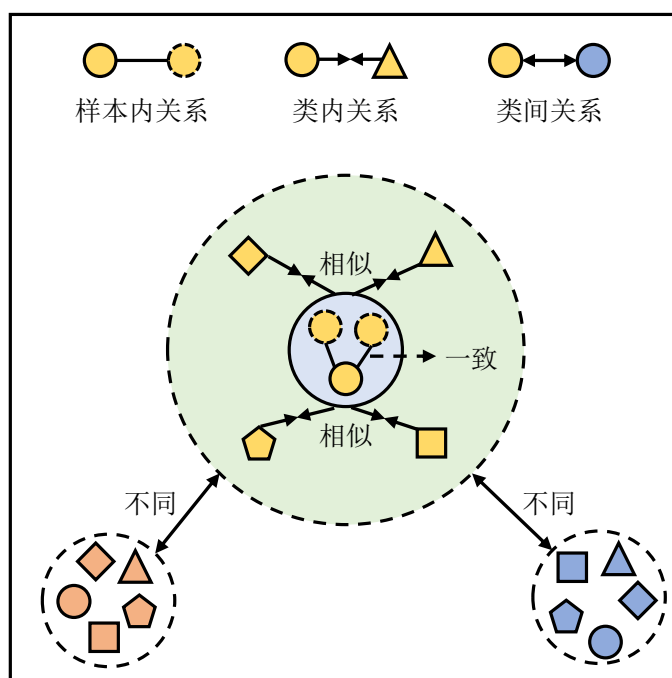


图 3.1 样本关系示意图。在该图中，不同形状与颜色分别代表不同样本与类别。同一样本的不同变换由相同的颜色和形状表示。样本关系包括三种类型：样本内关系、类内关系和类间关系。本章所提方法约束同一样本不同变换版本在语义内容上保持一致，同类样本保持相似，非同类样本保持不同。

Fig. 3.1 Illustration of sample relations. In this figure, different shapes and different colors represent different samples and different classes, respectively. Different transformations of the same sample are represented by the same color and shape. The sample relations contain three types: intra-sample relation, intra-class relation and inter-class relation. The approach proposed in this chapter enforces different transformations to be consistent in semantic content, homogenous samples to be similar, and inhomogeneous samples to be different.

3.1.2 方法概述

3.2 基于多粒度样本关系对比学习的少样本特征学习算法

在本节中，首先对少样本分类任务及其符号定义进行介绍；然后对所提出的基于多粒度样本关系对比学习的少样本特征学习模型进行简要介绍；接下来详细介绍了所提模型的各个模块及其损失优化；最后介绍了模型总体优化目标以及模型推理过程。

3.2.1 符号定义

在本章中，少样本分类任务的基类数据集和新类数据集分别表示为：

$$\begin{aligned}\mathcal{D}_{base} &= \{(x, y) | x \in X^{base}, y \in Y^{base}\}, \\ \mathcal{D}_{novel} &= \{(x, y) | x \in X^{novel}, y \in Y^{novel}\}.\end{aligned}\tag{3.1}$$

其中， \mathcal{D}_{base} 所包含的类别 \mathcal{C}_{base} 和 \mathcal{D}_{novel} 所包含的类别 \mathcal{C}_{novel} 不相交。另外， x 、 y 分别表示样本图像和样本标签； X^{base} 、 Y^{base} 和 X^{novel} 、 Y^{novel} 分别表示基类数据和新类数据的样本图像集合和标签集合。

\mathcal{D}_{base} 用于在预训练阶段训练一个具有良好泛化性能的模型， \mathcal{D}_{novel} 用于测试过程采样大量 N -way K -shot 少样本分类任务并计算平均准确率来评估模型性能。每个少样本分类任务 \mathcal{T} 包括一个支持集 $\mathcal{S}_{\mathcal{T}}$ 和一个查询集 $\mathcal{Q}_{\mathcal{T}}$ ，

$$\mathcal{T} = \{\mathcal{S}_{\mathcal{T}}, \mathcal{Q}_{\mathcal{T}}\}.\tag{3.2}$$

其中， $\mathcal{S}_{\mathcal{T}}$ 包含来自 N 个类别的 $N \times K$ 个标注样本，而 $\mathcal{Q}_{\mathcal{T}}$ 包含来自相同 N 个类别的 $N \times Q$ 个样本，并且 $\mathcal{S}_{\mathcal{T}}$ 和 $\mathcal{Q}_{\mathcal{T}}$ 中的样本是没有交集的。在测试阶段，针对每个采样的少样本分类任务使用 $\mathcal{S}_{\mathcal{T}}$ 重新训练一个分类器，使用 $\mathcal{Q}_{\mathcal{T}}$ 来评估分类器性能。

3.2.2 整体框架

本章重新审视了对比学习中的样本关系，并根据样本关系粒度的不同将其划分为三种类型：同一样本在不同变换下的样本内关系（intra-sample relation）、同类样本的类内关系（intra-class relation），以及不同类样本的类间关系（inter-class relation）。基于此，本章提出了一种新颖的多粒度样本关系对比学习方法（Multi-Grained Sample Relation Contrastive Learning，简称 MGSRCL），通过对少样本分类中不同粒度的样本关系进行建模从而获得了一个强大的特征提取网络。如图3.2所示，MGSRCL 模型包含三个主要部分：基础特征学习网络（Base Feature Learning Network，简称 Base）、变换一致性学习（Transformation Consistency Learning，简称 TCL）模块和类对比学习（Class Contrastive Learning，简称 CCL）模块。具体而言，基础特征学习网络是通过一般图像分类任务训练的神经网络。TCL 模块旨

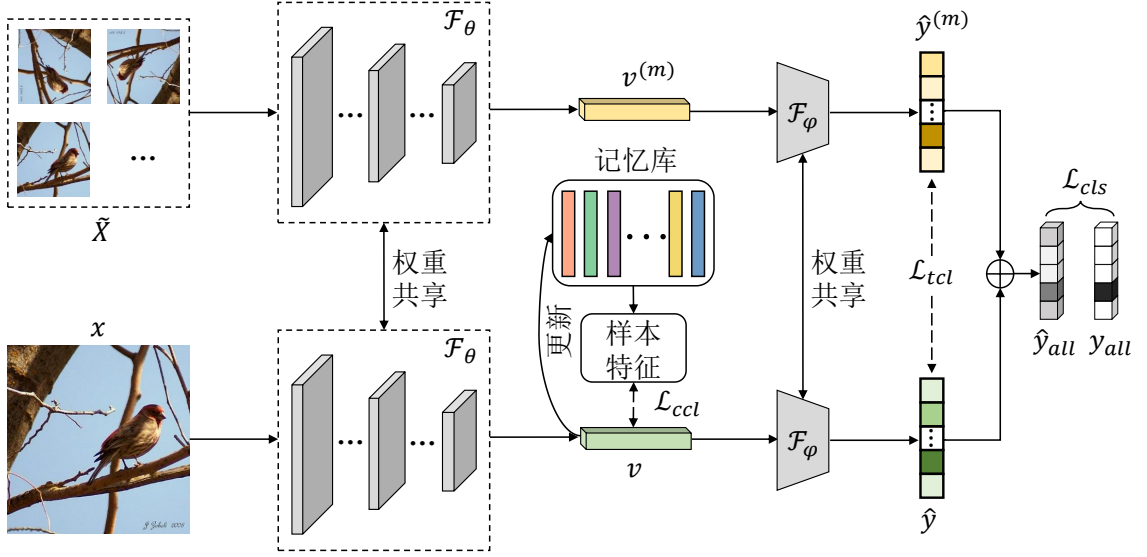


图 3.2 多粒度样本关系对比学习模型 (MGSRL) 示意图。它包含一个特征提取网络 \mathcal{F}_θ 和一个分类器 \mathcal{F}_φ 。在此图中, v 和 $v^{(m)}$ 代表原始图像 x 及其第 m 个变换版本 $x^{(m)}$ 的特征, 其中 $x^{(m)} \in \hat{X}$ 。 \oplus 是一个连接操作, 用于对原始图像的预测输出 \hat{y} 与 M 个变换的预测输出 $\hat{y}^{(1)}, \dots, \hat{y}^{(m)}, \dots, \hat{y}^{(M)}$ 进行连接。记忆库 (Memory Bank) 用于存储特征。 \mathcal{L}_{cls} , \mathcal{L}_{tcl} 和 \mathcal{L}_{ccl} 分别是分类损失、变换一致性学习 (TCL) 损失和类对比学习 (CCL) 损失。为了便于阅读, 此图中没有展示自监督模块。

Fig. 3.2 Illustration of Multi-Grained Sample Relation Contrastive Learning (MGSRL) model. It contains a feature extraction network \mathcal{F}_θ and a classifier \mathcal{F}_φ . In this figure, v and $v^{(m)}$ represent the features of the original image x and its m -th transformed version $x^{(m)}$, where $x^{(m)} \in \hat{X}$. \oplus is a concatenation operator for the predicted output \hat{y} of the original image and the predicted outputs $\{\hat{y}^{(1)}, \dots, \hat{y}^{(m)}, \dots, \hat{y}^{(M)}\}$ of M transformations. Memory bank is used to store the features. \mathcal{L}_{cls} , \mathcal{L}_{tcl} , and \mathcal{L}_{ccl} are the classification loss, transformation consistency learning (TCL) loss, and class contrastive learning (CCL) loss, respectively. For the sake of legibility, the self-supervised module is not shown in this image.

在确保同一样本的不同变换版本具有一致的语义内容。而 CCL 则用于确保同类样本具有相似的语义内容, 以及非同类样本具有不同的语义内容。接下来, 本节将对 MGSRL 方法的每个部分进行更为详细的阐述。

3.2.3 基础特征学习网络

如图3.2所示, 特征提取网络, 表示为带有参数 θ 的 \mathcal{F}_θ , 被用于提取图像特征。设 $(x, y) \in \mathcal{D}_{base}$ 表示从 \mathcal{D}_{base} 中采样的图像及其对应的标签。图像 x 的特征向量 v 可以通过 \mathcal{F}_θ 获得: $v = \mathcal{F}_\theta(x)$ 。然后, 使用参数为 φ 的分类器 \mathcal{F}_φ , 将特征向量 v 投影到标签空间, 以获得预测的置信度分数 p : $p = \mathcal{F}_\varphi(v)$ 。最后, 通过在 p 上应用 Softmax 函数, 可以得到预测概率输出 \hat{y} : $\hat{y} = \text{Softmax}(p)$ 。基础特征学习网络的参数 θ 和 φ 通过最小化整个基类数据集 \mathcal{D}_{base} 上的分类损失 \mathcal{L}_{cls} 来进行优化, 其可以

表示为以下公式，

$$\mathcal{L}_{cls} = -\frac{1}{|\mathcal{D}_{base}|} \sum_{\{x,y\} \in \mathcal{D}_{base}} y \log \hat{y}. \quad (3.3)$$

为了防止在训练集上过拟合，许多方法^[20-22]引入了变换样本参与训练，并使用自监督学习技术预测在训练过程中对图像执行了哪种变换以增强网络的特征提取能力。遵循这些方法，本文也添加了一个由多层感知机（Multilayer Perceptron，简称 MLP）构成的自监督（Self-Supervised，简称 SS）模块。设 $\tilde{X} = \{\tilde{x}^{(1)}, \dots, \tilde{x}^{(M)}\}$ 为一张图像的变换版本集合，其中 M 表示变换样本的总数， $\tilde{x}^{(m)}$ 表示图像的第 m 个变换版本。 \tilde{X} 可以通过在图像上应用一系列变换（如裁剪、调整大小、旋转等数据增强操作）获得。变换后的图像 \tilde{X} 和原始图像 x 同时输入模型，用于分类和自监督任务。自监督任务的目标是识别图像进行了哪种变换，其损失 \mathcal{L}_{ss} 表示为以下公式，

$$\mathcal{L}_{ss} = -\frac{1}{|\mathcal{D}_{base}|} \frac{1}{M+1} \sum_{x \in \mathcal{D}_{base}} \sum_{m=0}^M s^{(m)} \log \hat{s}^{(m)}, \quad (3.4)$$

其中 $\hat{s}^{(m)}$ 和 $s^{(m)}$ 分别表示自监督任务中第 m 个变换版本的预测概率输出和真实标签。 $s^{(0)}$ 是原始图像 x 的自监督标签。此外，增加了变换样本之后的分类损失可以重新定义为以下公式，

$$\mathcal{L}_{cls} = -\frac{1}{|\mathcal{D}_{base}|} \frac{1}{M+1} \sum_{x \in \mathcal{D}_{base}} \sum_{m=0}^M y^{(m)} \log \hat{y}^{(m)}, \quad (3.5)$$

$\hat{y}^{(m)}$ 表示分类任务中的预测概率输出， $y^{(m)}$ 表示分类任务的真实标签。

最后，基础特征学习网络的损失 \mathcal{L}_{base} 可以写为分类损失 \mathcal{L}_{cls} 和自监督损失 \mathcal{L}_{ss} 之和，

$$\mathcal{L}_{base} = \mathcal{L}_{cls} + \mathcal{L}_{ss}. \quad (3.6)$$

3.2.4 多粒度样本关系对比学习算法

(1) 变换一致性学习

一个样本图像与其变换版本包含完全相同的对象和背景，仅因为进行了数据增强而使得图像在旋转角度、明暗、颜色等方面发生变化，但其内在的类别属性和语义内容应保持不变。为了实现这一目标，本文设计了一个变换一致性学习（Transformation Consistency Learning，简称 TCL）模块，以约束同一样本不同变换版本的样本内关系。TCL 模块通过约束一个样本和其变换版本的预测输出相同来确保它们具有一致的语义内容。这是因为预测输出反映了样本在每个类别中的预测概率，这些概率不仅表示了模型对于样本属于各个类别的置信度，而且深入地揭示了样本的本质属性——语义内容。

本章方法将一个样本与其变换版本同时输入网络，并在预测标签输出层面计

算它们的 TCL 损失。这里, 本文使用 Jensen-Shannon 散度^[23,24] 作为 TCL 损失, 它能够衡量两个概率分布的差异, 通过最小化两个预测标签的输出, 可以使其概率分布一致, 从而达到使样本和其变换版本具有一致语义内容的目的。TCL 损失可以写为以下公式,

$$\mathcal{L}_{tcl} = \frac{1}{|\mathcal{D}_{base}|} \sum_{x \in \mathcal{D}_{base}} \frac{1}{M} \sum_{m=1}^M JS(\hat{y}_{\tau_1}, \hat{y}_{\tau_1}^{(m)}), \quad (3.7)$$

其中 \hat{y}_{τ_1} 和 $\hat{y}_{\tau_1}^{(m)}$ 分别是原始图像和第 m 个变换图像的平滑标签输出。它们通过以下公式获得,

$$\hat{y}_{\tau_1} = \text{Softmax}(p/\tau_1), \quad (3.8)$$

此公式中 $p = \mathcal{F}_{\varphi}(\mathcal{F}_{\theta}(x))$, τ_1 是一个温度参数, 本文在实验中将其设置为 4.0。使用平滑标签输出的原因在于不同变换的输出不仅需要在最大预测概率的类别上保持一致, 而且需要在所有其他类别上也保持一致, 以确保它们具有完全相同的语义内容, 而平滑标签输出可以提供更多关于概率分布差异的信息。

(2) 类对比学习

同类样本虽然图像内包含了同一个类别的物体, 但物体及其背景与同一图像不同变换版本相比差异性较大, 因此其预测概率输出之间差异也会较大。如果强行将其预测输出进行对齐, 可能会使得网络为了学习此种强关系而导致模型崩塌。但在另一方面, 同类样本间距离比不同类样本间距离更近是毋庸置疑的。因此, 本文采用类对比学习 (Class Contrastive Learning, 简称 CCL) 以一种相对距离的形式约束同类样本的类内关系和不同类样本的类间关系。CCL 模块通过最大化同类样本特征的相似性, 同时最小化不同类样本特征的相似性来在特征空间拉近同类样本, 推远不同类样本。

与之前对比学习不同, CCL 模块为了将样本和其他每个不同类间的距离推远, 对于每张图像都需要该图像的一个同类样本以及其他每个类别的不同类样本 (之前对比学习通常随机采样, 这使得每个批次计算损失时不同类样本可能仅来自部分不同类别)。为了实现这一目标并加快训练速度, 本文使用了一个记忆库 (Memory Bank) 来存储和从中采样图像特征, 记忆库存储了所有图像的特征。在一个批次中, CCL 模块从记忆库中为每类图像随机采样一个样本的特征。CCL 损失可以定义为,

$$\mathcal{L}_{ccl} = \frac{1}{|\mathcal{D}_{base}|} \sum_{x \in \mathcal{D}_{base}} -\log \frac{\exp(\frac{\cos(v, v')}{\tau_2})}{\sum_{i=1}^{|\mathcal{C}_{base}|} \exp(\frac{\cos(v, v_i)}{\tau_2})}, \quad (3.9)$$

其中 $|\mathcal{C}_{base}|$ 和 $|\mathcal{D}_{base}|$ 表示基类的类别数量和样本数量, v 和 v' 分别是某个样本及其同类样本的特征, v_i 代表来自第 i 类的样本的特征。这里 v' 和 v_i 是从记忆库中采样的。 $\cos(\cdot)$ 是余弦相似度, $\exp(\cdot)$ 为以 e 为底的指数函数。而 τ_2 是一个温度参

数，本文按照^[1,3]的实验设置将其设为 0.1。此外，记忆库的更新方式为，

$$v_k = r \times v_k + (1 - r) \times v_q, \quad (3.10)$$

v_q 和 v_k 分别代表在当前小批次中获得的图像特征以及在记忆库中存储的相同图像的特征， r 用于调整记忆库的更新速度，按照 IER 方法^[20]的实验，本文将其设置为 0.99。在训练阶段，记忆库每一轮训练过程都会完全更新一遍。

3.2.5 模型优化

结合公式3.6、3.7和3.9，本章提出的 MGSRL 模型总体损失函数可以表示为以下公式，

$$\mathcal{L}_{total} = \mathcal{L}_{base} + \alpha \cdot \mathcal{L}_{tdl} + \beta \cdot \mathcal{L}_{ccl}, \quad (3.11)$$

其中 α 和 β 是用于平衡不同损失的超参数，分别表示 TCL 模块和 CCL 模块的损失权重。

MGSRL 模型通过在整个基类数据集上最小化上述损失函数对模型参数进行联合优化。通过建模多个粒度的样本关系，可以有效地增强模型的特征提取能力和泛化能力，帮助模型捕获更具判别性的特征，从而提高模型在新类 \mathcal{D}_{novel} 上的分类性能。

3.2.6 模型推理

模型在基类数据集 \mathcal{D}_{base} 训练完成之后，在测试阶段，将会冻结 MGSRL 模型特征提取网络的所有参数，并通过解决来自新类 \mathcal{D}_{novel} 的大量少样本分类任务来评估模型性能。在每个任务 \mathcal{T} 的推理过程中，本文使用特征提取网络 \mathcal{F}_θ 来获得支持集 $\mathcal{S}_\mathcal{T}$ 和查询集 $\mathcal{Q}_\mathcal{T}$ 的图像特征。然后，本文使用 $\mathcal{S}_\mathcal{T}$ 的样本特征训练一个逻辑回归分类器 LC ，并对 $\mathcal{Q}_\mathcal{T}$ 中的样本进行分类，最后将在多个少样本分类任务上的准确率平均值作为模型的评价指标。MGSRL 模型的推理过程如图3.3所示。

3.3 实验设置及结果分析

在本节中

3.4 本章小结

本章

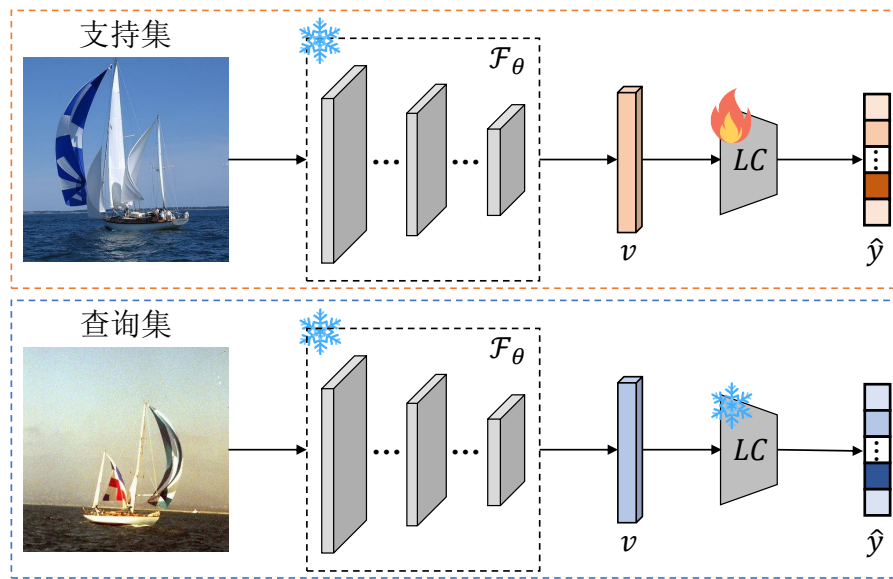


图 3.3 MGSRL 模型推理过程示意图。推理过程中，使用冻结参数的特征提取网络 \mathcal{F}_θ 提取支持集与查询集的图像特征。其中，支持集特征被用来训练一个逻辑回归分类器 LC ，查询集特征则是用来测试分类器性能。

Fig. 3.3 Illustration of MGSRL model inference process. During the inference process, the feature extraction network \mathcal{F}_θ with frozen parameters is used to extract image features from both support set and query set. Herein, the support set features are utilized to train a logistic regression classifier LC , while the query set features are used to assess the classifier's performance.

4 基于语义-视觉多空间关系建模的少样本分类研究

上一章研究了基于多粒度样本关系建模的少样本特征学习算法，通过充分挖掘不同粒度的样本关系提高了网络所提取特征的质量。然而，该算法仅在视觉空间对多种样本关系进行了建模，忽略了数据集中所隐含的丰富语义信息，限制了模型通过基类数据进行训练来学习新类数据知识的能力。因此，在上一章的基础上，本章主要研究基于语义-视觉多空间关系建模的少样本特征适配算法，通过引入语义信息并与视觉信息进行建模从而丰富模型所获得的信息，增强模型的泛化能力。本章内容共分为四节，第一节介绍研究动机和方法概述；第二节介绍本章提出的基于语义-视觉多空间关系建模的少样本特征适配算法；第三节给出实验设置和结果分析；第四节对本章进行小结。

4.1 引言

4.2 基于语义-视觉多空间关系建模的少样本特征适配算法

4.3 实验设置及结果分析

4.4 本章小结

本章研究基于语义-视觉多空间关系建模的少样本特征适配算法，针对少样本分类中仅根据少量视觉特征无法捕获类别代表性特征的缺点，引入语义信息作为视觉信息的补充，通过对语义-视觉多空间关系进行建模，提出了语义-视觉多空间映射适配模型（Semantic-Visual Multi-Space Mapping Adapter，简称 SVMSMA），以丰富样本特征的信息来源，利用语义特征对视觉特征进行补充与修正，从而提升模型在新类上的泛化能力。SVMSMA 模型使用单/多模态映射网络将样本语义特征映射到视觉空间获得单/多模态映射特征，并通过跨模态分类（CMC）模块与跨模态特征对齐（CMFA）模块对映射网络进行优化，以使得语义特征与视觉特征建立联系。测试过程中，本章方法将支持集的视觉特征、单模态映射特征、以及多模态映射特征共同作为分类器的训练数据，达到了较仅使用单一特征时更好的分类结果。在 miniImageNet、tieredImageNet、CIFAR-FS 和 CUB-200-2011 数据集的大量实验表明了 SVMSMA 方法的有效性。

综上所述，本章提出的基于语义-视觉多空间关系建模的少样本特征适配算法通过对语义-视觉多空间关系进行建模，充分利用语义信息对视觉信息进行了补充，丰富了样本特征信息来源，提升了模型的泛化能力。

5 总结与未来展望

本章内容共分为两节，第一节对本文研究内容与方法进行总结；第二节介绍本文所提方法的局限性并对未来研究方向进行展望。

5.1 总结

少样本分类致力于模拟人类的知识迁移能力，期望模型在具有大量标注数据的基类数据上训练之后，能够将所学知识迁移至新类别，实现用少量标注样本进行有效学习。目前，少样本分类问题已取得一系列研究成果，但仍存在一些问题与挑战：特征提取网络迁移能力不够强；样本数量极少情况下无法捕获类别代表性特征。针对这些问题，本文分别从多粒度样本关系建模与语义-视觉多空间关系建模两个角度出发，对少样本分类中的多元关系进行了深入挖掘与研究。

(1) 本文首先从多粒度样本关系建模的角度出发，开展了基于多粒度样本关系建模的少样本分类研究。针对少样本分类模型特征提取能力不足的问题，提出了一种基于多粒度样本关系对比学习的少样本特征学习算法：多粒度样本关系对比学习（Multi-Grained Sample Relation Contrastive Learning，简称 MGSRCCL）模型，旨在通过对不同粒度的样本关系进行建模以提升模型的特征提取能力。MGSRCCL 使用变换一致性学习来约束同一样本不同变换版本之间的样本内关系，通过使其预测概率分布相同令它们在语义内容上保持一致；使用类对比学习来约束同类样本的类内关系和不同类样本的类间关系，通过对其特征进行建模使同类样本语义内容相似、不同类样本语义内容不同。通过对多种粒度的样本关系细致地建模，MGSRCCL 提升了模型的特征提取能力，达到了优异的少样本分类结果。在 miniImageNet、tieredImageNet、CIFAR-FS 和 CUB 四个少样本基准数据集上的大量实验证明了 MGSRCCL 的有效性。另外，通过将 MGSRCCL 模型作为预训练模型与其他方法结合，证明了所获得特征提取网络的可迁移性。

(2) 上述提出的 MGSRCCL 方法虽然达到了优异的结果，但仍存在没有利用样本语义信息的问题。因此，以 MGSRCCL 方法为基础，本文进一步进行了基于语义-视觉多空间关系建模的少样本分类研究。针对少量样本的视觉特征无法捕获类别代表性特征的问题，提出了一种基于语义-视觉多空间关系建模的少样本特征适配算法：语义-视觉多空间映射适配（Semantic-Visual Multi-Space Mapping Adapter，简称 SVMSMA）模型，旨在引入语义信息对视觉信息进行补充，丰富样本特征的信息来源以提升其多样性与代表性。SVMSMA 使用语义-视觉多空间映射网络将语义特征映射到视觉空间，并通过跨模态分类模块对单/多模态映射特征执行分类任务使其与视觉特征建立联系，以及跨模态特征对齐模块将映射特征与视觉特征原型进行对齐以获得更接近类别原型的特征。通过对语义-视觉多空间关系进行建

模，SVMSMA 丰富了样本特征的信息来源，提升了模型的泛化能力。在四个基准数据集上的实验证明了 SVMSMA 方法能够有效利用语义信息，在 MGSRL 的基础上进一步提升少样本分类结果。

本文使用 MGSRL 模型和 SVMSMA 模型分别对数据间的多种样本关系以及多种空间映射关系进行了建模，有效利用了数据中的多元关系，通过多粒度样本关系建模提升了视觉特征提取网络的特征提取能力，通过语义-视觉多空间关系建模提升了模型的泛化能力，从而取得了较好的少样本分类结果。

5.2 未来展望

本文分析了少样本分类面临的挑战，以数据中的多元关系为切入点，从多粒度样本关系建模与语义-视觉多空间建模两个角度入手，提出了多粒度样本关系对比学习模型和语义-视觉多空间映射适配模型来解决少样本分类问题，并取得了一定成果。但仍存在一定不足，后续可从以下几方面进一步研究：

(1) 本文提出的 MGSRL 方法中产生变换样本时使用的多是一些弱数据增强方法，其对特征提取网络性能提升产生的作用较为有限。目前诸如 Mixup、CutMix、以及 AugMix 等强数据增强已被证明了能够提高模型泛化能力，但由于其会将不同图像融合形成一张新的图像，这使得图像类别不再是单一标签，无法应用于 MGSRL。因此，后续工作可以探讨如何将强数据增强方法引入所提出的方法，或者对方法进行改进以提高其适用性。

(2) 本文通过使用 CLIP 模型的文本编码器作为语义特征提取网络，引入语义信息对视觉信息进行补充并取得了优异结果。在将类别名称输入文本编码器时，使用了 CLIP 原论文中提出的提示文本。但使用的提示文本是固定的，并不一定能够让模型输出对少样本分类任务来说最优的语义特征。因此后续可进一步研究其他提示文本或者将提示文本换成可学习参数，以获取最优的语义特征。

(3) 本文中特征提取网络使用卷积网络，并没有使用近年来在很多视觉任务上取得良好表现的 Transformer 模型。这是因为 Transformer 模型一般需要大量的数据才能得到一个具有强大特征提取能力的预训练模型，而在少样本分类任务中，仅有 tieredImageNet 数据集规模较大，因此如何将 Transformer 模型引入少样本分类任务并取得像在其他任务上超越卷积网络的效果也是后续研究方向之一。

参考文献

- [1] 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.
- [2] 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.
- [3] Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning[J]. Advances in Neural Information Processing Systems, 2020, 33: 18661-18673.
- [4] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [A]. 2013.
- [5] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26.
- [6] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014: 1532-1543.
- [7] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Vol. 1. 2019: 4171--4186.
- [8] Reimers N, Gurevych I. Sentence-bert: Sentence embeddings using siamese bert-networks[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019: 3982-3992.
- [9] Liu Y, Ott M, Goyal N, et al. Roberta: A robustly optimized bert pretraining approach[A]. 2019.
- [10] Lan Z, Chen M, Goodman S, et al. Albert: A lite bert for self-supervised learning of language representations[C]//International Conference on Learning Representations. 2019.
- [11] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//International Conference on Machine Learning. 2021: 8748-8763.
- [12] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[J]. Advances in Neural Information Processing Systems, 2016, 29.
- [13] Ren M, Triantafillou E, Ravi S, et al. Meta-learning for semi-supervised few-shot classification [C]//International Conference on Learning Representations. 2018.
- [14] Bertinetto L, Henriques J F, Torr P H, et al. Meta-learning with differentiable closed-form solvers [C]//International Conference on Learning Representations. 2019.
- [15] Wah C, Branson S, Welinder P, et al. The caltech-ucsd birds-200-2011 dataset[M]. California Institute of Technology, 2011.
- [16] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//

- Proceedings of the IEEE/CVF International Conference on Computer Vision. 2009: 248-255.
- [17] Ravi S, Larochelle H. Optimization as a model for few-shot learning[C]//International Conference on Learning Representations. 2017.
- [18] Triantafillou E, Zemel R, Urtasun R. Few-shot learning through an information retrieval lens[J]. Advances in Neural Information Processing Systems, 2017, 30.
- [19] Tian Y, Wang Y, Krishnan D, et al. Rethinking few-shot image classification: A good embedding is all you need?[C]//European Conference on Computer Vision. 2020: 266-282.
- [20] Rizve M N, Khan S, Khan F S, et al. Exploring complementary strengths of invariant and equivariant representations for few-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 10836-10846.
- [21] Ma J, Xie H, Han G, et al. Partner-assisted learning for few-shot image classification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10573-10582.
- [22] Chen D, Chen Y, Li Y, et al. Self-supervised learning for few-shot image classification[C]//IEEE International Conference on Acoustics, Speech and Signal Processing. 2021: 1745-1749.
- [23] Endres D M, Schindelin J E. A new metric for probability distributions[J]. IEEE Transactions on Information Theory, 2003, 49(7): 1858-1860.
- [24] Fuglede B, Topsøe F. Jensen-shannon divergence and hilbert space embedding[C]//IEEE International Symposium on Information Theory. 2004: 31.

附 录

A. 作者在攻读硕士学位期间的论文目录

- [1] ***, ***, He T, et al. Mirrored EAST: An Efficient Detector for Automatic Vehicle Identification Number Detection in the Wild[J]. IEEE Transactions on Industrial Informatics, 2023. (中科院 SCI 一区)
- [2] ***, Wang Y, Zhang Y, et al. Adversarial Bidirectional Feature Generation for Generalized Zero-Shot Learning Under Unreliable Semantics[C]//Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Cham: Springer Nature Switzerland, 2022: 639-654. (CCF-C)
- [3] ***, Huangfu L, ***, et al. Rethinking the Sample Relations for Few-Shot Classification[J]. Image and Vision Computing. (中科院 SCI 三区, 返修中)

B. 作者在攻读硕士学位期间参与的科研项目

- [1] 国家自然科学基金面上项目, 少样本学习特征生成与鲁棒性关键技术研究
- [2] 重庆市自然科学基金面上项目, 文本描述协同的双向生成式少样本学习研究

C. 学位论文数据集

关键词		密级	中图分类号	
少样本分类; 关系建模; 对比学习; 语义信息表示		公开	TP	
学位授予单位名称	学位授予单位代码		学位类别	学位级别
***	***		学术学位	硕士
论文题名		并列题名	论文语种	
基于多元关系建模的少样 本分类算法研究		/	汉语	
作者姓名	***		学号	***
培养单位名称			培养单位代码	
***			***	
学科专业	研究方向		学制	学位授予年
软件工程	计算机视觉		3 年	***
论文提交日期	***		论文总页数	24
导师姓名	***		职称	教授
答辩委员会主席			A	
电子版论文提交格式				
文本 (✓) 图像 <input type="radio"/> 视频 <input type="radio"/> 音频 <input type="radio"/> 多媒体 <input type="radio"/> 其他 <input type="radio"/>				