

# 小样本图像识别研究综述

孙景浩, 聂凯

(中国人民解放军 91550 部队, 辽宁 大连 116023)

**摘要:** 随着大规模数据集的发展, 基于深度学习的神经网络模型在人脸识别、智能驾驶、工业质量检测、医疗诊断等图像识别领域取得了优异的表现; 然而在实际场景的推广应用, 由于诸多因素的限制, 研究人员无法获取大量的满足要求的样本数据, 难以达到满意的识别效果, 因此进行小样本情形下的图像识别研究是十分有意义的, 文章系统地梳理了近年来小样本学习在图像识别领域的应用进展, 主要从基于数据增强、基于表征学习以及基于学习策略等 3 个方面对相关工作进行了介绍分析, 并根据目前的研究情况, 展望探讨了未来的研究方向。

**关键词:** 深度学习; 小样本; 图像识别; 数据增强; 迁移学习

## Research on the Application of Few-Shot Learning in Image Recognition

SUN Jinghao, NIE Kai

(Unit. 91550 of PLA, Dalian 116023, China)

**Abstract:** With the development of large-scale datasets, deep learning-based neural network models have achieved excellent performance in image recognition fields such as face recognition, intelligent driving, and medical diagnosis. However, in practical applications, due to various limiting factors, researchers are unable to obtain a large number of samples that meet the requirements. Therefore, It is of great meaningful to study image recognition under small sample conditions. This paper systematically reviews recent few-shot learning advances in the field of image recognition, introduces and analyzes related work from three aspects: data augmentation-based methods, representation learning-based methods, and learning strategy-based methods, and explores and discusses future research directions by current research status.

**Keywords:** deep learning; few-shot learning; image recognition; data augmentation; transfer learning

## 0 引言

得益于大规模数据集的发展, 基于深度学习的神经网络模型, 由于其强大的拟合能力, 在人脸识别、智能驾驶、工业质量检测、医疗诊断等图像识别领域取得了出色的表现, 有的模型在识别效率和准确率上甚至已经超过了人类。然而在实际场景的推广应用, 人们发现困难重重, 主要原因是缺乏大量的标注数据。在某些场景下, 如医疗、军事等涉及隐私、保密等相关领域, 获取大量数据是十分困难的。而在某些场景下, 如故障诊断等领域, 纵使可以获取大量数据, 但由于设备运行情况等原因, 获取的数据往往呈现长尾分布, 即故障样本量较少, 大部分是正常运行状态样本, 由于数据分布不均衡, 不适合直接用于训练。此外, 即便能够获取满足要求的数据, 也需要耗费大量的人力物力成本进行标

注, 因此, 我们在实际应用场景中面临的更多的情况是, 没有足够的满足要求的标注数据。而在训练数据匮乏的情况下, 深度学习模型极易因过拟合而导致性能下降。

相反地, 人类比较擅长从较少的数据样本中学习到新知识, 并做出准确的判断推理, 这种高效的数据利用能力正是机器学习所不具备但又非常需要的能力。为解决因训练样本过少而带来的过拟合问题, Li 等人提出小样本学习的概念, 并基于变分贝叶斯框架, 提出通过生成模型和层次先验的方法解决样本量过少的问题。早期的研究方法由于视觉表征能力不足及需要给定大量监督信息, 限制了模型的训练效果。随着深度学习神经网络的发展, 小样本图像识别由机器学习转向深度学习阶段。基于度量学习、迁移学习、元学习等思想的方法相继被提出, 在图像表征学习及模型优化训练方面有非常

收稿日期: 2025-10-15; 修回日期: 2025-11-25。

基金项目: 预研基金(50901020101)。

作者简介: 孙景浩(1990-), 男, 大学本科, 助理工程师。

引用格式: 孙景浩, 聂凯. 小样本图像识别研究综述[J]. 计算机测量与控制, 2026, 34(1): 1-8.

明显的进步。相较于传统的数据扩充方法，数据合成、特征增强等新方法能有效丰富样本多样性、提升合成样本质量，进而增强模型的泛化能力。小样本学习旨在通过提升模型的学习能力，实现对有限数据的高效利用，提升模型的推理能力，进而提升模型的泛化性能。

受人类学习模式启发，目前小样本学习问题的解决思路主要是，充分利用已有的大规模数据集，学习到可共享的先验知识，然后泛化到新类别小样本任务上。根据目前的相关研究进展看，本文从 3 个方向介绍小样本图像识别问题的解决方法：首先是通过数据增强，实现训练数据集的扩充；其次是对样本进行高效表征学习，提取样本类别的本质特征；最后是设计合理的训练策略，充分利用有限的训练数据。

1 基于数据增强的方法

基于数据增强的方法主要解决样本数据量由小变大的问题，通过扩充样本量，将小样本问题转化为一般问题，缓解训练过程中的过拟合现象。传统的数据增强方法主要包括两种：一种是在原图像基础上做简单的翻转、裁剪、缩放等几何变换操作进行数据扩充；一种则是通过添加噪声、改变亮度等简单图像处理手段进行数据扩充。传统的数据增强方法简单易实现，计算成本低，但是生成样本质量受限，多样性不足，难以满足小样本学习任务需要。目前主要有基于无标签数据、数据合成以及特征增强的 3 种数据增强方法。

1.1 基于无标签数据的方法

基于无标签数据的方法，旨在充分利用现实世界存在的海量的无标签数据，根据不同的应用场景，主要采取半监督学习、直推式学习和无监督学习等 3 种不同的方法。此 3 种方法的适用数据场景及各自优势见表 1。

表 1 基于无标签数据方法的适用数据场景及优势

方法类型	适用数据场景	优势
半监督学习	适用数据呈现长尾分布和类别不平衡的场景,在中等样本量时效果最佳	通过无标签数据补充标注数据,弥补样本分布不足,有效减少过拟合情况
直推式学习	适用开放式数据集和增量学习场景,在小样本增量时效果最佳	利用查询集样本补充训练数据集,提前接触查询样本的内在结构,直接优化目标任务效果
无监督学习	适用模型初始化学场景,尤其是在零样本时效果最佳	从大量无标签数据中学习通用特征,对特征提取器进行初始优化

半监督学习是机器学习研究的重点问题，因为在现实世界中存在大量的无标注数据，将半监督学习方法应用到小样本问题已经取得了一些成果。文献 [1] 基于半监督学习思想，附加一个无监督元训练阶段，使多个顶层单元接触大量的无标注数据，进而学习无标注数据

中的低密度分离器的多样化集合。文献 [2] 利用现有标注数据训练模型，然后为无标注数据生成伪标签，再将所有数据训练模型，重复多次直至模型收敛。文献 [3] 采用标签传播的方式，依靠近邻样本的标注信息为无标注数据生成伪标签。文献 [4] 在原型网络基础上，添加无标签数据进行训练，提升模型推理的准确率。

直推式学习可以看作是半监督学习的子问题，因为这两种方法在训练模型时都有无标签样本的参与，两者的区别主要表现在：直推式学习中使用的无标签数据来自每个任务的查询集，而半监督学习使用的无标签数据并不是来自特定查询集。文献 [5] 基于直推式的思想，提出了转导传播网络，在特征嵌入阶段，将所有的标注数据和无标注数据都映射到了向量空间中，使模型取得了较强的泛化能力。文献 [6] 提出一个交叉注意力网络和转换推理算法，迭代利用未标记的数据以增加查询集，从而使类别特性更具鉴别性。

无监督小样本学习是指辅助数据集由无标注样本组成，且模型不需要标注样本。这种学习方式更符合实际的应用场景，因为数据的本身语义远比人工标注提供的信息更为丰富，而且标注的信息通常只适用于解决某个特定任务，为解决不同任务而对样本进行多次标注，成本非常高。文献 [7] 基于聚类的思想，为不同的簇生成伪标签。文献 [8] 对每个无标注样本进行数据增强，丰富样本的分布多样性，转为有监督小样本学习问题来解决。

1.2 基于数据合成的方法

基于数据合成的方法，主要对现有的标注数据，通过添加外部信号，生成可利用的伪真实数据，实现数据集扩充。生成对抗网络<sup>[9]</sup>（GAN，generative adversarial network）基于博弈论的思想，由生成模块和鉴别模块组成，两个模块互相对抗，最终达到纳什均衡，生成以假乱真的样本。在此基础上，文献 [10] 提出条件生成对抗网络，通过在两个模块中加入条件信号，实现指定类别图像生成。文献 [11] 提出 Wasserstein 生成对抗网络，使用 Wasserstein 距离作为损失函数，优化了训练中的模式崩溃问题，同时提升了样本的多样性。文献 [12] 提出 infoGAN，通过最大化两个模块之间的互信息，使生成编码模块能够学习到潜在的特征表示，生成更可控的样本。

除了基于 GAN 的数据合成方法，文献 [13] 基于源数据与目标数据具有相同的数据分布的假设，将源数据中的方差信息和类内的变化差异迁移到目标任务的数据上，丰富样本数据的类内多样性。文献 [14] 将元学习与数据合成相结合，采用端到端的训练方法，将现有图像的一些属性和特征迁移到小样本上，从而生成具有不同变化的新样本数据。文献 [15] 对两张图像进行线

性组合生成新的图像, 并对两张图像的向量进行线性组合作为新图像的向量。

通过数据合成使样本量和多样性实现了扩充, 但是合成样本的数据分布和可解释性方面尚有欠缺, 同时合成样本的置信度需要进一步研究提高。

1.3 基于特征增强的方法

基于特征增强思想, 文献 [16] 提出一种属性引导增强模型学习合成数据的映射, 先将样本映射到特定空间, 然后使用训练好的编解码器生成多姿态的样本数据。为了解决姿态轨迹的离散问题, 文献 [17] 提出特征迁移网络, 用以描述随物体姿态变化引起的运动轨迹变化。

此外, 文献 [18] 提出 Delta 编码器, 提取同类样本之间的类内变形, 然后将类内变形应用到小样本类别上, 但此方法无法显著改善分类边界。文献 [19] 提出一个基于语义空间的 TriNet 模型, 将样本多级深度特征映射到语义空间, 由于特征映射的语义空间是和标签的语义空间对齐的, 因此模型在语义空间寻找和输入语义特征最相似的另一个类别的语义特征, 并加入一定范围的高斯噪声, 通过解码器输出伪数据, 补充到少量样本类别中。文献 [20] 在模型的表征层提出判别变分表征算法, 通过约束表征空间的先验分布, 使得表征分布迁移具有良好的类无关性。

综合来说, 基于数据增强的方法是解决小样本问题最直观的一种思路, 它直接解决了样本短缺的困境, 而且该方法比较灵活。表 2 梳理了 3 种数据增强方法的优缺点对比。但是由于实际样本数目较少, 容易导致生成的伪数据多样性不足, 并出现知识偏移情况。此外, 生成的伪数据并不是越多越好, 因为生成伪数据时, 不可避免的引入噪声及干扰特征, 会影响模型的训练效果。但是这种解决问题的方法具有普遍意义, 通常能与其他算法结合使用, 进一步提升模型性能。

表 2 不同数据增强方法的优缺点

方法类型	优点	缺点
基于无标签数据的方法	辅助数据易获取, 降低人工标注成本, 更能满足现实世界场景	无标签数据本身可能存在噪声, 且数据的分布需要满足任务数据要求, 否则效果提升有限
基于数据合成的方法	合成样本可以控制生成条件, 有效弥补个别种类样本稀少的问题	合成样本质量不稳定, 有可能出现模式崩溃, 难以满足复杂的数据分布
基于特征增强的方法	直接优化特征表示, 减少低层次的噪声干扰	依赖于特征提取器的质量, 可能会破坏特征之间的空间关系

2 基于表征学习的方法

表征学习旨在通过将样本数据映射到低维空间, 学

习到能够有效描述其本质特征的表征向量。常用的表征学习骨干网络是卷积神经网络。卷积神经网络可以通过卷积层自动提取图像的层次化特征, 通过堆叠多个卷积层, 可以学习到从低级到高级的复杂特征。虽然该方法取得了良好的应用, 但是其由于局部感受野的限制, 对全局上下文感知能力较弱。文献 [21] 提出在标准卷积核的元素间插入空洞, 在不增加核参数数量和计算量的前提下, 可以实现感受野膨胀。文献 [22] 提出特征金字塔网络, 通过自上向下融合高层语义特征和底层空间细节, 扩大不同尺度目标的感受野。文献 [23] 通过在标准卷积核中插入偏移量预测分支, 动态调整卷积核的采样位置, 根据目标形状灵活扩大感受野。

注意力机制也是提升表征学习效果常用的方法。基于人类在处理信息时的启发, 注意力机制能够通过精心设计的权重分配, 有针对性地关注图像中的重要特征, 忽略不太相关的部分, 减少不必要的计算, 而且能有效地捕捉到不同位置之间的依赖关系。但是, 注意力机制因权重分配设计, 导致计算复杂度变高, 且权重分配机制设计需要紧贴样本实际, 才能取得理想效果。文献 [24] 设计的关联感知全局注意力机制, 结合了选择性注意力和自注意力的优点, 提升了对图像语义的理解能力, 降低了图像识别的错误率。

基于自注意力同时具有最短最大路径长度和并行计算两项优势, 谷歌团队提出 Transformer 架构, 完全基于注意力机制, 没有任何卷积层或循环神经网络层, 主要是由多个包含多头自注意力和前馈网络的层叠加而成。2020 年, 谷歌团队将此架构应用于计算机视觉领域, 提出 ViT 模型, 突破卷积神经网络对局部感受野的依赖。针对 ViT 模型对大规模标注数据的依赖, Facebook AI 与索邦大学的研究团队共同提出 DeiT 模型, 在 Transformer 基础上引入蒸馏 token, 优化训练策略, 显著提升模型训练效率。

为有效聚合邻居样本信息, 文献 [25] 引入特征记忆模块, 用于存储训练数据特征, 以支持每个训练实例将邻居特征信息聚合到自身的表征学习中。图神经网络也是使用较多的聚合邻居信息的方法。其基于样本间距离, 构建节点关系网络图, 通过挖掘节点间关系, 更新边缘特征, 聚合邻居信息, 使样本特征更具判别性。文献 [26] 提出一种动态相关学习模型, 通过构建动态图卷积网络, 捕捉支持集图像和查询集图像特征间的依赖关系, 获取更精确的样本间关系及更具判别性的样本表征。

针对不同查询对象, 为了兼顾样本与任务的长期依赖性, 文献 [27] 提出交叉非局部神经网络, 来捕获样本内部及样本与特定任务之间的依赖性。针对不同的查询对象, 其可以综合考虑特定查询对象与支持集之间的

关系以及样本本身的全局信息，使得模型能够关注特定样本的不同区域，同时提取上下文感知的特征，并提出一种缩放因子来处理多尺度和上下文感知的特征。相较于其他基准模型，该方法对于样本中包含多个对象时，处理效果更为优秀，因为它能很好地感知特定查询任务的信息，提高了表征学习的精度。

此外，在表征学习中通过使用正则化技术也可以提升学习性能，提高对类内变化的鲁棒性。文献 [28] 提出使用正则化技术和自监督技术相结合，学习到特征流形的正则化，显著提升了样本的表征学习效果。

综上，基于表征学习的方法旨在尽可能提取样本类内的本质特征和类之间的区别特征，使得表征更具判别性，模型能获得较强的泛化能力，达到降低大数据量需求的目的，从实验结果看，提升表征学习能力可以使模型性能获得阶段性的提升<sup>[29]</sup>。表 3 梳理了不同表征学习方法的优缺点。

表 3 不同表征学习方法的优缺点

	优点	缺点
卷积神经网络	层次化特征提取，具有良好的平移不变性和局部不变性	局部感受野受限，长距离依赖捕捉不足
注意力机制	能针对性关注重要信息，擅长捕捉全局上下文依赖	局部特征提取能力偏弱，需设计合适的注意力模块才能优化特征提取效果特征记忆模块
图神经网络	通过聚合邻居信息，获得更具判别力的样本特征	当样本量变多时，计算和存储资源需求会明显变大
交叉非局部神经网络	既能捕获样本内部的上下文信息，又能捕获样本与特定任务之间的依赖	计算复杂度高，训练成本大
正则化	通过约束训练过程，使得模型对数据分布的微小变化具有鲁棒性	需采取合适的正则化方法，否则可能会导致无法捕捉关键特征、表征区分度低等问题

3 基于学习策略的方法

基于学习策略的方法主要分为基于度量学习的、基于迁移学习的和基于元学习的 3 种方法。

3.1 基于度量学习的方法

基于度量学习的方法，是指将样本映射到一个统一的特征空间，通过训练使得在该空间内的样本呈现类内集中而类间发散的特点，利用查询样本与训练样本之间的相似性度量来确定样本的所属类别。基于度量学习的常见网络主要有原型网络、孪生网络、匹配网络和关系网络等四种网络模型。

文献 [30] 于 2017 年提出原型网络，其主要思想

是在特征空间中，为每个类别找到一个类别原型，然后将待查询样本与每个类别原型进行相似度度量，从而得到类别结果。类别原型是以该类别所有样本表征的平均值计算的，相似度度量采用的则是欧氏距离。在训练过程中，通过不断最小化损失函数，使得同一类别的样本在空间上更为集中，不同类别的样本相互更加疏远。该方法在思想和实现上都比较简单，但是在训练数据较少、数据分布不均匀及噪声较大的情况下，非常容易导致类别边界偏差较大。针对这个问题，文献 [4] 利用半监督的思想，在训练数据集中加入了无标签数据，对类别的原型进行了修正。文献 [31] 认为在类别原型的计算中，不同样本的重要程度是不一样的，提出了基于样本级的注意力和特征级的注意力的原型网络，对样本特征进行加权计算，改进了原型的计算方法。

文献 [32] 提出的孪生网络，是由两个参数共享的卷积神经网络组成。在训练阶段，由训练数据集两两成一组，同类别打标签为 1，不同类别打标签为 0，然后每组数据经由卷积神经网络提取特征，映射到向量空间，再通过交叉熵函数计算损失，使得同类别样本损失最小化，不同类样本损失最大化。之后，文献 [33] 提出三元组孪生网络，将样本组扩展为 3 个样本，其中包含两个来自同一类别的样本和一个来自其他类别的样本，并设计了一个 hinge 损失函数，使得同一类别样本之间的距离远小于不同类别样本之间的距离，在每次训练中可以同时使同类别样本集中而不同类别样本发散。孪生网络设计思路比较清晰，同时由于其训练样本的配对策略，实际上相对于原训练数据集的样本数量进行了扩充，缓解了过拟合现象的产生。

文献 [34] 将匹配网络应用于小样本学习中，通过随机构建小批量数据进行情节式训练，提升了模型的泛化能力。匹配网络将待查询样本与支持集样本映射到嵌入空间，并使用余弦相似度度量完成样本分类。为了更好的建立支持集中样本之间的联系，匹配网络使用双向的长短期记忆网络（LSTM，long short-term memory）对每个样本在整个支持集的上下文中进行编码，以获得更加鲁棒的特征。在计算相似度时，采用注意力机制计算待查询样本与支持集样本之间的相似度关系，再将注意力值作为权重，以加权求和的方式获得测试样本的类别。文献 [35] 将嵌入函数替换为 4 层的卷积神经网络，采用双向 LSTM 和基于注意力机制的 LSTM 编码支持集样本和待查询样本的特征，最后采用平方欧氏距离进行分类识别，在类别数更多而样本数较少的场景下具有明显的优势。

文献 [36] 将关系网络应用于小样本学习中，将非参数度量方法替换为参数化的度量方法。关系网络主要由特征编码模块和关系模块组成。在特征编码模块中，

对支持集样本和待查询样本分别进行表征学习。在关系模块中, 将支持集样本和待查询样本的表征向量进行直接拼接, 然后通过回归的方式进行相似度计算, 完成分类识别任务。

小样本任务中, 因样本数据较少、分布不均衡且表征学习能力不足, 少量样本难以表征同一类别的特征分布, 容易出现类内距离较大, 类间距离较小, 进而导致类间重叠的情况, 严重影响度量网络效果。文献 [37] 从特征优化角度, 提出一种类间距均衡方法, 实现特征表示优化, 并以类间隔损失为样本类别保留足够的距离空间, 进而减少了类间重叠。文献 [38] 在计算类级特征时, 考虑到不同样本的属性不同且特征属性的多层次, 在特征融合阶段设计了权重网络, 提高了模型对类别特征的计算能力。

综上, 度量学习网络结构简单, 灵活性较强, 在小样本学习任务上常常可以取得较好的效果。但由于其在利用辅助数据训练完成后往往会直接应用于小样本任务中, 对不同任务的适应能力较差, 而且目前使用的度量函数改进空间较少。

### 3.2 基于迁移学习的方法

迁移学习旨在将一个任务上学到的先验知识迁移到另一个相关任务上, 从而弥补样本数据不足的问题。这一点在人类的学习中也有对应体现, 例如在学习“虎”这个类别之前, 如果已经学了“猫”这个类别, 那么学起来会比较容易。但是, 不同源域数据训练出的基础模型在目标任务上呈现的迁移能力是不同的, 很大程度上取决于源域数据与目标域数据的关联度, 即数据的分布差异和任务相似性。在这方面, 文献 [39] 提出域差异度量的概念, 认为当源域与目标域分布 H-divergence 度量差异越大时, 迁移后的泛化误差上界越高, 指出减小域差异是提升迁移效果的关键方法。文献 [40] 从特征层面分析了数据的特征关联度对迁移效果的影响, 认为底层特征关联度越高, 迁移效果越好; 而高层特征的关联度依赖两域的任务相似性, 任务相似性越高, 高层特征迁移效果越好。在两域关联度降低的情况下, 即使增加训练数据量, 迁移效果仍然会明显下降。

根据迁移方式的不同, 迁移学习可分为基于样本的、基于特征的、基于模型的和基于关系的四种迁移方式。目前应用最广泛的是基于模型的迁移学习方法, 该方法认为基于源域训练的模型包含了许多通用性的结构, 可以被迁移到目标域任务中。具体而言, 在模型迁移前先用源域数据进行预训练, 然后使用目标域数据对预训练的模型进行参数微调。以卷积神经网络说明, 卷积神经网络随着层数的叠加, 从通用特征逐步学习到特定特征。这样网络的底层神经网络一般是与任务无关

的, 可以在不同任务之间共享, 而网络的高层特征以及全连接层等需要在目标域数据上进行微调。

文献 [41] 提出一种简单的微调方法, 固定预训练得到的特征提取器参数, 只对分类器权重进行微调, 同时为了确保稳定性, 将分类器的向量积运算改为了向量夹角余弦运算。文献 [42] 在微调阶段, 将待查询数据的香农熵作为正则项融入损失函数中, 使得网络模型预测趋于概率集中。为了减少在目标域上泛化的计算复杂度, 文献 [43] 提出一种参数印刻方法, 通过前向传播的方式对分类器权重进行初始化, 加快了在小样本类上的泛化速度。文献 [44] 引入参数映射网络, 学习了一个从视觉特征到分类器权重的非线性映射。在此基础上, 文献 [45] 探索了不同类之间的相似度对新类的分类器参数的影响。

迁移学习基于相似任务之间的关联性进行, 容易理解。但是随着神经网络的深度加深, 模型的编码信息越来越缺乏可解释性, 使得模型参数调整变得更加困难。

### 3.3 基于元学习的方法

元学习旨在解决模型如何学习的问题, 相比于迁移学习侧重知识迁移, 元学习更注重通用的学习策略, 通过在大量的元任务中学习得到元知识, 可以加快在新任务上的收敛速度。元学习采用基于片段的训练模式, 每个片段包含从训练数据中提取的一组支撑集和查询集, 而且每个片段中的支撑集类别和对应数量是和目标任务一致的, 可以看成是一种二阶的学习机制, 如图 1 所示。



图 1 元学习方法示意图

文献 [46] 最早将元学习的思想应用于小样本学习中, 提出未知模型的元学习 (MAML, model agnostic meta learning) 模型, 通过优化的方法得到一个任务无关的初始参数, 在新任务上只需要很少的梯度下降步数就可以快速收敛。基于一阶的 MAML 模型, 文献 [47] 提出 Reptile 算法, 降低了计算成本和内存需求。更进一步, 文献 [48] 提出 MAML++ 算法, 针对 MAML 存在的不足进行了较全面的改进, 通过多步损失优化的方法改善训练中的不稳定问题, 采取导数退火加快训练速度, 通过学习内层优化学习率提升模型灵活性, 通过每一步相关信息的统计改善批量归一化效果。

不同于 MAML, 文献 [49] 提出一种 Meta-Learner LSTM 模型, 通过借助 LSTM 的长短期记忆特性, 学习一个可以直接输出梯度更新的元优化器, 相较传统的随机梯度下降优化方法, 具有明显的优势。文献

[50] 则基于神经图灵机的思想, 提出记忆增广网络 (MANN, memory-augmented neural network) 模型, 在元学习中引入外部记忆模块, 存储和利用历史任务的信息, 能对少量的样本类别进行记忆增强。但记忆模块设计较为复杂, 模型在训练中需要与记忆模块进行多次交互, 效率较低。

此外, 元学习与其他算法具有良好的结合性。文献 [51] 将元学习与迁移学习相结合, 提出元迁移学习, 利用元学习的思想, 优化迁移学习的训练过程, 为神经网络的权重参数学习缩放和移动策略, 提升对新任务的快速适应能力。

基于元学习的方法通过对元任务的学习, 使模型具备了一定的自动学习能力, 能够在解决小样本问题时更加灵活。但是其模型复杂度较高, 训练难度较大, 对任务数据分布要求也较高, 这导致元学习可能在某些任务下表现很差, 甚至难以收敛。不同学习策略的优缺点见表 4。

表 4 不同学习策略的优缺点对比

	优点	缺点
度量学习	模型简单, 易于实现, 只需通过合适的距离度量计算样本间相似度	对模型的表征学习能力要求较高; 受数据分布影响, 可能导致类别边界偏差较大
迁移学习	性能较好, 在源域数据预训练后仅需调整部分参数	要求源域与目标域关联度高, 否则可能导致负迁移
元学习	通过元任务学习可以快速在目标任务上适应, 泛化能力强	模型复杂度高, 训练时间长, 且模型效果仍有很大的改进空间

此外, 在实际场景中, 样本信息在不同的空间维度存在不同的语义表现形式, 如图像、文本等多种形式。而这些多模态数据蕴含着丰富的、可互相拓展补充的信息, 可以帮助模型提高对新样本的学习能力, 因此实现多模态信息融合是提升小样本学习效果的必要手段。文献 [27] 提出一种基于图像—文本对来识别视觉对象的模型, 该模型利用低秩多模态融合方法对多模态特征进行融合, 获得模态的共有特征, 对图像—文本对进行多尺度特征提取作为模态的特有特征, 最后通过多视图度量函数处理共有和特有两类特征。

4 未来研究展望

总体来说, 针对小样本场景下的图像识别问题, 研究人员提出了多种解决方法, 取得了一定成果, 但是仍有一些问题和方向需要我们继续努力研究。

1) 降低对大量数据的依赖性。目前的小样本学习工作大多是建立在已有相当规模的辅助数据集的前提

下, 但是在部分应用场景中, 由于数据隐私或专业领域等原因, 很难获得大量的辅助数据集。即使有足够的辅助数据集, 但是其数据分布可能与目标任务存在一定差距, 导致任务难以完成。因此, 无监督学习将是未来研究的重要方向。无监督学习能够在降低标注成本的前提下, 充分利用海量的无标签数据, 使模型学习到图像的底层通用特征, 进而在小样本数据上快速适应。

2) 提升模型的学习效率。现有的小样本学习算法在识别的准确率上已经取得了很大进步, 但是由于在表征学习阶段严重依赖深度学习神经网络, 因此其模型算法的复杂度依然很高。针对于此, 可以将强化学习和主动学习引入小样本学习中。在表征学习阶段, 通过强化学习奖励策略, 引导图像特征的筛选提取, 可有效降低计算规模。在半监督学习中, 引入主动学习范式, 选择有价值的样本进行标注, 达到减少计算需求的目的。

3) 模型的轻量化部署。小样本学习面对的目标数据集样本量有限, 但是其在训练阶段使用的辅助数据集却是十分庞大, 加之模型参数量巨大, 导致其对硬件平台的存储和计算能力有很高要求。在实际应用中, 很多场景模型需要在移动端部署且需要满足离线工作能力, 因此尽量减少模型的参数量和计算量就十分必要。当前的轻量化方法如高效轻量网络设计、量化、网络剪枝、知识蒸馏等虽然能够实现对网络的压缩, 但是模型精度也受到了影响。因此, 如何平衡模型的压缩率和精度将是十分有意义的研究。针对于此, 可以先根据模型精度指标要求与部署硬件性能, 明确压缩率和精度的阈值, 然后在不显著降低模型精度的前提下, 实现最大化压缩。而在提高压缩率任务上, 可以综合采取多种方法, 如先用剪枝降低冗余, 再进行量化、蒸馏等渐进式压缩, 避免激进操作导致模型精度崩溃。

4) 增强模型的鲁棒性。小样本学习研究通常是基于理想的数据假设, 如公开数据集或专门构建的领域数据集, 但在实际应用场景中是难以满足的, 采集的数据易受背景噪声干扰且有可能包含多个对象, 影响图像特征提取质量。针对这种情况, 文献 [52] 提出通过对训练集同步添加噪声来提高模型对实际环境数据的适应性, 但是效果不明显。对此, 可以考虑在表征学习阶段, 加入语义提示, 通过注意力机制, 聚焦图像中要提取的目标对象, 可有效抑制背景噪声的影响。

5) 脑神经科学的启发应用。深度学习主要通过神经网络来实现, 也就是通过模拟人类大脑的结构和功能处理数据。以卷积神经网络为例, 其卷积核的设计, 借鉴了人脑初级视觉皮层的边缘和方向特征提取功能, 其深层特征提取功能, 则借鉴了人脑高级视觉皮层的高效物体识别功能。而迁移学习和元学习, 则借鉴了人脑在复用已有知识和快速学习方面的认知功能。随着脑神经

功能更深入的研究,深度学习神经网络将获得更多的启示,提出更先进的模型理论,进一步提升其视觉任务处理能力。

## 5 结束语

由于实际图像识别场景中样本数据稀缺的问题,导致传统的深度学习神经网络算法难以取得理想效果,因此小样本场景下的图像识别越来越受到重视。本文对近年来小样本学习在图像识别领域的研究进展进行了分析阐述,同时结合当前的研究情况,提出了未来的研究方向。

## 参考文献:

- [1] WANG Y X, HEBET M. Learning from small sample sets by combining unsupervised Meta-Training with CNNs [C] //29th Proceedings of Neural Information Processing Systems, 2016: 244 - 252.
- [2] GRANDVALET Y, BENGIO Y. Semi-supervised learning by entropy minimization [C] //17th Advances in Neural Information Processing Systems, 2005.
- [3] ISCAN A, TOLIAS G, AVRITHIS Y, et al. Label propagation for deep Semi-Supervised learning [C] //2019 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5070 - 5079.
- [4] Ren M Y, TRIANTAFILLOU E, RAVI S, et al. Meta-Learning for Semi-Supervised Few-Shot classification [EB/OL]. <https://arxiv.org/abs/1803.00676>.
- [5] LIU Y, LEE J, PARK M, et al. Learning to propagate labels: transductive propagation network for Few-Shot learning [J]. International Conference on Learning Representations, 2019.
- [6] HOU R B, CHANG H, MA B P, et al. Cross attention network for Few-Shot classification [C] // 32nd Conference on Neural Information Processing Systems, 2019.
- [7] HSU K, LEVINE S, Finn C. Unsupervised learning via Meta-Learning [C] //Proceedings of IEEE International Conference on Learning Representations, 2019: 235 - 246.
- [8] QIN T, LI W, SHI Y, et al. Diversity helps: unsupervised Few-Shot learning via distribution Shift-Based augmentation [EB/OL]. <https://arxiv.org/abs/2004.05805>.
- [9] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [C]. 14th Proceedings of Neural Information Processing Systems, 2014: 2672 - 2680.
- [10] MIRZA M, OSINDERO S. Conditional generative adversarial nets [J]. Computer Science, 2014: 2672 - 2680.
- [11] ARJOVSKY M, SOUMITH C, LEON B. Wasserstein GAN [C] //International Conference on Machine Learning, 2017.
- [12] CHEN X, DUAN Y, I LOU THOOF R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets [C] //Proceedings of the Advances in Neural Information Processing Systems, Barcelona, 2016: 2172 - 2180.
- [13] HARIHARAN B, GIRSHICK R. Low-Shot visual recognition by shrinking and hallucinating features [C] //Proceedings of 2017 IEEE Conference on Computer Vision, IEEE, 2017: 3037 - 3046.
- [14] WANG Y X, GIRSHICK R, HEBERT M, et al. Low-Shot learning from imaginary data [C] //Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018: 7278 - 7286.
- [15] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization [C] //International Conference on Learning Representations, 2018.
- [16] DIXIT M, KWITT R, NIETHAMMER M, et al. AGA: attribute guided augmentation [C] //Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2017: 3328 - 3336.
- [17] LIU B, WANG X, DIXIT M, et al. Feature space transfer for data augmentation [C] //Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018: 9090 - 9098.
- [18] SCHWARTZ E, KARLINSKY L, SHTOK J, et al. Delta-Encoder: an effective sample synthesis method for Few-Shot object recognition [C] //18th Proceedings of Neural Information Processing Systems, 2018: 2845 - 2855.
- [19] CHEN Z T, FU Y W, ZHANG Y D, et al. Multi level semantic feature augmentation for One-Shot learning [J]. IEEE Transactions on Image Processing, 2019: 4594 - 4605.
- [20] 周琳钧. 小样本学习理论方法研究 [D]. 北京: 清华大学, 2021.
- [21] FISHER Y, VLADLEN K. Multi-Scale context aggregation by dilated convolutions [C] //International Conference on Learning Representations, 2016.
- [22] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE International Conference on Computer Vision, IEEE, 2017: 2117 - 2125.
- [23] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks [C] // Proceedings of the IEEE International Conference on Computer Vision, IEEE, 2017: 764 - 773.
- [24] 张直政. 神经网络的注意力机制研究 [D]. 合肥: 中国科学技术大学, 2021.

- [25] 李岁缠. 面向小样本的图像识别研究 [D]. 合肥: 中国科学技术大学, 2021.
- [26] LIU W, LI H, YU S, et al. Dynamic relevance learning for Few-Shot object detection [EB/OL]. <https://arxiv.org/abs/2108.02235>.
- [27] 赵佳宝. 基于多源信息的小样本算法研究 [D]. 上海: 华东师范大学, 2021.
- [28] MANGLA P, KUMARI N, SINHA A, et al. Charting the right manifold: manifold mixup for Few-Shot learning [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, IEEE, 2020: 2218 – 2227.
- [29] 彭云聪, 秦小林, 张力戈等. 面向图像分类的小样本学习算法综述 [J]. 计算机科学, 2022, 49 (5): 1 – 9.
- [30] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for Few-Shot L [C] // 17th Proceedings of Neural Information Processing Systems, 2017: 4077 – 4087.
- [31] GAO T Y, HAN X, LIU Z Y, et al. Hybrid Attention-Based prototypical networks for noisy Few-Shot relation classification [C] // AAAI Conference on Artificial Intelligence, 2019: 6407 – 6414.
- [32] KOCH G, ZEMEL R, SALAKHUTDINOV R. Siamese neural networks for One-Shot image recognition [C] // ICML Deep Learning Workshop, 2015.
- [33] YE M, GUO Y. Deep triplet ranking networks for one shot recognition [EB/OL]. <https://arxiv.org/abs/1804.07275>.
- [34] VINYALS O, BLUNDELL C, LILLICPAP T, et al. Matching networks for one shot learning [C] // Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, 3637 – 3645.
- [35] JIANG L B, ZHOU X L, JIANG F W, et al. One-Shot learning based on improved matching network [J]. Systems Engineering and Electronics, 2019, 41 (6): 1210 – 1217.
- [36] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare: relation network for Few-Shot learning [C] // Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018: 1199 – 1208.
- [37] LI B, YANG B, LIU C, et al. Beyond Max-Margin: class margin equilibrium for Few-Shot object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7363 – 7372.
- [38] 张智辉. 基于特征融合的小样本学习 [J]. 工业控制计算机, 2021, 34 (1): 61 – 63.
- [39] BEN-DAVID S, JOHN BLITZER, et al. A theory of learning from different domains [J]. Journal of Machine Learning Research, 2007 (8): 151 – 175.
- [40] YOSINSKI, JASON, CLUNE, et al. How transferable are features in deep neural networks? [C] // Advances in Neural Information Processing Systems, 2014: 3320 – 3328.
- [41] CHEN W Y, LIU Y C, KIRA Z, et al. A closer look at Few-Shot classification [C] // International Conference Learning Representations, 2019.
- [42] DHILLON G S, CHAUDHARI P, RAVICHANDRAN A, et al. A baseline for Few-Shot image classification [C] // International Conference on Learning Representations, 2020.
- [43] QI H, BROWN M, LOWE D G. Low-Shot learning with imprinted weights [C] // Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018: 5822 – 5830.
- [44] QIAO S Y, LIU X C, SHEN W, et al. Few-Shot image recognition by predicting parameters from activations [C] // Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018: 7229 – 7238.
- [45] ZHOU L, CUI P, YANG S, et al. Learning to learn image classifiers with visual analogy [J]. IEEE, 2020.
- [46] FINN C, ABBEEL P, LEVINE S. Model-Agnostic Meta-Learning for fast adaptation of deep networks [C] // Proceedings of 34th IEEE International Conference on Machine Learning, IEEE, 2017: 1126 – 1135.
- [47] NICHOL A, ACHIAM J, SCHULMAN J. On First-Order Meta-Learning algorithms [EB/OL]. <https://arxiv.org/abs/1803.02999>.
- [48] ANTONIOU A, EDWARDS H, STORKEY A. How to train your MAML [EB/OL]. <https://arxiv.org/abs/1810.09502>.
- [49] RAVI S, LAROCHELLE H. Optimization as a model for Few-Shot learning [C] // Proceedings of 5th International Conference on Learning Representations, 2016: 458 – 469.
- [50] SANTORO A, BARTUNOV S, BOTVINICK M, et al. Meta-Learning with Memory-Augmented neural networks [C] // PMLR, International Conference on Machine Learning, 2016: 1842 – 1850.
- [51] SUN Q, LIU Y, CHUA TS, et al. Meta-Transfer learning for Few-Shot learning [C] // 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019: 403 – 412.
- [52] YAN, YAO, REN, et al. CAN: cascade augmentations against noise for image restoration [C] // IEEE Transactions on Image Processing, IEEE, 2025: 5131 – 5146.