

摘要

近年来，深度学习技术的蓬勃发展推动了视觉表征学习在各种视觉任务中的突破，在图像分类、目标检测、语义分割等领域取得了显著成就。通过深度神经网络，视觉模型能够从海量数据中有效地提取出特征表示，极大地提升了机器对视觉信息的理解能力。然而，视觉模型在实际应用中依旧面临着一系列严峻的挑战，限制了其在复杂、动态的真实环境中的表现。例如，当视觉模型在面对新的场景或数据分布时，模型在保持准确预测的泛化能力方面仍然存在明显的局限。在实际应用中，模型往往依赖于已知的数据进行训练，但一旦遇到来自不同领域或环境的新数据，模型性能会出现明显下降。因此，在视觉模型适配不同下游任务时，如何确保其泛化能力成为突出问题。

构建具有更强大泛化能力的视觉模型成为当前研究中的迫切需求。如何使模型在训练过程中学习到的规律和特征有效地迁移到未知的数据、任务或多样化场景中，提升其在跨域任务中的表现，已经成为提升视觉模型应用落地的关键。因此，针对这些挑战，提升视觉模型的泛化能力，探索更加有效的泛化学习方法，已成为学术界和工业界亟待解决的重要问题。

本文从深度学习中视觉模型的泛化性出发开展深入研究，针对视觉模型的领域泛化需求、多任务适配需求、多元化感知需求，以及面向高效适配的模型选择需求，着力解决上述各方面存在的问题与挑战，并从表征学习视角对模型泛化方法展开研究。本文的主要创新性贡献如下：

第一，提出了一种高斯分布先验引导下的表征不确定性建模方法，针对领域泛化中的特征统计分布偏移现象，引入一种特征扰动机制，有效刻画了模型学习过程中的不确定性；进一步，设计了一种不确定性嵌入式学习框架，利用重参数化优化技术，有效估计高斯先验分布核参数，构建了随机采样驱动的动态化表征，增强模型学习中应对领域分布偏移的泛化能力。实验结果表明，该方法显著提升了模型在多个视觉感知任务中的泛化能力，在多域图像分类任务上提升约 6.1% 准确率 (Acc)，在跨域语义分割任务上提升约 6.0% 平均交并比 (mIOU)，在实例检索任务中提升约 8.2% 平均精确度 (mAP)，在多种图像降质条件下的分类任务上提升约 2.8% 准确率 (Acc)。

第二，提出了一种基于图像自相关性的视觉单词软标签离散化方法，针对代理任务监督中存在的表征歧义性问题，通过挖掘视觉分词器的概率分布以及图像的自相关性，设计了视觉软标签集成策略，以进一步挖掘图像内部语义和结构信息。不依赖于特定数据和任务，模型可以从海量无标记数据中学习对下游任务更通用的视觉表征。实验结果表明，通过该方法在 ImageNet-1K 上进行自监督预训练后，模型在多个迁移

学习下游感知任务的性能得到了有效提升,在图像分类任务提升约 0.9% 准确率 (Acc),在目标检测任务上提升约 0.5% 平均精确度 (mAP),在分割任务上提升 1.4% 平均交并比 (mIOU)。

第三,提出了一种混合专家模型的多源视觉感知信息融合方法,针对视觉语言模型感知信息源受限的问题,构建了融合多视觉专家模型感知能力的上下文建模框架,提升了模型对视觉信息的整合能力。进一步设计了基于知识蒸馏的监督式感知信息融合方法,充分发挥各类视觉专家模型的能力。实验结果表明,该方法不仅有效提高了模型的多粒度视觉感知信息获取能力,还在多个语义分析理解任务展现了优异的泛化能力,在六个图文理解任务上提升约 3.6% 准确率 (Acc),在四个图文问答任务上提升了 2.7 % 准确率 (Acc)。此外,利用该感知信息融合方法,构建了百万级大规模图文对齐数据集 DenseFusion-1M,并展示了高质量的图文描述数据在视觉理解和问答方面的效果。

第四,提出了一种基于迁移学习过程特性模拟的模型泛化能力评估方法,构建了模型适配过程与势能变化模拟的对偶分析框架,提出了一种基于物理先验引导的迁移过程动力学建模方法,以模拟表征空间的动态变化,并利用动力学模型的对偶性验证了方法的合理性。该方法通过捕捉表征空间的变迁特性,实现了更有效的模型泛化能力度量。实验结果表明,所提方法在模型优选应用任务上,有效地提升 12% 的模型优选评估准确率 (Acc),评估效率提升了 2.5 倍。

本文围绕表征学习中的视觉模型泛化问题,从以上多个方面开展了方法设计与实验分析,从模型优化和选择两个模型泛化环节,针对受限数据扩增、海量数据利用、数据模态融合以及模型能力评价方面,递进式提出一系列表征泛化学习方法,有效解决模型泛化能力提升与评估的关键难题。所提方法在多种视觉任务基准测试数据集上,相比前沿方法展现出明显性能优势,验证了方法设计的有效性。

关键词: 领域泛化; 自监督学习; 多源感知信息融合; 模型泛化能力评估