

机器人视觉识别大数据增强策略与小样本学习结合

宋孟月¹, 张书伟¹, 万林炜¹, 刘子一², 朱耀晨¹, 张建刚³

¹河南工业大学人工智能与大数据学院, 河南郑州, 中国

²河南工业大学信息科学与工程学院, 河南郑州, 中国

³科大讯飞股份有限公司, 安徽合肥, 中国

【摘要】在机器人视觉识别这个领域当中, 小样本学习存在着因训练数据稀缺致使模型泛化能力较弱的情况, 而大数据提高策略尽管可扩充数据, 然而却很难直接适配小样本场景, 本文创新性地想出了把大数据提高策略和小样本学习结合起来的办法, 借助改进数据提高技术来适应小样本的特点, 引入元学习算法以便可快速适应新类别, 构建基于注意力机制的神经网络来强化特征提取, 并且在多个公开数据集以及机器人实际场景当中展开实验。结果显示, 该方法提升了小样本视觉识别性能, 在少样本情形下对于新类别的识别准确率相较于传统方法有了大幅提高, 为机器人在复杂多变环境里的视觉识别提供了更有效的解决办法。

【关键词】机器人视觉识别; 大数据增强策略; 小样本学习; 元学习; 注意力机制

1.引言

随着机器人技术于工业、服务、医疗等领域的广泛运用, 机器人对于复杂环境里目标物体的准确视觉识别需求变得日益急切, 在实际应用情形中, 机器人面临着不少挑战, 其中训练数据的获取以及处理成了关键难题, 一方面, 小样本学习场景广泛存在, 像在新的工作环境里, 机器人可能只有少量特定物体的样本用于学习识别, 传统的深度学习模型依靠大规模标注数据来训练, 在小样本状况下很容易出现过拟合, 致使模型泛化能力欠佳, 无法准确识别新碰到的目标物体[1]。另一方面, 大数据提高策略是依靠对现有数据进行变换、合成等操作来扩充数据集, 以此提升模型性能, 然而这些策略在小样本场景下难以直接起到作用, 甚至可能因过度提高致使数据失真, 影响模型效果, 怎样把大数据提高策略与小样本学习有效结合, 成为提升机器人视觉识别能力的关键研究方向。

2.相关理论基础

2.1 小样本学习概述

小样本学习也就是 Few Shot Learning, 简称为 FSL, 它主要关注的是在仅有极少量样本数据的情形下, 让模型可迅速实现学习, 并精准识别出新的类别, 其核心面临的难题在于, 要克服因样本稀缺所引发的信息不充足状况, 避免模型出现过拟合现象, 还得拥有良好的泛化能力, 以此来应对新的类别[2]。当下小样本学习方法主要囊括基于度

量学习、基于模型微调以及基于元学习等几个方向, 基于度量学习的方法着力于学习一个恰当的特征空间, 让同一类别的样本在这个空间里距离较为接近, 不同类别的样本距离相对较远, 就像孪生网络也就是 Siamese-Network, 它借助对比学习的方式来训练网络, 依靠计算样本之间的相似度开展分类工作[3]。基于模型微调的方法一般先是在大规模的基础数据集上开展预训练, 之后利用少量的新类别样本对模型进行调整, 以便使其适应新的任务, 不过在小样本的情形下, 微调容易致使模型出现过拟合, 基于元学习的方法则是学习怎样去学习, 也就是借助在多个小样本学习任务上进行训练, 让模型掌握可快速适应新任务的能力, 典型的比如 MAML, 即 Model Agnostic Meta Learning 算法, 它依靠在多个任务上快速更新模型参数, 使得模型可在新的小样本任务中快速达到收敛状态。

在机器人视觉识别领域, 小样本学习有着颇为关键的应用意义, 就拿机器人的物流分拣场景来说, 新类型的货物可能会持续不断地出现, 然而在实际开展操作时, 想要快速获取大量新货物样本用以模型训练存在一定难度, 在这样的情况下, 小样本学习技术可以让机器人依据少量新货物样本实现快速学习并且准确识别, 提升物流分拣的效率以及准确性。

2.2 大数据增强策略原理

大数据提高策略借助对原始数据开展各

类变换操作，生成数量众多的新样本，以此扩充数据集的规模以及多样性，常见的数据提高技术覆盖几何变换、颜色变换、裁剪与拼接以及生成对抗网络等方式，像旋转、平移、缩放这类几何变换，可模拟物体于不同视角和位置时的视觉效果，提高模型针对物体空间变化的鲁棒性。颜色变换包含调整亮度、对比度、饱和度等，可让模型适应不同光照与色彩条件下的图像识别，裁剪与拼接操作是针对图像实施局部裁剪或者不同图像部分的拼接，丰富图像的内容结构，生成对抗网络由生成器和判别器构成，生成器负责生成新的数据样本，判别器用于判断样本是真实数据还是生成的数据，依靠二者的对抗训练，生成高质量的合成数据，扩充数据集[4]。

这些大数据提高策略于深度学习里有着广泛的应用，切实提升了模型在大规模数据训练状况下的泛化能力以及鲁棒性，就像在图像分类任务当中，借助数据提高所生成的多样化样本，可使模型学习到更具普遍性的特征，在面对不同场景的图像时，依旧可准确地进行分类，然而在小样本学习场景里，鉴于样本基数较小，直接运用这些大数据提高策略或许会出现数据提高过度或者生成的数据与真实数据分布差异较大这类问题，影响模型的学习效果。

3.融合方法设计

3.1 改进的数据增强技术适配小样本

小样本场景有其特殊性，要改进传统大数据提高技术，在几何变换方面，采取自适应的旋转、平移以及缩放策略，分析少量样本的特征分布，借助主成分分析也就是PCA来提取样本的关键结构特征向量，依据这个确定合理的变换范围，防止过度变换让数据失真，比如说，对于有特定形状和结构的物体样本，像机械臂抓取的螺栓、螺母等工业零件，依照其形状的对称轴或者主要结构方向，凭借计算样本轮廓的最小外接矩形角度，把旋转角度限制在正负15度范围内，保证变换后的样本依然保留核心结构特征。在颜色变换中，利用小样本的颜色统计信息，像RGB三通道的均值和方差，构建颜色分布模型，采用高斯扰动方法做更精细的亮度、对比度和饱和度调整，让提高后的图像颜色变化幅度控制在原始样本颜色分布的95%置信区间内，更接近真实场景中可能出现的光照变化[5]。

在小样本情形下，引入基于生成对抗网络的条件生成模型，此条件生成模型与传统GANs不一样，它在生成器和判别器里引入类别标签以及样本特征向量当作条件约束，借助给定的少量样本的特征信息，生成和这些样本有着相似特征与语义的新样本，生成器把小样本的特征向量当作条件输入，联合100维随机噪声向量，经反卷积层逐步生成 $64 \times 64 \times 3$ 的图像样本。判别器采用多任务学习框架，要判断生成的样本是不是真实，还要借助额外的分类分支判断生成样本与给定条件的匹配程度，借助这种双约束机制，生成的样本在维持原始小样本关键特征之时，可模拟不同姿态与背景下的视觉变化，在扩充数据多样性之际，把生成样本与原始样本的特征余弦相似度提升到0.85以上，保证和原始小样本数据的强相关性。

3.2 引入元学习算法快速适应新类别

元学习算法于小样本学习领域呈现出独特优势，其“学习如何学习”的核心机制可使模型迅速适应新类别，把元学习算法同改进的数据提高技术相结合，构建出“提高-元训练”闭环框架，以此提高机器人视觉识别在小样本场景中的性能，运用基于模型无关元学习的改进框架，在500个多样化小样本学习任务上开展元训练[6]。每个小样本学习任务由一个含有少量样本的支持集以及一个用于测试的查询集构成，借助分层抽样保证任务间的类别分布存在差异。

在元训练阶段运用改进的数据提高技术对支持集作动态扩充，具体是每个样本生成3个提高样本，之后用扩充后的支持集对模型开展快速梯度下降更新，其中学习率设定为0.01且迭代3次，以此获得针对当前任务的模型参数，借助计算查询集在更新后模型上的损失值，并反向传播至元参数实施优化，让模型参数 θ 达成 $L(\theta') = \min \nabla \theta L_{\text{task}}(f_{\theta'}(\text{query}))$ 的优化目标。依靠于多个这样的小样本学习任务上展开训练，训练的总迭代次数为100000次，模型可学习到有强迁移性的元参数，当面对新的小样本视觉识别任务时，模型依据学习到的元参数，只需利用新任务的5个样本进行2次梯度更新便可快速收敛，在机器人抓取新物体的场景里，能将类别适应时间缩短至0.5秒以内。

3.3 构建基于注意力机制的神经网络

为能更妥善地处理经过提高的小样本数据，构建了融合通道注意力与空间注意力的

双通道注意力神经网络即 DANet，在神经网络结构里，注意力模块运用残差连接方式嵌入 ResNet18 的 conv3_x 和 conv4_x 阶段，可自动学习数据中不同特征的关键程度，给关键特征赋予更高的关注权重，以此强化特征提取效果。通道注意力分支借助全局平均池化把特征图压缩成 $1 \times 1 \times C$ 的全局特征向量，接着凭借两个全连接层（生成 $1 \times 1 \times C$ 的通道权重向量，空间注意力分支是对特征图进行通道维度的最大值池化和平均池化，将结果拼接后经 3×3 卷积生成 $H \times W \times 1$ 的空间权重图。

把得到的通道权重跟空间权重做矩阵相乘，生成 $H \times W \times C$ 的注意力掩码，之后同原始特征图开展逐元素加权相乘，如此一来模型就能更关注和目标物体有关的特征区域，像物体的轮廓边缘、纹理细节等，同时抑制无关背景信息的干扰，比如杂乱的桌面、阴影区域等，借助热以此来可视化分析可知，该注意力机制可把特征响应值的 70%以上集中于目标物体区域，在小样本数据经过提高后，可以有效过滤提高过程里引入的噪声特征，使得模型在少样本情形下对关键特征的提取能力提升 20%以上。

4.实验与结果分析

4.1 实验数据集与设置

为对融合方法有效性给予验证，借助多个公开数据集开展实验，这些数据集囊括 MNIST、CIFAR-10 以及 Caltech256 等，于 MNIST 数据集中，挑选部分数字类别当作小样本学习的新类别，每个新类别仅给出 5 个样本用作训练数据，其余样本用于测试，在 CIFAR-10 数据集中，随机挑选 5 个类别作为新类别，每个类别提供 10 个样本用于训练，测试集使用剩余样本。Caltech256 数据集有 256 个类别，从中选取 20 个类别作为小样本学习对象，每个类别提供 15 个样本作为训练数据。

搭建机器人实际场景实验平台，以此来模拟机器人于不同环境下的视觉识别任务，在该实验场景当中放置多种有不同形状、颜色以及材质的物体，并且设置不同的光照条件与背景干扰，机器人借助摄像头采集图像数据，把这些数据当作小样本学习的输入内容，实验设置了对比组，分别运用传统小样本学习方法以及未改进的大数据提高策略结合小样本学习的方法，与所提出的融合方法展开对比。实验环境是配备 NVIDIA

RTX3090 GPU 的工作站，利用 Python 编程语言以及 PyTorch 深度学习框架来进行模型的搭建、训练以及测试[7]。

4.2 实验结果与性能评估

在针对公开数据集所开展的实验当中，以 MNIST 数据集为例，传统孪生网络在小样本情形下的识别准确率大概只有 65%，借助模型微调的方法，准确率能达到约 70%，未经过改进的数据提高与小样本学习相结合的方法，可将准确率提升至 75%左右，而所提出的融合方法，其准确率已然超过了 85%。在 CIFAR-10 数据集上，传统方法的准确率处于 40%至 50%之间，未改进的方法准确率约为 55%，融合方法的准确率则提升到了 65%以上，Caltech256 数据集的实验结果说明，传统方法的准确率在 30%至 40%，未改进的方法为 45%左右，融合方法达到了 55%以上。

在机器人实际场景的实验当中，针对不同物体类别的识别准确率展开了统计工作，最终结果显示，传统小样本学习方法于复杂环境下的平均识别准确率大约为 45%，未经过改进的数据提高与小样本学习方法相结合后，平均准确率提升至 50%左右，而融合方法的平均识别准确率则达到了 60%以上。经过对比可发现，所提出的融合大数据提高策略以及小样本学习的方法，在不同数据集以及实际场景里，都要明显优于传统方法和未改进的结合方法，切实有效地提升了机器人在小样本情形下对新类别物体的视觉识别准确率[8]。

对模型的泛化能力做评估，于不同数据集里，把训练集与测试集按照不一样的比例去划分，观察模型在不同比例情形下的识别准确率变化情况，结果说明，融合方法在训练集样本数量比较少的时候，依旧可以维持较高的准确率，并且随着训练集样本数量不断增加，准确率提升变得日益较大，该方法有良好的泛化能力，可更为出色地适应不同规模的小样本学习场景。

5.结论

本文所提出的将大数据提高策略与小样本学习相融合的方法，借助改进数据提高技术、引入元学习算法以及构建基于注意力机制的神经网络，切实提升了机器人在小样本视觉识别场景中的性能，实验结果显示，该方法在多个公开数据集以及机器人实际场景里，相较于传统方法提高了识别准确率与泛

化能力，为机器人在复杂多变环境中准确识别目标物体给予了更有效的解决办法。

参考文献

- [1]张红斌, 蒋子良, 熊其鹏等.基于改进的有效区域基因选择与跨模态语义挖掘的图像属性标注[J].电子学报, 2020, 48 (04) : 790-799.
- [2]张红斌, 邱蝶蝶, 邬任重等.基于分层基因优选多特征融合的图像材质属性标注[J].自动化学报, 2020, 46 (10) : 2191-2213.
- [3]李军强.基于 CapsNet-PMR 的 BLE 定位算法与故障信标识别技术的研究[D].西安电子科技大学, 2024.



- [4]刘丽, 赵凌君, 郭承玉等.图像纹理分类方法研究进展和展望[J].自动化学报, 2018, 44 (04) : 584-607.
- [5]化春键, 张宏图, 蒋毅等.基于图卷积与多头注意力的图文跨模态检索[J].光电子·激光, 2024, 35 (09) : 925-933.
- [6]薛志杭.基于跨模态语义挖掘的文本生成图像技术研究[D].北京交通大学, 2023.DOI:10.26944/d.cnki.gbfju.2023.0015 62.
- [7]普健杰, 尹兆良.基于 PyTorch 深度学习库的物联网恶意行为检测框架[J].网络安全技术与应用, 2023 (06) : 43-45.
- [8]邱增帅.面向室外环境的移动机器人自主场景理解[D].大连理工大学, 2018.