

# 基于深度学习的图像识别技术在计算机视觉中的应用与优化

何南

桂林师范学院 广西桂林 541199

**摘要:** 本文对该技术发展的脉络做了系统的梳理,指出其在安防、医疗、自动驾驶等场景起到关键作用的同时,也面对着复杂环境识别的难题、训练数据短缺以及对高昂硬件依赖等实际困境,为应对这些困境,本文针对多尺度特征融合、知识蒸馏与模型轻量化、动态卷积自适应以及硬件感知网络架构搜索等关键技术的优化途径展开深入探讨,采用算法与工程层面的协同创新,可有力促进图像识别系统的普遍性、效率与实用化水平,为其更深度、更普遍地赋能各个行业筑牢根基。

**关键词:** 深度学习; 图像识别技术; 计算机视觉

## 引言

如今,图像识别技术已从实验室走向产业前沿,大幅改变着社会生产与生活样式,当技术走进真实、开放、复杂的应用场景时,其内在固有的局限性逐渐显现出来了,就像应对动态复杂环境时适应能力不足、对大规模标注数据依赖显著,以及出现对算力资源高消耗这样的问题,共同阻碍了技术的进一步普及与深化发展,在肯定已有的成就之际,怎样有系统性地分析这些应用困境,并探求切实可行的优化途径,成为助力该技术持续健康进步的核心课题。

## 一、基于深度学习的图像识别技术发展现状

随着计算能力的高速提升与大数据的大量积累,基于深度学习的图像识别技术实现了质的突破性进展,已成为计算机视觉领域的核心推进动力,其发展现状体现出从基础模型取得突破到开展多元化、精细化应用探索的清晰脉络<sup>[1]</sup>。

技术革新的核心标志为神经网络,尤其是卷积神经网络(CNN)架构持续革新与性能实现突破,早期的LeNet、AlexNet等模型奠定了CNN在图像分类任务中的基础地位,VGG、GoogLeNet(Inception)、ResNet等系列模型凭借增加网络深度、引入残差连接与多尺度特征融合等创新方式,极大提高了模型的特征表达能力与识

别准确率,切实减轻了深度网络里梯度消失等问题,注意力机制(如SENet、CBAM)和Transformer架构在视觉任务中的成功迁移实践,进一步拉动图像识别技术朝着更高效、更具判别特性的方向前行,让其能够更高效地获取图像里的长程依赖与关键区域信息。

就应用情况而言,该技术已从实验室走向大规模的产业化落地,覆盖到社会生产和生活的各个层面,在安防监控领域,人脸识别、行为分析等技术已大规模应用;在医疗影像分析相关工作中,深度学习辅助疾病诊断(像肺部CT结节检测、眼底图像分析这些)明显提升了诊断效率与精度;在自动驾驶的范畴里,对车辆、行人及交通标志实施实时精准识别,是确保行车安全的关键技术依托;处于工业制造这个范畴中,智能质检利用图像识别自动去检测产品的缺陷,提升了生产线的智能化水平,在智能手机影像范畴、社交媒体内容审核事项、遥感图像解译业务甚至艺术创作等领域,深度图像识别技术都扮演着不可替代的角色。

## 二、图像识别技术在计算机视觉中的应用困境

### (一) 复杂背景下目标识别难度大

图像识别技术在实际场景中面临的第一个困境,是复杂背景和动态环境造成的识别挑战。现实世界并非像实验室那样的受控环境,目标物体频繁出现在杂乱、多变且具有干扰性的背景环境中。在实施自动驾驶期间,车辆要从光影杂乱无章的街道、人群密集的区域以及各种天气条件(如雨、雪、雾)中,精准辨认出行人、车辆和交通标志;在安防相关的监控工作中,目标人物也许处于拥挤的人群中、快速移动或呈现部分被遮挡状态;

**作者简介:** 何南(1974.12-),女,瑶族,湖南江华人,研究生,副教授,研究方向:图形图像处理、计算机应用。

在农业遥感的场景中，作物识别要克服不同生长期、杂草干扰和复杂地形地貌带来的难题<sup>[2]</sup>。这些复杂背景使模型面临多重困境：一是背景噪声的干扰，与目标不相关的纹理、颜色以及物体，容易导致模型产生误检或漏检；二是目标外观发生变化，同一目标在不同光照、角度、尺度、遮挡程度情况下，呈现的视觉特征差异极大，要求模型具备极强的鲁棒性与不变性特征提取能力；三是小目标和密集目标的检测难题，在广角或远景的图像中，小尺寸目标的有效特征极少，而密集放置的目标存在边界模糊、相互粘连的情况，导致精准定位与分类工作陷入困境。

### （二）样本数量有限影响模型训练

采用深度学习的模型，尤其是深度神经网络这类模型，一般被视为“数据饥渴”型模型，其卓越性能的发挥主要依赖大规模、高质量、带标注的训练数据集。在大量专业或特定的应用领域内，获取充足样本面临极大挑战，这构成了图像识别技术落地应用的第二个主要阻碍。这种困境主要有以下几个方面的形式体现：数据获取成本极高。在工业缺陷检测工作中，合格品数量远远超过缺陷品，导致缺陷样本数量稀少且形态各异；在医疗影像分析领域，涉及患者隐私和数据安全问题，且病灶图像标注工作需要领域专家（如放射科医生）耗费大量时间和精力，导致标注数据稀少且成本高。二是长尾分布问题，在实际的分类任务中，类别分布通常极为不平衡，大量样本集中在少数常见类别，而不少有意义但稀有的类别（如特定种类的稀有病症、特殊类型的机械故障）样本数量极少，致使模型难以掌握这些少数类别的有效特征，导致对其识别表现不佳。三是领域适应与泛化能力不足，通过特定数据集训练得到的模型，若应用于数据分布不同的新场景，性能通常会大幅下降。

### （三）对硬件设备和能源消耗要求高

图像识别技术，尤其是追求高精度的大型深度学习模型，其训练和部署过程对计算硬件和能源消耗要求十分苛刻，这是影响其普及、边缘应用及可持续发展的主要困境。在模型训练阶段，处理海量数据、优化数百万乃至数十亿参数需要强大的计算集群，通常会依赖价格昂贵的GPU或TPU阵列，且持续运行时间为数天甚至数周。此过程消耗了大量电量，产生了明显的经济开支，还引发了关于人工智能碳足迹的环境担忧<sup>[3]</sup>。在模型推理和部署阶段，挑战同样十分严峻。诸多先进的图像识别模型（如大型卷积神经网络或视觉Transformer）参数

数量大、计算复杂度高，若要实现实时或准实时推理，高性能处理器和大内存带宽必不可少，这给资源匮乏的边缘设备（如智能手机、嵌入式传感器、自动驾驶汽车的车载计算单元、无人机）带来了极大压力。在这些设备上，功耗、散热、计算能力和存储空间均受到严格限制。

## 三、基于深度学习的图像识别技术在计算机视觉中应用的优化建议

### （一）多尺度特征的融合优化

为了应对复杂背景下目标尺度多变、小目标及密集目标识别困难的挑战，多尺度特征融合优化已成为提升模型鲁棒性与精度的核心技术方向。传统卷积神经网络通常通过层叠卷积和下采样操作提取特征，但随着网络加深，特征图分辨率逐渐降低，虽然语义信息愈发丰富，但空间细节大量丢失，这对精确定位和小目标检测极为不利。多尺度特征融合的核心思想在于协同利用网络浅层、中层与深层特征。浅层特征具有高分辨率，蕴含丰富的纹理、边缘等细节信息，有利于目标定位和小尺度物体检测；深层特征则经过高度抽象，具有强大的语义信息，有利于物体分类与理解。通过有效的结构设计将这些不同层次、不同语义强度的特征进行融合，可以使模型同时具备“看得清”细节和“读得懂”内容的能力。典型的技术路径包括：其一，特征金字塔网络（FPN）及其变体，采用自上而下（Top-Down）的路径将高层语义特征上采样，并与浅层高分辨率特征进行横向连接与融合，从而在不同尺度上构建具有强语义信息的特征图，显著提升了对多尺度目标的检测性能，尤其是在处理尺度差异巨大的目标时效果明显<sup>[4]</sup>。其二，U-Net类型的编解码结构及其在通用视觉任务中的拓展，通过跳跃连接（Skip Connection）将编码器各阶段特征与解码器对应阶段特征融合，在保持高分辨率输出的同时注入深层语义，在图像分割、医疗影像分析等需要精细边界的任务中效果显著。其三，自适应空间特征融合（如ASFF、BiFPN），该技术进一步探索了不同尺度特征并非同等重要，通过引入可学习的权重，让网络动态地学习在不同位置、不同尺度上，哪些特征贡献更大，从而实现更智能、更高效的特征融合。这些多尺度融合策略，使模型能够更全面地“观察”图像，从容应对目标尺度变化、局部遮挡以及复杂背景干扰，是实现高精度、高鲁棒性视觉感知系统的基石。

### （二）知识蒸馏和模型轻量化

为了破解模型对高算力硬件的依赖及在边缘设备部

署的难题，知识蒸馏与模型轻量化技术构成了降低计算成本、提升推理效率的关键手段。其目标是在不显著牺牲模型性能的前提下，大幅减少模型的参数量、计算复杂度和内存占用，使之能够在手机、嵌入式系统、IoT设备等资源受限的环境中高效运行。知识蒸馏是一种“师生网络”学习范式。通常，一个预先训练好的、复杂而强大的模型（教师模型）将其所学到的丰富知识——不仅限于最终的输出标签概率（软标签），更包括中间层的特征表示、神经元之间的关系等——转移给一个结构简单、参数更少的小模型（学生模型）。软标签相较于硬标签包含了类别间的相似性关系（例如，猫和狗在某些特征上可能比猫和汽车更相似），为学生模型提供了更丰富的监督信息，从而帮助小模型获得接近甚至超越大模型的性能。此方法尤其适合数据有限的情形，由于学生模型可从教师模型泛化的知识中获益。模型轻量化涉及众多技术集合：网络剪枝：识别并去掉模型中那些冗余、贡献不大的连接，获得一个稀疏化且更紧凑的网络。权重量化：将模型的权重和激活值从高精度进行量化。量化不仅能降低模型的存储空间，还可借助支持低精度运算的专用硬件加快推理流程。紧致网络架构设计：直接设计高效的网络构建模块，如采用深度可分离卷积（Depthwise Separable Convolution, MobileNet 核心要点、ShuffleNet的通道混洗操作等），它们通过分解标准的卷积操作，在基本维持表达能力的同时，极大地减少了运算量。

### （三）动态卷积核自适应机制

传统CNN固有的特性之一是卷积核静态，当训练完成，其参数就维持固定不变，在应对复杂多样、目标外观不断变化的实际场景时，灵活性欠佳，动态卷积核自适应机制意在赋予模型凭借输入内容自适应调整卷积核参数或结构的能力，进而提升模型的特征表达水平与场景适应本领，这是处理复杂背景里目标外观大幅变动问题的前沿方法。该机制的核心所在是把卷积操作从“静态”变为“动态”模式，传统卷积核对各个输入样本都一视同仁，而动态卷积则允许核函数的参数（像权重、大小、感受野情况）按照当前输入的局部或全局上下文信息进行微调<sup>[5]</sup>，其主要的实现模式包含：实时动态滤

波器生成：采用一个轻量的子网络（诸如注意力模块或小型MLP）来分析输入特征图，动态生成与之匹配的卷积和权重。这说明针对图像的不同区域或不同样式的输入，模型会采用不同的“工具”（卷积核）以提取图像特征，以此更精准地抓取特定模式，条件卷积：就如CondConv，它把多个专家卷积核加以线性组合，输入样本的相关情况决定组合权重。这就好比给模型配备了一个能动态挑选的“卷积核工具包”，增强了模型的容量以及灵活度，而不像单纯拓宽网络宽度那样引发巨大的计算消耗，可按需变形的卷积：借助额外的卷积层来学习输入特征图里每个采样点的空间偏移量，让标准的规则采样网格可依据目标形状产生自适应形变，进而让卷积核的感知区域更贴合目标的实际几何轮廓，对处理非刚性变化、部分遮挡现象的效果较为明显。

### 结语

综上所述，基于深度学习的图像识别技术虽已取得显著成就，但其从实验室精度到工业级鲁棒性的道路上依然充满挑战。复杂场景、数据瓶颈与资源约束是横亘于前的现实障碍。面对这些困境，本文探讨了通过多尺度特征融合增强模型感知能力、借助知识蒸馏与轻量化技术促进高效部署、利用动态机制提升模型灵活性，以及依托硬件感知搜索实现算法与平台协同优化等一系列策略，希望能够推动图像识别技术向着更智能、更可靠、更普适的方向演进，从而在更广阔的维度上释放其潜在价值。

### 参考文献

- [1] 孙何明生. 基于深度学习的计算机图像识别技术研究[J]. 中国宽带, 2024, 20(11): 132-134.
- [2] 侯红英. 深度学习图像识别技术研究[J]. 信息记录材料, 2023, 24(12): 92-94+98.
- [3] 董毅, 汪安祺. 基于深度学习的图像识别技术探讨[J]. 现代工业经济和信息化, 2021, 11(06): 154-155.
- [4] 程锦甫, 张万贺. 深度学习在图像识别中的研究及应用[J]. 电子世界, 2020, (19): 48-49.
- [5] 吴剑岚. 物联网系统中图像识别技术的研究[D]. 北京邮电大学, 2020.