

目 录

摘要	I
ABSTRACT	III
1 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.3 研究内容	5
1.4 本文组织结构	6
1.5 本章小结	7
2 Logo 检测相关理论概述	8
2.1 深度学习卷积神经网络	8
2.2 目标检测算法	10
2.2.1 传统目标检测算法	10
2.2.2 基于深度学习目标检测算法	10
2.3 目标检测相关技术	12
2.3.1 特征融合机制	12
2.3.2 视觉注意力机制	13
2.4 本章小结	14
3 基于解耦的跨层双向引导融合网络的多尺度 Logo 检测研究	15
3.1 引言	15
3.2 DCFNet 检测方法	16
3.2.1 引导聚合和分布特征金字塔	16
3.2.2 解耦分支模块	20
3.2.3 损失函数	20
3.3 实验结果及分析	20
3.3.1 实验设置和评价指标	20

3.3.2 实验细节	21
3.3.3 对比实验	21
3.3.4 消融实验	25
3.3.5 可视化结果.....	26
3.4 本章小结	27
4 基于多阶段渐进细化和 RoI 上下文增强网络的小 Logo 检测研究	28
4.1 引言.....	28
4.2 MPRRCENet 检测方法.....	29
4.2.1 多阶段渐进细化模块	30
4.2.2 RoI 上下文增强模块	31
4.2.3 损失函数	32
4.3 实验结果及分析	32
4.3.1 实验设置	32
4.3.2 对比实验	33
4.3.3 消融实验	37
4.3.4 可视化分析.....	38
4.4 本章小结	38
5 总结和展望.....	40
5.1 本文工作总结	40
5.2 未来工作展望	41
参考文献	42
攻读硕士学位期间论文发表及科研情况	49
致谢	50

摘要

Logo 是一种视觉化的信息表达方式，对特定信息具有形象化的视觉传递效果。随着时代的不断发展，越来越多的企业、组织、研究机构等拥有代表自身内涵和形象的 Logo，Logo 的多样性日益明显。Logo 检测作为目标检测的重要组成部分，在现实中拥有广泛的应用前景，例如知识产权保护、Logo 版权追踪、智能交通、商业广告推荐、准确识别可信品牌、保护消费者权益等。在现实应用场景中，Logo 检测可能面临着多尺度、大长宽比、小目标、低分辨率、扭曲和变形、复杂背景以及数据的长尾分布等问题。由于 Logo 的尺度变化大导致不同尺度 Logo 的检测精度差异明显，Logo 的多尺度检测是 Logo 检测的难点，而小尺度 Logo 检测是多尺度 Logo 检测中的难点。深度学习技术近年来发展迅速，在目标检测领域中取得了显著的进展，本文对基于深度学习的通用目标检测算法做相应的改进，针对 Logo 检测中的难点，设计出性能更加优秀的 Logo 检测算法。本文工作主要包括以下内容：

(1) 针对 Logo 检测中的多尺度问题，提出一种解耦的跨层融合网络模型。首先使用跨层连接和双向引导的方式在相邻层和非相邻层促进高层语义特征和低层细节信息之间的融合和交互，然后分两个阶段采用特征平均和再增强策略并以两条信息流动路径来增强特征之间的交互，增强特征表示。此外，将分类和边界框回归分支进行简单的分离，避免分支间的干扰，使分类分支更多地关注 Logo 的分类，而边界框回归分支可以更多地关注于 Logo 的定位，提高 Logo 的分类和定位性能。

(2) 小尺度 Logo 通常在图像中占据很少的像素，因此特征信息不足，此外还容易受到复杂背景、压缩、模糊或其他形式失真的影响，使得检测算法难以获取到可辨别性的信息，因此小尺度 Logo 的检测是多尺度 Logo 检测的难点。针对这个问题，本文提出了一个多阶段渐进细化和 RoI 上下文增强网络模型。首先采用渐进式学习方式，通过多个阶段去逐步细化和增强特征的上下文信息，同时在每个阶段，不同尺度特征图的交互和融合范围逐渐扩大，从而可以产生更强的多尺度特征表示。之后，通过以可学习的加权方式来重用之前经过细化和增强的特征图，向 RoI 特征中引入更多的上下文信息，使用注意力机制和特征变换进一步增强 RoI 特征。

综上，本文针对 Logo 检测中的多尺度 Logo 和小 Logo 的检测问题设计了两个算法模型，在公开的不同规模的 Logo 数据集上构建了大量的实验，实验结果和可视化结果证明了所设计算法模型的有效性。

关键词：深度学习；Logo 检测；特征融合；上下文信息

中图分类号：TP391

ABSTRACT

Logo is a visual means of conveying information in an imagistic way, with the effect of visually representing specific information. With the continuous development of the times, more and more enterprises, organizations, research institutes and so on have logos representing their own connotation and image, and the diversity of logos is becoming more and more obvious. Logo detection, as a crucial component of object detection, holds extensive application prospects in reality, such as intellectual property protection, logo copyright tracking, intelligent transportation, commercial advertising recommendations, accurate brand recognition, consumer rights protection, and more. In practical application scenarios, logo detection may face challenges like multi-scale detection, large aspect ratios, small objects, low resolution, distortion and deformation, complex backgrounds, and the long-tail distribution of data. The detection accuracy of logos varies significantly due to the large scale variation, making multi-scale logo detection a challenge. Among multi-scale logo detection, detecting small-scale logos is particularly difficult. The rapid development of deep learning technology in recent years has made remarkable progress in the field of object detection. This thesis makes corresponding improvements to general object detection algorithms based on deep learning, targeting the difficult points in logo detection, and designs a more efficient logo detection algorithm. The main contributions of this thesis include the following:

(1) For the multi-scale issue in logo detection, a decoupled cross-layer fusion network model is proposed. It first uses cross-layer connections and bidirectional guidance to promote the fusion and interaction between high-level semantic features and low-level detail information across adjacent and non-adjacent layers. Then, in two stages, it adopts an averaging and re-enhancement strategy with two information flow paths to enhance the interaction between features, thereby enhancing feature representation. Moreover, by simply separating the classification and bounding box regression branches, interference between branches is avoided, allowing the classification branch to focus more on logo categorization, while the bounding box regression branch can pay

more attention to logo localization, improving both classification and localization performance for logos.

(2) Small-scale logos often occupy very few pixels in an image, making feature information insufficient. Furthermore, they can easily be affected by complex backgrounds, compression, blurring, or other forms of distortion, making it difficult for detection algorithms to obtain discernible information. Therefore, small-scale logo detection is a challenging aspect of multi-scale logo detection. To address this issue, a multi-stage progressive refinement and RoI context enhancement network model is proposed. It adopts a progressive learning approach, refining and enhancing contextual information of features through multiple stages. At each stage, the interaction and fusion scope of different scale feature maps gradually expand, producing a stronger multi-scale feature representation. Subsequently, by reusing the previously refined and enhanced feature maps in a learnable weighted manner, more contextual information is introduced into RoI features, using attention mechanisms and feature transformations to further enhance RoI features.

In summary, this thesis designs two algorithm models to tackle the issues of multi-scale logo detection and small logo detection within logo detection. Extensive experiments conducted on publicly available logo datasets of various sizes have demonstrated the effectiveness of the designed algorithm models through experimental and visual results.

Key words: Deep learning; Logo detection; Feature fusion; Context information

Classification: TP391

1 绪论

1.1 研究背景与意义

Logo 是一种视觉化的信息表达方式，它的形状具有简洁、优美、形象的特点，容易识别，易记忆，对特定信息具有很好的视觉传递效果。对于企业来说，品牌的 Logo 是用于企业品牌识别和提供相关的服务等商业应用非常合适的表达载体，它可以最有效地表示和传达某个企业的文化与理念、向消费者传达其产品理念、塑造企业形象与品牌的力量，同时也是区分不同企业的一种简单有效的方法。对于某些机构来说，Logo 可以形象地传达其使命以及提升公众对机构的认知，例如教育机构可以通过 Logo 传达教育理念和学术氛围，团队和俱乐部利用 Logo 传达团队精神和特色，宗教组织则可以利用 Logo 传达其信仰和认知价值观，慈善机构使用 Logo 来传达其使命和提升公众对其活动的认知等。可见 Logo 的使用已经广泛深入到各行各业，并发挥着重要的作用。

随着经济全球化的不断发展和全球市场一体化程度的加深，越来越多的企业出现并根据企业自身特点设计出独特的 Logo 来建立其品牌识别度。此外，视觉 UI 设计软件和集成工具包的发展使得 Logo 设计变得更加容易和可行，降低了设计和更新 Logo 的成本和门槛，让更多企业参与到品牌建设中。企业为了适应时代的变化与发展，紧跟时代潮流，吸引更广泛的消费者群体，以及企业某时刻发展战略的转换，用户需求的需要，其品牌 Logo 的设计也需要不断进行更新迭代，以适应这些变化并保持与时代的一致性。因此，Logo 的数量随之急剧增多，且每个品牌的 Logo 也具有多样性的特点。随着 Logo 使用场景的增多以及应用越来越广泛的同时也面临着一些问题，例如不法商家利用假冒品牌商品来误导消费者使他们认为购买的是正品；非法使用知名公司的品牌 Logo 虚假认证和宣传，提高自己的可信度和吸引力；将未授权品牌的 Logo 用于商业目的宣传或商品上来获取商业利润；设计相似 Logo 混淆市场，误导消费者等。Logo 检测作为目标检测算法的一个重要应用，旨在从复杂的视觉环境中准确识别到特定的品牌 Logo^[1]，可以很有效的对这些不正当的商业行为进行检测。此外，Logo 检测对提升企业发展的核心竞争力和实现关键的业务增长等方面起到重要的作用，例如企业可以使用 Logo 检测技术来监控和追踪 Logo 在各种互联网

媒体中的出现情况去评估品牌的曝光率和市场影响力；Logo 检测可以分析广告、新闻报道中的 Logo 的出现频率，为企业提供市场观察与分析，有助于制定合理的营销策略；Logo 检测在各种交通场景中的应用可以赋予交通智能化等。Logo 检测技术已经服务于社会中的多个方面，在知识产权保护^[2-3]、智能交通^[4-6]、品牌合规性验证^[7-8]等多个领域发挥着重要的作用，促进数字经济的健康发展，是现代商业活动中不可或缺的关键工具，Logo 检测算法的研究有着重要的意义和应用价值^[9-12]。而随着计算机视觉和人工智能技术的不断进步，Logo 检测算法的准确性将进一步提升，应用范围预计将进一步扩大。

早期 Logo 检测算法主要是采用传统手工特征设计的机器学习算法，随着 Logo 的数量急剧增多和检测场景的复杂化，传统 Logo 检测算法已经无法满足现实需要。随着近些年深度学习技术迅速发展，以深度学习为基础的目标检测算法表现出优秀的性能，深度学习的特点是需要大量的数据中学习特征表示，为大规模 Logo 的检测提供了技术可行性。Logo 检测主要涉及两个方面，一是对待检测的 Logo 进行正确的分类，另一方面是对待检测的 Logo 进行准确的定位。由于 Logo 本身的特点以及所在场景的不同，Logo 检测中也面临着一些挑战，例如多尺度 Logo、相似 Logo、复杂背景以及大长宽比 Logo 等。

Logo 在不同的应用场景中可能会以极其不同的尺寸出现，尺度上的巨大变化使目标检测模型难以准确的对不同尺度大小的 Logo 都有着优秀的检测性能。而小尺度的 Logo 占用的像素非常少，细节特征不丰富，也很容易被复杂背景中的噪声所干扰，增加了 Logo 检测模型的检测难度，因此小尺度 Logo 是多尺度 Logo 检测的难点。针对多尺度 Logo 检测的难点，本文以基于深度学习的目标检测算法为基础，并结合 Logo 本身以及所在场景的特点对通用目标检测算法进行改进，设计出性能优秀的 Logo 检测算法实现对 Logo 的正确分类和精确定位。

1.2 国内外研究现状

目标检测是计算机视觉和图像处理领域的一个快速发展的领域，旨在对图像和视频中感兴趣的物体进行正确分类和精准定位^[13]。得益于深度学习技术的大数据处理能力和强大的特征抽取能力，近年来目标检测技术取得了显著的进展。目前目标检测在智能交通、面部识别、人群计数、自动驾驶、智能监控、智慧农业、医疗诊断等多个方面都有着广泛的

应用。Logo 检测作为目标检测中的一个重要应用任务，在知识产权保护、智能交通、品牌合规性验证、多媒体品牌推荐以及运动赞助等方面有着广泛的应用。Logo 检测经历了从传统方法到深度学习方法的转变，这种变革带来了在准确性和效率方面的显著提升，模型的鲁棒性和泛化能力也得以进一步提高。随着 Logo 检测算法的不断更新和迭代，为了更好地评估和比较不同算法的性能，近年来大规模 Logo 检测数据集也随之提出。这些数据集包含了大量不同品牌、不同尺寸、不同背景的 Logo 图片，覆盖了从简单到极其复杂的各种场景，为算法的训练和测试提供了丰富的样本。这些数据集的建立使得研究者能够更好地评估他们设计的算法的检测性能、泛化能力和鲁棒性等，推动了 Logo 检测技术的发展。

在 Logo 检测算法发展的早期，主要是传统方法为主，例如利用方向梯度直方图^[14]和尺度不变特征变换^[15]等手工设计特征方法，然后使用传统机器学习算法如支持向量机、最近邻算法等进行 Logo 检测或者利用模板匹配算法，计算 Logo 相似度的方法等。1993 年 Doermann 等人提出了一种多层次的 Logo 识别方法，该方法首先使用全局不变量对数据库进行修剪，然后使用局部仿射不变量来获得更加精细的匹配^[16]。2003 年 Hollander 等人提出一种利用从视频中的 Logo 提取的字符串和从原始 Logo 图片提取的模板字符串进行匹配的方法来对 Logo 进行正确的分类^[17]。2006 年 Bai 等人提出了一种 Logo 自动检测方法，首先通过动态规划和无监督聚类去自动学习 Logo 的转换模板并在这个学习过程中提取关键帧，然后将提取到的关键帧和学习到的 Logo 模板联合来用于体育视频中的 Logo 检测^[18]。2007 年 Sanyal 等人为了解决自然照片中以任何方向或比例出现并可能受到不同程度遮挡的 Logo 检测难点，提出了一种基于关键点和密度聚类的方法来匹配自然图像中 Logo 的方法^[19]。2009 年 Gao 等人对常见 Logo 的关键特征分析的基础上，进一步提出了一种基于空间频谱显著性和部分空间上下文的两阶段检测方案。首先空间频谱显著性加快了 Logo 的定位，避免了杂乱背景的影响，而部分空间上下文利用局部不变点的空间一致性来过滤虚假的匹配结果。两个模块的集成实现了对 Logo 更快的定位和更高的精度^[20]。2010 年 Arafat 等人提出了一种使用首先利用 SIFT 关键点为每个查询 Logo 和正在处理的视频帧创建特征集，然后采用了随机样本一致性算法来抑制双向匹配过程中的异常值，最后利用新的阈值方案来最终决定是否是否存在 Logo 的方法^[21]。2012 年 Chen 等人提出一种直接从梯度方向、梯度幅度通道和灰度强度通道这三方面计算的哈尔特征来对各种 Logo 图像的可

判别显著特征进行有效提取的 Logo 检测方法^[22]。

虽然以上这些传统方法主要依靠手工设计的特征和比较初级的机器学习算法。尽管在一定场景下效果显著，但面对复杂多变的背景、多样化的 Logo 形态以及大规模 Logo 数据时，它们的适应性和泛化能力受到限制。随着深度学习技术的兴起和快速发展，尤其是卷积神经网络在图像处理领域的广泛应用，基于深度学习的目标检测方法表现出强大的特征信息抽取能力，能够更准确、更高效地识别和定位图像或视频中的物体，因此逐渐替代了以往的传统目标检测方法。基于深度学习的目标检测算法分为单阶段的算法（例如 YOLO^[23]系列、SSD^[24]、TOOD^[25]、DTER^[26]、Deformable DETR^[27]、DINO^[28]等）和两阶段的算法（Faster RCNN^[29]、Mask RCNN^[30]、Cascade RCNN^[31]等）。目前，研究者已经提出了一系列利用深度学习技术进行 Logo 检测的方法，以解决 Logo 检测中的难题。2014 年 Thubsang 等人提出一种结合了卷积神经网络和金字塔梯度直方图特征的 Logo 检测方法，能够准确定位和识别车辆的 Logo^[32]。2015 年 Eggert 等人提出了一种使用预先训练的深度卷积神经网络进行特征提取，然后使用一组支持向量机对 Logo 特征进行分类的 Logo 检测方法^[33]。2016 年 Bao 等人比较了一系列 R-CNN 检测框架，目的是研究 R-CNN 用于 Logo 检测的适当结构和设置^[34]。2016 年 Su 等人提出了一种能够显著提高 Logo 检测性能的模式训练图像合成方法，通过合成上下文的 Logo 训练图像来增加模型对未知背景干扰的稳健性，从而实现更优秀的 Logo 检测性能^[35]。2017 年 Eggert 等人提出了一个用于小 Logo 检测的生成锚提案方案^[36]。2019 年 Bhunia 等人利用一次性学习方法开发了一个基于查询的 Logo 检测和定位系统^[37]。2021 年 Hou 等人提出了一个大规模的食品 Logo 检测数据集 FoodLogoDet-1500，以及一种解决大规模分类和相似性 Logo 识别的方法^[38]。2021 年 Leng 等人发现模型性能受到长尾数据集不同类别的梯度不平衡的限制，提出了一种梯度平衡方法，通过重新调整每个类别的梯度，来引导训练过程中所有类别的平衡。此外，他们还设计了一系列数据增强策略，并提出了渐进式数据增强训练策略以处理对抗性样本^[39]。2021 年 Xu 等人使用重采样，均衡损失 v_2 ，数据增强和加权框融合方法来解决 Logo 检测中的三个挑战：小 Logo 检测，长尾分布和对抗性样本^[40]。2022 年 Wang 等人提出一个大规模 Logo 数据集 LogoDet-3K 并提出一个新的 Logo 检测方法 Logo-Yolo 用于解决样本不平衡问题^[41]。2022 年 Jiang 等人为了提高复杂背景下车辆 Logo 检测的准确性，首先引入

CSPDenseNet 对骨干特征提取网络进行改进，然后采用可变形卷积来适应不规则 Logo 形状，此外还提出了一种新型 Logo 检测头去减少复杂背景对车辆 Logo 检测的影响^[42]。2023 年 Li 等人针对 Logo 检测中的多尺度、大长宽比以及扭曲变形的问题，提出了一种新的长程依赖对合检测方法。他们通过构造多级表示神经结构搜索算法和新的多径拓扑结构来检测多尺度和大长宽比 Logo 目标，实现的自适应 RoI 池模块主要解决 Logo 的变形问题^[43]。2023 年 Zou 等人提出了一种通过增加一个验证阶段来消除分类歧义的 Logo 检测方法。他们将重叠的 Logo 建议分组，并通过学习更具有强烈类间差异的辨别特征，在每个组内来验证每个类别的预测，并且他们还提出了名为小 Logo 保留拼接数据增强策略^[44]。

Logo 数据集为 Logo 检测算法的研究提供了必要的训练和测试样本，也用于评估和比较不同检测算法模型的性能表现，因此在 Logo 检测研究中扮演着至关重要的角色。在深度学习技术成为主流之前，早期的 Logo 数据集相对较小，通常只包含几百到几千个图像，而且 Logo 通常出现在简单且控制的背景中，这些数据集可满足测试传统的基于特征工程的方法的需要。早期发布的 FlickrLogos-27^[45]、FlickrLogos-32^[46]、MICC-Logos^[47]数据集 Logo 种类在十几到几十类之间，样本数量在几百到几千之间，用于初步研究 Logo 检测算法的开发和评估。然而随着 Logo 检测算法的不断更新和迭代，早期 Logo 数据集的规模和多样性无法满足训练深度神经网络所需的大量数据的要求。为了满足算法研究的需要，Logo 数据集经历了从早期的小规模、单一场景集合，发展到如今的大规模、多样化和复杂的数据集。QMUL-OpenLogo^[48]、OpenBrands^[3]、FoodLogoDet-1500^[38]和 LogoDet-3K^[41]等中大规模数据集包含数百到数千类个类别以及数十万到数百万个标注好的 Logo 目标，覆盖了从简单到极其复杂的各种场景，使 Logo 检测算法的研究得到进一步发展。

1.3 研究内容

本文总结了 Logo 检测算法的发展和研究现状，分析了目前 Logo 检测中存在的挑战并调研了最新的 Logo 检测算法。通过大量调研，目前 Logo 检测中主要存在着如下挑战：由于 Logo 的数量急剧增多，以及非法 Logo 的出现，导致出现了非常多相似的 Logo，为 Logo 检测算法正确辨别造成了困难；大长宽比 Logo 为检测网络的准确定位带来挑战；多尺度 Logo 的存在使 Logo 检测算法很难兼顾不同尺度 Logo 的检测，导致不同尺度 Logo 的检测

性能差距很大，尤其是在小尺度 Logo 的检测上性能很差。虽然一些工作对 Logo 检测中存在的问题，例如相似 Logo 以及大长宽比 Logo 有了较好的针对性的改善，但在多尺度 Logo 方面仍有很多不足之处需要进一步研究和改进。多尺度 Logo 是 Logo 检测中的难点，尤其是小尺度 Logo 的检测，因此本文针对多尺度 Logo 检测面临的问题，分析 Logo 本身的特点和复杂的场景，在基于深度学习的目标检测算法中的两阶段算法进行研究和针对性改进，最终实现算法模型对多尺度 Logo 的准确分类和精确定位。本文主要包括如下内容：

(1) 针对多尺度 Logo 检测，本文提出了一个解耦的跨层融合网络模型。引导聚合和分布特征金字塔模块包括一个特征跨层引导模块用于把高层语义信息更好地与低层细节信息之间融合，还包括一个两阶段聚合与分发模块采用双向信息流动路径加强信息之间的融合和交互并在特征信息流动过程中捕捉全局上下文信息细化特征图。此外，解耦分支模块简单的将分类和边界框回归分支解耦为两个相同的全连接头，从而避免它们之间的相互干扰，提高对 Logo 分类和定位的性能。

(2) 小尺度 Logo 的检测是多尺度 Logo 检测的难点，针对小 Logo 的检测本文提出了一个多阶段渐进细化和 RoI 上下文增强网络。首先通过一个多阶段渐进细化模块来渐进式进行特征融合和交互，在信息流动以及各阶段的递进过程中捕捉全局上下文信息的范围也逐渐扩大，然后利用 RoI 上下文增强模块重用上个模块中经过细化和增强的特征图，来为 RoI 特征中引入更多的多尺度和上下文信息，丰富 RoI 特征表示，使用注意力机制和特征变换来进一步捕捉 RoI 特征之间的关系。

1.4 本文组织结构

本文的研究工作共分为五个章节，各个章节的主要内容安排如下：

第一章是绪论。本章节首先介绍了 Logo 检测研究背景、意义与应用价值以及 Logo 检测算法和 Logo 数据集的发展和研究现状，之后分析了 Logo 检测中面临的挑战并介绍了本文的研究内容。

第二章是 Logo 检测的相关理论概述。这一章节首先介绍了深度学习卷积神经网络的主要组成部分和各部分在网络中起到的作用，然后对传统的目标检测算法的优缺点以及基于深度学习的目标检测算法的分类和构成做了介绍，最后还对目标检测中常用到的特征融

合方法和视觉注意力机制做了介绍。

第三章是解耦的跨层双向引导融合网络 Logo 检测方法研究。这个章节首先对多尺度 Logo 检测问题的难点做了介绍，然后针对多尺度检测难点提出了改进方案，对模型改进的各个部分作了详细分析，然后分别在小规模数据集 FlickrLogos-32, 中等规模数据集 QMUL-OpenLogo、大规模数据集 FoodLogoDet-1500 上进行了大量的对比实验以及消融实验进行验证，并通过可视化分析，证明算法模型在检测多尺度 Logo 的有效性。

第四章是多阶段渐进细化和 RoI 上下文增强网络 Logo 检测方法研究。小尺度 Logo 的检测是多尺度 Logo 检测中的难点，针对此问题首先分析了小 Logo 难以检测，检测性能低的原因，然后提出一个用于小 Logo 检测的多阶段渐进细化和 RoI 上下文增强网络模型，旨在探索如何更好地利用多尺度信息和上下文信息之间的相互作用，使检测器能够更好地学习小 Logo 的特征表示。同时在四个不同规模数据集上进行了大量实验证明模型的有效性。

第五章是总结与展望。本章对本文的研究工作进行了总结和回顾并指出本文中存在的不足之处，以及对未来工作做出展望。

1.5 本章小结

本章为绪论章节，重点介绍了 Logo 检测的研究背景和意义并详细阐述了 Logo 的国内外研究现状、存在的问题以及面临的挑战。最后对本文研究内容和整体脉络做了简单介绍。

2 Logo 检测相关理论概述

2.1 深度学习卷积神经网络

深度学习的历史可以追溯到上个世纪 50 年代，早期的研究主要集中在比较简单的神经网络模型上，由于计算机硬件性能的限制和相关理论的局限，因而进展缓慢。到了 1980 年代和 1990 年代，随着计算机计算能力的提升和大量相关理论的提出，神经网络的研究获得了新的动力，尤其是反向传播算法的引入，为训练多层网络提供了有效的方法。进入 21 世纪后，随着大数据时代的到来和计算机计算性能的快速提高，尤其是图形处理单元 (GPU) 可以并行训练神经网络模型，GPU 在耗费功率更低、占用资源更少的情况下能够比单纯使用 CPU 时的效率高几十到上百倍，因此 GPU 在神经网络训练中被广泛应用，深度学习技术也开始经历飞速发展的时期，各种有价值的研究层出不穷，开始取得显著的成就^[49]。这一时期，卷积神经网络^[50](CNN)、循环神经网络^[51](RNN)、长短期记忆网络^[52](LSTM)、Transformer^[53]等多种深度学习模型各个不同领域展现出优秀的性能。深度学习受益于从海量数据中进行训练和学习，可以从原始数据中端到端地直接学习高级的特征表示，无需手动设计特征提取器，且具有很强的泛化能力和鲁棒性。

卷积神经网络中在图像处理中表现出优秀的性能，它主要由卷积层、池化层和全连接层等组成。下面对卷积神经网络的基本知识做介绍：

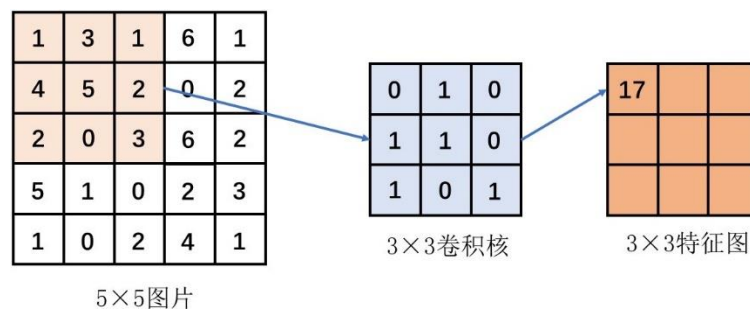


图 2-1 卷积操作过程示意图

(1) 卷积层。卷积层是卷积神经网络的核心部分，它的功能主要是负责执行卷积操作，以自动从输入数据中学习特征，每个卷积层可以包含多个不同的卷积核，每个卷积核通过在输入数据上进行滑动，捕捉到输入数据中的不同特征。卷积层具有平移不变性，即无论目标在输入中的具体位置在哪里，卷积核都能检测到相同的特征，这使得神经网络更好地

处理平移性质弱化的图像特征，提高了模型的鲁棒性。卷积层可以处理任何尺寸的输入数据，使得模型能够适应不同尺寸的输入图像，此外还可以在整个输入数据上共享卷积核的参数，减少模型的参数量降低了过拟合的风险并提高了模型的泛化能力。卷积操作过程如图 2-1 所示，卷积核中的参数是通过卷积神经网络的训练过程中得到，对于输入数据中与卷积核大小相符的局部区域与卷积核进行逐元素相乘，然后将所有乘积结果进行相加求和来得到相对应位置上的输出值。

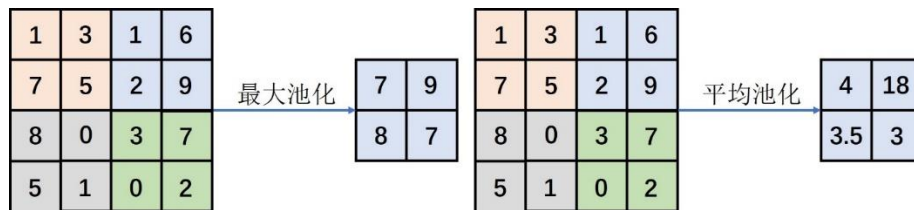


图 2-2 池化操作示意图

(2) 池化层。池化层的作用主要是用于减小特征图的空间维度（高度和宽度），从而减少模型的参数量和计算复杂度，并且可以增强模型的鲁棒性，防止过拟合来提高模型的泛化能力。池化操作示意图如图 2-2 所示，最常用的主要有最大池化和平均池化，最大池化就是在给定的池化窗口内只保留一个最大的特征值，这种方式能够突出最显著的特征，而平均池化就是在给定的池化窗口内计算所有值的平均值，这种池化方式有助于平滑特征图，但也可能会模糊一些重要的特征细节。池化操作通过保持主要特征并减少冗余信息的方式来提取出更加重要的特征，在卷积神经网络中发挥着重要的作用。

(3) 全连接层。全连接层的作用主要是把之前卷积层和池化层提取到的特征连接起来进行汇总和整合，为复杂任务如分类或回归提供决策基础。在全连接层中，每个神经元都与上一层的每个神经元相连，使得信息可以在网络中自由传递，它可以将输入特征与每个神经元之间的连接权重进行矩阵乘法和偏置加法操作，从而得到输出结果。全连接层的输出结果可以看成是对输入特征的一种非线性变换，将输入特征空间映射到输出结果空间，从而增加模型的非线性拟合能力。

(4) 激活函数。激活函数在卷积神经网络中是实现非线性映射的关键部分，因此发挥着重要作用。无论卷积神经网络的层数有多深，如果在没有激活函数的情况下整个网络只能表示线性函数，这大大限制了模型表达能力。由于实际的数据和决策边界往往是非线性的，激活函数使得卷积神经网络能够捕捉这些复杂的非线性关系，从而学习和模拟更复杂的

的函数，此外激活函数还可以促进梯度的有效传播。目前常用的激活函数有 ReLU 函数、Sigmoid 函数、Tanh 函数等。

2.2 目标检测算法

在本小节中主要对传统目标检测算法和基于深度学习的目标检测算法做介绍。

2.2.1 传统目标检测算法

在深度学习成为主流技术并得到广泛应用之前，目标检测算法有相当长的一段时期依赖于传统手工特征的方式实现。这类算法通常依靠专门设计的特征提取器来识别图像中的某种关键元素和模式，其核心思想是通过分析图像的颜色、纹理、形状等基本属性来提取有意义的信息，进而用于后续的分类和检测任务。典型的传统特征包括 HOG^[14]和 SIFT^[15]等。HOG 算法核心在于将图像划分为若干小块，对每块内的像素点进行梯度方向及大小的计算，并对每个小块内的梯度方向进行统计汇总，形成梯度方向直方图。这一方法能够有效地描绘出图像区域内的形状与纹理特征。SIFT 算法能够识别图像中的关键点，并为这些点创建一个独立于图像旋转、缩放和光照变化的唯一描述符，这一能力使得 SIFT 特别适合进行图像匹配和物体识别等任务。SIFT 通过在不同的图像尺度上识别特征点，并计算这些点周边的梯度信息来构建特征描述符，可以有效地捕获图像局部的纹理特性，使其在处理各种尺寸的图像目标时保持稳定和准确。

尽管基于传统手工特征的检测算法在计算机视觉早期阶段取得了显著的成果，但它们也存在一些局限性。传统的手工特征依赖于先验知识和人为设定的特征提取规则，这使得它们在特定任务上表现良好，但泛化到新的图像数据时性能可能会显著下降，另外在复杂的视觉场景下，容易受到外部因素的干扰，影响检测或识别的准确性。此外传统算法通常需要多个阶段的处理流程，每个阶段都需要手动设计和优化，缺乏端到端的学习能力。

2.2.2 基于深度学习目标检测算法

随着深度学习特别是卷积神经网络的发展，基于深度学习的特征提取方法逐渐取代了传统的手工特征方法，尤其是在需要高度泛化能力的复杂视觉任务中。深度学习方法通过

从大量数据中自动学习特征，不仅减轻了人工设计特征的负担，而且在许多任务中都展示了优越的性能。基于深度学习的目标检测算法主要包括单阶段和两阶段算法，单阶段算法主要包括骨干网络部分、特征融合部分和检测分支部分，而两阶段算法比单阶段算法多了一个区域候选建议网络。骨干网络的作用主要是提取输入模型的图片或者视频的特征，从而为后续的目标分类和定位提供必要的信息。特征融合部分主要是融合骨干网络提取出的多尺度特征从而生成更有效的特征表示。检测分支主要包括分类分支和边界框回归定位分支两个部分，利用模型前面部分生成的特征信息实现对目标的正确分类和精确定位。在这个小节中将对单阶段算法和两阶段算法分别做介绍。

(1) 单阶段算法。

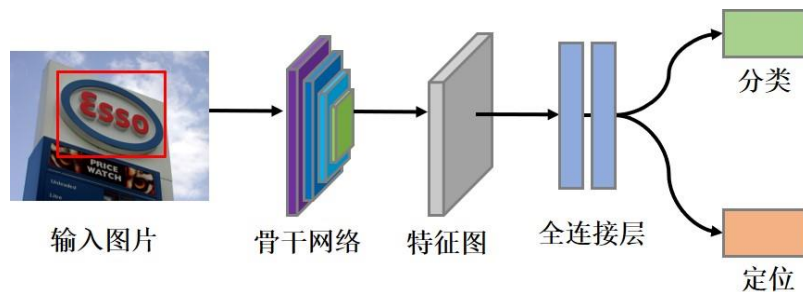


图 2-3 YOLO 模型示意图

单阶段目标检测算法直接从图像中预测对象的位置和类别，避免了两阶段方法中生成候选区域的步骤，使得处理速度显著提高，特别适用于对实时性要求高的场景下，比如视频监控和自动驾驶系统。单阶段算法实现了从图像直接到物体类别和位置信息的端到端学习，但相较于两阶段算法，单阶段算法在精度上可能有所牺牲，尤其是在复杂背景或小目标检测场景中。单阶段算法中的代表性算法是 YOLO^[23]，YOLO 模型示意图如图 2-3 所示，其算法核心思想是将目标检测任务转化为一个整体的回归问题，通过将图像分成多个小网格并针对每个网格同时预测边界框的位置和包含的物体类别。YOLO 通过将图像划分成多个小格子，并对每个格子进行边界框预测和类别概率计算，然后应用非极大值抑制技术消除重叠较多的预测框，从而实现对物体的准确识别和定位。目前 YOLO 系列已经从最初版本升级到了 YOLOv9^[54]版本，其准确度得到很大程度提升的同时，训练迭代数和计算复杂度也进一步降低。除了 YOLO 系列，还有一些其他的单阶段算法，例如 ATSS^[55]、FSAF^[56]、GFL^[57]、TOOD^[25]、DW^[58]等。最近 Transformer^[53]模型在计算机视觉中的任务中表现出优秀的性能，基于 Transformer 的目标检测算法也相继出现，例如 DETR^[26]、Deformable

DETR^[27]、DINO^[28]等。DTER 是基于 Transformer 的目标检测器的里程碑工作，它把目标检测看做一个集合预测问题，并采用二分图匹配的形式强制模型输出一组独一无二的预测框，此外它还去掉了非极大值抑制和目标框的生成部分，减少了超参数的数量。

(2) 两阶段算法。

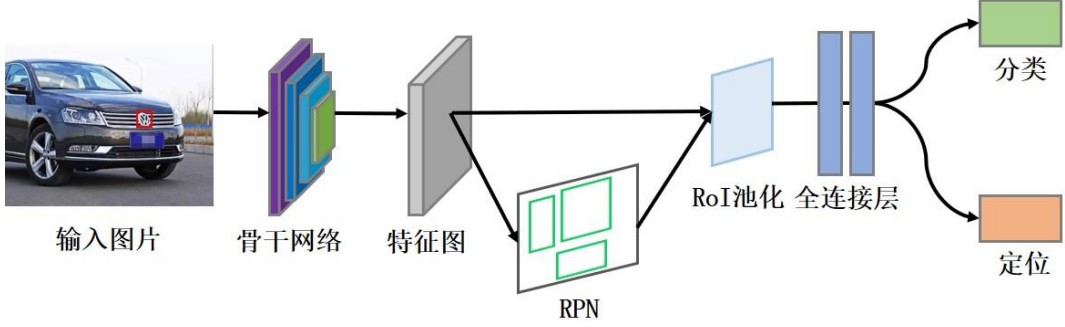


图 2-4 Faster RCNN 模型示意图

两阶段算法相对于单阶段算法来说多了一个生成区域建议的步骤，代表性算法是 Faster RCNN^[29]。Faster RCNN 算法的示意图如图 2-4 所示，该算法首先利用骨干网络提取输入的图像特征，然后利用区域建议生成预先的建议框，之后通过感兴趣区域池化部分将不同的候选区域映射成固定大小的特征图，最后利用分类分支和边界框回归定位分支对目标正确分类和精确定位。Faster RCNN 算法中的区域建议网络使用滑动窗口在特征图上滑动，对于每个窗口位置都预测多个区域候选建议的目标得分，表示该区域是否包含目标，之后利用非极大值抑制去除重叠度较高的候选区域，仅保留得分最高的候选框。目前已经提出了一系列的两阶算法，例如 Cascade RCNN^[31]、Libra R-CNN^[59]、Dynamic R-CNN^[60]等算法，其检测性能得到了进一步提高。

2.3 目标检测相关技术

2.3.1 特征融合机制

目标检测网络利用骨干网络对输入的数据进行特征提取后往往会生成多个尺度大小的特征图，低层特征图的特点是分辨率更高，包含更多位置线索和边缘、纹理等细节信息，但是由于经过的卷积层更少，其语义性更低，噪声也更多，而高层特征图特点是分辨率较低，对细节的感知能力较差，但是感受野比较大，因此包含更强的语义信息。多尺度特征

图都对检测起到至关重要的作用，因此采用特征融合方法可以很好的结合不同层次特征图的优势，更好加强浅层和深层之间信息的交互，进而提升各种尺度目标的检测准确率，可以改进多尺度目标的检测性能尤其是对小目标的检测可以起到很好的效果，特征融合方法可以使高分辨率特征表示同时具有丰富的语义和精细定位信息。

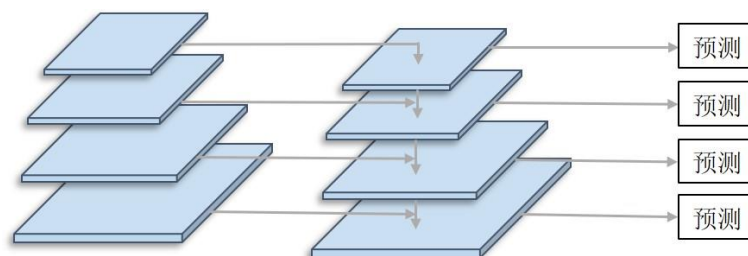


图 2-5 特征金字塔示意图

特征金字塔^[61]（FPN）是特征融合方法的代表性工作，示意图如图 2-5 所示，其算法的核心思想是同时利用低层次的高分辨率特征和高层次的语义特征，来提高对不同尺度目标的检测性能，首先通过 1×1 卷积调整原始多尺度特征图的通道数，然后通过自顶向下和横向连接实现特征图之间的特征融合。之后也有一些方法被提出进一步改进特征融合机制，例如 PAFPN^[62]、BFP^[59]、HRFPN^[63]等。

2.3.2 视觉注意力机制

对于人类而言，可以很快地浏览图片并定位到含有重要信息的区域，视觉注意力机制的提出正是受到人类视觉感知的启发。视觉注意力机制可以帮助模型有选择性地聚焦于图像的不同区域，可以学习在不同位置或通道上分配不同权重的能力。目前比较常用的注意力机制有通道注意力^[64]、空间注意力^[65]、混合注意力机制^[66]、非局部注意力^[67]等。由于不同的通道可能会捕捉到不同类型的信息（如边缘、颜色、纹理等），通道注意力可以对这些通道分配不同的权重使模型更有效地利用这些信息。空间注意力在计算机视觉中被用来关注图像中的某些特定区域，而忽略其他不相关或不重要的区域。混合注意力机制的结构如图 2-6 所示，它的思想是通过把通道注意力和空间注意力以某种方式组合，使它们可以更好地结合各自的优点。非局部注意力对于给定的输入特征图，每个位置的输出是通过对所有位置的特征进行加权求和来计算的，权重由特征之间的相似度确定，因此意味着每个输出位置的计算都考虑到了输入中的所有位置，使模型能够捕捉图像中的远程依赖关系，有

助于模型理解图像的全局上下文信息，提升在复杂场景下的处理能力。

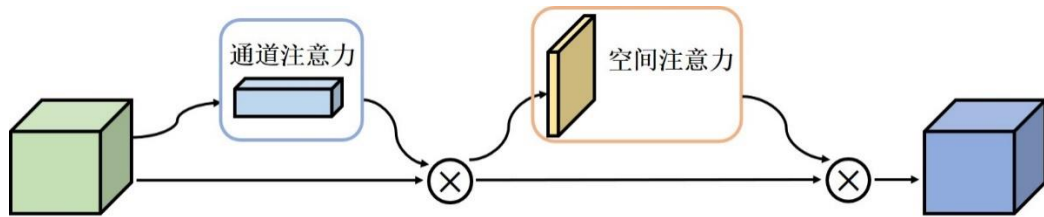


图 2-6 混合注意力机制示意图

2.4 本章小结

本章节介绍了 Logo 检测相关理论，主要内容包括卷积神经网络、目标检测算法的发展和分类以及目标检测中常用的技术。卷积神经网络部分主要介绍了其相关组成部分以及各部分起到的作用；目标检测算法部分主要阐述了传统目标检测算法和基于深度学习的目标检测算法的特点以及简单原理；目标检测相关技术部分主要介绍了常用的特征融合机制和视觉注意力机制两方面。

3 基于解耦的跨层双向引导融合网络的多尺度 Logo 检测研究

3.1 引言

由于经济全球化、品牌国际化的不断发展以及互联网大数据、社交媒体、广告等传播媒介的盛行，Logo 的存在形式及尺度变化变得多种多样。Logo 检测的多尺度问题是一项具有挑战性的任务，Logo 既可能以很小的形态出现在产品上，也可能作为巨幅广告形式出现在建筑物的外墙上，同样一张图片中也可能存在着不同尺度大小的 Logo，检测模型往往难以同时适应尺度变化大的 Logo，因此不能很好的准确识别各种尺度的 Logo，导致不同尺度 Logo 的检测性能有很大差距。尺度变化不仅影响了 Logo 的可视化特征，如形状和纹理，还可能导致检测过程中的背景信息和干扰因素发生变化，进一步增加了检测的难度。由于小尺度 Logo 的像素覆盖范围有限，细节不清楚，导致用于模型检测的特征信息不足，因此很容易被复杂的背景混淆，并且对边界框的扰动容忍度较低，与中等尺度和大尺度的 Logo 相比，小尺度 Logo 更难被检测到，因此多尺度 Logo 检测问题中的难点是小尺度 Logo 的检测。如图 3-1 所示，直观的展示了 Logo 的多尺度问题，在每个 Logo 类别中存在着大中小三种尺度的 Logo，这些 Logo 大多数还面临旋转、遮挡、扭曲变形以及复杂背景等情况。



图 3-1 多尺度 Logo 示例

目标检测模型通过骨干网络生成多个不同尺度大小的特征图，这些特征图包含不同的特征信息，在低层特征图中往往含有比较多的细节信息，高层特征图的感受野更大，因此在高层特征图上往往含有比较多的语义信息，这两种信息对检测都有重要作用。如果可以

更好的融合这些特征图并使特征信息之间得到充分交互就可以有效的改善多尺度 Logo 的检测性能。在本文中，针对多尺度 Logo 检测的问题，提出了一种解耦的跨层融合网络（DCFNet），该网络增强了 Logo 的特征表示，从而获得更加优秀的检测性能。具体而言，所提出的模型首先采用双向跨层连接机制来捕获不同层之间的互补特征信息，接下来利用两阶段的特征平均和再增强策略来进一步增强特征的表示，在双向特征信息流动与交互的同时捕捉全局上下文信息进一步细化特征图。在检测阶段，模型网络将分类和边界框回归定位分支进行解耦，通过避免分支之间的相互干扰来提高多尺度 Logo 分类和定位的准确性。在三个公开可用的 Logo 数据集上进行的大量实验表明，模型网络在检测多尺度 Logo 方面实现了最先进的性能。

3.2 DCFNet 检测方法

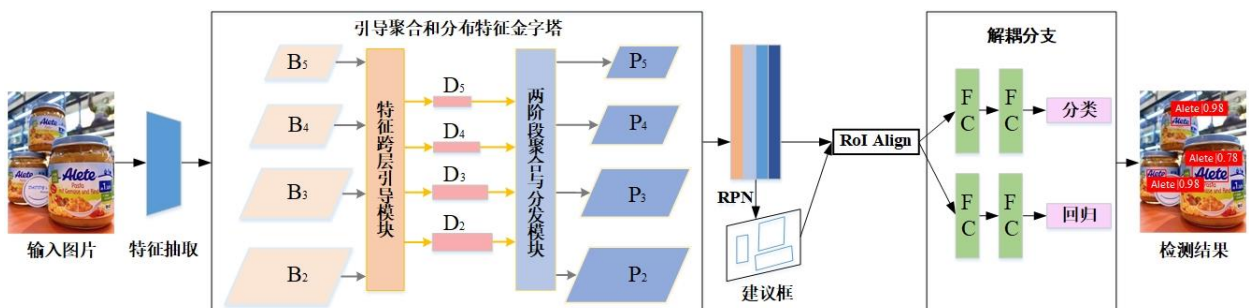


图 3-2 DCFNet 模型总体框架图

针对多尺度 Logo 检测问题，本文提出了解耦的跨层融合网络（DCFNet）来提高多尺度 Logo 的检测性能，模型框架图如图 3-2 所示。DCFNet 首先需要接收输入的图片，对图片进行特征提取，生成 4 个不同尺度的特征图，然后利用引导聚合和分布特征金字塔对这些多尺度特征图进行特征融合并促进特征信息之间的交互的同时进一步增强特征信息表示，之后再通过 RPN 网络生成预先的建议框并通过 RoI Align 生成 RoI 特征图，最后经过解耦的分类分支和回归定位分支判断 Logo 的具体类别并对 Logo 进行精确的定位。

3.2.1 引导聚合和分布特征金字塔

在本小节中，将对引导聚合和分布特征金字塔（GADFP）中的特征跨层引导模块（FCG）和两阶段特征聚合与分发模块（TAD）进行详细描述。

(1) 特征跨层引导模块

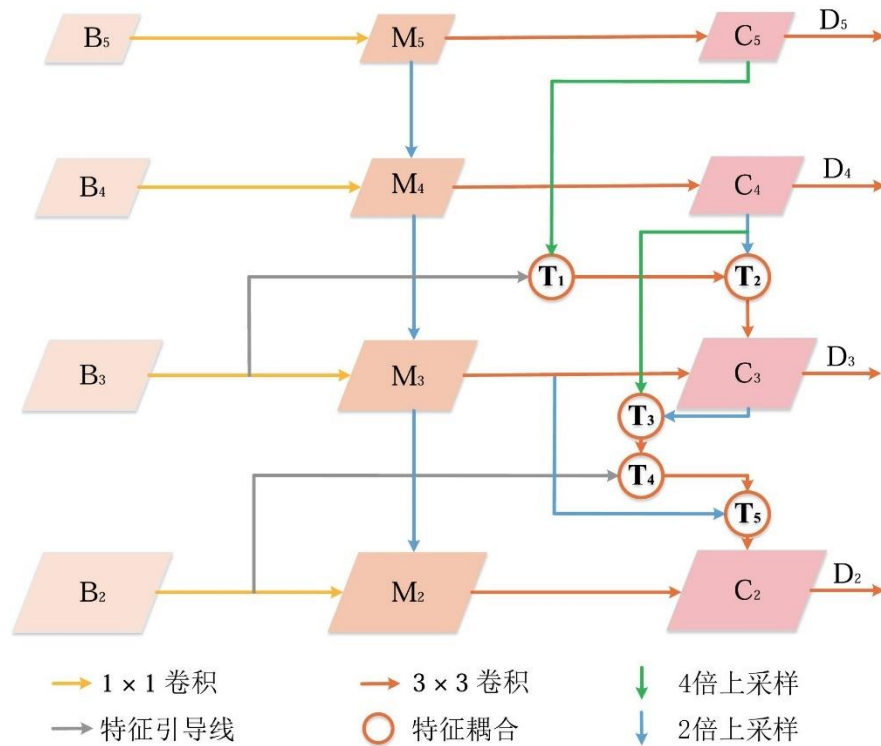


图 3-3 特征跨层引导模块示意图

由于 Logo 在图像中存在着多个尺度的变化，因此对特征提取网络生成的多级特征图进行有效的特征融合提升检测模型的性能是非常有必要的。以往的特征融合方法大多数在相邻层进行融合，对特征信息的利用不足，而非相邻层之间的特征融合和信息交互对检测同样有着重要作用。为了改善多尺度 Logo 检测中小 Logo 的检测性能，高层语义信息和低层细节信息之间合理的融合方式也尤为重要。基于以上考虑，本文设计出特征跨层引导模块，旨在引导和融合不同层次的特征的同时加强特征信息之间的交互，以增强 Logo 的多尺度特征。它利用跨层连接和特征融合来获得增强后的多尺度 Logo 特征表示，特征跨层引导模块的示意图如图 3-3 所示。

在提供具体的公式化描述之前，对以下公式中所使用符号的含义做出解释： B_i 表示从特征提取网络中提取的第*i*个阶段的特征图， M_i 表示在特征融合的中间过程中生成的特征图， C_i 表示经过特征跨层引导和融合后的特征图， D_i 表示在 C_i 特征图上执行 3×3 卷积运算操作后的最终特征图。特征跨层引导模块将特征图的集合 B_i 作为输入，并输出经过跨层连接和特征融合后的特征图集合 D_i 。

下面是生成特征图集合 M_i ($i=2,3,4,5$) 过程的详细描述：特征图 M_5 由经过 1×1 卷积运

算的特征图 B_5 生成；对于特征图集合 M_i ($i=2,3,4$)，它是通过将 2 倍上采样的特征图集合 $M_{(i+1)}$ 与在特征图的集合 B_i 上执行 1×1 卷积运算操作后得到的特征图相加得到的。该过程可以公式化如下：

$$M_i = \begin{cases} Conv_{1 \times 1}(B_i) & i = 5 \\ Conv_{1 \times 1}(B_i) + UpSampling(M_{(i+1)}) & i = 2,3,4 \end{cases} \quad (3-1)$$

为了减少上采样产生的混叠效应，将 3×3 卷积运算操作应用于中间特征图 M_4 和 M_5 从而生成特征图 C_4 和 C_5 。此过程可以公式化为：

$$C_i = Conv_{3 \times 3}(M_i) \quad i = 4,5 \quad (3-2)$$

以下是用于生成特征图 C_3 过程的详细描述。首先，特征图 T_1 是由经过 1×1 卷积操作的特征图 B_3 与经过 4 倍上采样的特征图 C_5 相加得到。接下来，特征图 T_2 是由经过 3×3 卷积运算操作的特征图 T_1 和经过 2 倍上采样的特征图 C_4 相加得到。最后，特征图 C_3 是由分别经过 3×3 运算操作后的特征图 M_3 和特征图 T_2 相加得到。此过程可以公式化为：

$$\begin{cases} T_1 = Conv_{1 \times 1}(B_3) + UpSampling(C_5) \\ T_2 = Conv_{3 \times 3}(T_1) + UpSampling(C_4) \\ C_3 = Conv_{3 \times 3}(M_3) + Conv_{3 \times 3}(T_2) \end{cases} \quad (3-3)$$

以下是用于生成特征图 C_2 过程的详细描述。首先，特征图 T_3 由经过 4 倍上采样的特征图 C_4 和经过 2 倍上采样的特征图 C_3 相加得到；而特征图 T_4 由经过 1×1 卷积运算操作的特征图 B_2 和经过 3×3 卷积运算操作的特征图 T_3 相加得到。接下来，特征图 T_5 是由经过 3×3 卷积的特征图 T_4 和经过 3×3 卷积运算操作与 2 倍上采样的特征图 M_3 相加得到。最后，特征图 C_2 由分别经过 3×3 卷积运算操作的特征图 M_2 和 T_5 相加得到。此过程可以公式化为：

$$\begin{cases} T_3 = UpSampling(C_4) + UpSampling(C_3) \\ T_4 = Conv_{1 \times 1}(B_2) + Conv_{3 \times 3}(T_3) \\ T_5 = Conv_{3 \times 3}(T_4) + UpSampling(Conv_{3 \times 3}(M_3)) \\ C_2 = Conv_{3 \times 3}(M_2) + Conv_{3 \times 3}(T_5) \end{cases} \quad (3-4)$$

为了保持一致性，对特征图集合 C_i ($i=2,3,4,5$) 施加一个 3×3 卷积运算操作，以减少由于上采样产生的混淆效应。

(2) 两阶段聚合与分发模块

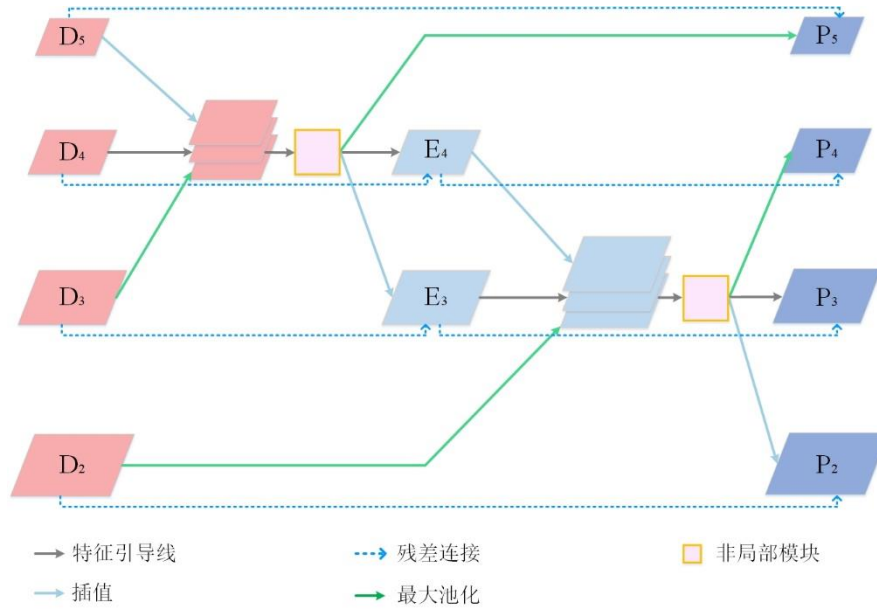


图 3-4 两阶段特征聚合与分发模块示意图

Libra R-CNN^[59]提出了平衡特征金字塔方法，它依赖于集成的平衡语义特征来增强原始特征，将多层级的特征图 D_2 、 D_3 、 D_4 和 D_5 调整到与特征图 D_4 相同的中间尺寸，并通过简单的平均操作获取平衡的语义特征。然而，由于不同特征图尺寸之间存在显著差异，仅通过调整尺寸到特征图 D_4 的大小，无法充分利用所有层次的特征。为了匹配特征跨层引导模块，本文设计了两阶段聚合与分发模块，如图 3-4 所示。由于尺度差异，这些特征图无法直接相加，因此需要将它们转换为相同的尺度大小。在第一阶段中，通过最大池化和插值操作，将特征图 D_3 和 D_5 调整到中间的特征图 D_4 的尺度，得到调整后的特征图 D'_3 和 D'_5 。这个过程可以表示为：

$$\begin{cases} D'_3 = \text{Max-pooling}(D_3) \\ D'_5 = \text{Interpolation}(D_5) \end{cases} \quad (3-5)$$

通过对特征图 D'_3 、 D_4 和 D'_5 进行平均操作，得到平衡的语义特征 D ，然后使用非局部神经网络^[67]来捕捉远程依赖关系，并对 D 中的特征进行细化处理。这个过程得到的特征图再调整回每个特征图的原始尺寸，并与原始图特征相加，得到增强的特征图 P_5 、 E_4 和 E_3 ，它们包含了增强的语义信息和上下文信息。在第二阶段中，特征图 E_4 、 E_3 和 D_2 生成特征图 P_4 、 P_3 和 P_2 的过程类似于第一阶段，在这里不做过多阐述。经过一二阶段，得到最终的一组特征图 $\{P_5, P_4, P_3, P_2\}$ ，这些特征图在引导和增强后包含了更强的语义和细节信息。在信息

流动过程中，不仅有自上而下的信息流动路径，还有一条自底向上的信息流动路径，两种信息流动路径互相配合最大程度上促进了不同特征信息之间的融合和交互。

3.2.2 解耦分支模块

小 Logo 的特征信息不足，对边界框扰动的容忍度也较低，与中大尺度的 Logo 相比，小 Logo 分类和定位将更加困难。以前的 Logo 检测方法通常在分类分支和边界框回归分支之间共享参数，在这里本文设计出解耦分支模块（Decouple FC）简单地将分类分支和边界框回归定位分支解耦为两个相同的全连接头，来避免两个分支之间的相互干扰，使它们专注于各自的任務，提高 Logo 检测中的分类和定位性能。此外，与在检测头部的边界框回归定位分支中使用卷积的方法^[68]相比，解耦两个分支为全连接头部的实现形式更简单，计算复杂度也更低。

3.2.3 损失函数

解耦的跨层融合网络的损失 L 定义如下：

$$L = L_{rpn} + L_{cls} + L_{reg} \quad (3-6)$$

其中， L_{rpn} 、 L_{cls} 和 L_{reg} 分别是 RPN 的损失、分类的损失和定位的损失。 L_{cls} 和 L_{reg} 分别采用交叉熵损失和 L1 损失。

3.3 实验结果及分析

3.3.1 实验设置和评价指标

本文在三个不同规模的数据集上进行了实验，以评估所提出的方法的有效性，数据集包括大规模数据集 FoodLogoDet-1500^[38]、中等规模数据集 QMUL-OpenLogo^[48]和小规模数据集 FlickrLogos-32^[46]。这些数据集的详细说明如表 3-1 所示。在每个数据集中，根据图像中的 Logo 大小（以像素为单位）分为三组： AP_S 是小 logo 目标（面积 $< 32^2$ ）的 AP； AP_M 是中型 logo 目标（ $32^2 < \text{面积} < 96^2$ ）的 AP； AP_L 是大型 logo 目标（面积 $> 96^2$ ）的 AP。在评估时使用平均精度均值(mean Average Precision, mAP)指标，交并比(IoU)的阈值为 0.5。

表 3-1 三个 Logo 数据集的统计信息

数据集	类别数量	图片数量	目标数量	训练集和验证集数量	测试集数量	小 Logo 目标数量
FlickrLogos-32 ^[46]	32	2,240	3,405	1,478	762	185
QMUL-OpenLogo ^[48]	352	27,083	51,207	18,752	8,331	11,841
FoodLogoDet-1500 ^[38]	1,500	99,768	145,400	80,280	19,488	16,463

3.3.2 实验细节

为了确保与其他目标检测模型进行公平比较，本文使用开源的目标检测工具箱 MMDetection^[69]构建了所有的实验。本文采用 ResNet-50^[70]作为特征提取网络，使用单 GPU 策略训练所提出的模型，其中采用 FoodLogoDet-1500 数据集的模型在 NVIDIA A30 GPU 上训练和测试，batch size 的大小为 2，而采用 QMUL-OpenLogo 数据集和 FlickrLogos-32 数据集的模型在 NVIDIA 3090 GPU 上训练和测试，batch size 的大小同样设置为 2。本文使用 SGD 优化器，权重衰减率和动量参数设置分别为 0.0001 和 0.9。训练计划采用 12 个迭代轮数，学习率初始化为 0.0025，并在第 9 轮时下降 10 倍，以帮助模型更好地收敛。

3.3.3 对比实验

在本小节中，将介绍本文提出的 DCFNet 在三个 Logo 检测数据集上的主要结果，并将其性能与其他流行的目标检测器模型进行比较。首先，本文在 FoodLogoDet-1500 数据集上评估了提出的方法，检测结果如表 3-2 所示。从实验结果可以看出，与单阶段和两阶段检测器相比，本文提出的方法在 mAP 、 AP_S 、 AP_M 和 AP_L 指标上都取得了显著的性能提升。具体来说，本文的方法在 mAP 、 AP_S 、 AP_M 和 AP_L 指标上的性能分别达到了 88.5%、82.2%、85.8% 和 92.1%，证明本文提出的方法在多尺度 Logo 的检测上有着良好的性能表现。此外，与单阶段方法中的 ATSS、FSAF、GFL、TOOD 和 DW 模型相比，本文的方法在 AP_S 指标上分别提高了 4.5%、5.1%、5.0%、3.8% 和 0.9%。与其他两阶段方法的 Faster R-CNN、PANet、Cascade R-CNN、Libra R-CNN、Complete IoU、SABL、Double Heads、Dynamic R-CNN、Detectors、Sparse R-CNN 和 CTDNet 相比，本文提出的 DCFNet 模型在 AP_S 指标上分别提高了 3.8%、2.8%、4.1%、1.1%、4.6%、3.7%、2.9%、2.4%、3.0%、5.3%、2.4%，突出了本文的方法对比其他检测方法有着显著优势。FoodLogoDet-1500 数据集上的实验结果显示

本文的方法在大规模数据集上有着良好的性能表现。

表 3-2 各模型在 FoodLogoDet-1500 数据集上的检测结果

模型	骨干网络	mAP(%)	AP _S (%)	AP _M (%)	AP _L (%)
单阶段模型:					
ATSS ^[55]	ResNet-50-FPN	85.0	77.7	81.9	88.2
FSAF ^[56]	ResNet-50-FPN	84.0	77.1	81.3	87.7
GFL ^[57]	ResNet-50-FPN	84.5	77.2	81.7	88.1
TOOD ^[25]	ResNet-50-FPN	85.7	78.4	83.1	89.1
DW ^[58]	ResNet-50-FPN	86.2	81.3	83.4	89.7
两阶段模型:					
Faster R-CNN ^[29]	ResNet-50-FPN	86.4	78.4	83.8	90.0
PANet ^[62]	ResNet-50-PAFPN	86.8	79.4	83.6	90.8
CascadeR-CNN ^[31]	ResNet-50-FPN	86.0	78.1	82.6	90.0
Libra R-CNN ^[59]	ResNet-50-BFP	86.7	81.1	84.4	89.7
Complete IoU ^[71]	ResNet-50-FPN	85.9	77.6	83.0	89.8
SABL ^[72]	ResNet-50-FPN	86.7	78.5	83.6	90.7
DoubleHeads ^[68]	ResNet-50-FPN	87.8	79.3	85.0	91.6
Dynamic R-CNN ^[60]	ResNet-50-FPN	87.7	79.8	85.0	91.5
Detectors ^[73]	ResNet-50-FPN	86.6	79.2	82.7	90.6
Sparse R-CNN ^[74]	ResNet-50-FPN	81.7	76.9	80.2	85.1
CTDNet ^[75]	ResNet-50-FPN	85.6	79.8	83.1	88.8
DCFNet	ResNet-50-GADFP	88.5	82.2	85.8	92.1

为了验证 DCFNet 的有效性，本文在另一个 Logo 数据集 QMUL-OpenLogo 上进行了实验，实验结果如表 3-3 所示。从实验结果中可以看到，两阶段方法相对于单阶段算法而言普遍取得的性能较高。本文的方法在 mAP 、 AP_S 、 AP_M 和 AP_L 指标上取得了最佳性能，分别为 59.7%，38.7%，60.9% 和 73.6%，证明本文的方法在中等规模数据集上同样有着良好的性能表现。

表 3-3 各模型在 QMUL-OpenLogo 数据集上的检测结果

模型	骨干网络	mAP(%)	AP _S (%)	AP _M (%)	AP _L (%)
单阶段模型:					
ATSS ^[55]	ResNet-50-FPN	48.1	31.0	48.8	60.4
FSAF ^[56]	ResNet-50-FPN	46.5	31.5	48.9	87.7
GFL ^[57]	ResNet-50-FPN	47.6	28.2	49.2	59.1
TOOD ^[25]	ResNet-50-FPN	51.9	33.3	52.7	63.6
DW ^[58]	ResNet-50-FPN	54.9	34.8	55.0	68.3
两阶段模型:					
Faster R-CNN ^[29]	ResNet-50-FPN	54.0	32.9	54.5	66.8

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/098046112122007012>