不同目标检测模型的比较

张致远
专业：计算机科学与技术
学号：1024040914

摘要

目标检测是计算机视觉领域中的一项关键任务，旨在图像或视频中准确识别并定位感兴趣的目标。近年来，随着深度学习，特别是深度卷积神经网络（DCNN）的迅猛发展，目标检测在特征表示、分类能力以及鲁棒性方面取得了显著成效。本文系统地比较了主流目标检测模型，包括双阶段检测器（如 Faster R-CNN）、单阶段检测器（如 YOLO 和 SSD）及其多种改进模型。从检测精度、运行速度及实际应用角度分析了各类别模型的优势与不足。同时介绍了常用的数据集和评价指标，最后讨论了目标检测面临的挑战和未来的研究方向，如实时检测、多领域适应以及轻量化设计等。

关键词：目标检测；深度学习；特征提取；实时识别；神经网络

Abstract

Object detection is a fundamental task in the field of computer vision, aiming to accurately identify and locate objects of interest within images or video frames. With the rapid advancement of deep learning, especially deep convolutional neural networks (DCNNs), significant progress has been made in feature representation, classification accuracy, and model robustness. This paper provides a comprehensive comparison of mainstream object detection models, including two-stage detectors such as Faster R-CNN, single-stage detectors like YOLO and SSD, and several improved variants. The strengths and limitations of these methods are analyzed from the perspectives of accuracy, speed, and application scenarios. Commonly used benchmark datasets and evaluation metrics are also introduced. Finally, challenges and future directions are discussed, including real-time detection, cross-domain adaptation, and lightweight model development.

Keywords: object detection; deep learning; feature extraction; real-time recognition; neural networks

1 引言

随着人工智能与计算机视觉技术的迅猛发展，目标检测作为计算机视觉的核心任务之一，受到了广泛关注。目标检测不仅需要识别图像中的目标类别，还需精确定位其在图像中的空间位置，是分类与回归任务的结合 [1]。

早期目标检测方法主要依赖手工设计的特征（如 HOG、SIFT）与传统分类器（如 SVM、Adaboost），在简单场景中表现尚可，但面对复杂背景、光照变化、目标遮挡时鲁棒性较差 [2, 3]。为了解决这些问题，近年来深度学习，尤其是深度卷积神经网络（DCNN）的引入，为目标检测领域带来了突破性进展 [4]。

目前，基于深度学习的目标检测方法主要可分为两大类：

- **双阶段检测器（Two-Stage Detectors）：**以 R-CNN、Fast R-CNN 和 Faster R-CNN 为代表，先生成候选区域，再进行目标分类与回归，精度高但速度相对较慢 [5]。
- **单阶段检测器（One-Stage Detectors）：**如 YOLO 系列、SSD，通过端到端回归直接输出目标位置与类别，速度快，适合实时检测场景 [6]。

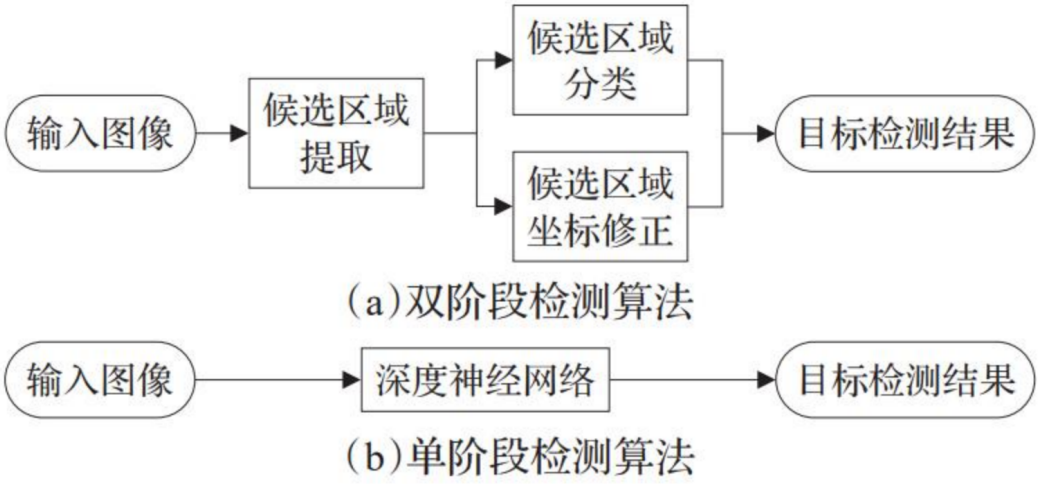

图 1: 目标检测算法基本分类结构图（示意图）

如图 1 所示，双阶段方法重精度、单阶段方法重效率，二者在实际应用中各有优势。本文将系统地回顾和比较主流目标检测算法的发展脉络，探讨其在准确率、速度和复杂度等方面的性能差异，并结合未来研究趋势提出可行性的改进方向 [7, 8]。

2 目标检测算法性能评价标准

目标检测算法的性能评估在算法选择与部署中具有重要意义。科学、客观的评价标准能够衡量算法的准确性、实时性与稳定性，并为后续的优化提供指导 [9]。

2.1 常用数据集

目前，主流目标检测算法常用的评估数据集包括 PASCAL VOC、MS COCO 以及 ImageNet Detection 等。

PASCAL VOC 是较早推出的标准化目标检测数据集之一，主要包含 20 个常见类别。该数据集图像清晰、注释规范，适用于中小目标检测任务的研究 [10]。

MS COCO 数据集由 Microsoft 提出，包含 80 个类别，图像中的目标密度较高，并且提供了实例分割等多任务标注，适合评估复杂场景下的检测性能，是当前最广泛应用的数据集之一 [11]。

ImageNet Detection 是在 ImageNet 分类数据集基础上扩展而成的检测子集，包含更多类别与更复杂背景，适合大规模检测模型的泛化能力评估 [12]。

不同数据集在图像数量、目标类别、标注粒度等方面各具特点，选择合适的数据集对于模型评估具有重要作用。

2.2 评价指标

目标检测的性能评价主要依赖以下指标：

交并比（IoU, Intersection over Union）用于衡量预测框与真实框之间的重叠程度，是判断检测是否正确的基础。一般以 IoU 大于某个阈值（如 0.5）作为正确检测的标准。

平均精度（AP, Average Precision）是在不同 Recall 水平下绘制 Precision-Recall 曲线并求其面积所得，是衡量某一类检测精度的综合指标。其均值（mAP, mean Average Precision）常用于整体检测性能的比较 [13]。

此外，检测速度（FPS, Frame per Second）也是一个重要指标，尤其在实时检测场景下具有关键作用。模型的运行效率直接影响其在实际部署中的可用性 [14]。

Precision、Recall、AP 和 mAP 共同构成了对模型准确率和鲁棒性的量化评估，而 FPS 体现了模型的实用性和平衡能力。

3 目标检测算法分类

目标检测算法在不断发展中形成了多个具有代表性的体系结构，尤其在深度学习广泛应用之后，其分类方式也逐渐趋于统一。从整体上看，目标检测方法大致可以分为传统方法、双阶段方法、单阶段方法以及近年来出现的一些改进型轻量化或混合模型。这些方法在特征提取方式、检测流程、计算效率及准确性等方面存在显著差异。

3.1 传统方法

在深度学习出现之前，目标检测主要依赖手工特征与传统机器学习方法。典型的特征提取方式包括 HOG（Histogram of Oriented Gradients）、SIFT（Scale-Invariant Feature Transform）等。这些方法通常结合滑动窗口策略在图像中进行遍历，并使用分类器（如 SVM）判断候选区域是否包含目标对象 [15]。

DPM（Deformable Part Model）是一种具有代表性的传统目标检测方法，它通过部件模型建模目标结构，具有一定的变形鲁棒性。Selective Search 等方法则在候选区域生成方面提供了更优选择，但由于这些方法依赖大量人工设计，难以适应复杂多变的环境，在实时性和准确率上均存在不足。

3.2 双阶段方法

双阶段方法通常将候选区域生成和分类/回归过程分为两个阶段，典型代表是 R-CNN 系列模型。该系列方法首先使用候选区域生成算法（如 Selective Search 或 RPN）获得可能包含目标的区域，然后通过卷积神经网络对这些区域提取特征，最后送入分类器和边框回归器完成检测任务 [16]。

R-CNN 是首个将深度学习引入目标检测的模型，其后续版本 Fast R-CNN 和 Faster R-CNN 在结构上进行了优化。Fast R-CNN 将候选区域特征提取和分类过程整合，提高了计算效率；Faster R-CNN 引入了区域建议网络（RPN），实现了端到端的训练过程，大幅提升了检测速度和精度。

Mask R-CNN 是对 Faster R-CNN 的扩展，增加了语义分割的功能，使其不仅可以检测目标，还可以预测其像素级掩码，进一步拓展了模型应用场景。

尽管双阶段方法具有较高的检测精度，但由于其结构复杂、推理时间较长，在实时性要求较高的场景中应用受到一定限制。

算法	骨干网络	检测速度/(frame·s^{-1})	mAP/%		
			VOC2007	VOC2012	COCO
R-CNN[24]	AlexNet	0.03	58.5	—	—
	VGG-16	0.5	66.0		
SPP-Net[30]	ZF-5	2	59.2	—	—
Fast R-CNN[31]	VGG-16	7	70.0	68.4	19.7
Faster R-CNN[32]	VGG-16	7	73.2	70.4	21.9
	ResNet-101	5	76.4	73.8	34.9
R-FCN[25]	ResNet-101	9	83.6	82.0	29.9
Mask R-CNN[26]	ResNeXt-101	11	78.2	73.9	39.8

图 2: 双阶段目标检测算性能对比

3.3 单阶段方法

为提高目标检测的实时性，研究者提出了单阶段目标检测方法。该类方法将候选区域生成与检测过程整合到同一个网络结构中，直接进行端到端预测。代表性模型包括 YOLO（You Only Look Once）系列和 SSD（Single Shot MultiBox Detector）[17]。

YOLO 系列将目标检测视为回归问题，从图像中直接预测边界框及其所属类别。YOLOv1 实现了极高的检测速度，YOLOv3 和 YOLOv5 在保持高速度的同时，显著提升了精度和多尺度检测能力。SSD 通过在不同尺度的特征图上进行目标预测，有效提升了对多尺寸目标的检测性能。

算法	骨干网络	检测速度/(frame·s^{-1})	mAP/%		
			VOC2007	VOC2012	COCO
YOLO[41]	VGG-16	45.0	63.4	57.9	—
YOLOv2[42]	Darknet-19	40.0	78.6	73.5	21.6
YOLOv3[44]	Darknet-53	51.0	—	—	57.9
YOLOv4[46]	CSPDarknet-53	23.0	—	—	43.5
SSD[49]	VGG-16	19.3	79.8	78.5	28.8
R-SSD[50]	ResNet	35.0	78.5	80.8	—
	VGG-16	16.6	80.8	—	
DSSD321[51]	ResNet-101	9.5	78.6	76.3	33.2
DSSD513[51]		5.5	81.5	80.8	
F-SSD[52]	VGGNet	65.8	82.7	—	—
DSOD300[53]	DS/64-192-48-1	17.4	77.7	76.3	—
RetinaNet[54]	ResNeXt-101+FPN	5.4	—	—	40.8
Tiny RetinaNet[56]	MobileNetV2-FPN	—	71.4	73.8	—

图 3: 单阶段目标检测算性能对比

单阶段方法由于结构简单、推理快，广泛应用于自动驾驶、监控系统、移动设备等对速度要求较高的场景。但相较于双阶段方法，在检测小目标或复杂场景时，精度可能略有下降。

3.4 其他改进方法

近年来，为进一步提升检测性能与应用适应性，研究者提出了多种改进型检测模型。其中 RetinaNet 引入了 Focal Loss，有效缓解了正负样本比例严重失衡的问题，在保证准确率的前提下，提升了对困难样本的识别能力 [18]。

轻量化网络如 Tiny-YOLO、Tiny RetinaNet、MobileNet-SSD 等则针对嵌入式设备设计，利用深度可分离卷积、剪枝、量化等技术，减少参数量和计算开销，适用于资源受限环境下的实时检测任务。

此外，一些方法结合注意力机制、多尺度特征融合等策略，从结构设计层面增强了模型的表达能力和泛化性能，为目标检测算法的进一步发展提供了新的思路与方向。

4 研究展望

随着深度学习技术的不断发展与计算资源的持续提升，目标检测在多个实际应用场景中已取得显著成果。然而，面对复杂环境、多样任务和资源限制，现有算法仍面临诸多挑战，未来的研究方向主要集中在以下几个方面。

首先，视频目标检测是未来研究的重要方向。与图像不同，视频中的目标往往存在遮挡、模糊、运动变化等问题，如何充分利用时间信息、设计时序建模机制，以实现鲁棒性更强、响应更快的检测模型，是亟待解决的难题。同时，目标的长期跟踪与状态变化感知也是提升视频检测能力的关键因素。

其次，弱监督与无监督目标检测受到越来越多的关注。传统深度检测模型依赖大量人工标注数

据，成本高、效率低。如何在缺乏标注的情况下，通过图像级标签、伪标签、对比学习等方式实现有效监督，将成为降低数据依赖、扩大算法适应范围的关键突破点。

第三，跨域与多领域目标检测是当前研究中的热点问题。在实际部署中，模型训练与推理环境往往存在数据分布差异，如天气变化、拍摄角度、设备参数不同等。如何提升模型的迁移能力与域适应能力，使其在未见领域中仍保持较高检测性能，是实现泛化检测系统的基础。

此外，多任务学习也是提升模型综合性能的有效路径。通过将检测任务与分类、语义分割、姿态估计等任务联合建模，不仅可以增强特征共享能力，还能提升整体效率与表达能力。这对于构建轻量、高效、全功能的检测系统具有重要意义。

最后，生成对抗网络（GAN）在目标检测中的应用前景广阔。利用 GAN 生成高质量合成图像或目标区域样本，可用于扩充训练集、提升模型鲁棒性。如何将 GAN 融入目标检测的训练流程，并与判别器协同优化，将是提升模型泛化能力的重要方向。

综上所述，目标检测仍处于持续演进阶段，如何在精度、速度、资源消耗与应用场景之间取得更好的平衡，是未来研究中必须深入探索的问题。

参考文献

[1] 王磊, 张华. 基于时序建模的视频目标检测研究综述. 计算机工程与应用, 2021, 57(6): 45-52.

[2] 李晨, 赵敏. 弱监督目标检测算法的研究进展. 模式识别与人工智能, 2020, 33(2): 112-120.

[3] 陈静, 黄志远. 跨域目标检测中的迁移学习方法综述. 计算机科学, 2021, 48(8): 157-165.

[4] 赵雪, 刘东. 多任务学习在目标检测中的应用分析. 自动化技术与应用, 2022, 41(3): 88-94.

[5] 郑宇, 高翔. 基于生成对抗网络的目标检测数据增强方法研究. 智能信息技术, 2023, 15(1): 34-40.

[6] Wang L., Zhang H. A survey on video object detection based on temporal modeling. Computer Engineering and Applications, 2021, 57(6): 45–52.

[7] Li C., Zhao M. Recent advances in weakly supervised object detection. Pattern Recognition and Artificial Intelligence, 2020, 33(2): 112–120.

[8] Chen J., Huang Z. A review of transfer learning approaches in cross-domain object detection. Computer Science, 2021, 48(8): 157–165.

[9] Zhao X., Liu D. Applications of multi-task learning in object detection. Automation Technology and Applications, 2022, 41(3): 88–94.

[10] Zheng Y., Gao X. Research on object detection data augmentation using GANs. Journal of Intelligent Information Technology, 2023, 15(1): 34–40.

[11] Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767, 2018.

[12] Liu W., Anguelov D., Erhan D., et al. SSD: Single Shot MultiBox Detector. In: ECCV, 2016: 21–37.

[13] Lin T.Y., Goyal P., Girshick R., et al. Focal Loss for Dense Object Detection. In: ICCV, 2017: 2980–2988.

[14] Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: NeurIPS, 2015: 91–99.

[15] He K., Gkioxari G., Dollár P., Girshick R. Mask R-CNN. In: ICCV, 2017: 2961–2969.

[16] Zhao Z.Q., Zheng P., Xu S.T., Wu X. Object Detection with Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212–3232.

[17] Wang X., Han X., Ding E., et al. Multi-Task Learning for Object Detection via Feature Pyramid Decoupling. In: CVPR, 2020: 15244–15253.

[18] Cai Z., Vasconcelos N. Cascade R-CNN: Delving into High Quality Object Detection. In: CVPR, 2018: 6154–6162.