电路与系统

·封面文章·

基于深度学习的目标检测研究与应用综述

吕 璐，程 虎，朱鸿泰，代年树
（中科芯集成电路有限公司，江苏 无锡 214072）

摘 要：基于深度学习的目标检测算法相较于传统的目标检测算法，对复杂场景的稳健性更强，是当前研究的热点方向。根据基于深度学习的目标检测算法的流程特点将其分为两阶段目标检测算法和单阶段目标检测算法，着重介绍了部分经典算法所解决的问题及其优缺点，梳理了其在工业界的应用情况，对其存在的问题进行了讨论，对未来可能的发展趋势进行了展望。

关键词：计算机视觉；深度学习；目标检测；工业应用
中图分类号：TP391.4 **文献标志码**：A **文章编号**：1681-1070（2022）01-010307
DOI: 10.16257/j.cnki.1681-1070.2022.0114

中文引用格式：吕璐，程虎，朱鸿泰，等. 基于深度学习的目标检测研究与应用综述[J]. 电子与封装，2022，22(1)：010307.
英文引用格式：LYU Lu, CHENG Hu, ZHU Hongtai, et al. Progress of research and application of object detection based on deep learning[J]. Electronics & Packaging, 2022, 22(1): 010307.

Progress of Research and Application of Object Detection Based on Deep Learning

LYU Lu, CHENG Hu, ZHU Hongtai, DAI Nianshu
(*China Key System & Integrated Circuit Co., Ltd., Wuxi* 214072, *China*)

Abstract: Compared with traditional object detection algorithms, object detection algorithm based on deep learning is more robust to complex scenes, and is currently a hot research direction. It is divided into two-stage detection algorithm and one-stage detection algorithm according to the process characteristics of the object detection algorithm based on deep learning. The problems solved by some of the classic algorithms and their advantages and disadvantages are introduced. Its application in the industry is sorted out. The remaining problems are discussed, and the possible future development trends are further prospected.

Keywords: computer vision; deep learning; object detection; engineering application

1 引言

目标检测技术的主要目的是在输入图像中找到目标的位置，同时判断目标的类别属性。随着计算机软硬件和深度学习理论的发展，基于深度学习的目标检测已广泛应用于智能安防、智慧医疗、无人驾驶等领域。在深度学习出现之前，传统目标检测算法都是以手工设计特征为主，如 Sobel[1]边缘检测特征、Haar[2]特征、Hog[3]特征等，这些特征的泛化能力较弱，在复杂场景中性能表现较差。基于深度学习的目标检测算法使用的是卷积神经网络（Convolutional Neural Networks,CNN）学习特征的方式，这种特征学习方式能自动发现检测及分类目标所需要的特征，同时通过

收稿日期：2021-06-21
E-mail: 吕璐 693491925@qq.com；程虎（通信作者）chhu1989@163.com

卷积神经网络将原始输入信息转化成更抽象、更高维的特征，这种高维特征具有强大的特征表达能力和泛化性，所以其在复杂场景下的性能表现较好，可满足工业界的大部分应用需求。

基于深度学习的目标检测算法根据其算法流程特点大致可以分为两类：两阶段 (Two-Stage) 目标检测算法和单阶段 (One-Stage) 目标检测算法。Two-Stage 目标检测算法的主要代表是 Regions with Convolutional Neural Networks Features (R-CNN)[4]系列，此类检测算法检测精度较高，但是检测速度较慢。One-Stage 目标检测算法的代表有 Single Shot MultiBox Detector (SSD)[14]系列、You Only Look Once (YOLO)[13]系列和 Anchor-Free 系列，此类检测算法精度一般，但是检测速度很快，在工业界应用广泛。

本文概述了深度学习类目标检测算法的发展史，并将其分为两大类来对基于深度学习的目标检测算法进行综述，文中对其中的典型算法进行了指标对比和分析，概述了基于深度学习的检测算法应用领域，并对算法当前存在的问题和其可能的解决路径进行了分析。

2　基于深度学习的目标检测算法

随着 2014 年 Two-Stage 目标检测算法 R-CNN 的提出，目标检测算法正式进入深度学习时代，但是算法耗时严重，无法实际应用。通过对算法的深入理解和改进，SPP-Net[5]、Fast R-CNN[6]及 Faster R-CNN[7]等算法被相继提出，算法速度提升上百倍，基本达到应用需求。随着算法应用增多，算法缺点也随之暴露出来，对小目标的检测效果不理想是其中之一，研究人员通过对特征融合方法的研究，提出了 FPN[8]、Cascade R-CNN[11]、M2Det[12]等算法，大幅改善了小目标的检测效果，大大提升了算法精度。

为了从方法论上解决 Two-Stage 目标检测算法的耗时问题，以 YOLO 系列为基础的 One-Stage 目标检测算法被提出，YOLO v2[15]、YOLO v3[16]、YOLO v4[17]逐步解决了 YOLO v1[13]算法的检测框定位不准、小目标检测效果差及算法精度低等问题。但其从 YOLO v2 开始引入了 Two-Stage 目标检测算法的 Anchor 机制，为算法带了新的问题，如 Anchor 参数设置麻烦、正负样本比例严重失衡等。为了解决这些问题，研究人员提出了 Anchor-Free 系列算法，如 ConerNet[29]、CenterNet[30]、FCOS[32]、FoveaBox[33]等，通过设计更加合适的特征表达形式，同时提升目标检测算法的精度和速度。这两类目标检测算法的发展史如图 1 所示。

2.1　Two-Stage 目标检测算法

基于深度学习的 Two-Stage 目标检测算法，第一阶段的主要任务是生成一组目标候选区域，然后将这些目标候选区域送入第二阶段中进行坐标回归和类别分类，其基本流程如图 2 所示。

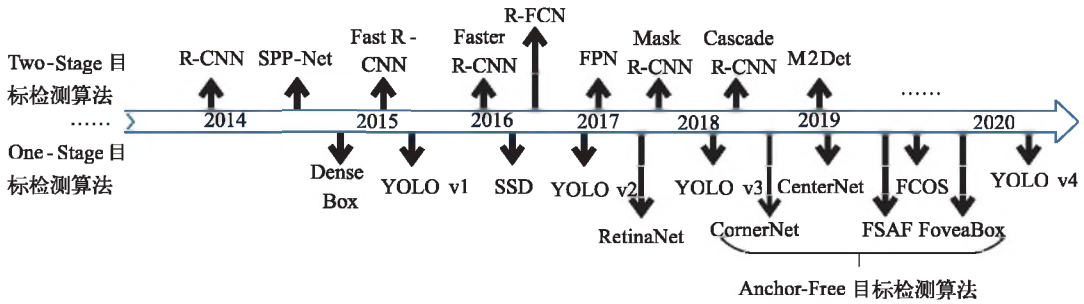

图 1　基于深度学习的目标检测算法发展史

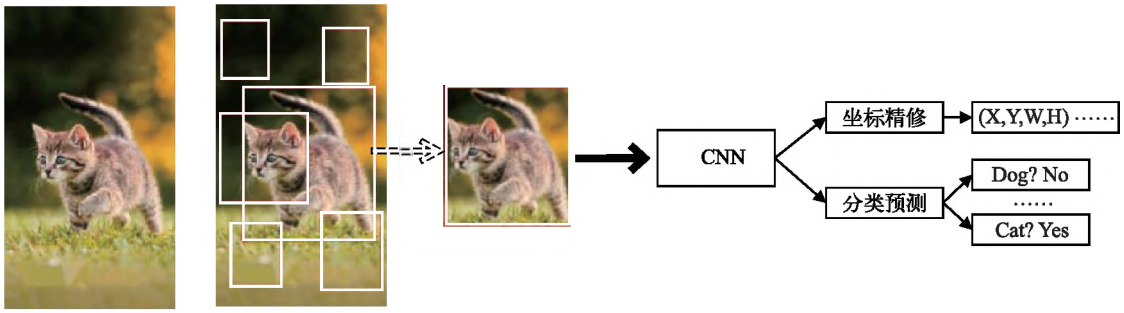

图 2　Two-Stage 目标检测算法的基本流程

2.1.1　R-CNN 算法

GIRSHICK 等人[4]在 2014 年提出了 R-CNN 算法，其算法流程是首先采用候选区域提取算法，提取输入图像上可能包含目标的大约 2000 个不同的图像块，再将这些图像块归一化到固定尺寸后送入到 CNN 中进行特征提取，然后基于支持向量机 (Support Vector Machine，SVM) 算法对这些提取到的特征进行目标分类，最后再通过坐标回归模型对目标的位置进行进一步精修。R-CNN 算法是早期将深度学习融入到目标检测任务中的算法之一，为后面一系列的目标检测算法提供了方法基础。其缺点也很明显，首先是候选区域提取算法耗时严重；其次是对每个候选区域都要进行一次 CNN 特征提取，造成了大量重复运算，浪费计算资源，最终导致平均每幅图像的检测时间高达 34 s，无法实际应用。

2.1.2　SPP-Net 算法

SPP-Net 算法是由何凯明等人[5]于 2014 年提出，主要是为了解决 R-CNN 重复对图像进行特征提取的问题，所以该算法是首先通过 CNN 对原始图像进行特征提取，随后将候选框坐标映射至 CNN 最后一层特征图上，直接获取候选区域的特征图并将其送入第二阶段的分类器训练和坐标精修。由于分类器的训练要求输入的特征尺寸必须统一，所以该算法提出了空间金字塔池化层，候选区域的特征图经过空间金字塔池化操作后，会生成一个统一尺寸的特征向量，从而解决了这一问题。SPP-Net 只需要对输入图像进行一次特征提取操作，大幅节约了计算资源，相较于 R-CNN 算法，其特征提取速度提升了约 100 倍。但 SPP-Net 算法仍存在一些缺点，候选区域提取仍耗时严重，且候选区域提取、图像特征提取、分类器和回归模型训练等流程仍是分离的。

2.1.3　Fast R-CNN 算法

2015 年由 GIRSHICK 等人[6]提出的 Fast R-CNN 算法是在 SPP-Net 的基础上做的进一步改进，首先该算法将空间金字塔池化层改进为 ROI 池化层，可将任意尺寸的特征图池化至统一的固定尺寸特征图；还引入了多任务学习的模式，第二阶段的网络使用两个全连接分支，其中一个使用 Softmax 分类器代替 SVM 作为类别预测分支，另一个使用 Smooth L1 预测坐标偏移来作为坐标精修分支，同时还将分类损失和坐标回归损失整合在一起作为最终的网络损失函数来对网络模型进行训练。Fast R-CNN 算法相较于 R-CNN 算法和 SPP-Net 算法，大幅提升了检测精度，同时也提升了检测速度，平均每幅图像的检测耗时约 2 s。但此算法仍在候选区域提取上耗时严重，无法满足检测的实时性要求，也没有实现真正的端到端的训练和测试。

2.1.4　Faster R-CNN 算法

为了实现真正的端到端的训练和测试，REN 等人[7]在对 Fast R-CNN 做了进一步改进，该算法将候选区域提取工作整合到了检测网络中，提出了区域生成网络 (RPN) 用来提取候选区域的方法。RPN 通过设置多种尺寸的 Anchor，以滑动窗口的方式在特征图上生成多个候选区域，随后对这些候选区域进行坐标粗略估计和属于背景或前景预测的分类。在经过 RPN 计算之后，将属于前景的候选区域筛选出来，送入第二阶段进行坐标精修和具体类别预测。Faster R-CNN 算法比 Fast R-CNN 算法在精度方面有一定的提升，在检测速度方面则有较大提升，平均每张图像的检测时间缩短至 0.2 s 左右。该算法基本上解决了 R-CNN 算法的缺点，并且真正实现了端到端的训练和测试。但其仍有不足之处，由于网络经过了多次的下采样操作，其对小尺寸目标的检测效果不佳。

2.1.5　FPN 算法

当前基于深度学习的检测算法都是通过 CNN 来进行特征学习和提取的，但如 R-CNN 系列算法都是从 CNN 的最后一层特征图上提取特征，该层经过多次的下采样操作，虽具有很强的特征表达能力，但其分辨率已经变得很小，小尺寸目标在此层上已基本无信息保留，所以仅使用此层特征会导致对小目标的检测精度较差。若使用浅层分辨率较大的特征层来进行目标检测，虽然浅层保留了小尺寸目标的信息，但浅层特征表达能力较差，小目标检测召回率增加的同时也会导致误检率大幅增加。为了解决以上问题，提升小尺寸目标的检测效果，LIN 等人[8]在 2017 年提出 Feature Pyramid Networks for Object Detection (FPN) 算法。该算法最主要的核心是在检测网络中加入 FPN 结构，该结构特点有：1) 自底向上连接 CNN 下采样通道，可以得到不同分辨率的特征图；2) 自顶向下连接，不同分辨率的特征图无法直接进行特征融合，需要进行上采样操作，使需要融合的特征层保持一致的分辨率；3) 侧向连接，将上采样后的深层特征层和浅层特征层进行融合，增强浅层特征的表达能力。具体连接方式如图 3 所示。

FPN 结构采用了多尺度特征融合的方式，增强了浅层特征层的特征表达能力。随后，通过对浅层大分辨率特征层进行特征提取和目标检测，大幅提升了算

法对小尺寸目标的检测能力。

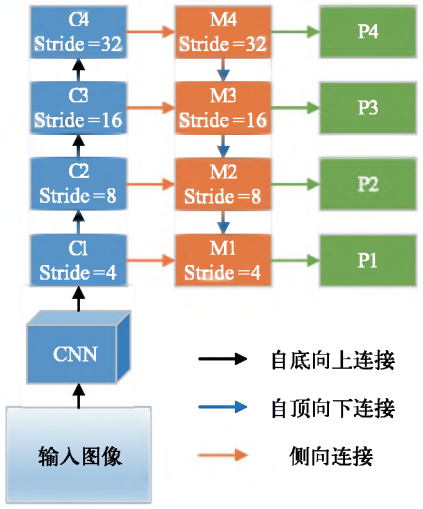

图 3　FPN 网络连接结构

2.1.6　Two-Stage 目标检测算法小结

R-CNN 系列算法和 FPN 算法是 Two-Stage 目标检测算法中较为经典的几个，剩下的一些算法都是基于这些进行的小改进。Two-Stage 目标检测算法生成目标候选区域，对候选区域进行位置精修和类别分类，该类检测算法对大、小尺寸的目标都拥有较好的检测精度，常见于各大目标检测算法竞赛和多个目标检测公开数据集的榜单前排。在 Faster R-CNN 算法和 FPN 算法出现之后，还有很多基于这两个算法的改进算法[9-12]，该类算法检测流程复杂，虽然算法精度高，但检测速度较慢，无法满足工业界实时应用的需求。

2.2　One-Stage 目标检测算法

为了提升基于深度学习的目标检测算法的效率，One-Stage 目标检测算法应运而生。该类算法通过 CNN 提取输入图像特征之后，直接对目标坐标和类别进行预测，其主要流程如图 4 所示。

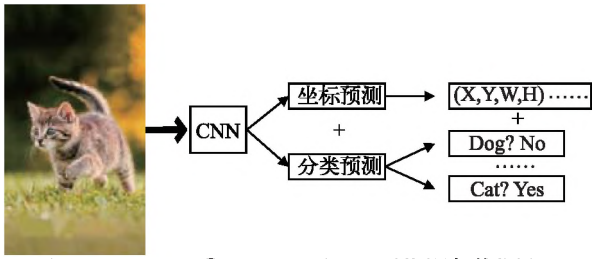

图 4　One-Stage 目标检测算法流程

2.2.1　YOLO v1 算法

YOLO v1 算法是由 REDMON 等人[13]在 2015 年提出的一种采用了回归思想的端到端的目标检测算法。该算法通过在输出特征图上划分网格的方法直接进行目标坐标回归和类别分类，省略掉了显式的提取候选区域过程，这种设计方式大幅降低了检测流程的耗时，平均每幅图像的检测耗时约 0.02 s，同时还有低误检率的优点。

但 YOLO v1 算法也存在着明显的缺点，首先，其对小目标、临近目标、群体目标的检测效果较差，这是由于其网格的设计方法导致的，YOLO v1 的 CNN 最后一层特征图输出的张量尺寸是 $SS[C+B(4+1)]$，其中 S 为特征图的长宽，C 为训练目标类别总数，B 为每个网格预测的目标框个数。以 448 输入为例，其中 C=20，B=2，则最终输出张量为 7×7×30，其意味着每个网格只预测 2 个目标，且最终只保留一个目标作为最终预测结果。然而当有临近目标或群体目标存在时，多个目标的中心可能会落在同一个网格，这就导致了只有一个目标会被检出，其他的都不会被检测到。其次，其对边界框的定位精度不是很好，这是由于在其损失函数中，大目标和小目标的交并比（指预测框和真实框的相交面积除以预测框和真实框的并集面积，缩写为 IoU）误差对网络训练时损失贡献相当，从而影响了该算法的定位精度。

2.2.2　SSD 算法

2016 年 LIU 等人[14]结合 Faster R-CNN 算法和 YOLO v1 算法的优点提出了 SSD 算法。与 YOLO v1 相比，SSD 使用了多个不同分辨率的特征层，显著提升了对小尺寸目标的检测能力。同时还提出了类似于 Faster R-CNN 算法中 Anchor 机制的默认框方法，以 300×300 大小网络输入为例，总共产生 8732 个默认框，最后对这些默认框进行坐标回归和类别预测。在训练时，SSD 算法采用了线上困难样本挖掘的方法，对困难的负样本进行采样，同时只对满足条件的默认框进行训练，保持了训练时的正负样本均衡，提升了算法的检测精度。

与 YOLO v1 算法和 Faster R-CNN 算法相比，SSD 算法保证了检测的精度，也提升了检测的速度。但其仍有一些不足，SSD 算法使用了浅层高分辨率的特征层，该层特征表达能力不足，导致其在小尺寸目标上的检测效果仍一般，且背景误检率也有一定提升。

2.2.3　YOLO v2 算法

为了改善 YOLO v1 算法，REDMON 等人[15]在2017 年提出了 YOLO v2 算法，其在检测类别扩展到 9000 类后也叫 YOLO9000。YOLO v2 算法从多个方面对 YOLO v1 算法进行了提升：1）引入 Anchor 机制，并且使用 Kmeans 算法聚类产生的 Anchor 代替 Faster

R-CNN 和 SSD 手工设计的 Anchor；2）使用卷积层代替 YOLO v1 中的全连接层，同时在主干网络的所有卷积层后面使用 Batch Normalization (BN) 层，这有利于网络模型训练的收敛速度和参数的优化；3）引入多尺度训练策略，有效提升了网络模型对不同尺寸目标的感知能力；4）优化了损失函数中位置回归损失部分，增强了网络模型训练时的稳定性；5）使用了特征提取能力更强的 Darknet-19 作为基础网络。

YOLO v2 算法主要是在 YOLO v1 算法上做了一些改进，在保证检测速度的同时，显著提升了检测精度，但其仍有小尺寸目标检测精度较低的缺点。

2.2.4　YOLO v3 算法

REDMON 等人[16]在 2018 年又对 YOLO v2 算法进行了进一步改进，提出了 YOLO v3 算法。主要的改进点有以下几个方面：1）采用更大更深的 Darknet-53 网络作为基础特征提取网络，提升了算法的整体性能；2）在算法中引入 FPN 结构，分别提取了 3 个不同分辨率的特征层作为最终预测的特征层，提升了算法对小尺寸目标的检测能力；3）分类器不再使用 Softmax，而是采用二分类交叉损失熵，主要是为了解决训练集目标可能存在着重叠类别标签的问题。YOLO v3 算法提升了 YOLO 系列算法的小尺寸目标检测精度，同时拥有很快的检测速度和较低的背景误检率，但其对目标坐标的预测精准性较差。

2.2.5　YOLO v4 算法

ALEXEY 等人[17]在 2020 年提出了 YOLO v4 算法，其算法本质上是将当前多种先进方法融合到 YOLO v3 算法中，对 YOLO v3 算法的性能进行进一步提升。YOLO v4 算法性能提升主要是通过以下几个方面的改进，首先是图像增强方面，YOLO v4 算法在训练时使用了 Mosaic 数据增强方法，此方法是基于 CutMix[18] 数据增强方法进行的改进，CutMix 数据增强方法使用了 2 张图片进行拼接，而 Mosaic 数据增强方法则采用了 4 张图片，并且以随机缩放、随机裁剪、随机排布的方式进行拼接，该数据增强方法不仅丰富了训练集，而且使用单个或少量 GPU 训练就可以达到较好的效果。其次，YOLO v4 算法在基础网络里加入了 CSP[19] 模块、Mish[20]激活函数、SPP 模块、PAN[21]结构等，这些方法的加入进一步提升了网络的特征提取能力。而在损失函数方面，YOLO v4 算法加入了比 IoU Loss [22]和 GIoU Loss[23]优化能力更强的 CIoU Loss[24]，使得预测框回归的速度和精度更高。在 COCO 测试集上输入尺寸为 608×608 的 YOLO v4 算法达到了 43.5%的平均精度均值 (Mean Average Precision, mAP) 指标，而同样输入的 YOLO v3 算法在相同测试集上的 mAP 为 33.0%，并且在同一型号的 GPU 上 2 种算法的检测速度相差不大。综合来看，YOLO v4 算法不仅在保证检测速度的同时达到了更高的检测精度，而且其对目标坐标的回归精度也有较大的改善。

2.2.6　CornerNet 算法

YOLO 系列自 YOLO v2 算法开始，引入了 Two-Stage 目标检测算法的 Anchor 机制，大幅提升了算法的检测精度，但是其算法中的 Anchor 机制仍存在着一些不足，与 Anchor 相关的一些超参数的设置 (如尺寸、比率、IoU 阈值等) 会对检测效果产生影响，而且大量的 Anchor 不仅会增加算法运算复杂度，还会导致训练时正负样本比例严重失衡。为了解决 Anchor 机制带来的问题，大批研究人员开始了基于 Anchor-Free 的目标检测算法研究。早期探索的算法有 DenseBox[25] 和 YOLO v1，它们拥有较快的检测速度，但检测精度较低。随后，研究者们在前人的基础上对基于 Anchor-Free 的目标检测算法进行了改进[26-29]，CornerNet[29]算法是其中的代表之一。

CornerNet 算法主要是借鉴热力图的思想，使用左上角点和右下角点来定义目标框，以预测一组角点 (Corner) 的方式来预测目标坐标。为了使角点计算更加精确，算法使用了 Corner Pool 的方法，该方法融入了更多目标边的信息，但也不可避免地导致网络对边更加敏感，从而忽略了更多内部细节。

2.2.7　CenterNet 算法

CenterNet[30]算法在 CornerNet 的基础上增加了中心信息作为其中一个预测标准，从而使网络能够获取到目标的内部特征。该算法还在 Corner Pool 的基础上进行改进，提出了 Cascade Corner Pool，使得角点也能编码一些内部的信息，增强了网络的表征力。与此同时，提出了 Center Pool，获得水平方向以及竖直方向上的最大值，也能够表示更多的信息，从而提升了算法的检测精度。

与 CornerNet 算法相比较，CenterNet 算法通过对特征表达形式进行了修改和优化，从而获得了更高的算法精度。而后的一些基于 Anchor-Free 的目标检测算法[31-33]都对目标坐标的特征表达形式进行重新定义。事实上，对各个目标检测算法来说，设计合适的特征表达形式，是提升目标检测精度和加快目标检测速度的关键举措。

2.2.8　One-Stage 目标检测算法小结

One-Stage 算法具有较大的检测速度优势，在精度方面还有所欠缺，但是随着该类算法的深度改进，如 YOLO v4 算法及 Anchor-Free 系列最新的算法在精度方面都有所改善，已经接近 Two-Stage 系列算法了。One-Stage 目标检测算法的快速发展，大大推进了基于深度学习的目标检测算法在工业界的应用。

2.3　基于深度学习的目标检测算法小结

Two-Stage 目标检测算法通过基础网络、Anchor 机制、高分辨率特征层、损失函数等方向的优化改进，逐步提升了算法效果，同时进一步降低算法耗时，在各类算法竞赛和服务器端应用广泛。

One-Stage 的目标检测识别算法中，YOLO 系列算法通过改进基础网络、加入 FPN 结构、使用更强大的数据增强策略、添加新型损失函数等，大幅提升算法精度，同时还保持了算法速度优势，在工业界颇受欢迎。基于 Anchor-Free 的系列算法，除了使用关键点进行目标预测的算法外，还有最新的基于密集预测的 Anchor-Free 系列算法，其通过使用更加先进的基础网络、使用密集预测进行分类和回归以及设计更合适的特征表达形式，同时提升了算法的精度和速度，如 FSAF[31]、FCOS[32]、FoveaBox[33]等。

除上述算法外，D2Det [36]、OHEM [37]、Focal Loss、GHM[38]等算法也较为经典，解决了基于深度学习的目标检测算法发展过程中出现的部分问题，如小目标检测效果差、正负样本不均衡等问题。以后，还可以基于生成式对抗网络 (Generative Adversarial Networks，GAN)系列算法[39-41]对现有的目标检测算法进行效果提升，也可以进一步研究小样本学习的相关方法[42]，扩大目标检测算法的应用场景。

3　基于深度学习的目标检测算法的实验结果分析

为了进一步介绍基于深度学习的目标检测算法，对以上算法在目标检测经典测试集上的精度表现进行对比，经典测试集有 PASCAL VOC[34]系列和 MS COCO[35]测试集，其中 PASCAL VOC 系列包含 20 个类别，COCO 测试集包含 80 个类别。各算法在测试集上的精度如表 1 所示[8,11,13-17,33]。

表 1　各个算法在测试集上的效果对比

	算法	基础网络	速度 / (frame·s^{-1})	VOC2007 (mAP@IoU=0.5)	VOC2012 (mAP@IoU=0.5)	MS COCO (mAP@IoU=0.5:0.95)
Two-Stage 目标检测算法	R-CNN	—	0.02	58.5%	53.3%	—
	SPP-Net	—	0.43	60.9%	59.1%	—
	Fast R-CNN	—	3	70.0%	68.4%	19.7%
	Faster R-CNN	VGG16	5	78.8%	75.9%	21.9%
	FPN	ResNet-101	5.8	—	—	36.2%
	Cascade R-CNN	ResNet-101	7	—	—	42.8%
	M2Det	VGG16	12	—	—	44.2%
	D2Det	ResNet-101	6	—	—	45.4%
One-Stage 目标检测算法	YOLO v1	—	45	63.4%	57.9%	—
	SSD512	VGG16	22	76.8%	75.9%	26.8%
	YOLO v2	Darknet-19	40	78.6%	73.4%	21.6%
	YOLO v3	Darknet-53	20	—	—	33.0%
	YOLO v4	CSPDarknet-53	33	—	—	43.5%
	CornerNet	Hourglass-104	4	—	—	42.2%
	CenterNet	Hourglass-104	3.7	—	—	41.6%
	FSAF	ResNeXt-101	2.8	—	—	44.6%
	FCOS	ResNeXt-64x4 d-101-FPN	10	—	—	44.7%
	FoveaBox	ResNeXt-101	7	—	—	43.9%

mAP 是目标检测算法的主要评估指标，目标检测模型通常会用速度和精度指标描述优劣，速度越快，mAP 值越高，表明该目标检测算法在给定的数据集上的检测效果越好，表 1 中的 IoU 代表预测框和真实框的交并比，其中 @ IoU=0.5 指 IoU=0.5 时 mAP 的具体值，@IoU=0.5:0.95 则是指 IoU 为 0.5~0.95 时，每隔

0.05 测试一次 mAP 的值，最后取平均。

从表 1 可以看到，随着对基于深度学习的目标检测算法研究的深入，Two-Stage 的目标检测算法在速度和精度上都在提升。One-Stage 的目标检测算法中，YOLO 系列算法经过发展和完善，目前在精度和速度上都有很大的优势。而 Anchor-Free 系列算法当前在精度上表现不错，但在速度上还有较大提升空间。

综合来看，Two-Stage 类的目标检测算法在精度上较好，所以可以应用于计算资源较为丰富的 GPU 服务器端，One-Stage 类的目标检测算法在速度上较快，适合应用于计算资源较为匮乏的嵌入式前端。

4　基于深度学习的目标检测算法应用领域

4.1　安防领域

基于深度学习的目标检测算法最先在安防领域落地，其主要应用于雪亮工程、平安城市等项目，主要体现在人脸检测、行人检测、车辆检测等方面，可以为打击犯罪、管理交通等提供强有力的支撑。同时，在家庭安防方面，可以进行入侵检测、烟火检测等，一旦发现有陌生人入侵或者发生火灾的情况，能及时报警，尽可能地为户主挽回损失。

安防领域的公司，如海康、大华、宇视等，研发了很多款应用于安防工程的小型化智能摄像头，大多使用了华为海思系列的 AI 前端芯片，如海思 3559A、海思 3519A 等，由于此类芯片计算资源较为紧张，且算法要求实时检测，所以智能摄像头中的人脸检测、行人检测等目标检测功能大多是基于 One-Stage 类的目标检测算法开发，可保证检测算法的实时运行。

4.2　自动驾驶领域

在自动驾驶领域，基于深度学习的目标检测算法能对人、车辆、障碍物等进行实时检测，同时配合车辆雷达系统，可以为驾驶员提供更加轻松、智能的驾驶体验，还可以为驾驶员、乘车人以及车外行人提供更加安全的出行保障。

自动驾驶需要通过雷达、摄像头来收集大量的路况、障碍物、交通标识等信息，并对这些信息进行实时处理和决策，由于车载环境的限制，传统 GPU 由于能耗高、发热大无法应用，当前新型智能汽车大多使用专门的自动驾驶芯片，其计算力强且能效比高。但是，应用于自动驾驶上的目标检测算法对精度和实时性都有较高的要求，所以在应用时往往会根据实际需求和硬件平台性能，去选择 One-Stage 类或者 Two-Stage 类的目标检测算法。

4.3　医疗领域

目前基于深度学习的检测算法通过对大量的医学影像和诊断数据的学习，可以使模型具有检测病变区域和预测病因的能力。通过算法的辅助诊断，大大提升了医生的诊断效率，为患者“早发现，早治疗”提供了有力的帮助。

由于医疗领域对算法精度要求较高，且没有实时性和小型化等要求，所以应用于医疗领域的检测算法大多基于 Two-Stage 类的目标检测算法开发而来，可直接部署于资源丰富的 GPU 服务器端，为医生提供精准的疾病预测结果。

5　总结与展望

本文对基于深度学习的目标检测算法做了概述，并对其中一些典型的算法进行了介绍，同时还对非常热门的 Anchor-Free 系列算法进行了分析，随后对各算法的效果进行了对比，最后对检测算法目前的热门应用领域进行了阐述。综上所述，基于深度学习的检测算法在精度和速度方面相较于传统检测算法都有较大的提升，并且目前还处在快速发展的阶段，但是其仍然存在着一些问题尚未解决：1）对小尺寸目标、遮挡目标的检测精度仍然不够；2）训练时正负样本不够均衡，对算法性能会产生负面影响；3）部分领域训练样本获取难度较高，而训练集样本数量较少会导致算法模型的效果不佳。

针对上述这些依然存在的问题，本文对基于深度学习的目标检测算法以后的发展进行了一些展望。

对于小尺寸目标检测精度低的问题，基于 Anchor 的算法主要是由于设置的 Anchor 跟小目标无法较好匹配，导致无法较好地提取到小目标的特征。D2Det 算法通过引入密集局部回归的方法提升了小目标与 Anchor 的匹配效果，从而提升了小目标的检测效果，未来还可以在此基础上对密集局部回归进行改进，进一步提升小目标的检测精度。对于训练时正负样本不均衡的问题，OHEM、Focal Loss、GHM 等算法通过手工抑制负样本损失的方式来平衡正负样本给网络带来的损失反馈，未来可以通过线上自适应的方式更加合理地平衡正负样本损失，进一步提升算法效果。部分应用领域由于保密或难获取等原因，导致训练样本量较少，可以通过模拟场景的方式增加样本量，也可以通过 GAN 系列算法对样本做一些增强和扩充，还

可以进一步研究小样本学习的相关方法，提升小样本集的训练效果。

参考文献：

[1] ROBERTS L G. Machine perception of three-dimensional solids[D]. Massachusetts: Massachusetts Institute of Technology, 1963.

[2] LIENHART R, MAYDT J. An extended set of haar-like features for rapid object detection[C]// Proceedings International Conference on Image Processing. IEEE, 2002, 1: I.

[3] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 1: 886-893.

[4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[5] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[6] GIRSHICK R. Fast R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]// Advances in Neural Information Processing Systems, 2015: 91-99.

[8] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[9] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]// Advances in Neural Information Processing Systems, 2016: 379-387.

[10] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.

[11] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.

[12] ZHAO Q, SHENG T, WANG Y, et al. M2det: A single-shot object detector based on multi-level feature pyramid network[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 9259-9266.

[13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]// European Conference on Computer Vision. Springer, Cham, 2016: 21-37.

[15] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.

[16] REDMON J, FARHADI A. Yolo v3: An incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.

[17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: Optimal speed and accuracy of object detection [J]. arXivPreprint arXiv:2004.10934, 2020.

[18] YUN S, HAN D, OH S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features [C]// Proceedings of the IEEE International Conference on Computer Vision. 2019: 6023-6032.

[19] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390-391.

[20] MISRA D. Mish: A selfregularized non-monotonic neural activation function [J]. arXivPreprint arXiv: 1908.08681, 2019.

[21] YANG J, FU X, HU Y, et al. PanNet: A deep network architecture for pan-sharpening [C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 5449-5457.

[22] YU J, JIANG Y, WANG Z, et al. Unitbox: An advanced object detection network[C]// Proceedings of the 24th ACM international conference on Multimedia, 2016: 516-520.

[23] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 658-666.

[24] ZHENG Z, WANG P, LIU W, et al. Distance-IoU Loss: Faster and better learning for bounding box regression[C]// AAAI, 2020: 12993-13000.

[25] HUANG L, YANG Y, DENG Y, et al. Densebox: Unifying landmark localization with end to end object detection [J].

arXivPreprint arXiv:1509.04874, 2015.

[26] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points [J]. arXiv preprint arXiv:1904.07850, 2019.

[27] ZHOU X, ZHUO J, KRAHENBUHL P. Bottom-up object detection by grouping extreme and center points [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 850-859.

[28] LIU W, LIAO S, REN W, et al. High-level semantic feature detection: A new perspective for pedestrian detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 5187-5196.

[29] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 734-750.

[30] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]// Proceedings of the IEEE International Conference on Computer Vision, 2019: 6569-6578.

[31] ZHU C, HE Y, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 840-849.

[32] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully convolutional one-stage object detection[C]// Proceedings of the IEEE International Conference on Computer Vision, 2019: 9627-9636.

[33] KONG T, SUN F, LIU H, et al. Foveabox: Beyound anchor-based object detection [J]. IEEE Transactions on Image Processing, 2020, 29: 7389-7398.

[34] VICENTE S, CARREIRA J, AGAPITO L, et al. Reconstructing pascal VOC[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 41-48.

[35] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]// European Conference on Computer Vision. Springer, Cham, 2014: 740-755.

[36] CAO J, CHOLAKKAL H, ANWERR M, et al. D2Det: Towards high quality object detection and instance segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11485-11494.

[37] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 761-769.

[38] LI B, LIU Y, WANG X. Gradient harmonized single-stage detector [C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 8577-8584.

[39] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Advances in neural Information Processing Systems, 2014: 2672-2680.

[40] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2223-2232.

[41] CHOI Y, UH Y, YOO J, et al. Stargan v2: Diverse image synthesis for multiple domains [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 8188-8197.

[42] FU K, ZHANG T, ZHANG Y, et al. Meta-SSD: Towards fast adaptation for few-shot object detection with meta-learning[J]. IEEE Access, 2019(7): 77597-77606.

作者简介：

吕 璐 (1992—)，男，河南信阳人，硕士，工程师，主要从事图像处理、深度学习、目标检测领域的工作。