

DOI:10.19651/j.cnki.emt.2208773

基于深度学习的电子元器件快速检测算法研究^{*}

张志杰 顾寄南 李静 余雪飞

(江苏大学 镇江 212001)

摘要: 针对电子元器件装配过程中由于元器件体积小、外观相似,导致工人在长期高强度的工作下极易误识别、误装配元器件的问题,提出了一种基于深度学习的检测算法 ETS-Net 实现电子元器件的快速准确检测。算法引入深度可分离卷积,减少模型参数量和运算量,降低了模型的复杂度。提出一个轻量化高性能卷积神经网络提取具有分辨力的抽象特征,采用 K-Means 聚类并微调得到适合本场景的锚框,使用高效率的区域提议网络获取高质量的预选框。其次利用两个并联的全连接层预测类别并再次调整预选框,采用非极大抑制排除冗余检测结果。实验结果表明,该算法在电子元器件装配机器人视觉检测任务上具有较高的鲁棒性和实时性。

关键词: 目标检测;机器视觉;智能装配;深度学习

中图分类号: TH166 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Research on fast electronic component detection algorithm based on deep learning

Zhang Zhijie Gu Jinan Li Jing Yu Xuefei

(Jiangsu University, Zhenjiang 212001, China)

Abstract: Aiming at the problem that workers are easy to identify and assemble components by mistake under long-term and high-intensity work due to the small volume and similar appearance of components in the assembly process of electronic components, a detection algorithm ETS-Net (Efficient Two-Stage Network) based on deep learning is proposed to realize the rapid and accurate detection of electronic components. The algorithm introduces depthwise separable convolution to reduce the amount of model parameters and computation, and eliminate the complexity of the model. A lightweight and high-performance feature extraction network is proposed to extract discriminative features, K-Means clustering and fine-tuning are adopted to obtain a set of anchor boxes suitable for the scene, an efficient regional proposal network is introduced to obtain high-quality proposals. And then, two sibling fully-connected layers are used to predict classes and adjust proposals again, and non-maximum suppression is introduced to reduce redundant detection results. The experimental results show that the algorithm has high robustness and efficiency in the visual detection task of electronic component assembly robot.

Keywords: object detection; machine vision; intelligent assembly; deep learning

0 引言

随着物联网的兴起,各个行业都致力于万物互联的产业化应用,这需要依靠大量的电子设备实现算法部署、数据处理、通信等功能,电子设备依靠基础的电子元器件实现具体的功能。而现阶段电子设备主要依靠人工进行装配,由于电子元器件体积小、外观相似度高,需要工人高度集中注意力识别元器件并完成元器件的装配。但是人体的注意力无法长期保持在一个较高的水平,长时间高强度的工作难

免会导致误识别、误装配等问题的出现,在一些对安全性要求较高的电子设备中,这样的错误会造成极大的生命和财产损失。因此亟需研究电子元器件的智能装配方法来减轻繁重的人工劳动,提高制造业的智能化水平。

人工装配电子元器件的过程可大致分为识别元器件种类、确定其位置、抓取元器件、移动至插装位置并进行插装等4个步骤。其中后两个步骤可以使用机器人代替人的手臂完成动作,前两个步骤则需要构建一套机器视觉系统代替人眼完成电子元器件的识别与定位。机器视觉系统由工

收稿日期:2022-01-05

^{*} 基金项目:国家自然科学基金(51875266)项目资助

业相机、光源、计算机设备组成,通过工业相机采集图像数据,计算机运行视觉识别与定位算法从而确定电子元器件的类别与位置信息。基于视觉系统输出的类别与位置信息,机械手即可完成相应抓取和装配的动作。

近年来,基于深度学习的目标检测算法在大规模公共数据集上取得巨大的成功,这些算法可以很好地识别行人、动物、车辆等自然场景中的目标,输出其类别标签,并用矩形框表示其位置。基于深度学习的目标检测算法大致分为 one-stage 和 two-stage 两种,其中 one-stage 以 YOLO 和 SSD 系列^[1-4] 为代表,two-stage 算法以 R-CNN 系列^[5-7] 为代表。One-stage 算法将目标的类别预测与位置直接归结为回归问题^[8],在特征提取过程之后利用大量的样本进行训练,简化了网络结构,使得检测速度大幅提升。Two-stage 算法先从图像中找到有可能是目标的区域,以目标区域的特征作为主要样本进行训练,参与训练的样本质量较高,因此该类算法通常具有较高的精度,但在检测速度上低于 one-stage 方法。在这两个框架下,研究人员进行了大量的探索,并持续在公共数据集上刷新精度和速度的纪录。

目标检测算法在公共数据集上的成功为许多研究工作提供了灵感^[9-11],张静等^[12]通过减少 YOLOv3 特征提取网络中卷积层的层数,改进网络激活函数实现了工业场景下零件的实时检测,精度提高了 1.52%,速度提升了 7.25 ms,但是对类间差距不明显的零件仍存在误检的情况。杜学丹等^[13]采用 ZFNct^[14]作为特征提取网络,在 two-stage 算法框架下实现了目标零件的检测,然而检测速度仍无法突破 30 fps(frames per second),达不到实时性的要求。余永维等^[15]将 Inception^[16]结构引入 SSD,以不同尺度的卷积核提取不同尺度的特征,丰富了模型的特征表达维度,并通过将非极大值抑制加权,提升了网络的后处理效果,在较为复杂的工业场景中达到了 97.8% 的检测精度,并且检测速度得到了保证,达到了 41 fps。

以上研究人员的成果在相应的应用场景下取得了较好

的检测效果,但是应用现有的算法仍然难以取得可用于工业场景的检测效果,主要难点如下:

1)电子元器件各个类别尺寸差距较小,颜色极为相似,在使用深度卷积神经网络提取特征时,难以生成区分度较高的特征,因此会造成大量误检;

2)在公共数据集上取得精度提升的算法在特征提取部分堆叠了过多的特征提取层,虽然精度持续提升,但是随着网络深度的增加,检测效率难以得到提高,这阻碍了目标检测算法在实际生产中的应用;

3)在特征提取之后,许多网络通过一个特征融合模块拼接不同尺度的特征图^[17],以预测不同尺度的目标,但这进一步增加了网络的复杂度和计算量,限制了在低算力设备上的迁移部署。

为解决以上难点,本文提出一种轻量化的高精度 two-stage 检测模型(efficient two-stage network, ETS-Net),基于深度可分离卷积,构建轻量化的特征提取网络和区域提议网络以加速特征提取和区域提议,通过分辨性高效收缩模块(DS Module)在模型浅层快速提取具有强分辨性的特征,在 two-stage 框架下构建目标检测模型,通过训练得到最优模型参数。通过实验与其他算法对比,实验结果表明本文提出的方法具有最优的精度与速度的平衡。

1 ETS-Net

通过模型结构的对比,本文选择在 two-stage 框架下构建检测模型。One-stage 算法虽然具有较高的推理速度,但是由于其只进行一次区域修正,精度难以得到保证。Two-stage 算法在特征提取之后先进行一次目标区域筛选,再通过一次提议区域的类别预测和位置回归,在结构上具有实现较高精度的可能,因此本文选择 two-stage 算法框架作为研究基础,提出性能优异且高效的特征提取网络,快速的区域建议模块,在提高模型效率的同时保持较高的精度,图 1 所示为该算法的结构。

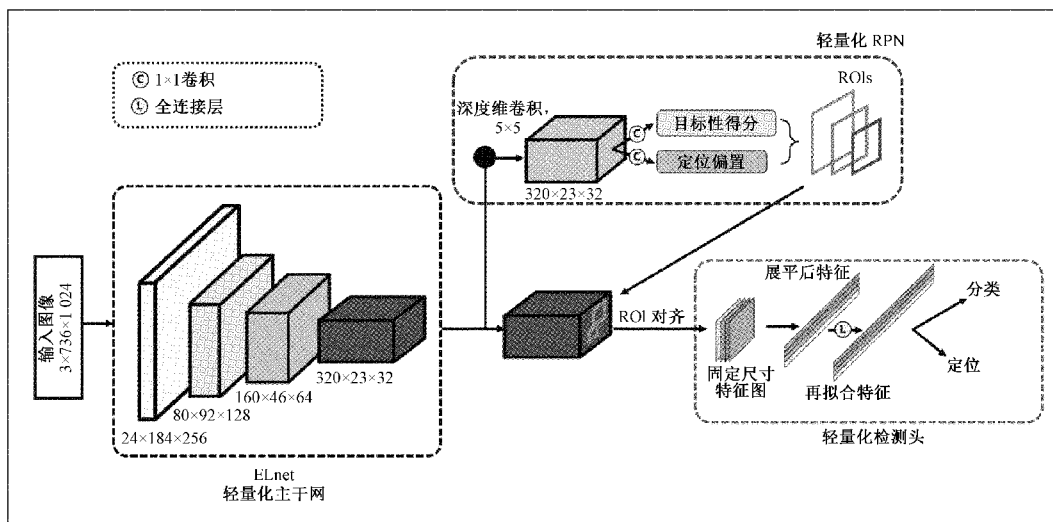


图 1 ETS-Net 算法框架

1.1 高效特征提取网络 (lightweight backbone)

特征提取作为目标检测模型最关键的一部分,其推理速度决定着整个网络的效率。特征提取模块以一张经过预处理的三通道图像作为输入,经过多层卷积计算得到数百个通道的特征图,每个通道到的特征图都代表了目标的深层抽象信息。许多研究者通过在特征提取模块堆叠更多的卷积层来提高模型的精度^[17],部分网络的深度达到了数百层,这导致模型推理速度的下降,限制了在算力有限的设备上应用这些先进的研究成果。此外,在每个卷积层中每组相同参数的卷积核与全部的通道进行矩阵运算,假定输入该层的特征图通道数为 N ,输出特征图通道数为 M ,以大小为 K 的卷积核进行卷积运算,则该层计算产生的参数量为:

$$C_c = K^2 \times N \times M \quad (1)$$

通常 N 和 M 随着网络层数的增加而增长至 512、1 024 等,整个网络的参数量也随之剧增^[17-18]。为解决以上两个影响模型效率的问题,需要构建一个轻量化高性能的特征提取网络。

当前轻量化特征提取网络的进步主要集中在图像分类任务上,通过应用深度可分离卷积和合理的网络结构,这些算法在单 GPU 上实现了实时分类。深度可分离卷积由深度维卷积和点维卷积组成,深度维卷积每个卷积核只与一个通道进行卷积运算,点维卷积则是以卷积核大小为 1×1 的标准卷积,用于调整通道数。与标准卷积不同,给定相同的通道数和卷积核大小,深度可分离卷积在每一层产生的参数量为:

$$C_D = K^2 \times N + N \times M \quad (2)$$

因此,深度可分离卷积与标准卷积产生的参数量比值为:

$$P = \frac{K^2 + M}{K^2 \times M} \quad (3)$$

在输出通道数达到 128 以上时, K^2 远小于 M ,因此 P 可以简化为 $1/K^2$,以最常用的卷积核尺寸 3 为例,深度可分离卷积的参数量仅占标准卷积的 $1/9$ 。由于深度可分离卷积可以大幅减少参数量,本文选择在特征提取网络的部分层使用深度可分离卷积以加速模型推理。

但是由于目标检测和图像分类对图像特征的要求不同,直接将图像分类算法用于目标检测的特征提取并不能得到最好的检测效果。目标检测算法更需要提取具有目标分辨性的局部特征,而图像分类算法则需要提取具有图像整体分辨性的全局特征。Tan 等^[17]的研究表明,在特征提取过程中可以通过提高输入网络的图像分辨率,以及对低层特征的使用可以提升目标检测的效果。高输入分辨率在在网络下采样的过程中包含了更多的目标信息,因此局部特征更明显,更利于目标的检测。而图像分类网络往往会采用较低的图片分辨率以减少运算量。ShuffleNetV1/V2^[19-20]将输入图像尺寸限制在 320×320 pixels,这导致在下采样的过程中目标信息丢失过多,难以区分视觉差别较小

的不同目标。Xception^[18]由于在网络构建块中使用了大量的特征拼接,导致网络推理效率下降。

基于以上分析,本文选择从 ShuffleNetV2 为基础,构建一个高性能的轻量化特征提取网络 ELNet,实现电子元器件的快速特征提取。首先,将网络的输入图像分辨率提升至 $736 \times 1\,024$,使网络可以提取更加细粒度的特征以增加目标间特征的区分性。其次,在网络的浅层用 DS (Discriminative Shrink) 模块快速收缩特征图尺寸,并且以最高效的方式保留具有分辨性的目标特征:给定一张输入图像 $U_i \in \mathbb{R}^{C \times H \times W}$,则该模块的输出为 $U_o \in \mathbb{R}^{C' \times H' \times W'}$,该模块的计算过程可表示为,

$$U_o = F_c(P_m(U_i)) \quad (4)$$

其中, F_c 和 P_m 是步长为 2 的卷积和最大池化。最大池化保留了特征图 S 中 $k \times k$ 区域内的最大值,假设 S 代表输入特征图第 i 个通道,则该层的最大池化操作可以表示为,

$$P_m^i(U_i^i) = \begin{bmatrix} \max_{\substack{1 \leq x \leq k \\ 1 \leq y \leq k}} S_{(1,1)} & \cdots & \max_{\substack{1 \leq x \leq k \\ 1 \leq y \leq k}} S_{(1,W/r)} \\ \vdots & \ddots & \vdots \\ \max_{\substack{1 \leq x \leq k \\ 1 \leq y \leq k}} S_{(H/r,1)} & \cdots & \max_{\substack{1 \leq x \leq k \\ 1 \leq y \leq k}} S_{(H/r,W/r)} \end{bmatrix} \quad (5)$$

式中: x 和 y 是 S 的行、列索引。其中, r 是下采样率,本文采用的下采样率为 2。最大池化输出了尺寸更小的特征图,保留了图像中最具分辨性的特征。之后应用步长为 2 的卷积增加通道数同时提取更加抽象的特征。如果在 DS 模块将卷积和最大池化以相反的顺序构建,则卷积操作会增加通道数,使浅层的特征先抽象,下一步的最大池化则在抽象特征通道上下采样,这不仅增加了运算负担,而且降低了目标特征的区分性。在 DS 模块之后,分 3 个阶段提取特征,每个阶段第一层进行一次下采样。如图 2(a) 所示,每层的特征图分为两组分别送入两个通路提取不同层次的特征^[20],该部分结构在每个阶段只出现一次,图 2(b) 所示的结构重复分别重复 3、7、3 次。表 1 展示了整个特征提取网络的结构参数,最后一层输出通道数为 320。

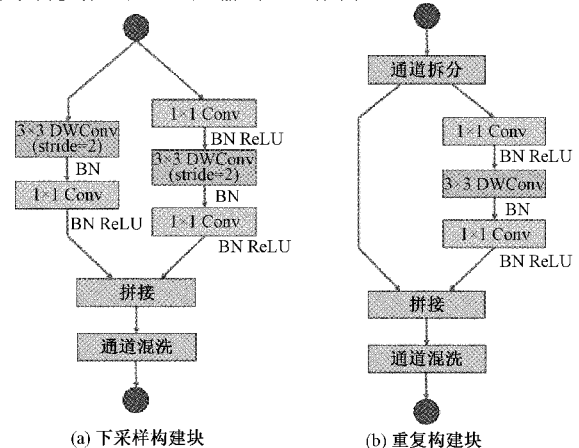


图 2 ELNet 构建块

表 1 ELNet 网络结构参数

	输出尺寸	步距	层数	输出通道
原始图像	736×1 024			3
最大池化卷积	368×512	2	1	24
	184×256	2	1	
第一阶段	92×128	2	1	80
	92×128	1	3	
第二阶段	46×64	2	1	160
	46×64	1	7	
第三阶段	23×32	2	1	320
	23×32	1	3	

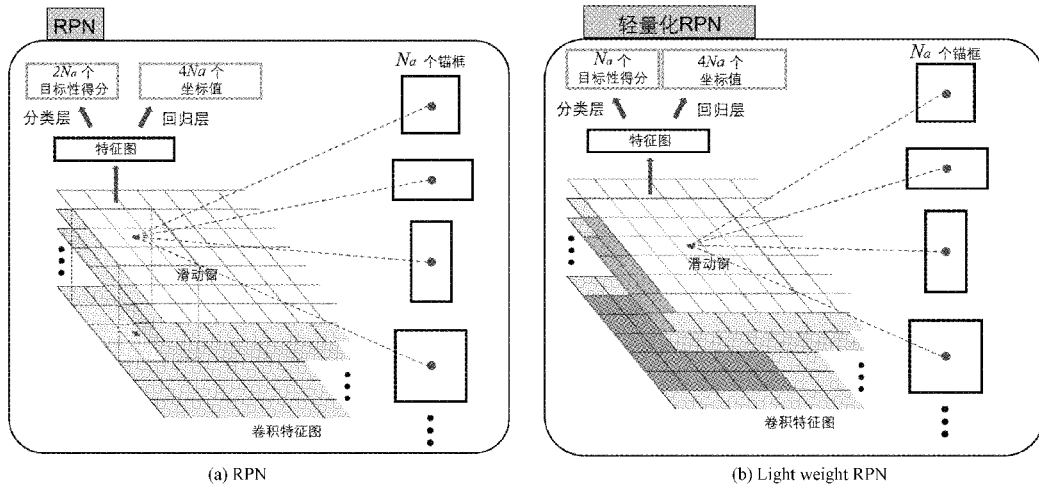


图 3 区域提议网络结构对比

判断某个区域是否存在目标这样简单的分类任务,这一步会产生额外的计算。此外,以特征图的每个位置为锚点(anchor)都会向原图映射一组锚框(anchor box),这些框有 5 个尺寸(32、64、128、256、512)和 3 个长宽比(1:2、1:1、2:1),因此一组锚框的数量达到了 15 个,并且大多数锚框都不能正确包围目标所在区域。

基于以上分析,提出一个轻量化的区域提议网络降低计算开销,以快速得到提议区域。首先,使用一个通道维滑动窗编码由特征提取网络输出的特征。如图 3 所示,图 3(a)为原 RPN 结构,图 3(b)为轻量化 RPN,每个窗口在一个特定的通道上滑动,而不是所有的通道。给定一个大小为 $n \times n$ 的窗口和输出通道数 M ,通道维滑动窗的参数量仅是原来的 $1/M$ 。其次,使用更少的锚框进一步减轻计算负担。锚框的数量可以表示为:

$$N_a = N_s \times N_r \times H \times W \quad (6)$$

其中, N_s 和 N_r 分别表示锚框尺寸和比例的个数, H 和 W 则表示输入特征图的高和宽。为保证目标的尺度不变性, N_s 通常都被设定为 3 个值(1:2、1:1、2:1), N_r 则是 5 个经验值(32、64、128、256、512)。 N_r 所包含的 5 个尺度对电子元件的检测会造成大量冗余,并不能很好地表

1.2 轻量化区域建议网络

区域提议网络(region proposal network, RPN)以 ELNet 的最后一层输出作为输入,得到可能存在目标的区域提议。其中,引入了一个滑动窗和两个并联的全连接层^[7]。滑动窗直接作用于特征提取网络的输出,生成固定通道数的特征图。这部分特征图被送入两个并联的全连接层执行分类和回归子任务,分别得到该区域可能是目标的概率和锚框的偏移量,并以此为依据筛选样本。

然而这个简单的网络在两个方面引入了计算冗余:滑动窗和大量的锚框。由主干网输出的特征图通常可以达到数千个通道(如 ResNet-50 输出 2 048 通道的特征图),而滑动窗每移动一次需要在所有通道上进行编码,对于判

示电子元件的尺度范围。本文首先采用 K-means 聚类得出 6 个锚框尺寸,再经过微调保留至两个尺寸(100、128),此时 N_s 为 2,即锚框数量减少了 60%。

在本文中,通道维滑动窗通过一个 5×5 的深度维卷积实现,并设置两个锚框尺寸,在大幅减少计算量的同时保证了精度。

1.3 轻量化检测头

检测头接受来自 RPN 的区域提议,并将特征收缩至固定大小,通过展平和全局平均池化,然后送入两个串联的全连接层进一步提取特征,最终由两个并联的全连接层实现类别预测和位置偏移量预测。其中,两个串联的全连接层会由于特征提取网络的大量特征图而引入计算冗余。本文以轻量化的特征提取网络向检测头输出的特征图通道数为 320,远小于 ResNet 的 2 048 个特征通道^[21]。此外,本文采用单个全连接层拟合特征,进一步减轻了计算负担。

1.4 损失函数

ETS-Net 涉及分类和回归两个子任务,因此采用文献[7]中的多任务损失函数同时训练两个子网络的参数。损失函数定义如下:

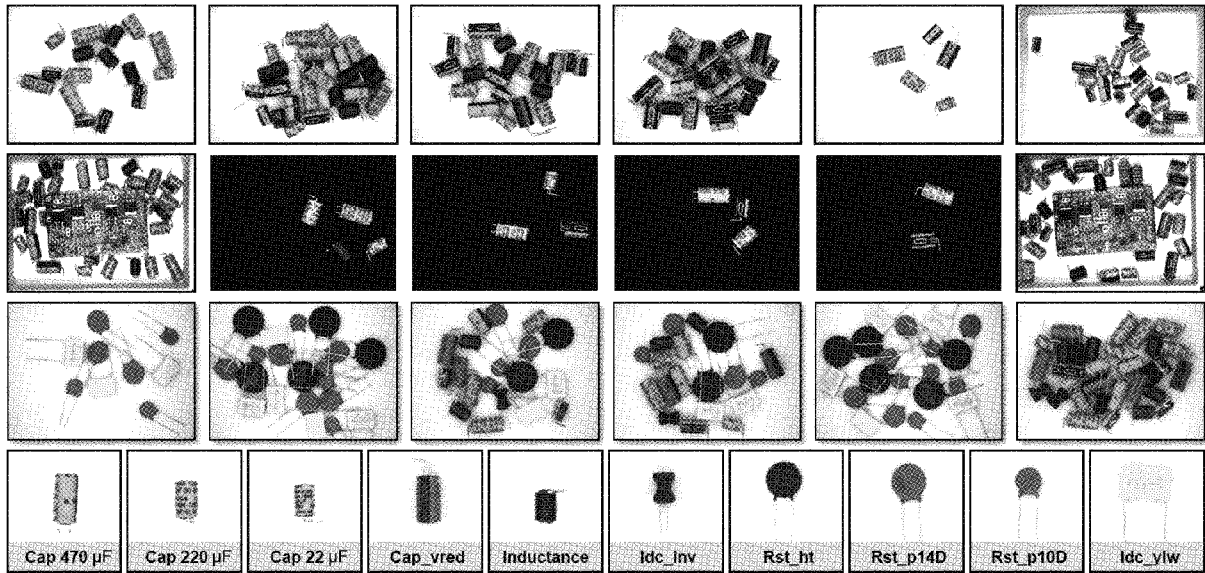


图4 电子元器件图像数据集

$$L(\{p_i\}, \{d_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{loc}} \sum_i p_i^* L_{loc}(d_i, d_i^*) \quad (7)$$

式中:加号前的一项代表分类误差,后一项代表位置误差。其中 i 是样本索引, p_i^* 是该样本对应的真实类别标签。 d_i 是边界框的预测结果, d_i^* 是与之对应的真实位置信息。 N_{cls} 和 N_{loc} 是一个批量中样本的数量以及样本位置信息的数量。为平衡两部分误差,引入 λ 控制两部分误差在总误差中所占权重。 L_{cls} 是一个对数损失函数,由交叉熵损失实现。采用 smoothL1 损失来计算预测的位置误差。需要注意的是, $p_i^* L_{loc}$ 只计算正样本的误差。

在 RPN 的训练过程中,根据锚框与真实标签的交并比设定样本标签为 0(即负样本,代表背景)或 1(即正样本,代表该样本区域存在目标)。正样本由以下两个条件确定:1)锚框与该标签具有最大的交并比(intersection of union, IOU);2)锚框与标签的 IOU 大于一个阈值(本文中,该阈值设定为 0.7)。类似地,若一个锚框与标签的 IOU 低于某个阈值即被认定为负样本(本文中,该阈值设定为 0.3)。由于 RPN 只判断某个区域内是否存在目标,而不判断具体的目标类别,因此 RPN 的分类损失函数使用二分类交叉熵计算预测误差。对于位置预测误差,仍采用 smoothL1 损失函数。

2 实验及结果分析

本文通过实验与其他性能优异的算法对比,验证 ETS-Net 的准确性和效率。对于准确性的验证,本文采用 mAP(mean average precision)作为评估指标。对于算法效率的验证,有许多研究者使用参数量、浮点运算量等参数评估模型的效率,但是这些指标并不能直观地反映算法的

速度表现。本文使用 FPS 作为主要验证指标,该指标可以直观地反映模型的运行效率。

2.1 实验环境

本文实验使用的操作系统为 64 位 Ubuntu 18.04, GPU 设备为 NVIDIA GeForce RTX 2080ti,显存容量为 11 GB。CPU 设备为 Intel Xeon(R) Gold 5118,内存容量为 128 GB。所有涉及的算法模型均由 PyTorch^[22]深度学习框架实现,并且为了客观地对比所有算法模型,均使用随机初始化的方式设定初始权重。

2.2 数据集介绍

为模拟电子元器件的装配场景,构建了用于目标检测的电子元器件图像数据集 Electronics,共包含 1 305 张图像及对应的标签数据。本数据集共涉及 10 类元器件,随机分配 80%用于训练,20%用于测试,图 4 是数据集中的部分图像。可以看到,有几类电子元器件的外观差距很小,颜色几乎相同且尺寸相近,这造成识别难度增加。此外,该数据集标注了 12 555 个目标,平均每张图像中至少包含 9 个元器件,与大型公共数据集相比,每张图像中的目标密度更大,造成识别结果的后处理时间变长,因此模型的检测效率也面临着挑战。

2.3 实验结果

本文对 ETS-Net 进行了 100 个周期的训练,采用混合精度法训练模型,使用随机梯度下降算法更新权重,起始学习率为 0.005,冲量为 0.9,权重衰减系数为 0.000 5。输入网络的图像分辨率为 736×1 024。RPN 训练过程中每张图像采样数量为 256,其中正样本比例为 0.5。计算网络最终的误差时,每张图像的采样数量为 512,正样本比例为 0.25。图 5 所示为算法在训练过程中,在验证集上的精度变化情况。由图可知,在第 30 个训练周期之后,mAP 再无大幅波动且持续提高,说明模型参数具有较

好的数值稳定性。

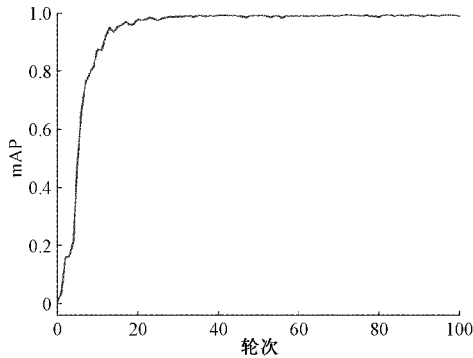


图 5 训练过程中验证集精度变化曲线

经过训练,模型在 Electronics 测试集上得到的精度与效率的结果如表 2 所示,其中非极大抑制的阈值均为

表 2 测试集实验结果

模型	主干网	输入尺寸	fps	mAP/%
SSD	ResNet50	300×300	21	90.55
YOLOv3	Darknet53	416×416	74	94.23
EfficientDet	EfficientNet-D2	768×768	15	79.56
YOLOv4	CSPDarknet53	416×416	41	96.88
Faster RCNN	ResNet50-fpn	736×1 024	26	97.44
Faster RCNN	MobileNet v2	736×1 024	35	97.83
ETS-Net(<i>ours</i>)	ELNet	736×1 024	78	97.36

为了评估算法在测试集上采样的样本准确率,本文统计了 ETS-Net 在各类别下的错误样本数量与标签数量的比值 F_p ,

$$F_p = \frac{N_{pred}}{N_{label}} \quad (8)$$

其中, N_{pred} 是预测结果的错误样本数量, N_{label} 是该类别真实标签数量。以该指标与其他算法进行对比,对比结果如表 3 所示。表格中括号内字母 R 和 M 分别代表特征提取网络 ResNet50 和 MobileNetV2。需要注意的是,由于 One-stage 和 Two-stage 算法结构的差异,前者通常在后处理部分会涉及大量的错误预测样本,错误样本的比例也相对较高,因此仅对比几个 Two-stage 算法的错误样本比例。

由表 3 中的数据可以看出,ETS-Net 的平均错误样本比例最低,证明该算法具有较高的准确度,可以较好地地区分视觉差距不明显的目标。虽然 Faster R-CNN(M) 的 mAP 最高,但是其在电阻 Rst_ht 这一类的错误样本率达到了近 50%,因此检测的稳定性并不能得到保证。Faster R-CNN(R) 检测电容 Idc_inv 的错误样本超过了真实样本数量,极易造成误识别。表 4 记录了以上各算法平均错误样本比率的标准差,进一步评估算法检测的稳定性。与其他算法相比,本文提出的算法具有最低的错误率标准差,证明该算法具有较高的识别准确度和稳定性。

0.3。从表中可以看出,ETS-Net 准确率几乎与使用 ResNet50-fpn^[23] 作为特征提取网络的 Faster R-CNN 持平,但是在速度上快了两倍。同样地,其速度比使用 MobileNetV2^[24] 为特征提取网络的 Faster R-CNN 快 123%。此外,ETS-Net 的准确率比 SSD 高近 7%,且速度几乎是其 4 倍。YOLOv3 在 Electronics 上达到了较好的检测精度和效率,但是 ETS-Net 仍然比 YOLOv3 的 mAP 高 3.13%。YOLOv4 的精度超过了 96%,但是其后处理过程耗时长,导致模型整体效率并不高,检测速度只能达到 41 fps,影响了其综合性能。另外,本文在相同的实验平台上训练了 EfficientDet 系列的 D2 版本,这是在该实验平台上能训练的该系列算法的最高版本,但是其速度只能达到 15 fps,且 mAP 未突破 80%。因此,ETS-Net 具有最好的精度与速度的平衡。

表 3 Two-stage 算法各类别错误样本比例

类别	Faster R-CNN (R)	Faster R-CNN (M)	ETS-Net (本文)
Cap 220 μ F	0.25	0.18	0.21
Cap 22 μ F	0.14	0.13	0.15
Cap 470 μ F	0.26	0.43	0.29
Inductance	0.10	0.09	0.14
Cap_vred	0.05	0.14	0.23
Rst_p14D	0.08	0.07	0.08
Rst_ht	0.76	0.49	0.20
Idc_ylw	0.06	0.11	0.06
Rst_p10D	0.09	0.14	0.05
Idc_inv	1.11	0.23	0.10
(F_p)	0.29	0.20	0.15

表 4 Two-stage 算法错误样本比例标准差

	Faster R-CNN (R)	Faster R-CNN (M)	ETS-Net (本文)
F_p 标准差	0.36	0.14	0.08

为更直观地展示算法检测效果,本文选取精度最高的 one-stage 算法 YOLOv4 与本算法对比,在测试集中任意

挑选一张图像,进行检测效果可视化,结果如图6所示。图6(a)是YOLOv4的检测效果图,可以看出在图像中间偏右部位及上部偏右部位,漏检了电容Cap 220 μ F,在图像中间偏左部位漏检了黄色电容Idc_ylw。并在图像左下和右下部位,错误地将电容Cap 470 μ F检测为与之及其相似的电容Cap_vred。从图中可以看出即使是正确检测出的元器件,其边界框也未紧密包围目标。图6(b)是ETS-Net的检测结果,所有元器件都以高于90%的置信度被正确检测出来,并且边界框以较高的精度包围着目标,具有

较高的定位精度。考虑到工业场景中元器件易出现混淆的情况,本文设计干扰目标检测实验,检验算法应对干扰目标的性能。实验结果如图7所示,图7(a)为YOLOv3的检测结果,从图中可以看出,除了几个漏检和误检之外,一个黑色的干扰元器件被错误地检测为电感Idc_inv。图7(b)是ETS-Net的检测结果,可以看出检测结果并未受到干扰项的影响,且无漏检和误检。以上的检测结果证明ETS-Net在实际的检测任务中具有较高的稳定性和精确度,以及较强的抗干扰性。

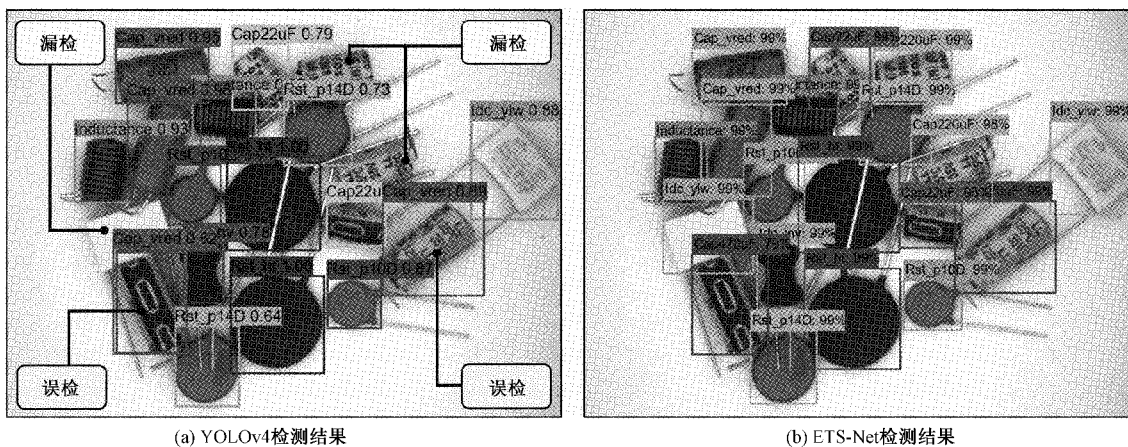


图6 YOLOv4与ETS-Net检测结果对比

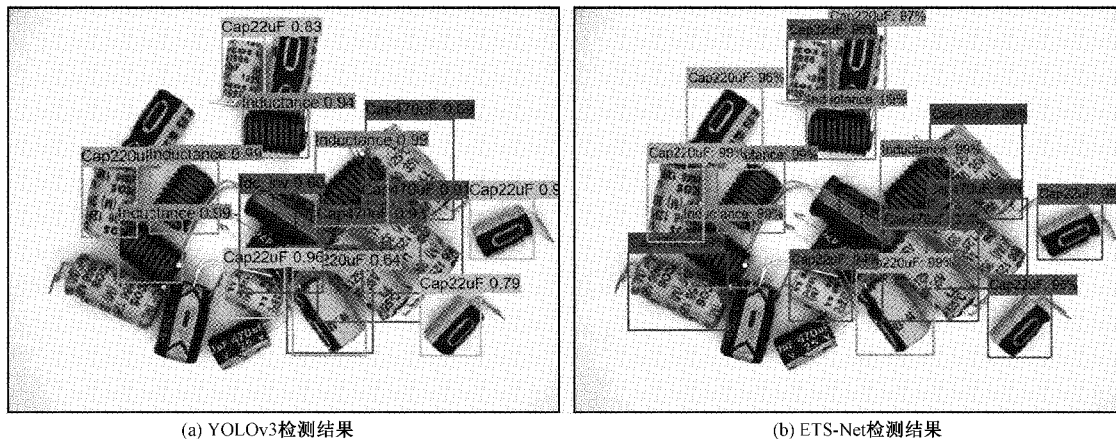


图7 干扰目标检测实验结果

在评估模型效率时,本文使用FLOPs和模型参数量两个指标来量化模型,通过分析模型的运算量和参数量解释模型的运行效率和实际使用价值。表5展示了各模型的浮点运算量(floating point operations, FLOPs)和参数量,运算量统计了乘法和加法操作的次数,以G为单位计量,1G=10⁹。参数量统计了权重参数的个数,以M为计量单位,1M=10⁶。表格中括号内字母R和M分别代表特征提取网络Res.Net50和Mobile.NetV2。

由表5可以看出,EfficientDet具有最少的参数量,但其实际效率确是最低的。而ETS-Net在保持较少参数量

的同时具有最低的运算量,且效率达到了最高。YOLOv3和YOLOv4虽然具有相近的参数量和运算量,但是YOLOv3的效率远高于后者。SSD的运算量约为YOLOv3的1/2,且仅有24.20M个参数,但实际效率只能达到21fps。而以ResNet50和MobileNet v2为主干网的Faster R-CNN虽然具有数倍于SSD的参数量和运算量,但是在速度上仍快于后者。

模型的运算量越大,说明模型的运算复杂度高。模型的参数量大,表明模型对设备的内存容量要求较高。以上的实验结果说明,ETS-Net具有最低的运算复杂度和较低

表 5 不同模型运算量和参数量

模型	FLOPs(G)	Params(M)	fps
SSD	31.18	14.18	21
YOLOv3	65.58	61.57	74
EfficientDet	20.28	8.02	15
YOLOv4	59.82	63.99	41
Faster RCNN(R)	204.32	39.54	26
Faster RCNN(M)	79.26	82.4	35
ETS-Net(<i>ours</i>)	18.24	16.51	78

的参数量,因此在算力更低廉的设备上部署时,本文提出的算法将更具有效率优势。

此外,将本文提出的特征提取网络 ELNet 与其他性能优异的轻量化特征提取网络对比,主要对比指标为浮点运算量 FLOPs(以 G 为度量单位,1 G=10⁹)以及检测时实际帧数表现 fps,对比结果如表 6 所示。从表中可以看出,虽然 ShuffleNetV1 具有最小的浮点运算量 0.28 G,但是实际效率并未达到最高。Xception 的乘加运算量达到了 17.74 G 次,其推理效率也仅有 15 fps,这在算力有限的设备上通常会有更低的帧数表现。ELNet 具有最高的推理速度,并且其消耗的运算资源也仅有 ShuffleNetV2 的一半。虽然 MobileNetV2 也具有较低的 FLOPs,但是其效率仅有 ELNet 的 40%。由于本文主要考虑实际效率表现,因此 ELNet 更适合应用于实际的检测任务中,并且 ELNet 消耗的计算资源更少,有利于向算力更低的设备部署。

表 6 特征提取网络效率对比

主干网	FLOPs (G)	fps
ShuffleNet V1	0.28	42
ShuffleNet V2	0.94	70
MobileNet V2	0.62	32
Xception	17.74	15
ELnet(<i>ours</i>)	0.48	78

以上实验结果说明,本文提出的 ETS-Net 算法具有较高的检测准确率,而且检测的稳定度高于其他 mAP 相近的 Two-stage 算法,同时所提出的算法还具有较强的抗干扰能力,能准确屏蔽干扰目标、准确区分外观极为相似的目标。实验最后一部分对比了其他一些轻量化特征提取网络,结果表明本文提出的 ELNet 的实际效率最高。因此,本文提出的 ETS-Net 算法具有较高的检测精度和稳定度,能够代替人工在电子元器件装配过程中进行识别和定位,为智能装配机器人提供必要的装配信息。

3 结 论

深度学习在诸多视觉任务中获得成功,这对于工业领域的零件识别、定位问题提供了新的解决方案。本文以智

能装配机器人视觉自主识别电子元器件为背景,应用深度学习技术在 two-stage 算法框架下构建轻量化的高性能目标检测模型,实现了电子元器件的快速视觉识别与定位。该算法可以识别具有极小视觉差异的目标,并具有较高的稳定性和效率。本文提出的算法还可以广泛应用于诸如缺陷检测、零件外观检测、识别等智能系统中的其他视觉任务。虽然该算法已经具备较高的精度和效率,但是在一些算力更低的嵌入式设备上仍然无法达到实时性要求,下一步还需要在网络结构和数据规模上继续优化,继续扩充数据集,以提升算法在实际生产中的抗干扰性和实用性。

参考文献

- [1] REDMON J, DIVVILA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Computer Vision and Pattern Recognition, Piscataway, USA: IEEE, 2016:779-788.
- [2] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017: 7263-7271.
- [3] REDMON J, FARHADI A. YOLOv3: An incremental improvement[C]. Computer Vision and Pattern Recognition Piscataway, USA: IEEE, 2018: 36, 1-6.
- [4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, Legender Str, Germany: Springer, 2016: 21-37.
- [5] GIRSHICK R, DONAHUE J, DARELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Computer Vision and Pattern Recognition Piscataway, American: IEEE, 2014: 32: 580-587.
- [6] GIRSHICK R. Fast R-CNN [C]. International Conference on Computer vision. Piscataway, USA: IEEE, 2015: 1440-1448.
- [7] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(6): 1137-1149.
- [8] 周晓彦,王珂,李凌燕.基于深度学习的目标检测算法综述[J].电子测量技术,2017,40(11):89-93.
- [9] 杜年丰,史泽地,陈伟,等.基于机器视觉的障碍物危险层次划分系统[J].电子测量与仪器学报,2021,35(6): 17-26.
- [10] 吕政阳,邓涛,张丽艳.一种基于机器视觉的飞机钣金件跨粒度识别方法[J].仪器仪表学报,2020,41(2): 195-204.
- [11] 刘雷霆,袁芬.机器视觉技术的包装缺陷在线识别系

- 统[J]. 国外电子测量技术, 2021, 40(5): 109-113.
- [12] 张静, 刘凤连, 汪日伟. 智能装配中基于YOLO v3的工业零件识别算法研究[J]. 光电子·激光, 2020, 31(10): 1054-1061.
- [13] 杜学丹, 蔡莹皓, 鲁涛, 等. 一种基于深度学习的机械臂抓取方法[J]. 机器人, 2017, 39(6): 820-828, 837.
- [14] ZEILER M, FERGUS R. Visualizing and understanding convolutional networks[C]. European Conference on Computer Vision, 2013: 818-833.
- [15] 余永维, 韩鑫, 杜柳青. 基于Inception-SSD算法的零件识别[J]. 光学精密工程, 2020, 28(8): 1799-1809.
- [16] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]. Computer Vision and Pattern Recognition, Piscataway, USA: IEEE, 2015: 1-9.
- [17] TAN M X, PANG R M, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Computer Vision and Pattern Recognition, Piscataway, USA: IEEE, 2020: 10781-10790.
- [18] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]. Computer Vision and Pattern Recognition, Piscataway, USA: IEEE, 2016: 770-778.
- [19] ZHANG X Y, ZHOU X Y, LIN M X, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]. Computer Vision and Pattern Recognition, Piscataway, USA: IEEE, 2018: 6848-6856.
- [20] MA N N, ZHANG X Y, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient CNN architecture design [C]. European Conference on Computer Vision, Piscataway, USA: IEEE, 2018: 116-131.
- [21] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. Computer Vision and Pattern Recognition, Piscataway, USA: IEEE, 2017: 1251-1258.
- [22] PASZKE A, GROSS S, MASSA F, et al. PyTorch: An imperative style, high-performance deep learning library[C]. Neural Information Processing Systems, 2019: 32, 8026-8037.
- [23] LIN, T Y, PIOTR D, ROSS G, et al. Feature pyramid networks for object detection[C]. Computer Vision and Pattern Recognition, Piscataway, USA: IEEE, 2017: 2117-2125.
- [24] SANDLER M, ANDREW H, ZHU M L, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Computer Vision and Pattern Recognition, Piscataway, USA: IEEE, 2018: 4510-4520.

作者简介

张志杰, 硕士研究生, 主要研究方向为基于深度学习的目标检测技术、图像分类、图像分割等。

顾寄南, 工学博士, 教授, 主要研究方向为智能制造技术及系统、智能机器人及机器视觉、复杂产品建模与创新设计理论、基于网络的设计制造及远程监控、机械 CAD/CAE 及计算机图形处理技术等。

E-mail: gujinan@tsinghua.org.cn

李静, 博士研究生, 主要研究方向为基于深度学习的目标检测、图像分割等。

余雪飞, 硕士研究生, 主要研究方向为基于深度学习的三维重建、目标识别等。