

DOI:10.19651/j.cnki.emt.2107119

特征增强的 SSD 算法及在仪表读数识别上的应用*

胡鑫 欧阳华 侯卓成 尹洋
(海军工程大学 电气工程学院 武汉 430033)

摘要: 针对传统 SSD 算法在不同尺度的特征缺少相互之间的特征信息的交流,不利于目标的定位和识别,使得传统 SSD 算法的检测结果往往无法达到实际要求精度,并且检测速度也无法满足实际需求的问题,因此对其进行改进,引入轻量化网络 MobileNet-v2 和特征融合模块,使得该算法(ENMobileSSDnet)在应对巡检环境下数字式仪表读数的识别问题检测效果相比于传统 SSD 算法性能有了显著性提高,检测精度(mAP)达到了 98.32%,检测速度(FPS)达到了 72 帧/s,对实际的工程项目有更好的应用价值。

关键词: SSD;轻量化网络;特征融合;目标识别

中图分类号: TP39 **文献标识码:** A **国家标准学科分类代码:** 510.1050

Feature enhanced SSD algorithm and its application in
meter reading recognition

Hu Xin Ouyang Hua Hou Zhuocheng Yin Yang

(College of Electrical Engineering, Naval University of Engineering, Wuban 430033, China)

Abstract: Aiming at the problem that the traditional SSD algorithm lacks the exchange of feature information between features at different scales, which is not conducive to target positioning and recognition, so that the detection results of the traditional SSD algorithm often can not meet the actual accuracy requirements, and the detection speed can not meet the actual requirements, so it is improved, and the lightweight network MobileNet-v2 and feature fusion module are introduced. This makes the algorithm (ENMobileSSDnet) deal with the identification problem of digital instrument indication in patrol environment, and the detection effect is significantly improved compared with the traditional SSD algorithm. The detection accuracy (mAP) has reached 98.32% and the detection speed (FPS) has reached 72 frames/s. It has better application value for practical engineering projects.

Keywords: SSD;lightweight network;feature fusion;target recognition

0 引 言

近年来,随着深度学习广泛应用,目标检测相关算法得到了巨大的发展。目前基于深度学习的目标检测算法主要包括两大类:第1类是两段式(two-stage)算法,基于候选区域的检测算法,首先生成一系列的样本作为候选框,获取目标所在的位置区域,然后再对样本进行分类,包括 R-CNN (regions with CNN features)、Fast R-CNN^[1]、Faster R-CNN^[2]等;第2类是一段式(one-stage)算法,通过卷积神经网络直接输出目标所在的位置以及类别,包括 SSD (single shot multibox detector),YOLO (you only look once)^[3]等。一般来说基于候选区域的两段式方法在准确率与精度上表现较好,端到端的一段式方法在速度上比较

有优势^[4-6]。为了保证对目标的实时性,本文选择一段式目标检测模型,选择 SSD300 作为巡检过程中对数字式仪表读数的检测模型^[7-10],在保证检测精度的同时,也可以满足检测的实时性。

SSD 目标检测模型属于一段式检测算法,运行速度较快,但是不同于 YOLO 等其他同类方法,SSD 模型通过提取不同尺度特征图、引入先验框等方法,弥补了 YOLO 算法难以检测小目标,定位精度低等不足^[11-12]。在 PASCAL VOC 2007 测试集上对不同的目标检测算法进行测试,该数据集一共包括 9 963 张图像,其中训练集 5 011 张,测试集 4 952 张,共计 20 个种类的目标。测试后得到结果:一段式的目标检测算法在 mAP(mean average precision)精度上略有优势,但耗时间长,FPS(frames per second)较低;

收稿日期:2021-06-30

* 基金项目:国家自然科学基金(41876222)项目资助

两段式算法检测速度非常快,但精确度稍低于 Faster R-CNN 系列算法。总体来说,SSD 模型在保持较高检测精度的同时,提供了更快的检测速度,并对小型目标的检测效率也具备良好的效果。对于巡检环境下仪表读数的检测识别,选用 SSD 模型是一个相对较为理想的检测器。本文针对传统 SSD 模型对数字式仪表读数识别速度慢、精度不高的问题,引入了轻量化网络和特征增强模块,研究了一种 SSD 改进算法。在保证识别速度的情况下,提升读数识别的精度。针对数字式仪表读数识别的整体设计,本文将其分为了两个阶段,一张经过采集的仪表图像,通过连续两次的 SSD 目标检测,以级联的方式将两个阶段相连,从而实现了更为准确连贯的读数识别。

1 轻量化网络

所引入的 MobileNet-v2 则属于从 MobileNet 网络衍生出的一个改进结构。其核心部分依旧是深度可分离卷积(depthwise separable convolution)。

假设给定的卷积核大小尺寸为 $D_K \times D_K \times M$,输入特征图的尺寸大小为 $D_F \times D_F \times M$,输出特征图尺寸为 $D_F \times D_F \times N$, M 为输入的通道数, N 为输出的通道数。当有 padding 并且步长为 1 时,标准卷积层的计算量设为 C_1 ,深度可分离卷积的计算量设为 C_2 ,那么可得:

$$C_1 = D_F \times D_F \times M \times D_F \times D_F \times N \quad (1)$$

即其计算量由输入通道 M ,输出通道 N ,卷积核大小 D_K ,输出特征图大小 D_F 决定。

$$C_2 = D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F \quad (2)$$

即深度卷积和 1×1 的逐点卷积的和。其通过将卷积分为滤波和组合的过程得到对计算量的缩减,将上述 C_1 、 C_2 进行比较,可得:

$$\frac{C_2}{C_1} = \frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_F \times D_F \times M \times D_F \times D_F \times N} = \frac{1}{N} + \frac{1}{D_K^2} \quad (3)$$

MobileNet 使用 3×3 的卷积核时,相较于标准卷积而言,在识别精度上相差无几,但网络计算量减少了 8~9 倍。深度可分离卷积的运用,使得神经网络开始变得轻量化。

而 MobileNet-v2 相比与最初的 MobileNet-v1 在 DW 卷积之前加入了一个 PW 卷积,用来进行升维,在更高层的空间提取特征;另外,网络为全卷积的形式,使得模型可以适应不同尺寸的图像,使用 ReLU6 激活函数,使得模型具有更强的鲁棒性。MobileNet-v2 还引入了残差结构,不同于一般的 ResNet 使用标准卷积提取特征,是一个先降维再卷积,最后再进行升维,类似于沙漏状;其通过构造反转残差模块(inverted residuals block)先进行升维再卷积,最后再进行降维,类似于纺锤状,这样也能够最小程度地抵消 ReLU 造成的信息丢失。

2 特征融合

不同尺度的特征缺少相互之间的特征信息的交流^[13-15],还是不利于目标的定位和识别,这样会导致传统 SSD 算法的检测结果往往无法达到实际要求精度。

本文是基于传统方法的不足介绍一种改进的特征增强策略。这里设计的特征融合模块就是对常用的单向融合的特征金字塔进行改进,浅层的特征分辨率高,结构信息完整,深层的特征分辨率低,但是包含丰富的语义信息。若将浅层和深层的特征信息结合起来,突出重要的部分,则可以达到增强图像特征的目的。结合引用的 MobileNet-v2 骨干网络,在其基础之上进行不同级别双向特征融合模块的构建,将不同级别的特征信息进行双向融合使之成为一个能够完成特征增强的 SSD 网络模型。网络结构如图 1 所示。

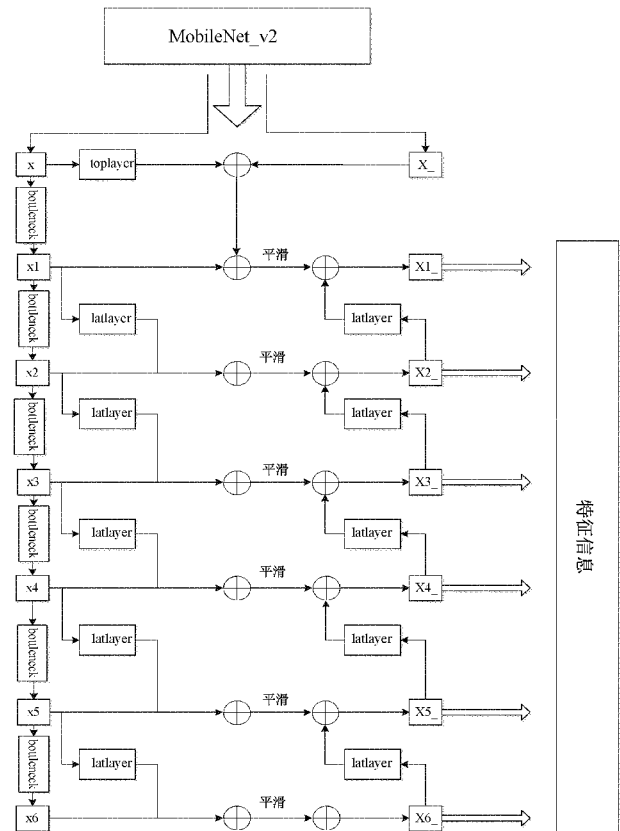


图 1 改进的网络结构

由图 1 网络结构可以看出,融合策略是以骨干网络 MobileNet-v2 的多个特征层为基础展开,通过上下采样,使得相邻层次之间的特征信息进行融合,并且每一层都能通过不断融合的方式使得其包含更多全局的特征信息,更精准地实现显著性目标的定位与识别。整个过程的计算公式如下:

$$X1_ = \text{Add}(\text{Cov}(X), X_ , X1, \text{Cov}(X2_)) \quad (4)$$

$$X2_ = \text{Add}(\text{Cov}(X1), X2, \text{Cov}(X3_)) \quad (5)$$

$$X3_ = \text{Add}(\text{Cov}(X2), X3, \text{Cov}(X4_)) \quad (6)$$

$$X4_ = \text{Add}(\text{Cov}(X3), X4, \text{Cov}(X5_)) \quad (7)$$

$$X5_ = \text{Add}(\text{Cov}(X4), X5, \text{Cov}(X6_)) \quad (8)$$

$$X6_ = \text{Add}(\text{Cov}(X5), X6) \quad (9)$$

综合网络结构和上述计算公式可知,融合算法是以 MobileNet-v2 的多个特征层 X1~X6 作为输入,将 X 经过上采样与 X_、X1 以及经过下采样的 X2_ 进行融合,从而得到新的金字塔特征层 X1_;将 X1 经过上采样与 X2 以及经过下采样的 X3_ 进行融合,从而得到新的金字塔特征层 X2_;将 X2 经过上采样与 X3 以及经过下采样的 X4_ 进行融合,从而得到新的金字塔特征层 X3_;将 X3 经过上采样与 X4 以及经过下采样的 X5_ 进行融合,从而得到新的金字塔特征层 X4_;将 X4 经过上采样与 X5 以及经过下采样的 X6_ 进行融合,从而得到新的金字塔特征层 X5_;将 X5 经过上采样与 X2 进行融合,从而得到新的金字塔特征层 X6_。最后特征金字塔 FPN 得到的 6 个特征层为 X1_~X6_,经过不同级别的特征进行双向融合,整个网络结构的输出端比之前能够包含更多的细节结构信息,实现更精准的显著性目标检测。

3 实验分析验证

为了证实本文方法的有效性,在这以一个完整工程实际应用的方式来进行实验验证。利用以 MobileNet-v2 作为骨干网络的 SSD 模型和改进方法进行对比,实现机器巡检过程中对数字式仪表表示数的识别。

loss 函数将损失函数分为位置误差和分类误差两部分,总体目标损失函数是定位损失 (loc) 和置信度损失 (conf) 的加权和,其整体表达式如下:

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (10)$$

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in (cx, cy, w, h)} x_{ij}^k \text{smooth}_{l,1}(l_i^m - \hat{g}_j^m) \quad (11)$$

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i) , \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \quad (12)$$

$$\text{smooth}_{l,1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{其他} \end{cases} \quad (13)$$

式中: x_{ij}^k 为第 i 个预测框与第 j 个真实框关于类别 k 是否匹配(值为 0 或 1); l_i^m 为预测框的大小; x_{ij}^p 为预测框与真实框关于类别 p 匹配; c_i^p 为类别为 p 的预测值,通过 SoftMax 转换; α 为权重系数,由于对目标所在图像中位置精度要求相对较高,因此设为 1; \hat{g}_j^m 为真实框的大小; N 为匹配的默认边界框的数量。如果 $N=0$,则将损失设为 0。定位损失是预测框 (l) 与真实框 (g) 参数之间的 Smooth L1 损失。类似于 Faster R-CNN,一般回归默认边界框 (d) 的中心偏移量 (cx, cy, cw, ch) 和其宽度 (w)、高度 (h) 的偏移量。

评判标准主要就是依靠这两个指标:采用每秒传输帧数 (frames per second, FPS) 来衡量比较经过这个网络结构

检测出目标的速度;采用平均精度均值 (mean average precision, mAP) 来评价检测多类目标时的精度,并且当只需探讨研究其中的某一类目标时,可利用平均精度 (average precision, AP) 来表达评估。

首先根据机器巡检识别这一具体应用环境将整个识别过程分为两个阶段。第 1 阶段完成在获取的图片信息中对数字显示区域的定位识别;第 2 阶段就在第 1 阶段的基础上对得到的数显区域的数字进行再次定位识别,最后输出的就是需要得到的示数。数字式仪表表示数识别算法的整个流程如图 2 所示。

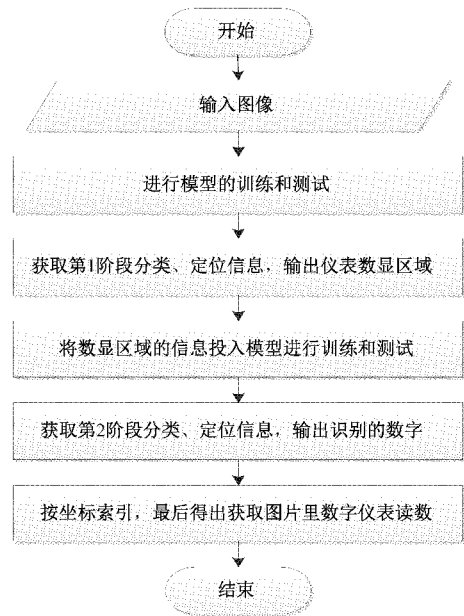


图 2 整体流程

因为是特殊应用对象,无法利用通用的标准数据集进行训练,需要根据应用对象进行数据集的制作。本次实验针对需要大量不同类别数字的需求选用一个小型数字式拉力计。根据数字的变化模拟仪表在巡检过程中示数的自然变换,得到了大约 2 000 张数据集样本。同时,考虑到光照在目标识别过程中可能会对精度造成影响,所以在制作部分数据集样本的时候通过改变图片的色调模拟不同强度光照环境下捕捉的仪表图像。第 1 阶段将图片分为两类,分别是 panel(数显区域)和 background(背景);第 2 阶段将得到的 panel(数显区域)分为 12 个类别,分别是数字 0~9、小数点和 background(背景)。并且数据集样本按照 9:1 的比例划分训练集和测试集,两个阶段的数据集均使用标注软件 labelImg 进行数据和类别的标注。首先在第 1 阶段中,模拟巡检过程中对仪表画面的采集,输入仪表的拍摄图片,利用之前用制作完毕的数据集训练好的训练模型之中进行识别;第 2 阶段的数据集为了保证数据来源的合理性,都是利用经过第 1 阶段识别之后得到的数显区域,通过得到的坐标、长和宽这几个信息将识别框从原图中裁剪出来。以这种方式制作的数据集使得两个阶段更具有连贯性、合

理性。第1、2阶段的数据集样本如图3(a)、(b)所示。

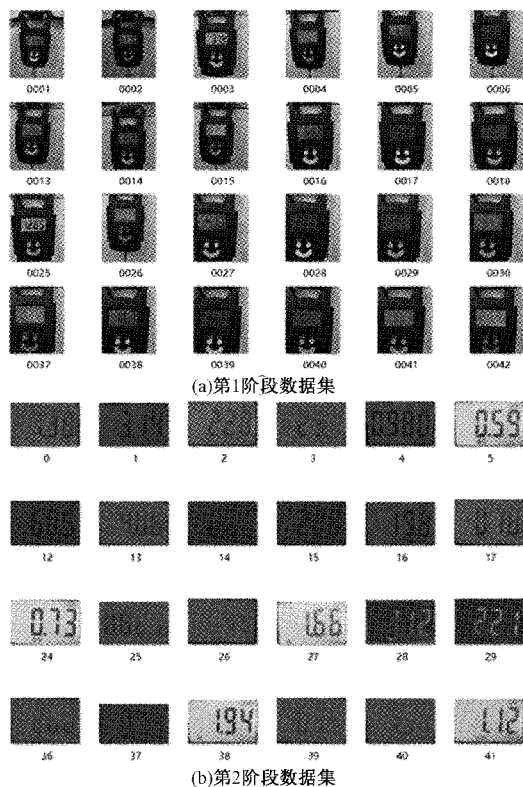


图3 两个阶段数据集

数字式仪表检测算法对训练学习率的选择也十分重要,因为学习率能够决定模型和算法的训练表现。当学习率过于小的时候,将会导致训练长时间无法收敛,当学习率过大的时候,将会导致利用梯度下降法无法找到最优的解。因此主要从数据集大小和成本函数两方面综合考虑选择的结果,根据训练迭代结果的收敛情况最终选定固定的训练学习率值为0.003。

由于第1阶段的任务较为简单,只是针对采集到的图片对数显区域进行一个识别,实则属于一个二分类问题,所以不进行特征增强的改进,利用的模型为MobileNet-v2—SSD(即将原来SSD的骨干网络VGG16换成轻量化网络MobileNet-v2)。

如图4所示,模型经过1600轮训练,训练的位置损失和分类损失最后都能逐渐趋于稳定,训练总的损失值在稳定后大致维持在1.37,结果处于一个较好的拟合状态。此阶段将图片处理分成数字显示区域和背景两大类,在经过1200轮训练之后,仪表数字显示区域的识别精度最终趋于稳定,并且大致维持在99.43%,处于一个很好的拟合状态。

本阶段的训练效果从迭代曲线来看,能够得到比较好的损失值和精度,分析原因主要是如下3点。

1) 选择了一个较为合适的学习率,使得模型经过多次训练迭代之后,能够得到一个对全局而言的相对最优解;

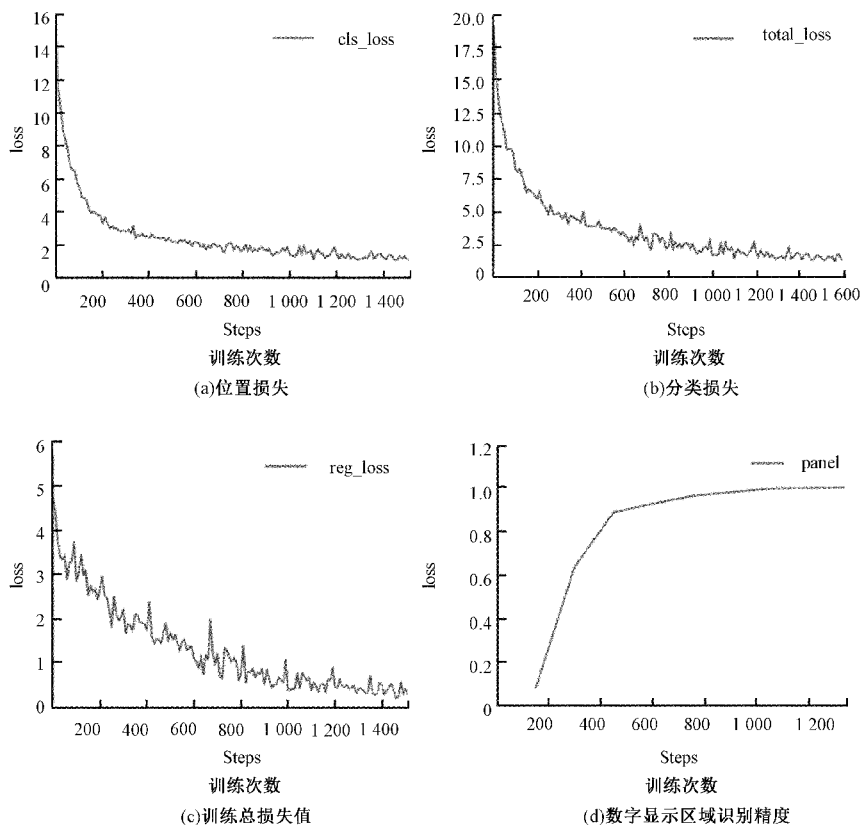


图4 第1阶段目标损失和精度

2)数据集较为充分,针对特定的仪表对象数字拉力计进行数据集的构建,并且将光照作为影响因素考虑进去;使得构造的模型训练更加充分;

3)识别任务较为简单,仅仅是针对数显区域进行一个识别,相当于一个二分类的用途,最主要的目的还是针对获取的图片对仪表的数显区域进行精准定位,为下一阶段的示数识别打好基础。

可以看到,数字式仪表第 1 阶段的识别定位效果较好,仪表经过检测之后,能够以比较高的精度得到要求的结果。

第 2 阶段,对数字式仪表示数的识别读取,也是整个项目中最为重要的部分。其所有部分都是建立在第 1 阶段的基础之上进行。改进得到的特征增强算法也将应用到第 2 阶段之中。

为了突出显示改进算法具有更好的性能,将之与未进行特征增强的 MobileNet 模型的结果进行对比实验,由于 MobileSSDnet 与原始模型相比,增强的是速度,精度几乎不改变,因此在进行 mAP 的仿真比较时,使用 MobileNet 和 ENMobileNet 两种模型对比即可。其仿真得到的损失和平均精度对比结果如图 5 所示。

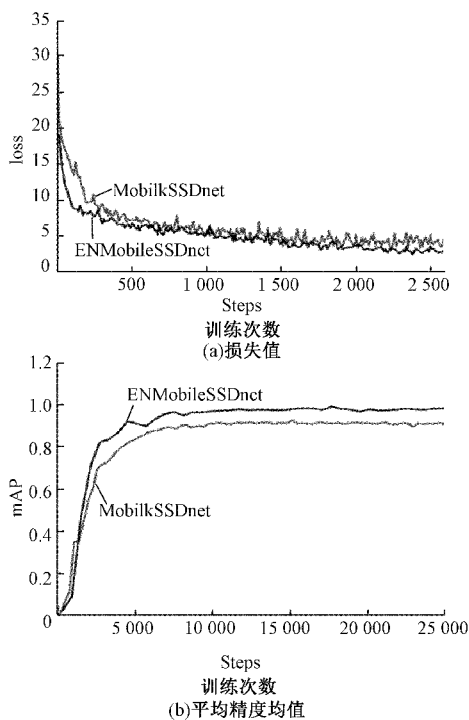


图 5 第 2 阶段损失和平均精度均值对比

分析图 5 可得,经过改进之后的 ENMobileSSDnet,在加入特征增强这个模块之后,整体的性能表现会有相当程度上的提高。经过至少 2 500 轮迭代之后,模型的损失值趋于稳定,并且相较之原来的 MobileSSDnet 模型,训练起始损失值会大大减小,趋于稳定之后,也会比原来减小 1.5 左右。同时平均精度均值与原来相比,经过 5 000 轮迭代

以后,精度曲线趋于稳定之后,原模型的 mAP 为 89.43%,改进之后的模型的 mAP 可以达到 98.32%,有了一个较大程度地提高。并且在此对各个类别的识别精度进行一个对比,如表 1 所示。

表 1 每个类别的精度对比

| 类别 | 平均精度 | |
|----|--------------|----------------|
| | MobileSSDnet | ENMobileSSDnet |
| 0 | 0.883 1 | 0.984 9 |
| 1 | 0.613 5 | 0.904 3 |
| 2 | 0.936 8 | 0.983 3 |
| 3 | 0.958 4 | 0.995 7 |
| 4 | 0.953 3 | 0.996 9 |
| 5 | 0.786 9 | 0.998 1 |
| 6 | 0.943 5 | 0.995 3 |
| 7 | 0.947 3 | 0.996 3 |
| 8 | 0.958 9 | 0.999 9 |
| 9 | 0.976 1 | 0.999 3 |
| . | 0.683 5 | 0.906 8 |

由表 1 可以看出,通过特征融合,使得每个类别的数字识别精度大幅度提高,这说明特征增强将浅层信息与高层信息进行互相融合,充分对捕捉的图像特征进行提取,使得模型在分类和定位上的精度都有着显著改进。

除了对 ENMobileSSDnet 的仪表检测精度进行分析之外,还将进一步探究此改进模型的检测速度,主要是根据每秒传输帧数(frames per second, FPS)这个指标进行评估,结果如表 2 所示。

表 2 检测速度对比

| 网络模型 | FPS/(帧·s ⁻¹) |
|--------------|--------------------------|
| SSD300 | 59 |
| MobileSSDnet | 79 |
| ENMobileNet | 72 |

将改进的模型 ENMobileSSDnet 与初始 SSD300(骨干网络为 VGG6)、MobileSSDnet(未进行特征融合)三者的检测速度进行对比,结果显示,MobileSSDnet 的检测速度最快,改进模型其次,SSD300 最慢。分析可得,对于检测速度这一指标而言,通过利用轻量化网络 MobileNet-v2 代替 vgg16,使得整体的检测速度大大提高,并且特征增强模块的引入会小幅度地降低检测速度。

可以看出,通过改进方法进行训练,数字拉力计的示数都能以高精度被识别出,和真实读数相比没有误差。最后将第 1 阶段和第 2 阶段以一个级联的方式串接成一个整体,完整实现对数字式拉力计这类数字式仪表的示数识别,其第 1 阶段、第 2 阶段和整体实验效果如图 6 所示,从

而可知本文的改进算法提升了模拟巡检情况对于数字式仪表读数的检测识别性能。

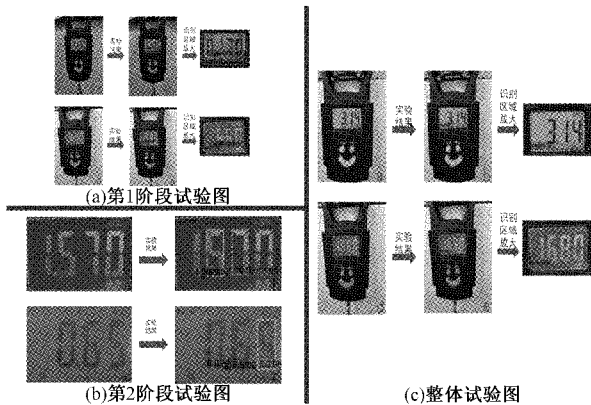


图6 实验效果

4 结 论

本文是基于特征融合的思想,以SSD模型为基础,通过选用轻量化的网络MobileNet-v2来提高检测速度,通过对传统的SSD算法的特征金字塔网络框架进行改进,通过上下采样、 1×1 卷积等操作,对浅层和深层特征信息进行融合,从而提高检测识别的精度。之后将改进算法引入到智能巡检环境下对数字式仪表读数的识别这一具体应用上进行验证其有效性。由实验结果可以看出,改进得到的算法在mAP和FPS两个指标上有显著提高,检测性能表现得更为均衡。本文的下一步工作将是通过优化模型,考虑更多因素,让应用对象具有更普遍的适用性,以达到让模型算法能够在以后具体项目上有更好的应用空间,从而推动巡检机器人进行智能读表有进一步的发展提高。

参考文献

- [1] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, Los Alamitos: IEEE Computer Society Press, 2015: 1440-1448.
- [2] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards realtime object detection with region proposal networks[C]. Proceedings of the 28th International Conference on Neural Information Processing Systems, Cambridge: MIT Press, 2015, 1: 91-99.
- [3] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, Los Alamitos: IEEE Computer Society Press, 2016: 779-788.
- [4] REINER A J, HOLLANDS J G, JAMIESON G A. Target detection and identification performance using an automatic target detection system [J]. Human Factors, 2017, 59(2): 242-258.
- [5] EVERINGHAM M, ESLAMI S M A, GOOL L V, et al. The pascal visual object classes challenge: A retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.
- [6] DAI J, LI Y, HE K. R-fcn: Object detection via region-based fully convolutional networks [J]. Advances in Neural Information Processing Systems, 2016: 379-387.
- [7] 何配林. 基于深度学习的工业仪表识别读数算法研究及应用[D]. 成都: 电子科技大学, 2020.
- [8] 董峻妃. 基于卷积神经网络的车牌区域检测和车牌字符识别研究[D]. 南充: 西华师范大学, 2018.
- [9] 张宇澄. 深度学习在车牌识别中的应用研究[D]. 南京: 南京大学, 2018.
- [10] 张珺. 基于机器视觉的仪表读数智能识别算法研究及应用[D]. 成都: 电子科技大学, 2019.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. 2015 14th European Conference on Computer Vision (ECCV), 2015: 1-17.
- [12] CAO G, XIE X, YANG W. Feature-fused SSD: Fast detection for small objects [C]. 9th International Conference on Graphic and Image Processing (ICGIP 2017), 2017.
- [13] 王飞, 张莹, 邵豪, 等. 多尺度残差网络模型的研究及其应用[J]. 电子测量与仪器学报, 2019, 33(4): 19-28.
- [14] 程淑红, 程彦龙, 杨镇豪. 基于手势多特征融合及优化 Multiclass-SVC 的手势识别[J]. 仪器仪表学报, 2020, 41(6): 225-232.
- [15] 范新南, 鹿亮, 史朋飞, 等. 结合 MSRCR 与多尺度融合的水下图像增强算法[J]. 国外电子测量技术, 2021, 40(2): 6-10.

作者简介

胡鑫, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 1810347286@qq.com

欧阳华, 副教授, 主要研究方向为信号处理。

E-mail: 78ouyang@163.com

侯卓成(通信作者), 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 2193513923@qq.com

尹洋, 讲师, 主要研究方向为控制工程。

E-mail: reeyan@163.com