



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331, CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目: 轻量化 YOLO-v7 的数显仪表检测及读数
作者: 章芮宁, 闫坤, 叶进
网络首发日期: 2023-07-19
引用格式: 章芮宁, 闫坤, 叶进. 轻量化 YOLO-v7 的数显仪表检测及读数[J/OL]. 计算机工程与应用.
<https://kns.cnki.net/kcms2/detail/11.2127.TP.20230718.1838.014.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

轻量化 YOLO-v7 的数显仪表检测及读数

章芮宁¹, 闫坤¹, 叶进²

1. 桂林电子科技大学, 广西 桂林 541004

2. 广西大学, 南宁 530004

摘要: 由于较大的参数体量和较高的计算复杂度, 通用检测及识别模型直接在移动端部署的难度较高。为解决这个困难, 本文研究了移动设备上使用计算机视觉的仪表检测及读数方法。针对实际工业生产环境下检测及识别的需求, 基于 YOLO-v7 重新设计了轻量化的仪表检测网络以及字符检测及识别网络。然后利用深度可分离卷积进一步降低计算复杂度, 压缩模型大小。接着采用 K-means++ 聚类算法加遗传算法自动产生初始锚框。最后使用通道剪枝, 再一次压缩模型。实验结果证明, 专用网络模型设计、深度可分离卷积以及通道剪枝对减少模型参数体量和降低算力需求具有显著效果。参数数量相较于原始 YOLO-v7 模型均下降了 99.67%, 模型算力需求均降至 0.3GFLOPs, 下降了 99.71%。实验中平均图片检测时间为 10.7ms。各网络的平均精准度 (mAP0.5) 达到了 99.63% 和 99.53%。系统整体读数精确度达 98.44%。

关键词: 数显仪表; YOLO-v7; 深度可分离卷积; 模型压缩; 通道剪枝

文献标志码:A 中图分类号:TP391 doi: 10.3778/j.issn.1002-8331.2304-0401

Lightweight YOLO-v7 for digital instrumentation detection and reading

ZHANG Ruining¹, YAN Kun¹, YE Jin²

1. Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

2. Guangxi University, Nanning 530004, China

Abstract: Due to the large parameter volume and high computational complexity, it is difficult to deploy generic detection and recognition models directly on mobile. To address this difficulty, a method for instrument detection and reading using computer vision on mobile devices is investigated. A lightweight meter detection network and a character detection and recognition network are redesigned based on YOLO-v7 to address the needs of detection and recognition in real industrial production environments. The depth-separable convolution is then used to further reduce the computational complexity and compress the model size. Then a K-means++ clustering algorithm plus a genetic algorithm is used to automatically generate the initial anchor box. Finally, channel pruning is used to compress the model once more. The experimental results demonstrate that the dedicated network model design, deep separable convolution and channel pruning have a significant effect on reducing the size of the model parameters and reducing the computational power requirements. The number of parameters both decreased by 99.67% com-

基金项目: 国家自然科学基金(62101147); 广西自然科学基金(桂科 2020GXNSFAA159146); 广西创新驱动发展专项(桂科 AA21077008)。

作者简介: 章芮宁(1997-), 男, 硕士研究生, CCF 学生会员, 研究方向为机器学习、图像处理; 闫坤(1983-), 通信作者, 女, 博士, 教授, 研究方向为机器学习、图像处理等, E-mail: yk5702@guet.edu.cn; 叶进(1970-), 女, 博士, 教授, 研究方向为网络协议优化、无线网络设计等。

pared to the original YOLO-v7 model, and the model arithmetic requirements were both reduced to 0.3 GFLOPs, a decrease of 99.71%. The average image detection time in the experiments equals to 10.7ms. The average accuracy (mAP0.5) of each network reached 99.63% and 99.53%. The overall system reading accuracy reaches 98.44%.

Key words: digital instrumentation; YOLO-v7; depth wise separable convolution; model compression; channel pruning

工业 4.0 时代, 向工业生产提出了新要求, 力求生产的每一阶段数字化, 企业纷纷提高数字化水平。但有些企业设备采购之初并未预置数字化接口, 而有些设备由于精确性和稳定性要求无法采用数字化仪表, 因此实时的准确的仪表检测及读数具有广大市场。

针对仪表的检测及读数, 国内外研究人员做了大量的相关研究, 给出了各种解决方案。文献^[1]针对室内的变电站数显仪表, 分析数字字符显示区域特性, 采用图像预处理的方式对显示区域进行定位切割, 最后采用模板匹配的方法进行读数转换。其运行效率虽然较同类图像处理仪表识别方法比较高, 但每秒也仅能处理近 10 帧的图像(Frames Per Second, FPS), 且牺牲了部分数字字符的识别率。这也反映了传统方法的弊端。传统方法对于环境背景干扰十分敏感, 需要较为纯净的背景图片输入识别网络才能获得较高的准确率。文献^[2]面对变压器直流电阻测试仪的数显仪表, 采用深度学习的方法。首先利用改进的 Faster R-CNN 搭建字符区域定位网络, 对感兴趣的字符区域进行检测切割。再将切割后的字符输入卷积神经网络进行字符识别。其中 Faster R-CNN 为两阶段目标检测网络。第一阶段寻找并生成大约 2000 个感兴趣目标可能出现的区域的候选框, 用以提高目标检测的准确率和召回率, 但是也增加了开销。第二阶段为了得出目标的种类和位置, 对每一个候选区域分别进行分类和回归。所以, 虽然作者的字符区域定位准确率达到了 96%。但是其总体的时间开销也来到了 200ms 每帧, 无法满足实时性的需求。2016 年, YOLO(You Only Look Once)^[3], 一种单阶段目标检测网络模型由 Joseph Redmon 等人提出, 并迅速应用于各种目标检测任务之中。其去除了区域建议生成网络, 将检测任务视为回归问题。并且由于输入网络的不是局部的建议区域, 而是整幅图像, 可以很好的联系上下文信息, 整合特征, 极大减少了两阶段网络容易产生的背景误差。文献^[4]针对工业环境下的数显仪表, 提出了基于 YOLO-v3 的仪表检测算法, 并进行了轻量化改进, 大幅度缩减了模型参数体量, 还提高了检测速度。但是其轻量化后的模型大小仍达 8MB, 算力需求为 1.83

千兆浮点运算次数(Giga Floating Point of Operations, GFLOPs)。并且准确率下降到 92.42%, 帧率也仅有 21FPS。继美团发布 YOLO-v6 之后, YOLO 系列的原作者也发布了 YOLO-v7。并且作者宣称 YOLO-v7 在速度和精度方面都超过了所有已知的目标检测算法。

本文将 YOLO-v7 目标检测网络应用在工业环境下的数显仪表检测和读数任务之中。凭借 YOLO-v7 出色的检测速度以及检测精度, 实现实时的准确的仪表检测及读数。为了压缩模型参数体量, 提高模型检测速度。本文针对工业环境采集图像的特性以及目标大小形体特点, 重新设计 YOLO 的骨干网络及头部网络模型。采用深度可分离卷积替换常规卷积。并用 K-means++ 聚类算法加速遗传算法找到合适的初始锚框数量及尺寸大小。最后, 为了便于后续将模型部署在移动端, 对模型进行进一步剪枝压缩, 在少量牺牲准确率的情况下, 将模型中冗余的通道进一步优化删减。

1 YOLO-v7

YOLO 是最先进的目标检测算法之一。它能够在图像中快速而准确地识别和定位多个物体。它将给定的图像分割成 $S \times S$ 个网格, 然后以每一个网格为中心, 目标的中心点落在哪一个网格内, 则由该网格负责检测。每一个格子分配并检测 B 个预测框, 并将目标存在的置信度 $Pr(\text{Object})$ (存在目标则为 1, 不存在则为 0)

以及预测框的置信度 $IOU_{\text{pred}}^{\text{truth}}$ (预测框与真实框 (ground truth) 的交并比) 的乘积作为置信度, 同时输出物体的对应每一类的类别概率。为了得到最有可能的预测结果, 对置信度和类别概率采用非极大值抑制。

YOLO-v7^[5]是新发布的 YOLO 版本, 其建立在传统的 YOLO 型号上, 如 YOLO-v4 和 YOLOR。作者通过调整参数的数量、计算量和处理密度来提高性能, 使其在实时应用方面具有很好的性能和效率。YOLO-v7tiny^[6]为官方发布的 YOLO-v7 简化模型, 其在卷积深度上较 YOLO-v7 做了大幅度缩减。其网络架构如图 1 所示。其中圆括号内为每一层输出的特征图的宽, 高, 通道数量 (W, H, C); 方括号内为卷积的输入通道数目, 卷积核大小, 步长, 填充数目

$[c_{in}, k, s, p]$; 粗箭头表示虚线框内的数据流向。

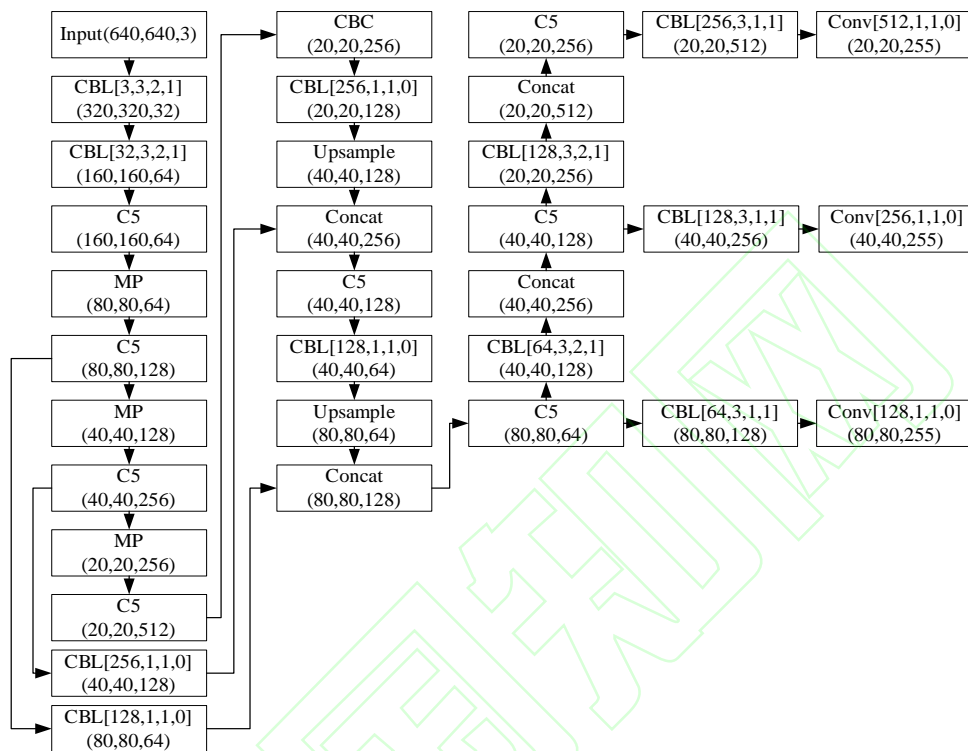


图 1 YOLO-v7tiny

Fig.1 YOLO-v7tiny

原始网络模型中预测框个数为 3, 检测类别为 80, 所以最后的输出特征个数为 $B \times ((x, y, w, h, confidence) + class) = 3 \times (5 + 80) = 255$ 。
(x, y, w, h) 表示预测框的坐标和尺寸信息。 x 和 y 表示中心点相对坐标, 以所在网格为参考系。 w 和 h 表示宽度和高度的比例, 对比整幅图像的尺寸大小。 $confidence$ 表示预测框的置信度, $class$ 表示类别概率。

图 2 为网络架构的细节补充。其中方括号内为卷积的 $[k, s, p]$ 。

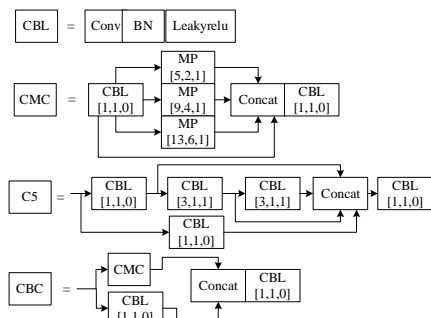


图 2 模型细节

Fig.2 Model's details

2 基于 YOLO-v7 的网络模型设计

2.1 网络模型设计

为了保证仪表检测和读数的稳定性并出于成本的考虑, 在实际工业生产环境中, 一般常用固定机位对仪表图像进行采集。本文采用车间监控摄像头拍摄的包含有仪表柜目标的图像作为最初的检测数据源。

首先将原始图像输入仪表检测网络, 进行仪表的检测定位以及分割。再将切割的仪表图像输入后级字符检测网络, 进行字符检测分类定位。这样极大的抑制了繁杂的背景干扰, 并且裁切放大了字符目标。本文利用字符检测网络对字符的分类进行数字识别, 同时利用对字符的定位, 进行数字的组合读数转换。

由于摄像头机位固定, 视野稳定, 并且拍摄的仪器对象也保持作业位置。虽然由于器件老化、机械振动甚至人为碰撞导致摄像头视野变动或者仪器位置变动, 但是在一定时间尺度上可以将摄像头视野内的目标仪器大小视作固定量级。因此输入仪表及字符检测网络的目标对象大小尺度稳定在一定范围。

YOLO-v7 模型中采用了三个检测头^[7]。对应的特

征图大小分别为 80×80 ， 40×40 以及 20×20 。在 YOLO 算法中，对应特征图尺寸越大的检测头负责检测更小的物体。具体来说，特征图尺寸越大，其细节就越丰富，可以有效检测物体在图像中占据较小区域的情况，但是对于大尺度物体，其冗余了大量的细节信息，对于检测不利。因此 YOLO 算法通过多尺度的检测头，在各个层级的特征图上同时进行检测，配合不同尺寸的预测框，可以检测到不同大小的物体。通过这一方法，不仅提高了目标检测的准确率还提高了检测的效率，更好地应对复杂情况。

而本文要检测的目标尺度大小情况相对单一，不

需要较多尺度的检测头进行检测。可以放弃多尺度检测，减少模型冗余，提高检测效率。所以，本文依据实际检测头的表现情况，对仪表检测网络模型以及字符检测网络模型的检测头进行删减。

同时，由于仪表盘以及字符的外形特征明显，不需要过深过宽的卷积特征。本文对整体网络模型进行删减，减少卷积核以及输出特征通道数量。图 3 为本文采用的基础网络模型结构。其中圆括号内为每一层输出的特征图的 (W, H, C) 。

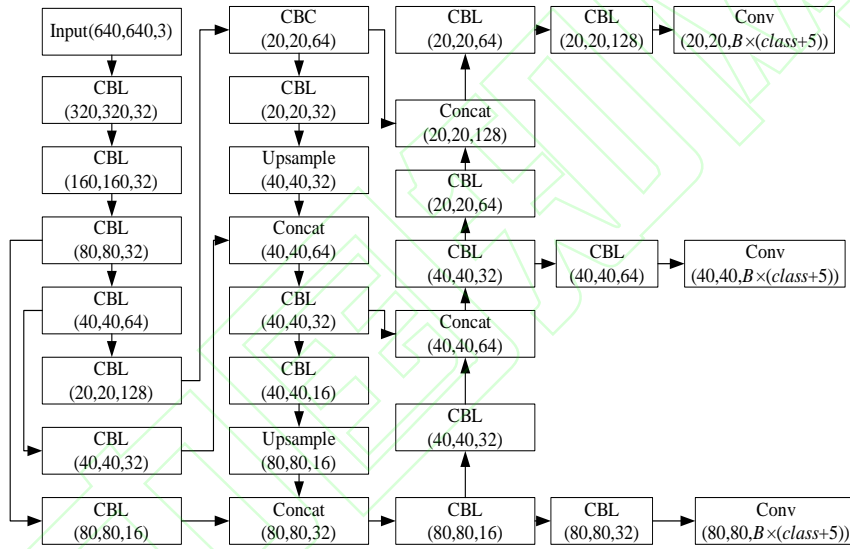


图 3 基础网络模型

Fig.3 Basic network model

2.2 快速金字塔池化

空间金字塔池化(Spatial Pyramid Pooling, SPP)^[8]结构，是一种自适应结构。它能自动将输入的任意尺寸的特征图转换成固定长度的特征向量输出。从而在特征图级别进行局部与全局的特征融合。CMC 结构即含有空间金字塔池化结构，只不过前后分别增添了一个 CBL 结构。而快速空间金字塔池化^[9](Spatial Pyramid Pooling - Fast, SPPF)通过统一规模的最大池化层级联实现 3 个不同规模的最大池化层并联的相同效果。实现了计算量的减少，变得更为高效。图 4 即为 CMC 结构更换 SPP 为 SPPF 后的模块图 CMCF。图 2 中的 CMC 为其原始模块图。其中方括号内为卷积的 $[k, s, p]$ 。

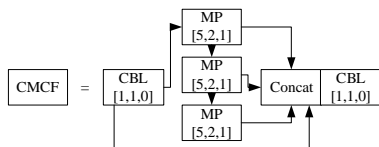


图 4 含 SPPF 结构的 CMC 结构图

Fig.4 CMC with SPPF

SPP 结构中 3 个池化的单次计算量为 $(5 \times 5 + 9 \times 9 + 13 \times 13) \times M = 275M$ ， M 为输入的通道数量。SPPF 结构中的池化的单次计算量为 $3 \times (5 \times 5) \times M = 75M$ 。可以看到在理论上 SPPF 内部的池化计算量相较于 SPP 降低了 70%。

2.3 深度可分离卷积

利用设定的不同参数的卷积核在图像上全方位的滑动，并求取对应元素乘积的和，可以提取图像中存在的不同方面的特征。初步卷积可以得到图像中存在的水平特征，垂直特征，边缘特征等浅层特征。再利用卷积对浅层特征进行组合，进而提取更抽象的深层语义特征或者全局特征。YOLO-v7 为了提高速度，采用了大量卷积运算。通过变换卷积的方式可以迅速达到提高检测速度，缩减参数体量的目的。

深度可分离卷积^[10] (Depthwise Separable Convo-

lution, DSConv)最初是由 Laurent Sifre 提出。其将标准卷积分解为逐深度卷积(depthwise convolution), 用来进行特征提取以及滤波和逐点 1×1 卷积(pointwise convolution), 用于对特征进行组合。通过分解运算, 在损失少量精度的情况下, 使得模型的计算量和参数数量明显下降, 便于应用在更多的移动端平台。图 5 分别展示了卷积大小为 3×3 的标准卷积和对应的深度可分离卷积的卷积结构, 包含其超参数设定。其中圆括号内为每一层输出的特征图的 (W, H, C) ; 方括号内为卷积的输入通道数目, 输出通道数目, 卷积核大小 $[c_in, c_out, k]$ 。

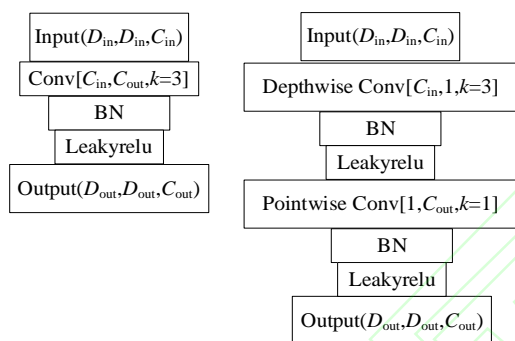


图 5 标准卷积和深度可分离卷积

Fig.5 Conv and DSConv

一个标准的 3×3 卷积层的计算代价为 $3 \times 3 \times C_{in} \times C_{out} \times D_{in} \times D_{in}$ 。在等效替代的深度可分离卷积中, 逐深度卷积的计算代价为 $3 \times 3 \times C_{in} \times 1 \times D_{in} \times D_{in}$, 逐点卷积的计算代价为 $1 \times 1 \times C_{in} \times C_{out} \times D_{in} \times D_{in}$, 故总代价^{[11][12]}为 $3 \times 3 \times C_{in} \times D_{in} \times D_{in} + C_{in} \times C_{out} \times D_{in} \times D_{in}$ 。深度可分离卷积的计算代价的减少量为公式(1)所示。

$$\frac{3 \times 3 \times C_{in} \times D_{in} \times D_{in} + C_{in} \times C_{out} \times D_{in} \times D_{in}}{3 \times 3 \times C_{in} \times C_{out} \times D_{in} \times D_{in}} = \frac{1}{C_{out}} + \frac{1}{9} \quad (1)$$

3 锚框聚类算法

在 YOLO 算法中, 使用了锚框(anchorbox)来对目标进行检测和定位。锚框是预先定义的一些框, 用于覆盖图像中可能存在的物体。模型训练中, 在待预测的特征图上, 每个网格内生成相同的 B 个不同尺寸大小和宽高比例的锚框进行预测。所以需要设定尺寸大小和宽高比例的初始锚框数目为检测特征图层级总数 (即检测头数目) 乘以预测框数目。并且在训练过程中, 只有锚框与真实框的 IOU 大于一定阈值, 这个锚

框才被视为正样本进行学习。因此, 为了获得更多的正样本, 更好的对检测目标特征进行提取, 以取得更好的检测效果, 需要对锚框的尺寸和宽高比进行优化, 使其更加适应目标的大小及形状特征。在正向推理, 也就是模型预测的过程中, 也是以锚框为基准对目标进行预测。

聚类是人们常用来生成指定数目的拟合目标数据集形状及大小锚框的方法。其中 K-means^[13]是非常经典且有效的聚类方法。K-means 算法首先在数据集中随机选择指定的数目 k 个锚框作为初始聚类中心, 然后将数据集内所有样本分配给最近的聚类中心, 并更新聚类中心的位置。重复这个过程, 直到聚类中心位置趋于稳定或者迭代次数达到上限为止。最终, K-means 算法会将最后得到的聚类中心返回, 即所需的锚框集合。图 6 即为 K-means 算法流程图。

由于 K-means 算法一开始就随机确定所有的初始聚类中心, 容易陷入局部最优。为了改进这一缺点, 提出了改进算法, K-means++^[14]。其只有最开始的第一个起始聚类中心是随机生成的, 后面的聚类中心点依次选取距已选聚类中心最远的样本点为新增聚类中心。然后再更新所有聚类中心位置, 得到聚类结果。其复杂度上较 K-means 多了一步聚类中心的选取优化, 除第一个聚类中心外, 每新增一个聚类中心都要对样本距离进行计算, 并选取距离最远的点。

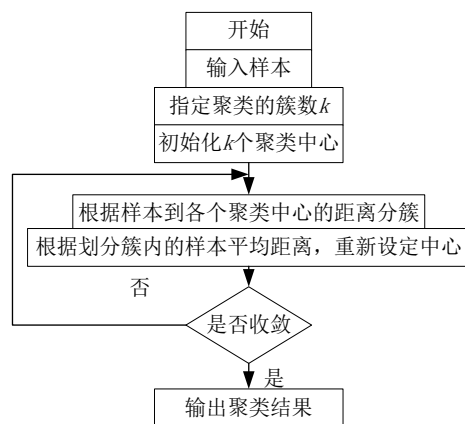


图 6 K-means 算法流程图

Fig.6 Flow chart of K-means algorithm

样本与聚类中心的距离计算方式有两种, 分别为欧氏距离(Euclidean Distance)和 IOU 。

为了得到更加符合当前数据集的锚框尺寸, 本文还使用了遗传算法, 对聚类的输出结果进行变异。其主要思想是对输出的锚框宽高进行一定次数的扰动, 每次扰动后, 计算新的锚框集合与样本间的平均拟合度, 如果平均拟合度提升则选取, 否则放弃, 以求得

到更加匹配的锚框尺寸和形状。这虽然增加了锚框聚类算法的复杂度,但是会得到更佳的锚框尺寸以及宽高比例。同时锚框的设定属于模型训练前的准备工作,不会增加模型的复杂度。

平均拟合度算法如下所示:

Algorithm: mean fitness

Input: Anchor, Box

Output: fitness

```

1. ratio=[];fitness=[]
2. for i in n:
3.     for j in k:
4.         ratio[j] =
            min(
                Box[i].weight/Anchor[j].weight,
                Box[i].height/Anchor[j].height,
                1/(Box[i].weight/Anchor[j].weight),
                1/(Box[i].height/Anchor[j].height,)
            )
5.     fitness[i] = max(ratio)
6. return mean(fitness)

```

除了尺寸外,锚框个数,即聚类中心数量也关系到模型的大小以及精度。锚框个数增加,拟合度也会增加,可能会增大正样本数目,增加提取特征,提升精度,但是随着锚框个数增加,预测框个数也会增加,输出特征数增加,模型体量增加。

4 模型剪枝压缩

移动端的算力有限,对于模型的参数体量也有着限制。而实际工业仪表抄录读数,对于实时性又有着严苛需求。因此,需要模型满足算力需求低,参数数量少,检测速度快的条件。本文对修改后的模型进一步采用模型剪枝的方法进行压缩。

模型剪枝是一种通过减少神经网络中冗余参数来压缩模型大小、提高计算效率和减少内存占用的技术。在模型剪枝中,通过移除神经网络中不必要的连接或节点来减少模型的规模,并且此方法不会显著影响模型的精确性。此外,模型剪枝还可以减少模型中的过拟合。模型可以从通道和卷积两个不同的维度进行剪枝。

本文选择通道剪枝^[15]。通道剪枝的基本思想是将卷积层的输出特征图分解为多个通道,然后将其中一些通道移除,以达到模型压缩的目的。首先需要确定要移除的通道。由于重新设计的模型中,基本为卷积结构,且每一个卷积层都标配一个批归一化层(Batch Normalization, BN),故而本文采用BN层中的缩放因子 γ 作为衡量通道重要性的指标^[16]。

本文通过对 γ 进行L1正则化,可以得到每个通道的重要性。L1正则化的目标函数定义为

$$L1_loss = \lambda \sum_{\gamma \in \Gamma} |\gamma|$$

后, γ 较大的通道对应的权重较大,这些通道在模型中具有更高的影响力。因此可以根据权重对通道进行排序,然后选择重要性较高的通道来压缩模型^[17]。将重要性低于阈值的通道进行裁剪,从而减小模型的参数量,提高模型的运行效率和泛化能力。

需要注意的是, λ 需要根据具体的任务和数据集进行调整,过大或过小都会影响到模型的性能。

5 实验与结果分析

5.1 数据集

5.1.1 原始数据集

本文采用分布在五个不同的实际工厂车间的监控摄像头拍摄的包含仪表柜的监控图像作为实验原始数据。一共1630张图像。分布情况为车间一628张;车间二269张;车间三299张;车间四223张;车间五211张。其车间抽样图像如图7所示。从左到右从上到下分别为车间一到五。



图7 原始图像数据集抽样
Fig.7 Sampling of original dataset

5.1.2 数据集扩充

为了防止模型过拟合,得到更好效果的训练模型。需要将大量数据喂入深度学习网络,用于模型训练。因此,本文采用数据集扩充的方法,对原始图像进行位移、错切、加噪、色域扭曲等方法,将原始数据集每一类型扩充至1千多张。扩充完成后获得标注好的数据图像5709张。分布情况为车间一1237张;车间二1125张;车间三1140张;车间四1152张;车间五1055张。其扩充后车间抽样图像如图8所示。



图8 扩充数据集抽样
Fig.8 Sampling of extended dataset

为了更好地评估模型、验证模型,按照 9:1 的比例将扩充后的车间图像数据集划分为训练加验证集 5138 张,测试集 571 张,然后再按 9:1 的比例将训练加验证集划分为训练集 4624 张,验证集 514 张。然后将数据喂入仪表检测网络进行模型训练。

对原始数据图像的仪表显示区域进行定位裁剪后,每种仪表挑选 200 张,作为字符检测网络输入的原始图像,其抽样图像如图 9 所示,并进行字符标记。将获得的图像同样进行数据扩充,扩充至每种类型 1000 张,其扩充后的抽样图像如图 10 所示。然后同样按照上述规则进行划分。最后将数据喂入字符检测网络进行模型训练。



图 9 原始字符数据集抽样

Fig.9 Sampling of original character dataset



图 10 扩充字符数据集抽样

Fig.10 Sampling of extended character dataset

5.2 模型定型

本文对模型进行训练的系统环境统一固定为 win10 操作系统,基于 Pytorch 深度学习框架,处理器为 12th Gen Intel® Core™ i5-12490F,显卡为 NVIDIA3080,显存 10GB,内存大小 16GB。采用同一个超参文件,迭代次数都为 100 次。通过测试集对模型进行评估。

利用模型平均检测精度 mAP0.5 代表模型的检测精度、平均检测一张图片所需时间代表检测速度、模型总的参数数量代表模型参数体量,以及模型所需的浮点运算数 GFLOPs 表示模型复杂度。

5.2.1 网络模型轻量化

以基础网络模型架构构建基础模型,对其进行进一步模型参数体量控制,以及降低模型复杂度。本文采用深度可分离卷积,对基础模型中的 3×3 标准卷积进行替换,并将 PPF 替换为更快的 SPPF。对其分别

采用标注并扩充后的原始数据集以及字符数据集进行仪表检测网络和字符检测网络的训练。利用测试集评估模型的性能。在各自测试集上,模型的性能表现如表 1 所示。

表 1 消融实验结果对比

模型名称	mAP0.5	检测速度	参数数量	GFLOPs
YOLO-v7(p)	0.9993	13.3ms	37.22M	105.2
Base(p)	0.9957	5.7ms	0.30M	1.2
Base DS(p)	0.9956	5.6ms	0.17M	0.4
YOLO-v7(c)	0.9988	12.7ms	37.25M	105.3
Bsae(c)	0.9956	4.4ms	0.30M	1.2
Base DS(c)	0.9954	4.6ms	0.17M	0.4

其中(p)表示仪表检测任务,对应的模型则称为仪表检测网络模型;(c)表示字符检测任务,对应的模型则称为字符检测网络模型。Base 为基础网络模型,DS 表示采用了深度可分离卷积以及快速特征金字塔结构。凸显颜色为不同任务中各指标的最佳值。

可以看到,本文采用的基础网络模型,无论在仪表检测任务还是字符检测任务中都取得了优异的表现。相较于原始 YOLO-v7 网络模型其精度下降控制在 0.005 以内,速度提高了一倍还多,参数体量以及模型复杂度下降了 2 个量级。在采用了 DSConv 以及 SPPF 后,在维持精度以及检测速度的情况下,模型复杂度降至基础模型的三分之一,参数体量下降了近一半。

5.2.2 检测头轻量化

为了进一步压缩模型参数体量,以及加速检测速度,本文同时采用不同层级的检测头组合,进行性能对比,保留性能最佳模型结构,剪除检测头冗余。以上一轮优化中的 BaseDS 为基础模型。

表 2 消融实验结果对比

模型名称	mAP0.5	检测速度	参数数量	GFLOPs
Base(p)	0.9956	5.6ms	0.17M	0.4
Base40(p)	0.9963	5.1ms	0.16M	0.3
Base80(p)	0.9955	5.4ms	0.15M	0.3
Base(c)	0.9954	4.6ms	0.17M	0.4
Base40(c)	0.9957	4.3ms	0.16M	0.3
Base80(c)	0.9953	4.2ms	0.15M	0.3

其中 40 表示采用 40×40 和 20×20 两个层级的特征尺度检测头组合的模型,80 表示采用 80×80 和 20×20 两个层级的特征尺度检测头组合。

对训练完成的模型进行测试评估,在各自测试集上,模型的性能表现如表 2 所示。可以看到后两个检测头组合的模型效果较好,检测精度最高,并且检测速度,模型复杂度,参数体量较基础模型都有所优化。

故而本文采用 Base40 模型,作为仪表以及字符检

测网络优化后的模型，并以此为基础进行后续的锚框优化以及模型剪枝。

5.2.3 锚框个数选定及尺寸初始化

YOLOv7 模型中采用手动锚框设定。原先模型有三个检测抽头，每个网格生成 3 个预测框，所以预置的 anchors 数量为 9 个。本文在进行检测头剪除冗余时，剪除哪一个检测头，则去除哪一个特征层对应的 anchors。但是不改变预测框数量。

为了找到更加适合目标对象的锚框个数，以及尺寸大小和宽高比例。提高模型的训练以及检测效率，去除锚框冗余。本文采用 K-means 配合欧氏距离以及 K-means++ 配合 IOU 距离对数据集进行自动锚框聚类，同时对聚类的结果尺寸进行遗传变异优化。利用平均拟合度评估生成的锚框的质量。同时利用平均拟合度随锚框个数的增量，寻找最佳锚框个数。

图 11、12 分别表示在扩充数据集和扩充字符数据集上平均拟合度及其增量随锚框个数的变化曲线。

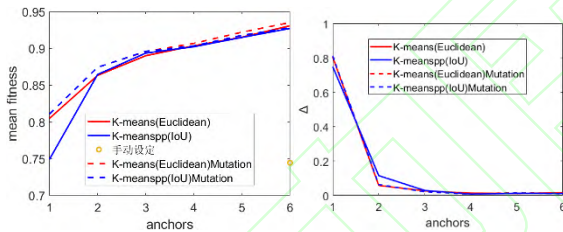


图 11 平均拟合度曲线及增量曲线

Fig.11 Mean fitness curve and Δ curve

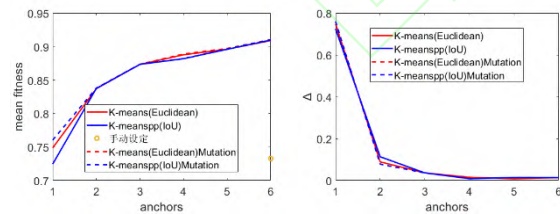


图 12 平均拟合度曲线及增量曲线

Fig.12 Mean fitness curve and Δ curve

从图中可以看出，采用了遗传变异后的锚框的平均拟合度提升，对于优化锚框尺寸具有一定效果。此外，从平均拟合度增量曲线图来看，在两个数据集中，当 anchors 数量超过 2 后平均拟合度增量均放缓，因此选择 anchors 的数量都为 2。对比 anchors 为 2 时，平均拟合度曲线，发现 K-means++ 配合 IOU 距离以及遗传算法的聚类方法和 K-means 配合欧氏距离以及遗传算法的聚类方法表现均为最优。通过计算两种算法的时间复杂度，发现前一种算法生成两个数据集共 4 个

锚框的平均时间复杂度为 2.79s，后一种为 4.72s。因此选用 K-means++ 配合 IOU 距离以及遗传算法的聚类方法设定锚框。

由此，确定仪表检测模型和字符检测模型的锚框数量为 2，预测框个数为 1，相较于上一轮的 Base40 模型中的 6 个锚框，3 个预测框，检测层输出特征个数降至三分之一。锚框平均拟合度分别提升了 13%，11%。生成的锚框尺寸分别为[[61, 50] [71, 71]]和[[50, 68] [91, 107]]。

以上一轮优化中的 Base40 作为基准模型，对锚框数量以及尺寸进行优化。

表 3 消融实验结果对比

Table 3 Comparison of ablation experimental results

模型名称	mAP0.5	检测速度	参数数量	GFLOPs
Base(p)	0.9963	5.1ms	0.16M	0.3
Base2(p)	0.9967	4.3ms	0.16M	0.3
Base(c)	0.9957	4.3ms	0.16M	0.3
Base2(c)	0.9957	3.5ms	0.16M	0.3

在各自测试集上，模型的性能表现如表 3 所示。可以看到采用两个锚框的网络的检测速度有所提升，网络模型的精确度均未见下降，并且仪表检测网络的精确度有所提高。

5.3 模型剪枝压缩

模型剪枝之前首先对模型进行稀疏训练。图 13 分别为采用不同正则化系数 λ 的情况下，迭代 100 次后，仪表检测网络模型尺度因子的分布直方图^[18]。

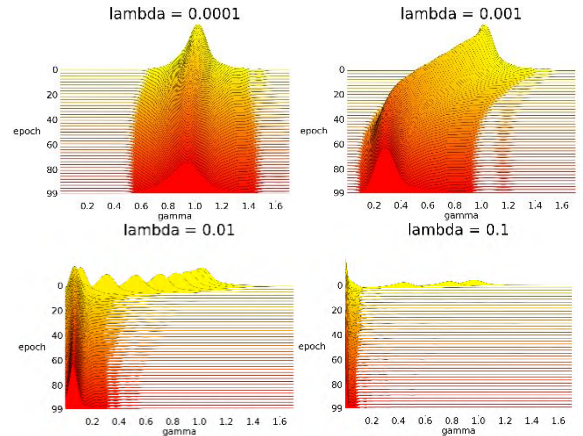


图 13 仪表检测网络模型尺度因子分布直方图

Fig.13 Instrumentation detection model's distribution histogram of γ

由图可以看出，随着 λ 的增大，模型的尺度因子从最初的正态分布越来越快速的被压缩。大量不重要的通道的尺度因子被压缩至接近于 0。

为了甄别出较优的 λ 下产生的模型，用于后续的通道剪枝操作，本文通过观察不同惩罚系数下的稀疏化训练后的模型在验证集上的表现来进行选择。图 14、

15 分别表示不同惩罚系数下稀疏化训练后的仪表检测网络模型和字符检测网络模型在各自验证集上的损失值随着训练次数的变化曲线。

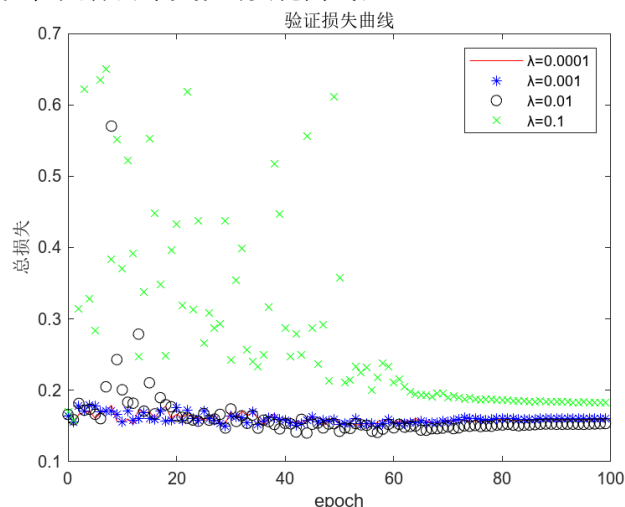


图 14 仪表检测网络模型验证损失随 λ 的变化

Fig.14 Instrumentation detection model's loss curve under different λ

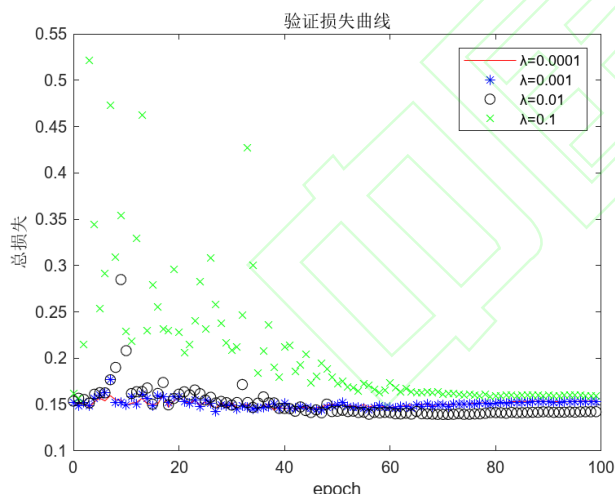


图 15 字符检测网络模型验证损失随 λ 的变化

Fig.15 Character detection model's loss curve under different λ

从验证集损失曲线可以看出,当 λ 为 0.01 时,模型表现最好,损失最小。所以仪表检测网络以及字符检测网络都采用此系数训练的模型进行通道剪枝。

5.4 实验结果分析

稀疏训练完成后,设定剪枝比例^[19]。本文采用的是整体比例设定,将所有的通道的尺度因子进行顺序排列,按照设定的剪枝比例计算尺度因子阈值,对低于阈值的通道进行剪除。为了避免阈值设置过大,导致有的卷积层的所有通道全被剪除,本文通过将每一

层的最大的尺度因子进行顺序排列,选择其中数值最小的尺度因子,以其为阈值,计算小于其数值的尺度因子占全部尺度因子的比例,以其作为最大剪枝比例,保证每一层至少有一个保留通道。由于本模型剪枝前就已经属于小模型,通道较为紧凑,比例不宜设置过大。所以本文将最大剪枝比例的一半并向下取整作为剪枝比例。因此本文采用 0.2 作为模型的剪枝比例,对模型进行剪枝。

对剪枝完成后的模型进行微调,设定 epoch 为 100。恢复模型检测的精确性。最终模型性能效果如表 4 所示。以上一轮中的 Base2 为基础对照模型。

表 4 检测网络模型的表现

Table 4 Detection model's performance

模型名称	mAP0.5	检测速度	参数数量	GFLOPs
Base(p)	0.9967	4.3ms	157146	0.3
Base L1(p)	0.9962	4.8ms	157146	0.3
Base F(p)	0.9963	6.3ms	116801	0.3
Base(c)	0.9957	3.5ms	157930	0.3
Bsae L1(c)	0.9957	3.7ms	157930	0.3
Base F(c)	0.9953	4.4ms	121566	0.3

L1 表示对基础模型进行正则化后的模型, F 表示对正则化稀疏训练后的模型进行剪枝并微调后的模型。

可以看到,剪枝微调后的模型,在损失少量检测精度及速度的情况下,参数数量有了明显减少。

以微调后的模型 Base F 作为最终模型 YOLO-v7(ours),同时还与当前主流的两阶段算法 Faster-RCNN,单阶段算法 SSD、YOLO 等进行对比。模型的性能评估如表 5 所示。其中 M_v3 表示采用 MobileNet_v3 骨干网络, R-50 表示 ResNet-50 骨干网络, V16 表示 VGG16 骨干网络, M_v2 表示 MobileNet_v2 骨干网络。

表 5 检测网络模型的表现

Table 5 Detection model's performance

model	mAP0.5	speed	parameter	GFLOPs
YOLO-v7(p)	0.9993	13.3ms	36.51M	103.2
YOLO-v7tiny(p)	0.9982	9.6ms	6.02M	13.1
YOLO-v7(ours)(p)	0.9963	6.3ms	0.12M	0.3
YOLO-v5s(p)	0.9950	11.5ms	7.03M	15.8
YOLOv5s M_v3(p)	0.9950	12.6ms	5.03M	11.3
Faster-RCNN R-50(p)	1.0000	46.3ms	28.33M	500.3
Faster-RCNN V16(p)	1.0000	32.2ms	136.79M	298.0
SSD300 V16(p)	1.0000	8.1ms	24.28M	61.3
SSD300 M_v2(p)	0.9998	8.6ms	4.21M	1.5
YOLO-v7(c)	0.9988	12.7ms	36.53M	103.3
YOLO-v7tiny(c)	0.9982	8.8ms	6.03M	13.1
YOLO-v7(ours)(c)	0.9953	4.4ms	0.12M	0.3
YOLO-v5s(c)	0.9950	6.1ms	7.04M	15.8
YOLOv5s M_v3(c)	0.9947	9.3ms	5.04M	11.3
Faster-RCNN R-50(c)	0.9996	43.3ms	28.37M	500.3
Faster-RCNN V16(c)	0.9997	26ms	136.87M	298.1
SSD300 V16(c)	0.9899	7.9ms	24.82M	61.7
SSD300 M_v2(c)	0.9821	8.3ms	4.74M	1.6

可以看到,本文所使用的最终模型相较于其他模型,在检测精度上与最佳模型相差不到 0.005。但是在检测速度,模型复杂度,参数体量方面全部领先。

对比 YOLO-v7 模型,最终的仪表检测模型,在精度上损失了 0.3%,但是检测速度减少了 52.63%,参数体量减少了 99.67%,模型复杂度减少了 99.71%。字符检测模型在精度上损失了 0.35%,但是检测速度减少了 60.35%,参数体量减少了 99.67%,模型复杂度减少了 99.71%。

本文从原始数据集中,每个车间图像各随机抽取 50 张,分别相继输入仪表检测网络和字符检测网络,观察系统的整体检测精度。

考虑仪表检测网络背景复杂,干扰较大,设网络的检测置信度阈值为 0.8,防止仪表误检;由于仪表间位置稀疏,重叠的概率低,所以设定非极大值抑制的 IoU 检测阈值为 0.25。仪表检测网络模型的检测精度如表 6 所示。

表 6 仪表检测精度

Table 6 Accuracy of instrumentation detection

车间	包含仪表个数	正确检测的个数	检测精度
1	50	50	100%
2	200	200	100%
3	50	50	100%
4	100	100	100%
5	50	50	100%
总计	450	450	100%

经过仪表检测网络之后,字符检测网络获得的输入图像较为干净,可以适当放松网络的检测置信度阈值,设为 0.7;但是字符排列较为紧凑,因此将非极大值抑制的 IoU 检测阈值设为 0.45。字符检测网络模型的检测精度如表 7 所示。

表 7 各类字符识别精度

Table 7 Accuracy of various character reading

字符	个数	正确识别个数	识别精度
0	662	661	99.85%
1	786	781	99.36%
2	87	86	98.85%
3	150	150	100.00%
4	184	184	100.00%
5	204	204	100.00%
6	262	262	100.00%
7	206	206	100.00%
8	127	127	100.00%
9	130	130	100.00%
总计	2798	2791	99.75%

在仪表检测阶段未正确检测到仪表盘的直接判负。通过仪表检测网络,但是在字符检测网络未正确检测到所有字符的也相应判负。最后系统的整体检测精度如表 8 所示。

表 8 系统整体精度

Table 8 Accuracy of the whole system

车间	仪表个数	完全识别正确个数	准确率
1	50	50	100.00%
2	200	198	99.00%
3	50	50	100.00%
4	100	96	96.00%
5	50	49	98.00%
总计	450	443	98.44%

6 结论

本文针对实际工厂中数显仪表的检测及读数问题。通过对 YOLO-v7 原始模型进行直接硬剪枝:直接删减卷积层,缩减卷积通道。利用深度可分离卷积替换常规卷积, SPPF 模块替换 PPF 模块以及剪除检测头等方式在最大程度保持精度的情形下对模型进行压缩,减少模型参数,降低模型复杂度,加快模型检测速度。利用 K-means++ 聚类算法加遗传算法优化找到适合数据集的初始锚框个数和尺寸。最后利用通道剪枝技术,进一步压缩模型。最终模型权重文件大小分别由原始的 71.3MB、71.4MB 降至 334KB 和 351KB。所需浮点计算量也由 105.2GFLOPs、105.3GFLOPs 降至 0.3GFLOPs。完全可以部署在算力有限的移动终端设备上。并且理论上每张仪表所需的检测时间为 $6.3+4.4=10.7\text{ms}$,满足实时性要求。各网络检测精度达到 99%以上,系统整体读数精确度达 98%以上,满足准确性需求。

参考文献:

- [1] 邓清男,石晓龙.变电站室内数显仪表的读数识别[J].工业仪表与自动化装置,2018,No.260(02):86-89.
DENG Q N, SHI X L. Reading identification of indoor digital instrumentation in substation[J].Industrial Instrumentation and Automation,2018,No.260(02):86-89.
- [2] 朱立倩.基于深度学习的数显仪表字符识别[J].计算机技术与发
展,2020,30(6):141-144.DOI:10.3969/j.issn.1673-629X.2020.06.027.
ZHU L Q. Character Recognition of Digital Display Instrument Based on Deep Learning[J].Computer Technology and Development,2020,30(6):141-144.DOI:10.3969/j.issn.1673-629X.2020.06.027.
- [3] 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.
- [4] 翁念标.面向工业环境的数显仪表识别技术研究[D].杭州电子科技大学,2022.DOI:10.27075/d.cnki.gzhdc.2022.000717.
WENG N B. Research on Digital Display Instrument Identi-

- fication in Industrial Environment[D]. Hangzhou Dianzi University, 2022. DOI:10.27075/d.cnki.ghzdc.2022.000717.
- [5] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv preprint arXiv:2207.02696, 2022.
- [6] 齐向明, 董旭. 改进 YOLOv7-tiny 的钢材表面缺陷检测算法[J/OL]. 计算机工程与应用: 1-13[2023-04-26]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20230329.0924.008.html>.
- QI X M, DONG X. Improved YOLOv7-tiny algorithm for steel surface defect detection[J/OL]. Computer Engineering and Applications: 1-13[2023-04-26]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20230329.0924.008.htm>.
- [7] X. Zhu, S. Lyu, X. Wang, et al. "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 2778-2788, doi: 10.1109/ICCVW54120.2021.00312.
- [8] 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.
- [9] Lu, X., Song, W. (2022). Improved YOLOv5s Model for Vehicle Detection and Recognition. In: Huang, DS., Jo, KH., Jing, J., Premaratne, P., Bevilacqua, V., Hussain, A. (eds) Intelligent Computing Methodologies. ICIC 2022. Lecture Notes in Computer Science(), vol 13395. Springer, Cham. https://doi.org/10.1007/978-3-031-13832-4_35.
- [10] Sifre L, Mallat S. Rigid-motion scattering for texture classification[J]. arXiv preprint arXiv:1403.1687, 2014.
- [11] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
- [12] 王艺霏, 贺利乐, 何林. 基于 YOLOv4 的轻量化口罩佩戴检测模型设计[J]. 西北大学学报(自然科学版), 2023, 53(02): 265-273. DOI:10.16152/j.cnki.xdxbzr.2023-02-012.
- WANG Y L, HE L L, HE L. Design of lightweight mask wearing detection model based on YOLOv4[J]. Journal of Northwest University(Natural Science Edition), 2023, 53(02): 265-273. DOI:10.16152/j.cnki.xdxbzr.2023-02-012.
- [13] 耿跃, 任军号, 吉沛琦. 基于 K-Means 变异算子的混合遗传算法聚类研究[J]. 计算机工程与应用, 2011, 47(29): 151-153+221.
- GEN Y, REN J H, JI P Q. Hybrid genetic algorithm clustering analysis based on K-Means mutation operator[J]. Computer Engineering and Applications, 2011, 47(29): 151-153+221.
- [14] 吴志高, 陈明. 基于改进 YOLO v7 的微藻轻量级检测方法[J]. 大连海洋大学学报, 2023, 38(01): 129-139. DOI:10.16535/j.cnki.dlhyxb.2022-338.
- WU Z G, CHEN M. Lightweight detection method for microalgae based on improved YOLO v7[J]. Journal of Dalian Ocean University, 2023, 38(01): 129-139. DOI:10.16535/j.cnki.dlhyxb.2022-338.
- [15] Liu Z, Li J, Shen Z, et al. Learning efficient convolutional networks through network slimming[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2736-2744.
- [16] 黄文斌, 陈仁文, 袁婷婷. 改进 YOLOv3-SPP 的无人机目标检测模型压缩方案[J]. 计算机工程与应用, 2021, 57(21): 165-173.
- Huang W B, CHEN R W, YUAN T T. Compression of UAV Object Detection Model Based on Improved YOLOv3-SPP[J]. Computer Engineering and Applications, 2021, 57(21): 165-173.
- [17] 杨国威, 许志旺, 房臣等. 融合剪枝与量化的目标检测网络压缩方法[J]. 计算机工程与应用, 2022, 58(22): 108-115.
- Yang G W, XU Z W, FANG C, et al. Object Detection Network Compression Method Based on Pruning and Quantization[J]. Computer Engineering and Applications, 2022, 58(22): 108-115.
- [18] 陈科峻, 张叶. 基于 YOLO-v3 模型压缩的卫星图像船只实时检测[J]. 液晶与显示, 2020, 35(11): 1168-1176.
- CHEN K J, ZHANG Y. Real-time ship detection in satellite images based on YOLO-v3 model compression[J]. Liquid Crystals and Displays, 2020, 35(11): 1168-1176.
- [19] Han-Li Zhao, Kai-Jie Shi, Xiao-Gang Jin, et al. Probability-Based Channel Pruning for Depthwise Separable Convolutional Networks[J]. Journal of Computer Science and Technology, 2022, 37(3): 584-600.