

控制与决策

Control and Decision

基于轻量化深度学习网络的工业环境小目标缺陷检测

叶卓勋, 刘妹琴, 张森林

引用本文:

叶卓勋, 刘妹琴, 张森林. 基于轻量化深度学习网络的工业环境小目标缺陷检测[J]. *控制与决策*, 2023, 38(5): 1231–1238.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1893>

您可能感兴趣的其他文章

Articles you may be interested in

[全局与局部图像特征自适应融合的小目标检测算法](#)

Small object detection algorithm based on adaptive fusion of global and local image features

控制与决策. 2023, 38(4): 935–943 <https://doi.org/10.13195/j.kzyjc.2021.1800>

[基于两阶段深度网络的输电线路异常目标检测方法](#)

Transmission line abnormal object detection method based on deep network of two-stage

控制与决策. 2022, 37(7): 1873–1882 <https://doi.org/10.13195/j.kzyjc.2020.1840>

[基于ResNet34_D改进YOLOv3模型的行人检测算法](#)

Pedestrian detection based on developed YOLOv3 with ResNet34_D

控制与决策. 2022, 37(7): 1713–1720 <https://doi.org/10.13195/j.kzyjc.2021.0136>

[多目标小尺度车辆目标检测方法](#)

Multi-target and small-scale vehicle target detection method

控制与决策. 2021, 36(11): 2707–2712 <https://doi.org/10.13195/j.kzyjc.2020.0635>

[基于双分支特征融合的场景文本检测方法](#)

A scene text detection based on dual-path feature fusion

控制与决策. 2021, 36(9): 2179–2186 <https://doi.org/10.13195/j.kzyjc.2020.0002>

基于轻量化深度学习网络的工业环境小目标缺陷检测

叶卓勋¹, 刘妹琴^{1,2†}, 张森林¹

(1. 浙江大学 电气工程学院, 杭州 310027; 2. 西安交通大学 人工智能与机器人研究所, 西安 710049)

摘要: 工业环境下表面缺陷检测是质量管理的重要一环, 具有重要的研究价值. 通用检测网络 (如 YOLOv4) 已被证实在多种数据集检测方面是有效的, 但是在工业环境的缺陷检测仍需要解决两个问题: 一是缺陷实例在表面占比过小, 属于典型的小目标检测问题; 二是通用检测网络结构复杂, 很难部署在移动设备上. 针对上述问题, 提出一种基于轻量化深度学习网络的工业环境小目标缺陷检测方法. 应用 GhostNet 替代 YOLOv4 主干特征提取网络, 提高网络特征提取能力及降低算法复杂度, 并通过改进式 PANet 结构增加 YOLO 预测头中高维特征图比例以实现更好的性能. 以发动机金属表面缺陷检测为例进行实验分析, 结果表明该模型在检测精度 (mAP) 提升 5.83% 的同时将网络模型参数量降低 83.5%, 检测速度提升 2 倍, 同时满足缺陷检测的精度和实时性要求.

关键词: 轻量化检测网络; 小目标缺陷检测; 主干特征提取网络; 改进式 PANet

中图分类号: TP273 文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1893

引用格式: 叶卓勋, 刘妹琴, 张森林. 基于轻量化深度学习网络的工业环境小目标缺陷检测 [J]. 控制与决策, 2023, 38(5): 1231-1238.

Small-scale defect detection in industrial environment based on lightweight deep learning network

YE Zhuo-xun¹, LIU Mei-qin^{1,2†}, ZHANG Sen-lin¹

(1. College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China; 2. Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Automated surface defect inspection in industrial environments is an important aspect of quality management and is of significant research value. Generic detection networks, such as YOLOv4, have proven to be effective in the detection of a wide range of datasets. However, defect detection in industrial environments still needs to address two issues: one is that the percentage of defect instances on the inspected surface is too small, which is a typical small-scale object detection problem; the other is that the structure of generic detection networks is complex and difficult to deploy on mobile devices. To address these problems, this paper proposes a small-scale defect detection method in the industrial environment based on the lightweight deep learning network. Firstly, we replace the YOLOv4 backbone feature extraction network with the GhostNet to improve the feature extraction capability and reduce the complexity of the algorithm. Secondly, the proportion of high-dimensional feature maps in the YOLO head is increased by the improved PANet structure to achieve better performance. The experimental results show that the model can improve the detection accuracy (mAP) by 5.83% while reducing the number of network parameters by 83.5% and improving the detection speed by 2 times, which meets the requirements of accurate and real-time detection.

Keywords: lightweight detection network; small-scale defect detection; backbone feature extraction network; improved PANet

0 引言

表面缺陷检测是机器视觉领域中非常重要的一项研究内容, 利用机器视觉设备获取图像以判断采集图像中是否存在缺陷, 而工业产品外观品质检

验是保障产品质量的关键环节. 工业产品的外观在消费者心理往往与产品质量等同, 外观缺陷检测成为质量控制的关键. 目前, 绝大多数企业的质检依赖于人工检验完成, 存在易主观错判、检测方式难标准

收稿日期: 2022-01-01; 录用日期: 2023-01-10.

基金项目: 科技创新 2030-“新一代人工智能”重大项目 (2020AAA0108302).

责任编辑: 杨涛.

†通讯作者. E-mail: liumeiqin@zju.edu.cn.

化、检测准确性和效率严重依赖于检测员经验等问题。近年来,随着以卷积神经网络(convolutional neural network, CNN)为代表的深度学习模型在计算机视觉领域的应用,基于机器视觉的缺陷检测技术取得重大突破。然而,实际工业场景下金属表面缺陷检测仍存在一些挑战性的问题:

1) 数据集问题。与真实工业场景中的几十张或几百张图片相比,传统的深度学习方法(如CNN)需要大量的标注数据进行训练,很难适应表面轮廓、光照条件、成像角度等不同环境的复杂工业场景。通常情况下,深度神经网络容易产生过拟合现象,需要网络被喂入尽可能多的数据^[1]。因此,最常用的方法是扩充送入网络的训练集,使网络获得更丰富的特征信息。传统数据扩增方法通常是对原缺陷样本使用模糊、镜像、随机遮挡、随机裁剪、对比度调整等多种基本图像处理方式来获取更多训练样本,但此方法生成的图片与原始图片有很强的关联性,导致神经网络过拟合问题缓解程度有限。相较于传统数据扩增方法,一些更为新颖的数据生成方法也被提出。Tao等^[2]提出了一种基于分割网络的图像生成方法,先将缺陷实例从自然背景中分割出来,后通过图像融合方法将其叠加到正常样本上。近些年,基于生成对抗网络(generative adversarial net, GAN^[3])的图像生成技术也在研究之中。Posilović等^[4]提出了一种新型的GAN结构,用于生成高质量和高分辨率的B超图像。Zhang等^[5]提出了一种基于Wassertein GAN^[6-7]的图像生成办法解决焊接缺陷图像中缺陷实例不平衡的问题。根据基于相对尺度的小目标定义,同一类别实例面积与图像面积之比在0.08%~0.5%之间。本文所使用发动机金属表面缺陷数据集,金属表面尺寸为 3548×3548 ,凹槽类缺陷实例尺寸均小于 64×64 ,相对面积为0.051%,划痕类缺陷实例形状不标准,但像素值不多于 177×58 ,相对面积为0.082%,均属于小目标检测问题。针对小目标检测可利用特征较少及样本不均衡等问题,Ye等^[8]提出了一种双通道GANDual-Path GAN(DPGAN),通过分别生成缺陷实例和背景图像再复制黏贴的方式,在生成高分辨率缺陷数据集的同时生成数据集标注文件,节省了大量时间与人力。Bosquet等^[9]基于GAN提出了一种全新的结构,通过由较大物体生成较小物体来扩充数据集。

2) 神经网络在获得更好的特征提取能力的同时规模越来越大,结构也随之越来越复杂,预测和训练所需的硬件资源也随之增多。通常情况下,移动装置

及嵌入式设备很难运行复杂的深度学习网络模型,因此在计算机视觉领域,人们也在努力推动神经网络的小型化,最大程度上减小模型复杂度,减少模型存储所需要的空间。使网络轻量化的常用方法是模型压缩,如模型剪枝^[10-11]、模型量化^[12]等,然而基于这些压缩方法得到的全连接层的矩阵结构是不规律的稀疏矩阵,与现代硬件加速平台如图形处理单元(GPU)并不匹配,导致即使网络参数减少了但运算时间仍相同,没有满足网络轻量化的要求^[13]。基于上述两种原因,轻量级网络开始被提出,旨在保持网络精度的基础上进一步增加特征提取能力、运行速度和减少模型参数量。MobileNetv1^[14]引入了深度可分离卷积替代传统卷积,大大减少了运算量和参数数量。MobileNetv2^[15]在此基础上进一步引入线性瓶颈和倒残差结构。MobileNetv3^[16]对倒残差结构进行简单改造,加入SE模块^[17]并更新了激活函数。GhostNet^[18]使用简单的线性运算方式生成相似特征图获得网络中的冗余信息。基于以上轻量化网络,研究者也开始使用此方法改造检测网络。Li等^[19]使用GhostNet改造Transformer以实现COCO数据集的检测。Zhang等^[20]运用MobileNetv3改造YOLOv3以实现无人机条件下的行人检测。Jiang等^[21]使用MobileNetv2替代SSD网络中的主干特征提取网络VGG-16以实现蛋鸡的检测系统。上述方法虽然均取得了很好的检测效果,但由于金属表面缺陷相较于行人和蛋鸡等目标更加小,样本数量更少,背景更加复杂,并不完全适用于工业环境下金属表面缺陷检测。

本文选择YOLOv4^[22]模型作为基本框架,针对工业环境下小目标缺陷检测、训练样本少、移动装置计算能力有限等问题进行一系列改造,具体包括:1) 使用轻量化网络GhostNet替代YOLOv4主干特征提取网络CSPDarkNet53,降低网络参数量,提升特征提取能力;2) 针对小目标易被错检漏检的问题,对YOLOv4加强特征提取网络进行改造,加大高维信息在最终检测时的比重,进一步提升检测精度,减少网络计算量。

1 基于Ghost模块的轻量级金属表面缺陷检测方法

1.1 基线检测网络YOLOv4

目前,在目标检测领域有双阶段网络和以YOLO为代表的单阶段网络,其中单阶段网络因其检测精度高、速度快,被广泛运用于工业产品的缺陷检测中。针对工业环境下小目标缺陷检测,本文使用YOLOv4作为基准模型。YOLOv4是高精度实时一阶段检测

算法,代表YOLOv1、YOLOv2、YOLOv3的持续改进,在检测速度和检测精度方面达到了目前的最优水平.YOLOv4延续YOLOv3整体结构不变,同时在以下方面进行了不同程度的优化:

1) 主干特征提取网络.

主干特征提取网络主要对输入图像的初步特征进行提取,运用主干特征提取网络可以输出3个初步的有效特征层,此部分也是本文后续主要改进之处.YOLOv4主干特征提取网络在YOLOv3的Darknet53主干特征提取网络的基础上融合跨阶段局部网络(cross stage partial networks, CSPNet^[23])的优势,解决了Darknet53等大型卷积神经网络主干特征提取网络中网络优化的梯度信息重复问题,将梯度变化集成到特征图中,在主干特征提取网络轻量化的同时增加了卷积神经网络的学习能力,保持了网络的准确性.YOLOv4主干特征提取网络包含5个CSP模块,每个CSP模块第1个卷积核的大小均为 3×3 , $stride = 2$,当输入特征图尺寸为 $416 \times 416 \times 3$ 时,输出的3个有效特征图分别为 52×52 , 26×26 , 13×13 ,3种大小的特征图分别代表低中高维信息,感受野分别为 8×8 、 16×16 和 32×32 ,用于后续小中大目标的检测.此类方法可以获得丰富的特征信息,但网络结构较为复杂,特征增强能力不足,限制了模型对小目标检测精度的进一步提升.

主干特征提取网络另一个较明显的改进在于采用了Mish激活函数^[24],有

$$Mish = x \cdot \tanh(\text{softplus}(x)) = x \cdot \tanh(\ln(1 + e^x)). \tag{1}$$

Mish激活函数具有无上界有下界的特点,在训练时不会存在梯度消失的问题,相较于之前使用的ReLU、LeakyReLU等激活函数在 $x = 0$ 处更加光滑,在实际求解和模型泛化性方面表现更加优良,收敛速度更快.

2) 加强特征提取网络.

加强特征提取网络对主干特征提取网络输出的3个有效特征图进行进一步特征提取.YOLOv4使用了SPP和PANet的特征金字塔结构,SPP结构将主干特征提取网络最后一层CSP结构生成的特征图分别进行4个不同尺度的最大池化处理,增加特征图感受野并分离出显著的语义信息;PANet结构对SPP结构及主干特征提取网络输出的3个有效特征图分别进行特征层串联实现特征融合,并将融合得到的特征层传输到YOLO预测头 Y_1 、 Y_2 、 Y_3 中进行预测.

YOLOv4的金字塔结构对有效特征层进行反复特征提取,在一定程度上提高了通用检测的精度,但忽略了直接融合不同维度信息带来的语义冲突问题,导致待检测小目标信息被冲突信息淹没.在轻量化方面,PANet结构已普遍使用深度可分离卷积替代PANet中的普通卷积,本文后续也将使用此类结构.

1.2 轻量化检测网络整体架构

图1展示了所使用的基于YOLOv4的轻量化检测网络总体结构,针对小目标检测问题,对YOLOv4主干特征提取网络及加强特征提取网络中的PANet进行改进.

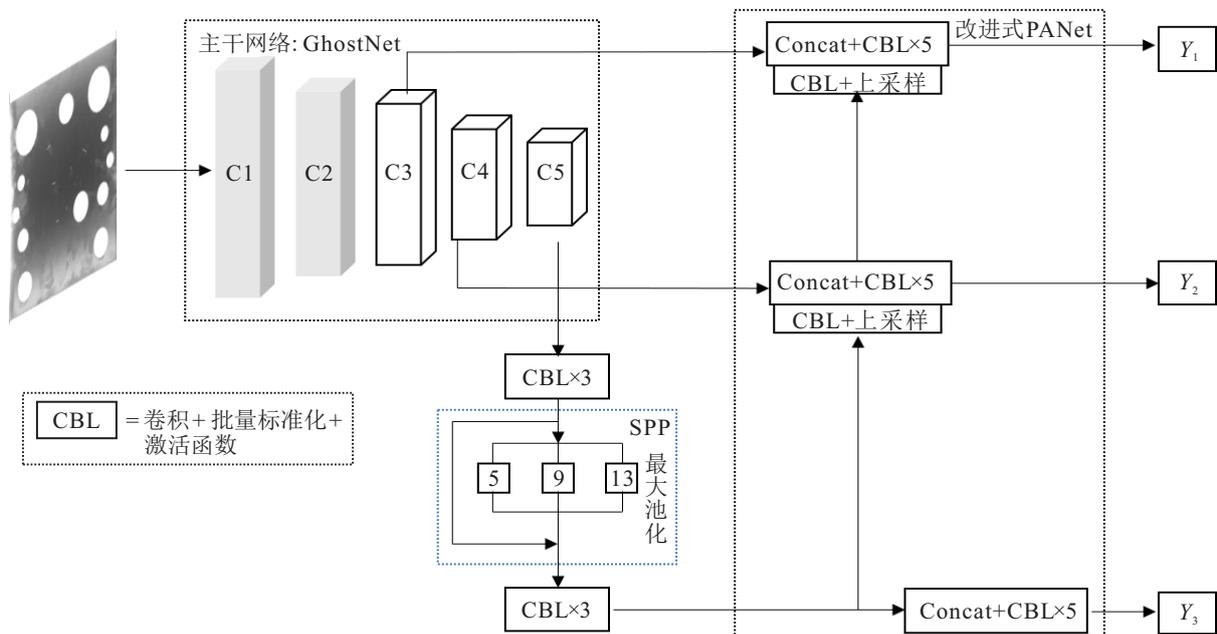


图1 基于YOLOv4的轻量化检测网络结构

1.2.1 轻量化主干特征提取网络 GhostNet

为了实现网络的轻量化及特征提取能力的提升,使用 Ghost 模块替代检测网络中的卷积操作,简单表示为

$$Y = X * f + b. \quad (2)$$

其中: $X \in R^{c \times w \times h}$ 为给定输入特征图, c 为输入特征图的通道数, h 和 w 分别为输入特征图的高度和宽度; $*$ 为卷积操作; b 为偏置项. 为了产生输出为 n 通道的输出特征图 $Y \in R^{n \times w' \times h'}$, 需要使用卷积核为 $k \times k$ 的卷积块, 即 $f \in R^{c \times k \times k \times n}$, 生成效果如图 2(a) 所示. 然而, 卷积神经网络为了得到丰富的特征表达, 残差经常需要进行多次卷积操作, 这意味着输出特征图 Y 的通道数 n 也会变得很大, 运用此方法进行一次特征提取需要的运算量为

$$\text{cost} = h' \times w' \times n \times k \times k \times c. \quad (3)$$

实际上, 通过此操作获得的大部分特征图是相似的, 意味着很多特征图可以通过更为简单的操作如旋转、平移等操作从另一组基础特征图中获得. 因此, 残差块中一些卷积操作是多余的, 网络同时进行了不必要的计算, 一定程度上浪费了计算和存储资源.

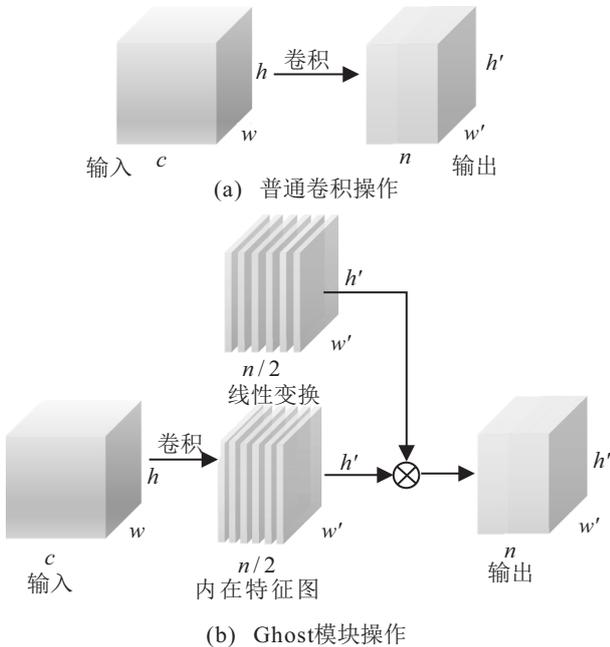


图2 普通卷积与 Ghost 模块对比

为了解决该问题, Ghost 模块提出了一种全新的卷积结构. 首先使用普通卷积块生成本征特征图 $Y' \in R^{m \times w' \times h'}$, 即

$$Y' = X * f' + b. \quad (4)$$

其中: $f' \in R^{c \times k \times k \times m}$, m 为本征特征图 Y' 的通道数. 然后为了生成与原输出特征图相同的通道数, Ghost

模块通过简单的线性运算方式增强特征并增加通道数, 有

$$y_{ij} = \Phi_j(y'_i), \forall i = 1, 2, \dots, m, j = 1, 2, \dots, s. \quad (5)$$

其中: s 为线性运算的次数, y'_i 为第 i 个本征特征图, Φ_j 为本征特征图实施第 j 个线性变换得到特征浓缩的相似特征图. 理论上可以得到 $m \times s$ 个相似特征图, 为了生成与原输出特征图 Y 相同数量的特征图, 只取前 n 个作为最终的输出结果. 运用 Ghost 模块后, 实现一次操作需要的运算量为

$$\begin{aligned} \text{cost}' &= h' \times w' \times \frac{n}{s} \times k \times k \times c + \\ &(s - 1) \times h' \times w' \times \frac{n}{s} \times k \times k. \end{aligned} \quad (6)$$

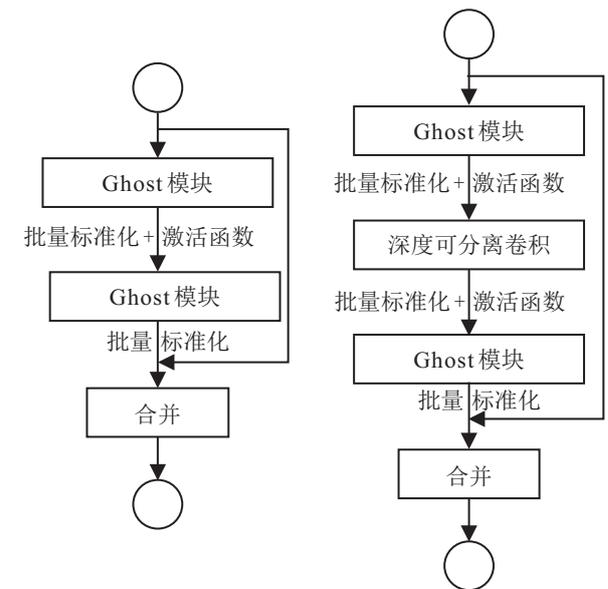
将前后方法进行对比, 由于 $s \ll c$, 运算提升率为

$$r = \frac{c}{\frac{c}{s} + \frac{s-1}{s}} = \frac{s \times c}{c + s - 1} \approx s. \quad (7)$$

比较可知: 使用 Ghost 模块替代传统卷积结构可以以更低的参数及更低的计算量获得相同甚至更优的特征提取效果.

图 2(b) 为 Ghost 模块较为简单且常用的一种情形, 即 $s = 2, m = n/2$, 后续也将主要使用此模型作为实现方法.

更进一步, 可以使用 Ghost 模块搭建瓶颈结构, 即使用 Ghost 模块替代传统瓶颈结构中的普通卷积, 如图 3 所示. 图 3(a) 主要实现增加输出特征图通道数的功能, 图 3(b) 主要实现改变特征图高和宽的功能.



(a) 步长为1的瓶颈结构 (b) 步长为2的瓶颈结构

图3 使用 Ghost 模块搭建的瓶颈结构

如表 1 所示, 使用 Ghost 瓶颈结构搭建 GhostNet 替代 YOLOv4 的主干特征提取网络 CSPDarknet53. SE 代表是否使用注意力模块, 0 表示

不使用注意力模块, 0.25表示SE模块对通道的压缩率. 根据MobileNetv3中的实验表明, 压缩率为0.25时既提高了精度, 又没有增加时间消耗.

表1 主干特征提取网络GhostNet整体架构

输入	操作	升维比例	输出通道	SE	步长
$416^2 \times 3$	卷积 3×3	—	16	—	1
$208^2 \times 16$	Ghost瓶颈	16	16	0	1
$208^2 \times 16$	Ghost瓶颈	48	24	0.25	2
$104^2 \times 24$	Ghost瓶颈	72	24	0	1
$104^2 \times 24$	Ghost瓶颈	72	40	0.25	2
$52^2 \times 40$	Ghost瓶颈	120	40	0	1
$52^2 \times 40$	Ghost瓶颈	240	80	0.25	2
$26^2 \times 80$	Ghost瓶颈	200	80	0	1
$26^2 \times 80$	Ghost瓶颈	184	80	0	1
$26^2 \times 80$	Ghost瓶颈	184	80	0	1
$26^2 \times 80$	Ghost瓶颈	480	112	0.25	1
$26^2 \times 112$	Ghost瓶颈	672	112	0.25	1
$26^2 \times 112$	Ghost瓶颈	672	160	0.25	2
$13^2 \times 160$	Ghost瓶颈	960	160	0	1
$13^2 \times 160$	Ghost瓶颈	960	160	0.25	1
$13^2 \times 160$	Ghost瓶颈	960	160	0	1
$13^2 \times 160$	Ghost瓶颈	960	160	0.25	1

GhostNet主干网络具体流程如下.

step 1: 输入为 $416 \times 416 \times 3$, 经过 3×3 卷积将输入特征图的通道数调整为16的同时进行第1次高和宽的压缩.

step 2: 将一次步长为2和一次步长为1的Ghost瓶颈结构构成一组初步特征提取模块, 步长为2的Ghost瓶颈结构对特征图进行高和宽的压缩, 步长为1的瓶颈结构进行加深通道数量, 丰富特征表达. 通过堆叠两组初步特征提取模块, 将输入图片降采样为 52×52 , 与YOLOv4主干特征提取网络第3个CSP输出的有效特征层尺寸相同.

step 3: 通过更多组Ghost瓶颈结构实现更深层次特征提取并生成 26×26 及 13×13 的特征图, 与

YOLOv4主干特征提取网络第4、第5个CSP模块输出的有效特征层尺寸相同.

step 4: 将YOLOv4主干特征提取网络生成的有效特征层替换为step 2和step 3生成的有效特征层, 完成主干特征提取网络的轻量化.

1.2.2 针对小目标检测的改进式PANet结构

工业环境下金属表面缺陷实例在表面占比极小, 属于典型的小目标检测问题. 针对这一现象, 对YOLOv4的加强特征提取网络进行进一步改进. 现有的YOLOv4使用PANet金字塔结构通过对SPP结构及主干特征提取网络输出的3个有效特征层进行一系列上采样、下采样及特征层串联合并操作, 实现输出到YOLO预测头中各维度特征层信息保持为 $1 : 1 : 1$. 这种方法可以很好地获得各个维度的特征信息, 实现大中小目标的有效检测, 使YOLOv4作为通用检测网络能灵活高效地应用在多种公共数据集上.

在本文所使用金属表面缺陷数据集中, 如图4缺陷实例细节所示, 凹槽与划痕之间有很多相似性特征, 划痕可分为连续线段和数个小凹槽并列两种类型. 此外, 划痕实例和金属表面的加工磨痕十分相似, 但划痕是缺陷, 而金属加工磨痕是在生产过程中不可避免的金属纹路, 常出现在孔洞周边及金属边缘, 属于正常现象. 如果网络感受野过小、过多关注边界信息则很难区分以上两种小目标易错检缺陷典型样例, 因此低维特征图所能提供的纹理特征作用在一定程度上低于中高维特征图所提供的语义信息. 基于此, 提出一种改进式PANet, 如图1改进式PANet部分所示. 通过保留原网络上采样部分减少降采样部分的方法, 将最终输出到YOLO预测头的各维度特征层信息改变为 $1 : 2.5 : 5.5$, 提升网络对高维信息获取的同时减少网络中冗余的下采样结构. 网络在保留低维信息的同时增加高维信息的比例, 使高维信息更好地指导划痕实例和凹槽实例的检测, 增加感受野, 跳出局部信息, 减少网络的误检率.

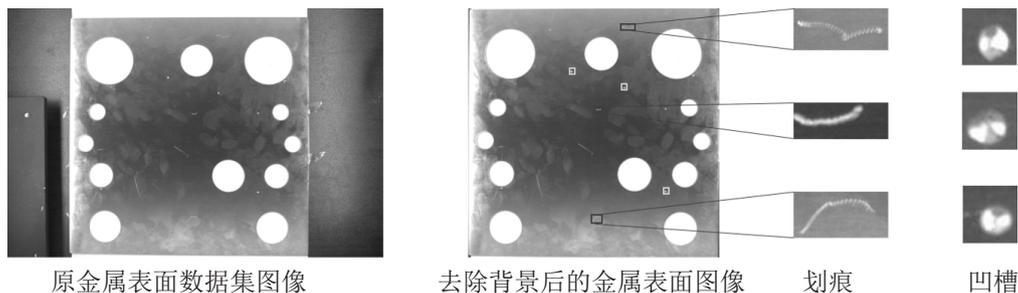


图4 真实金属数据集及缺陷实例细节

2 实验分析

2.1 数据集

本文使用发动机金属表面缺陷数据集评估方法的有效性. 该数据集并未公开且自行拍摄及标注, 共包括535张尺寸为 5472×3648 的图片, 金属表面尺寸约为 3548×3548 . 如图4所示, 数据集可划分为两个粗粒度类别, 分别为划痕和凹槽, 划痕类包含2014例, 凹槽类包含2106例. 凹槽类缺陷实例尺寸均小于 64×64 , 相对面积为0.051%; 划痕类缺陷实例形状不标准, 但像素值不多于 177×58 , 相对面积为0.082%, 根据相对尺寸的小目标定义方法, 数据集中待检测目标均属于小目标检测问题.

实验中使用DPGAN对数据集进行扩充, 缓解数据量少带来的模型过拟合问题, 提高模型的泛化能力, 通过调整缺陷实例与背景图像占比, 在一定程度上提升了小缺陷检测的准确率. DPGAN主要分为生成部分和复制粘贴部分: 生成部分使用两组生成器和判别器分别生成高分辨率的背景图像和缺陷图像; 复制粘贴部分首先将生成的缺陷实例图像与原始缺陷实例图像进行合并构成新缺陷实例数据集, 然后运用多种复制粘贴方法将新缺陷实例数据集均匀且正确地“生长”在背景图片上并同时生成含有缺陷位置的标注文件(.xml文件), 形成全新标注好的金属表面缺陷数据集. 通过此方法生成的数据集可以直接用于模型训练, 节省了重新标注的时间和精力. 数据增强前后的样本信息如表2所示, 其中训练集与测试集基于全图按照9:1的比例进行随机分配, 分配结果分别为1183张和132张.

表2 数据扩充前后总样本数

序号	类别	总样本数	
		数据增强前	数据增强后
1	划痕	2014	4351
2	凹槽	2106	4259
3	完整图片	535	1315

2.2 实验平台和参数设置

实验平台处理器为Intel(R) Xeon(R) E5-2660 v4CPU, 用于深度学习CNN计算的图像处理器为NVIDIA TITAN X(Pascal), 计算环境为1.1.0版本的PyTorch环境.

参数设置如下: 轻量化检测网络方面, 主干特征提取网络GhostNet参数设置如表1所示; 模型训练方面, 设置神经网络批量参数为16, 初始训练迭代步数为20, 冻结训练迭代步数为280, 冻结训练不仅可以

加快模型训练速度, 也可以在训练初期防止权值被破坏. 初始训练时学习率设置为0.001并采用余弦退火衰减; 初始训练结束后, 解冻后280迭代步数, 学习率为0.0001, 学习率的变化仍然采用余弦退火衰减.

在图像检测部分, 数据集本身尺寸为 5472×3648 , 首先通过边缘提取及裁剪方式得到金属表面 3548×3548 区域, 如果直接使用 3548×3548 图片检测, YOLOv4默认输入图片需要缩放为 416×416 , 则 Y_1 输入 52×52 尺寸特征图的1个像素在原图中为 $\frac{3548}{416} \times \frac{416}{52} \approx 68$, 很难学习到原图中 68×68 像素以下的特征信息网络, 而凹槽实例通常低于 64×64 . 因此, 在输入图像时运用图像重叠裁剪、检测后拼接的方法, 将一张图片裁剪为9张 1774×1774 大小的图像, 此时 Y_1 得到的 52×52 尺寸特征图的1个像素在原图中为 $\frac{1774}{416} \times \frac{416}{52} \approx 34$, 可以对数据集中缺陷进行有效检测, 但低维信息较易陷入局部信息, 进一步从侧面辅证了1.2.2节改进式PANet的有效性.

2.3 实验结果和分析

为了验证所提出轻量化检测网络模型算法的优越性, 在相同的实验条件下, 将该模型算法与基线算法进行比较. 为验证模型算法性能, 在识别精度方面使用全类平均精度(mean average precision, mAP)作为评价指标, 本实验中即为凹槽和划痕两类检测平均精度(AP)的平均值; 在检测速度方面使用网络参数量和每秒检测帧数(fps)作为性能指标, 具体效果如表3所示.

表3 基于YOLOv4的消融实验性能对比

网络及训练方式	mAP/%	fps	网络参数量
YOLOv4(基线)	80.53	8.9	64.04 M
YOLOv4+DPGAN	84.97	—	—
Ghost-YOLOv4+DPGAN	86.18	25	11.43 M
本文轻量化网络	86.36	26.1	11.26 M

从表3可以看出: 相比于基线网络YOLOv4进行金属表面缺陷检测, 表中数据增强方法及网络的轻量化改造在识别精度与检测速度上均有较大提升. 当使用DPGAN作为数据增强方式时, 在生成高分辨率缺陷图片的同时提高缺陷实例与背景的比例以带给网络更多的信息, 因此mAP有4.4%的提高. 当使用GhostNet替代YOLOv4主干特征提取网络时, 与YOLOv4网络相比, mAP从84.97%提升至86.18%, 提高了1.2%; 同时YOLOv4大幅金属图像(5472×3648)的检测速度为8.9 fps, 轻量化YOLOv4检测速度提升为25 fps, 提高了180%, 运算所需参数从64,040,001降低至11,428,525, 降低为原有的

17.8%。可以看出,与原YOLOv4相比获得了巨大提升。当继续使用改进式PANet结构时,模型精度进一步提升0.18%,检测速度提升1.1 fps,参数量减少1.5%。值得一提的是,改进式PANet结构并不需要重新训练,表中所展示数据的增长无随机性,且可即插即用至其余YOLO系列网络中,最终实现在精度提升5.83%的同时将参数量降低83.5%,检测速度提升2倍。

为验证所提出算法的性能,在相同的实验条件下,将该模型算法与工业场景经常使用的通用检测网络进行比较,具体效果如表4所示。YOLOv4-tiny是YOLOv4的官方轻量化版本,使用简易的CSPDarknet53和PANet结构;YOLOv5-s和YOLOv5-x是YOLOv5两个不同版本,前者为轻量化版本,检测速度快但精度较低,甚至不如YOLOv4原始版本,后者网络深度很高,参数量极大。通过对比可知,YOLOv4-tiny和YOLOv5-s均通过牺牲精度换取了超高的检测速度,然而工业场景下精度是第一要素,且工业相机并没有如此高的帧率。此外,本文所使用的网络在精度和检测速度上均优于YOLOv5-x,可以证实所提出方法在工业环境下小目标检测及模型轻量化方面效果显著,在保持精度的前提下提高了检测速度。

表4 各类方法性能对比

网络及训练方式	mAP/%	fps	网络参数量
YOLOv4-tiny	45.72	100.5	6.06 M
YOLOv5-s	73.59	52.7	7.27 M
YOLOv5-x	81.42	13.4	87.78 M
本文轻量化网络	86.36	26.1	11.26 M

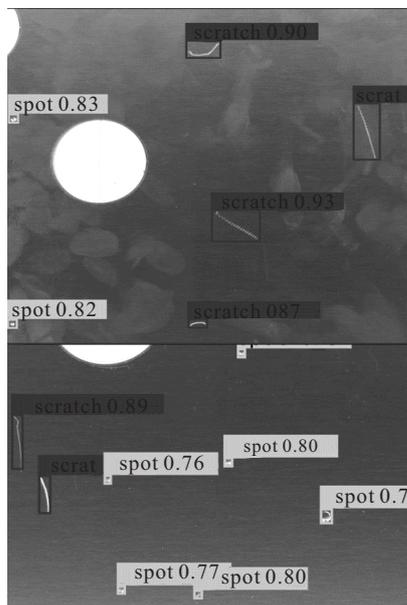


图5 金属数据集检测结果

检测可视化结果如图5所示,展示了金属表面两种缺陷的检测结果,从图5可以看出两类缺陷检测效果均很好,极少有漏检情况。

3 结论

当前工业环境下金属表面缺陷检测大多基于通用检测网络实验,识别精确度不高,运算量过于巨大,很难应用在真实工业场景中。为了提高工业环境下金属表面缺陷检测性能,本文提出了一种针对工业环境小目标缺陷的轻量化检测网络,该模型以YOLOv4为基线网络,使用GhostNet替代YOLOv4主干特征提取网络;并利用针对小目标检测的改进式PANet结构替换加强特征提取网络,在样本量有限的情况下,通过DPGAN等数据增强方式缓解过拟合现象,提升模型的泛化能力。通过在真实金属数据集上的实验验证了所提出模型的优越性。因为真实金属表面数据集较为缺乏,未来希望构建更丰富数据集,包括但不限于更复杂场景及更多种类缺陷。同时,所提出的针对小目标检测的轻量化网络不仅限于YOLOv4,理论上可以扩展到YOLO系列其他网络中,未来可以进行更丰富的实验。另外,所提出算法能够在减少网络参数量的同时提高特征提取能力,加强小目标检测精度,有效提升金属表面缺陷检测识别精度,未来希望将其扩展到航拍、小型生物、芯片检测等其他具有挑战性的目标识别领域。

参考文献(References)

- [1] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [2] Tao X, Zhang D P, Wang Z H, et al. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(4): 1486-1498.
- [3] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [4] Posilović L, Medak D, Subašić M, et al. Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans[J]. Neurocomputing, 2021, 459: 361-369.
- [5] Zhang H D, Chen Z Z, Zhang C Q, et al. Weld defect detection based on deep learning method[C]. IEEE 15th International Conference on Automation Science and Engineering. Vancouver, 2019: 1574-1579.
- [6] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C]. Proceedings of the 34th International Conference on Machine Learning.

- Sydney, 2017: 214-223.
- [7] Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks[J/OL]. 2017, arXiv: 1701.04862.
- [8] Ye Z X, Liu M Q, Zhang S L, et al. Dual-path GAN: A method for enhancing small-scale defect detection on metal images[C]. The 41st Chinese Control Conference. Hefei, 2022: 6292-6297.
- [9] Bosquet B, Cores D, Seidenari L, et al. A full data augmentation pipeline for small object detection based on generative adversarial networks[J]. Pattern Recognition, 2023, 133: 108998.
- [10] Han S, Mao H Z, Dally W J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding[J/OL]. 2015, arXiv: 1510.00149.
- [11] Frankle J, Carbin M. The lottery ticket hypothesis: Finding sparse, trainable neural networks[J/OL]. 2018, arXiv: 1803.03635.
- [12] Zhu F, Gong R H, Yu F W, et al. Towards unified INT8 training for convolutional neural network[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 1966-1976.
- [13] Zhu M, Gupta S. To prune, or not to prune: Exploring the efficacy of pruning for model compression[J/OL]. 2017, arXiv: 1710.01878.
- [14] Howard A G, Zhu M L, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[J/OL]. 2017, arXiv: 1704.04861.
- [15] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4510-4520.
- [16] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]. IEEE/CVF International Conference on Computer Vision. Seoul, 2020: 1314-1324.
- [17] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 7132-7141.
- [18] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 1577-1586.
- [19] Li S J, Sultonov F, Tursunboev J, et al. Ghostformer: A GhostNet-based two-stage transformer for small object detection[J]. Sensors: Basel, Switzerland, 2022, 22(18): 6939.
- [20] Zhang X X, Li N, Zhang R X. An improved lightweight network MobileNetv3 based YOLOv3 for pedestrian detection[C]. IEEE International Conference on Consumer Electronics and Computer Engineering. Guangzhou, 2021: 114-118.
- [21] Jiang Y X, Li L Z, Zhang F H, et al. Free range laying hens monitoring system based on improved MobileNet lightweight network[C]. IEEE International Conference on Artificial Intelligence and Computer Applications. Dalian, 2021: 487-494.
- [22] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J/OL]. 2020, arXiv: 2004.10934.
- [23] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, 2020: 1571-1580.
- [24] Mishra D. Mish: A self regularized non-monotonic activation function[J/OL]. 2019, arXiv: 1908.08681.

作者简介

叶卓勋(1998—),男,硕士生,从事工业品缺陷检测的研究, E-mail: qh.ye@zju.edu.cn;

刘妹琴(1972—),女,教授,博士生导师,从事人工智能理论与应用、水下信息感知与处理、多传感器信息融合等研究, E-mail: liumeiqin@zju.edu.cn;

张森林(1964—),男,教授,博士生导师,从事工业品缺陷检测、水下无人系统、多传感器网络等研究, E-mail: slzhang@zju.edu.cn.

(责任编辑: 郑晓蕾)