

结合数据增强和轻量化模型的 YOLOv3 木梁柱缺陷检测

王宝刚, 杨春梅, 夏鹏

(东北林业大学 机电工程学院 哈尔滨 150040)

摘要: 针对现有基于深度学习的木梁柱缺陷检测存在的泛化能力差、模型复杂、参数计算量大、实时性差导致的难以在线应用等问题,提出一种结合数据增强和轻量化模型的 YOLOv3 木梁柱缺陷检测算法。使用包括数据增强、区域删除和图像混合技术增强自制的 COCO 格式数据集,在不增加额外计算量的前提下使模型的泛化性能和鲁棒性增强。使用轻量化模型 MobileNetV3 替换原 YOLOv3 的骨干网络并更换激活函数,在减少模型参数量的基础上提升模型的预测速度。在制作的 COCO 数据集的测试结果表明:与原 YOLOv3 网络相比,所设计模型的平均精度(AP)提升了 6%, AP_{50} 达到了 92.3%, 参数数量减少了 62.52%, 预测速度达到了 53.47FPS。与 Faster-RCNN、Faster RCNN + FPN 和 SSD 等经典算法对比实验证明了所设计算法实现了兼顾精度和效率的木梁柱表面缺陷检测,并且模型参数量小更容易集成到工业应用环境。

关键词: 缺陷检测; 木梁柱; 数据增强; MobileNetV3; 实时检测

DOI: 10.15938/j.emc.2021.04.015

中图分类号: TP 273

文献标志码: A

文章编号: 1007-449X(2021)04-0123-10

YOLOv3-based wooden beam column defects detection combined with data enhancement and lightweight model

WANG Bao-gang, YANG Chun-mei, XIA Peng

(School of Electromechanical Engineering, Northeast Forestry University, Harbin 150040, China)

Abstract: For the wood beam column surface defects detection, an improved YOLOv3 algorithm based on data enhancement and lightweight model is proposed to solve the problems of poor generalization ability, complex model, large amount of parameter calculation, and poor real-time performance. The self-made coco format dataset was enhanced by data enhancement, region deletion and image mixing techniques. The generalization performance and robustness of the model were enhanced without additional computation. The lightweight model MobileNetV3 was used to replace the original backbone network of YOLOv3, and the activation function was changed, so the prediction speed of the model was improved on the basis of reducing the parameters of the model. Test results on the coco dataset show that the average precision of the proposed model is improved by 6%, and the AP_{50} is 92.3%; the number of parameters is reduced by 62.52%, and the prediction speed is 53.47 FPS compared with the original YOLOv3 network. Compared with fast RCNN, fast RCNN + FPN, SSD and other classical algorithms, experimental results show that the proposed algorithm can detect the surface defects of wood beam columns combination of accuracy and effi-

收稿日期: 2020-12-21

基金项目: 中央高校基本科研业务费专项资金项目(2572020DY12); 黑龙江省应用技术与开发计划项目(GA19A402)

作者简介: 王宝刚(1981—),男,博士研究生,研究方向为木材缺陷检测及智能优选加工;

杨春梅(1977—),女,博士,教授,博士生导师,研究方向为林业与木材加工装备智能化;

夏鹏(1986—),男,博士,副教授,研究方向为木工装备自动化加工与过程控制。

通信作者: 杨春梅

ciency, and the small amount of model parameters is easier to integrate into the industrial environment.

Keywords: defects detection; wood beam column; data enhancement; MobileNetV3; real-time detection

0 引言

木材缺陷的检测对有效降低有限森林资源的消耗,提高木材产品的商业价值有着重大意义。但是现阶段木材的缺陷检测主要是人眼检测,成功率在68%左右^[1]。随着传感器迅速发展,X射线检测、核磁共振技术成为木材缺陷检测的主流,但是成本高,复用率差。21世纪,出现了大量基于传统机器视觉的木材检测方法。Ruz等^[2]利用支持向量机对各类木材缺陷进行识别,准确率达到91%;Zhang等^[3]利用主成分分析和压缩感知来识别木材缺陷;Xie等^[4]针对木材本身纹理,采用灰度共生矩阵的方法来定位木材缺陷。传统的机器视觉方法通常采用人工设计的特征结合神经网络实现缺陷的检测,但是人工设计的特征总会有不足,对于新增的缺陷优化难以实现。

近年来,端到端的卷积神经网络在各种计算机视觉任务中取得了巨大的成功,包括图像分类^[5-6],目标检测^[7-9],语义分割^[10]。缺陷检测实质属于目标检测的工业应用,文献[11]使用神经结构搜索(neural architecture search, NAS)和Mask-RCNN结合的模型对单板表面缺陷进行检测,实现了95.31%的平均精确度,一批50张图片仅需2.5s,但采用的是组合模型的方法,模型十分复杂;文献[12]采用一种混合的全卷积神经网络对木材缺陷进行检测和定位;文献[13]使用Faster RCNN和迁移学习对木材表面缺陷进行检测,但是检测速度不能保证。

目前基于卷积网络的木材缺陷检测模型,往往存在模型复杂,参数计算量大,实时性差的问题,而且由于数据量的问题,模型泛化能力不能保证。YOLO(You look only once)算法由于检测速度和精度的优异,被广泛应用到水果质检^[14],自动驾驶^[15],CT图像识别^[16]等领域。本文在YOLOv3的模型基础上,使用区域删除和图像混合的数据增强方法增强模型的泛化能力和鲁棒性,同时使用轻量化模型MobileNetv3来替换原网络的backbone,减少模型的参数量,方便工业嵌入式使用。最后采用COCO数据集上的评价指标,综合评价模型的性能。

1 数据集制作

使用工业相机从木材加工现场拍摄得到木材缺陷图像,将感兴趣区域提取出来,得到结节、裂缝、虫洞和无损图像各300张,总计1200张。图像的像素为 200×200 ,每张缺陷图片包括一个或多个缺陷。这是一个相对小型的数据集,因此采用数据增强技术来增强模型的泛化能力。通过对数据集的分析发现,数据集中没有面积大于 96×96 的大缺陷,只有小于 32×32 的小缺陷和在 32×32 到 96×96 范围的中缺陷(缺陷大小的划分参考COCO数据集),因此木梁柱的表面缺陷大小一般都属于小和中。所要识别的缺陷图像和标签如图1所示。

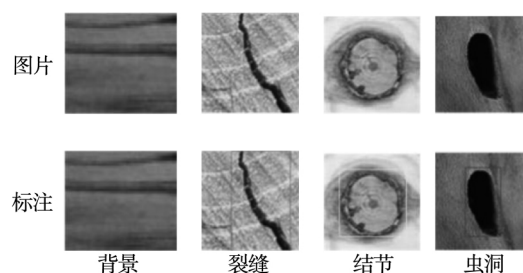


图1 部分缺陷图片和标签

Fig.1 Partial defect pictures and labels

将得到的数据集使用labellmg软件进行标注,得到缺陷位置的信息,然后将其制作成COCO格式的数据集。在相关研究中,往往选择 $\text{IOU} = 0.5$ 时的Average Precision(AP)作为最终的评价指标,实践表明其不能完全反映一个检测器的性能。因此文中使用 $\text{IOU} = [0.5:0.05:0.95]$ 的AP来评价模型,通过取10个IOU阈值,然后取AP的平均值,可以更全面的评价模型。

2 YOLOv3 算法

YOLO^[17-18]系列算法发展到现在被广泛应用的第三代YOLOv3,将目标检测作为一个端对端的回归问题,直接对锚框进行回归和分类,得到最终结果,检测速度有了大幅提升。YOLOv3的网络结构和多尺度预测如图2所示。Darknet53框架中每一个卷积部分都使用Conv2D结构,即图中的CBL部

分。每次卷积后,进行标准化(Batch Normalization, BN)和 Leaky ReLu 激活函数处理。使用多个残差块堆叠的方式构成骨干网络 Darknet53,从 3 个尺度预测模型输出。图 2 中以输入 416×416 为例,实际输入网络的大小从 32×10 至 32×19 中 10 个输入中随机选择。

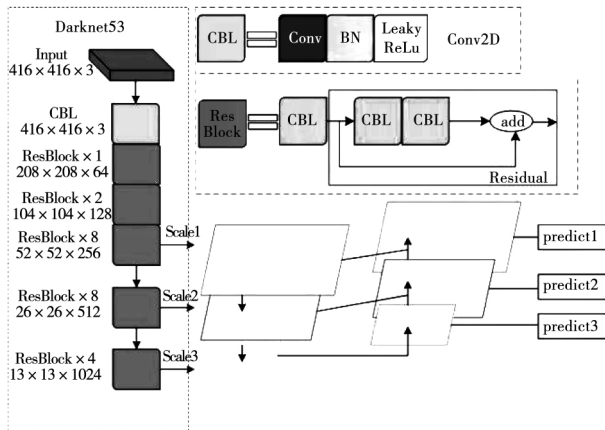


图 2 YOLOv3 结构

Fig. 2 Structure of YOLOv3

3 数据增强

在图像数量获取难以达到充足和完善的工业检测领域,为避免卷积网络的严重过拟合问题,数据增强技术是解决该问题最有效的方法。本文使用随机亮度、随机对比度、随机颜色扰动、随机翻转、随机旋转、随机裁剪和添加噪声七种基本数据增强作为基本数据增强。同时使用基于区域删除的数据增强技术 GridMask^[19]和基于图像混合的数据增强技术 MixUp^[20]进一步增强网络的泛化能力。

3.1 GridMask 数据增强方法

除了基本的数据增强方法,基于区域删除的数据增强方法得到了广泛应用。Random Erasing^[21]通过随机删除一部分区域来达到增强模型泛化能力,让模型学习到原本不敏感的信息。Cutout^[22]通过删除连续的正方块区域,有效增加了数据集和部分被遮挡的样本。GridMask 思想和前两种方法相同,但是前两种方法往往容易出现删除到整个目标区域的问题,所以 GridMask 技术通过产生一个均匀分布的掩膜来删除区域。如图 3 所示,通过生成一个和原图相同分辨率的掩膜(mask)图像,然后将该掩膜图与原图相乘,得到了特定区域信息删除的新图像。mask 图中黑色区域值为 0,表示删除区域。虚线框部分为基本的 mask 单元, (x, y) 表示第一个 mask

单元离图像边缘的距离, r 为保留图像的比例, d 为 mask 单元的边长。通过删除均匀分布的正方形区域,既避免了过度删除图像中的目标信息,又避免了没有删除到目标信息而不能起到增加网络泛化能力的作用。

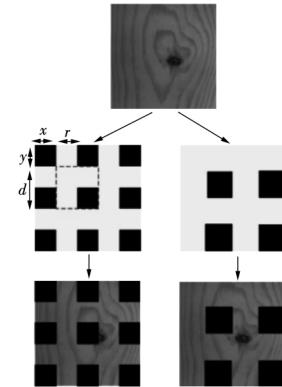


图 3 Gridmask 数据增强

Fig. 3 Gridmask data enhancement

3.2 图像混合增强

文献[23]中提出一种高效的图像混合方法 Sample Pairing,可以将训练集规模从 N 扩大到 $N \times N$,将随机两副图像的像素相加求平均,而样本的标签不变。Mixup^[20]从经验风险主义最小化和邻域风险主义最小化的角度出发,解释了为什么混合图像会对模型训练有效。Mixup 使用方法如下:

$$\begin{aligned} x' &= \lambda x_i + (1 - \lambda) x_j, \\ y' &= \lambda y_i + (1 - \lambda) y_j. \end{aligned} \quad (1)$$

其中: x', y' 表示混合后的图片和标签; x_i, x_j 代表从一个批次随机抽取的 2 种样本; y_i, y_j 代表对应的标签信息。 $\lambda = \text{beta}(\alpha, \alpha) \in (0, 1)$, α 为可以设定的参数。Mixup 之后的图像如图 4 所示。img1、img2 为批次中随机的两种图片,以 λ 比例混合,标签 label 以同样的方式混合。

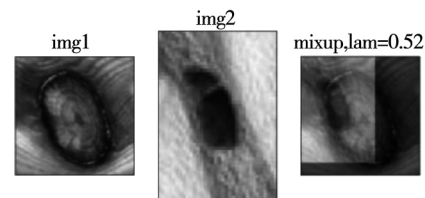


图 4 MixUp 示例

Fig. 4 MixUp example

4 轻量化模型设计

YOLOv3 由于其 Darknet53 结构多达 106 层的

网络深度和其多尺度的设计,使用大量堆叠的残差结构,参数运算量巨大,难以在移动端嵌入,所以对其网络模型进行轻量化是必要的。常见的轻量化方法包括网络剪枝^[24]、知识精馏^[25]等,除此之外还包括使用高效的卷积结构来减少模型的参数量。在 MobileNetV1^[26]中提出深度可分离卷积(Depthwise Separable Convolution),将传统卷积分为 DepthWise (DW) 和 PointWise (PW) 两步,同时使用 ReLU6 作为激活函数,如图 5 (b) 所示。标准卷积由卷积、批归一化和 ReLU 激活组成, MobileNetv1 使用 Depth-wise Separable Convolution 和 ReLU6 激活, MobileNetv2 使用逆残差的线性瓶颈结构,如图 5 中所示。

当卷积核尺寸为 $D_K \times D_K$, 输入特征图尺寸为 $D_F \times D_F$, 传统卷积的计算量为 $F_1 = D_K \times D_K \times M \times N \times D_F \times D_F$ (M, N 表示输入和输出的通道数), 而深度可分离卷积的计算量为 $F_2 = D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F$, 两者计算量之比 F_2/F_1 可由式 (2) 来描述。

$$\frac{F_2}{F_1} = \frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (2)$$

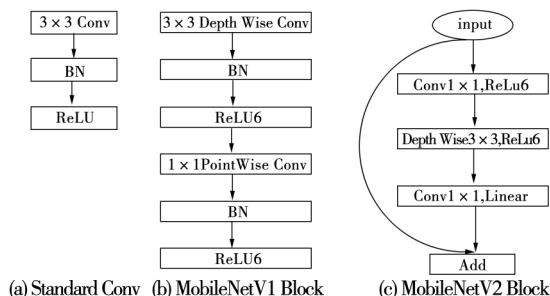


图5 标准卷积、MobileNetv1、MobileNetv1v2 的基本结构

Fig. 5 Basic structure of standard convolution, MobileNetv1, MobileNetv1v2

MobileNetv2^[27] 和 MobileNetv1 都是采用 DW (Depth-wise) 卷积搭配 PW (Point-wise) 卷积的方式来提取特征, 但为了避免 ReLU 对特征的破坏。 MobileNetv2 在 DW 卷积前面加入了一个 PW 卷积并且在第二个 PW 结构使用线性激活, 即线性瓶颈 (Linearbottlenecks) 结构。同时提出倒置残差 Inverted residuals 的结构, 使用 1×1 的卷积将 feature map 升维, 通过一个 DW 结构提取特征, 最后再通过一个 1×1 卷积降维。先进行扩张再进行压缩如图 6 (c) 所示。

MobileNetV3^[28] 为 2019 年提出, 结合了前两个版本的优点, 综合使用 V1 的深度可分离卷积和 V2 的具有线性瓶颈的逆残差结构, 在 V2 的 1×1 卷积之后加入 Squeeze-and-Excitation Networks (SE) ^[29] 的 attention 模块。同时提出使用 swish 激活函数, 可以有效地提高网络的精度, 其具有无上界有下界、平滑、非单调的特性。其在模型效果上优于 ReLU。 Swish 函数表示为: $Swish[x] = x \times sigmoid(\beta x)$ β 为一个常量或者可训练的参数。由于 swish 的计算量太大, 用图 6 中所示的 H-Swish 函数 (hard version of swish) 替代 Swish 激活函数, 其描述如下:

$$h-swish = x \times \frac{ReLU6(x+3)}{6} \quad (3)$$

ReLU 是深度神经网络常用的激活函数, 将负值置 0, 其余值保持不变; ReLU6 通过抑制最大值的方式, 来实现移动端高精度的应用。图 6 (b) 为 Swish 和 Hard-Swish 激活函数。 Swish 激活函数的运算成本是非零的, 所以用 Hard-Swish 来近似 swish 激活。

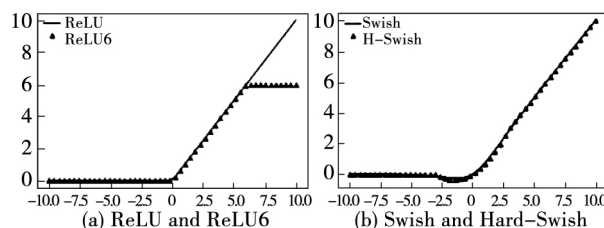


图6 激活函数的表示

Fig. 6 Representation of activation function

从图 6 (b) 可以看出 H-Swish 无限逼近 Swish 函数, 计算量更小, 速度更快。所以选择移动端网络 MobileNetv3 替换 Backbone, 达到减少模型参数量、提高模型预测速度的目的。将 YOLOv3 的 backbone 替换成 MobileNetv3 的结构如图 7 所示。 Exp size 为扩展系数, SE 表示是否使用 SE 模块, NL 为激活函数, HS 为 Hard-Swish, RE 为 ReLU 激活, Bneck 为 MobileNetv3 的 bottleneck, 使用 MobileNetv2 的线性瓶颈结构, 在特定层加入 SE 模块。

5 模型训练及实验

5.1 模型训练

所设计的木梁柱缺陷检测算法流程如图 8 所示, 主要包括数据增强、骨干网络特征提取和检测头的检测三部分。

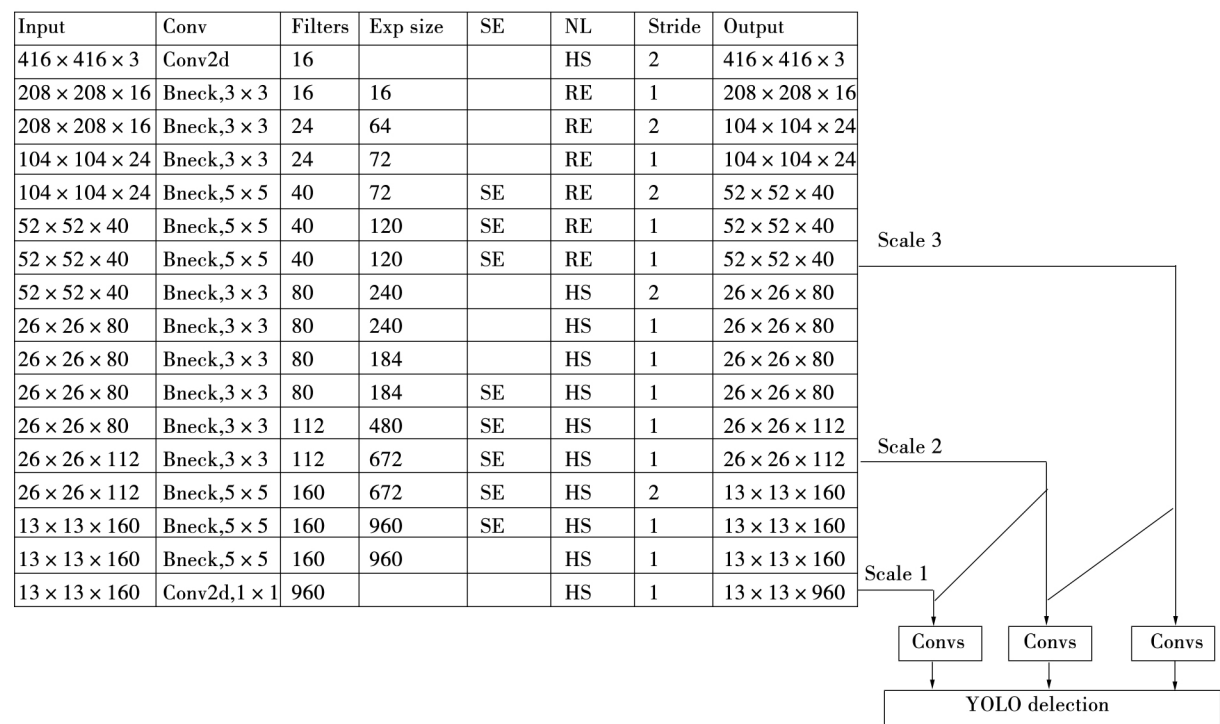


图 7 backbone 更换为 MobileNetv3 后的结构

Fig.7 Structure after the backbone is replaced with MobileNetv3

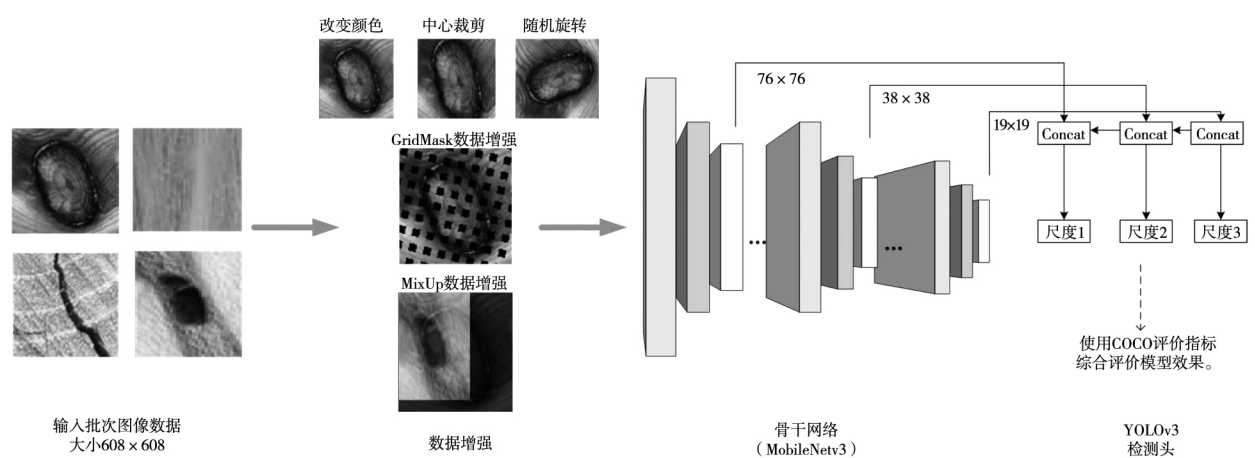


图 8 算法整体流程图

Fig.8 Algorithm overall flow chart

首先将收集到的缺陷图像按照 8:1:1 的比例划分训练集、验证集和测试集,制作成 COCO 格式。从三个尺度提取特征图像,进行融合,提取 76×76 , 38×38 , 19×19 三个尺度进行检测,使用 COCO 的评价指标来评价模型效果。采用基本的 YOLOv3 模型进行训练,训练时 batchsize 大小为 8,测试时 batchsize 大小为 1,训练时模型的输入从 32×10 至 32×19 ,从 10 个数值中随机选择大小,测试时输入大小为 608×608 ,每个批次中图像大小一致,使用

学习率预热的方法,在两百次迭代后将学习率上升到初始学习率 0.000 1 大小,如图 9(a) 所示。总迭代次数为 10 000 次,在迭代次数的 2/3 和 11/12 处分别将学习率下降 1/10,使用动量项为 0.9 的随机梯度下降方法。预热时学习率较小,模型可以慢慢趋于稳定,等模型相对稳定后在使用预定的学习率进行训练,可以使模型收敛的更快。图 9(b) 为训练过程的损失曲线,横轴 500 表示每 20 次迭代次数取一次数据,实线为训练集损失,虚线为验证集损失。

从图中可以看出,通过数据增强后,虽然使用的是一个相对较小的数据集,模型没有出现拟合现象,训练集和验证集的损失都是震荡下降。图9(c)、图9(d)分别为 AP 、 AP_{50} 在训练过程中的可视化,这里 AP 指的是选择 $IOU = [0.5:0.95]$ 时的 AP , AP_{50} 指选择 $IOU = 0.5$ 时的 AP 。可以看出YOLOv3网络 AP 最高达到0.41, AP_{50} 为0.83。

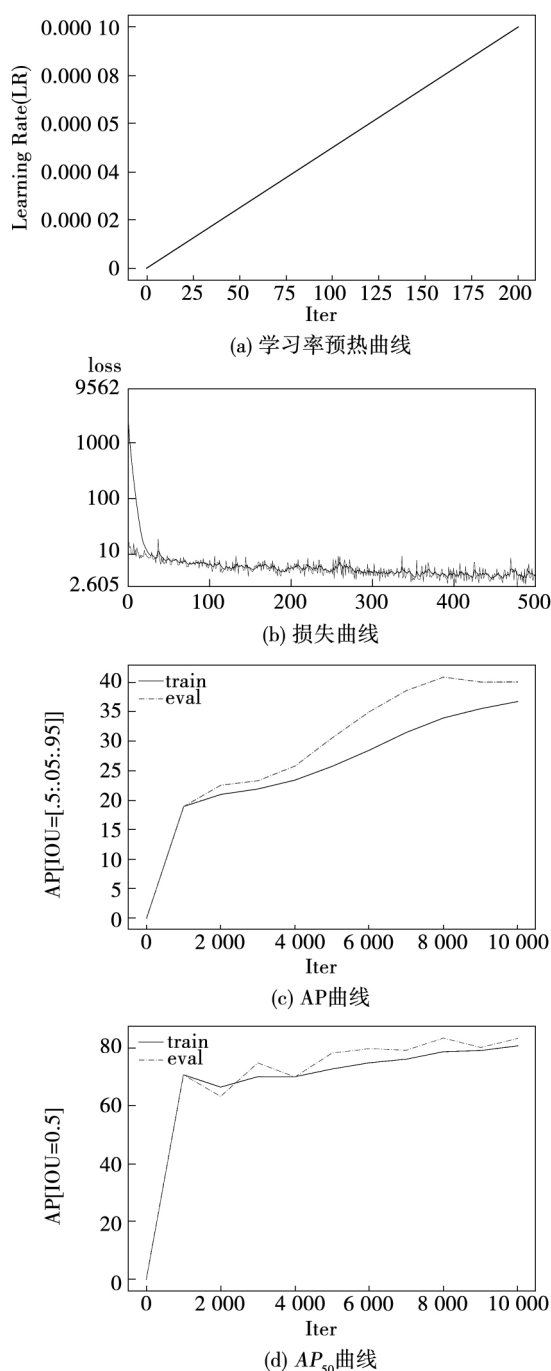


图9 训练过程的可视化

Fig.9 Visualization of the training process

5.2 改进后模型实验

使用5.1中的基线模型作为对照,加入数据增强和更换backbone,在相同的实验条件下进行测试。使用python language来实现,深度学习框架为PaddlePaddle 1.84,所有的实验都是在一台配备了16G内存的NVIDIA Tesla V100 GPU上进行,使用CUDA10.0和CUDNN7.4配合GPU加速。得到表现最好的迭代次数时结果如表1所示。COCO代表使用COCO的评价指标,表1模型中,“+”代表在YOLOv3基础上使用的技术,数字编号代替本行中所用技术。S代表small object,M表示medium object,AR为 $IOU = [0.5:0.95]$ 的Average Recall,Parameter为模型的参数数量。

编号1、2、3分别为加入了MixUp,GridMask和两个同时使用时的实验数据。设置MixUp的参数 $\lambda = \text{beta}(0.2, 0.2)$,GridMask设置概率 p 为0.7,随迭代次数的增加而变大,两种数据增强方法都只在前9000次使用,在最后1000次停止使用,让网络微调。可以看出,加入MixUp后, AP 没有变化, AP_{50} 增加了4.8%,在召回率方面均有提高,加入GridMask后, AP 提升了1.4%, AP_{50} 增加了3%,两种数据增强技术同时使用时, AP 增加了可观的3%, AP_{50} 达到了0.9,在 AP_s 、 AP_m 、 AR 、 AR_s 、 AR_m 均有不错的表现。

编号4为将原YOLOv3的backbone更换为MobileNet3后的实验数据,参数数量减少了62.52%,预测速度提高了10.6FPS,同时得益于MobileNet高效的网络结构和注意力模块,模型在8个指标上均有不错的提升,其中 AP_{50} 达到了0.91。不仅减少了参数数量,更提升了网络的性能。编号5、6是将数据增强技术使用到更换backbone之后的网络中,加入两种数据增强,实现了最好的模型性能, AP 提升了5.9%, AP_{50} 达到了0.924,实现了精度和速度的完美融合。综合来看,文中的改进效果提升显著,参数数量减少了62.52%,预测速度提高了10.6FPS, AP 提升了6%, AP_{50} 提升了近10%,精度达到了92.4%。

图10为利用改进后的网络的部分检测结果,包括缺陷类型和缺陷位置。可以看出改进后的网络可以准确的识别出三种缺陷的类型,而且置信度都超过了90%。

表 1 使用数据增强和更换 backbone 的实验结果

Table 1 Use data to enhance and replace the experimental results of the backbone

方法 编号	模型	COCO								参数 (M)	FPS (帧/秒)
		AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AR	AR _s	AR _M		
	YOLOv3	0.41	0.83	0.36	0.44	0.47	0.49	0.39	0.524	353.82	42.9
1	+ MixUp	0.41	0.878	0.32	0.25	0.45	0.50	0.4	0.548	352.82	42.9
2	+ Grid Mask	0.424	0.86	0.31	0.26	0.45	0.49	0.33	0.535	352.82	42.9
3	+ 1 + 2	0.44	0.90	0.41	0.44	0.47	0.50	0.45	0.526	352.82	42.9
4	+ MobileNet	0.447	0.91	0.41	0.45	0.47	0.52	0.46	0.554	133.22	53.5
5	+ 4 + 2	0.453	0.95	0.40	0.45	0.45	0.53	0.47	0.54	133.22	53.5
6	+ 1 + 2 + 4	0.469	0.924	0.32	0.40	0.5	0.54	0.43	0.58	133.22	53.5

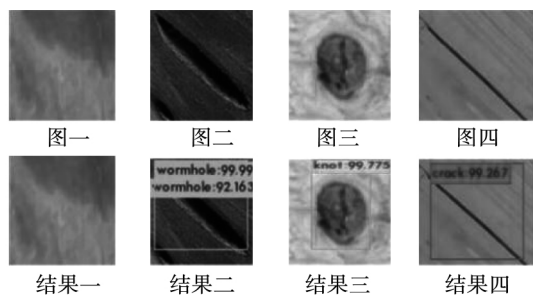


图 10 部分缺陷的检测结果

Fig. 10 Detection results of some defects

5.3 模型参数的确定

为了找到模型最佳时的参数,在相同实验条件下,使用 5.2 中编号 6 作为基线模型,分别控制学习率(Learning Rate, LR)、批次大小(Batch Size, BS)、动量项的大小(Momentum, M) 3 个变量进行学习,得出的模型检测效果。

1) 学习率(Learning Rate, LR),由于使用预训练模型,模型已经收敛,所以使用较小的学习率 0.000 1 取得了较好的结果,AP 为 0.47, AP₅₀ 为 0.92,而且采用了学习率梯度衰减的策略,后期学习率继续下降,以一个小的步长找到最优解,如图 11 所示(其中学习率 0.000 1 图例为最佳表现)。

2) 批次大小(Batch Size, BS),一般在合理范围内,BS 越大使得下降方向更加准确,但是过大的 BS 会出现局部最优的情况,太小的 BS 会使模型振荡,也不利于模型训练。将 BS 从 4 增加到 16,在 BS = 4、16 时模型效果明显不如 BS = 8、12,说明此时的模型 BS 太小或太大,而 BS = 8、12 时,检测效果 AP 相同为 0.47,但是 BS = 8 时,AP₅₀, AR 相对 BS = 8

时均表现更好,综合考虑选择 BS = 8 作为最优 BS。当 BS = 16 时,模型的 AP₅₀ 达到了 0.95,但是此时 AP 只有 0.42,说明仅仅只以 AP₅₀ 作为评判标准可能是不够准确的,此时模型可能出现了局部最优的情况,如图 12 所示。

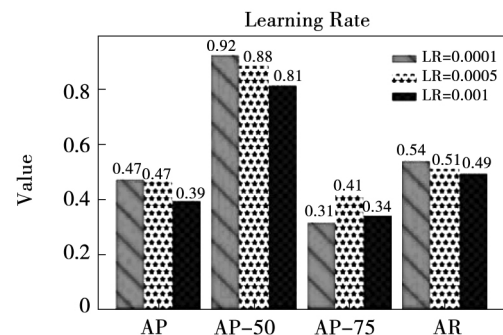


图 11 随学习率变化模型的测试结果

Fig. 11 Test result of the model with the learning rate

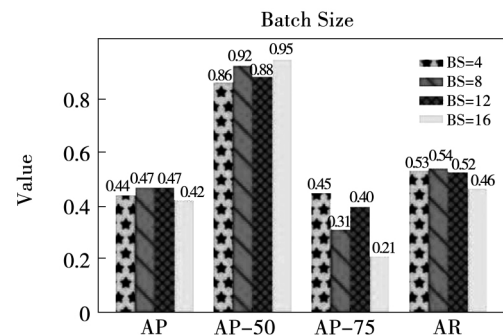


图 12 随 batchsize 变化模型的测试结果

Fig. 12 Test results of the model with batchsize

3) 动量项(Momentum, M)。优化器算法是深度

学习中必不可少的一部分 ,其中 Momentum 由于有效性被广泛采用 , M 越大 ,就越可能摆脱局部最优的束缚。动量项从 0.85 上升到 0.95 ,得到的结果如图 13 所示。当 Momentum 的值为 0.9 时 ,模型的性能最好 , $M=0.925$ 时 AP_{50} 达到了 0.95 ,进一步证明使用 AP 而不是 AP_{50} 来验证模型有效性 ,是很有必要的。

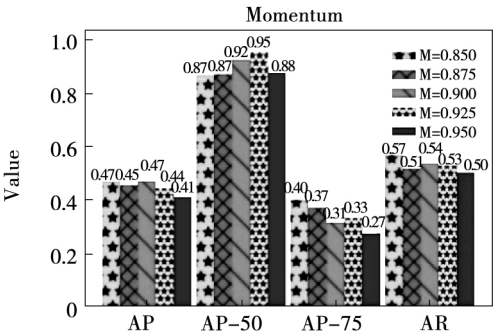


图 13 随动量项的值变化模型的测试结果

Fig. 13 Test results of the model with the M value

5.4 不同网络的对比实验

为验证所设计网络的优越性 ,选择经典的模型作为对照组 ,分别选择 Single Shot MultiBox Detector (SSD) [30] ,Faster-RCNN [7] ,Faster-RCNN + FPN [31] 作

为对照组。SSD 和 YOLO 系列都是单阶段算法的代表 ,SSD 分为 SSD300 ,SSD512 ,输入分别是 300×300 和 512×512 大小 ,使用 VGG16 作为 backbone ,在多层多尺度特征图上进行检测同时使用默认锚框的方式避免使用建议区域。

Faster RCNN 先使用 RPN (region proposal network) 找到目标区域 ,然后在目标区域上进行边框回归和分类 ,得到目标位置和类别信息。FPN (Feature Pyramid Networks) 采用金字塔结构提取融合特征 ,集成在 Faster RCNN 上实现了更好的效果。使用这四种算法和改进后的 YOLOv3 的测试结果如表 2 所示 ,括号里面为模型所使用的 backbone。

表 2 给出了目标检测的几个经典网络和本文模型在数据集上测试的结果以及参数数量和预测速度 ,其中 Faster-RCNN 和 Faster RCNN + FPN 的预测速度都为 20FPS 左右 ,不满足工业实时检测最低 30FPS 的条件 ,因此不适合作为木材缺陷的实时检测算法。而 SSD 算法在预测速度上满足工业实时需要 ,但是表现最好的 SSD300 ,AP 也低于本文提出的算法 ,因此本文的算法处于绝对优势。

表 2 不同网络的检测结果比较

Table 2 Comparison of detection results for different networks

模型	COCO								Parameter (M)	FPS
	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AR	AR _s	AR _M		
本文模型	0.469	0.924	0.31	0.391	0.50	0.537	0.43	0.58	133.22	53.5
SSD300(VGG16)	0.449	0.9	0.47	0.423	0.46	0.539	0.48	0.56	172.06	47.4
SSD512(VGG16)	0.406	0.864	0.35	0.378	0.43	0.484	0.43	0.51	172.06	43.8
Faster RCNN(ResNet50)	0.38	0.78	0.28	0.23	0.29	0.434	0.4	0.446	209.87	22
Faster RCNN + FPN(ResNet50)	0.427	0.854	0.40	0.49	0.42	0.508	0.52	0.51	245.87	20.4

6 实验结果分析

与使用其他方法在木梁柱表面缺陷检测方面的识别效果的比较见表 3。与其他文献中的方法相比较 ,文中的方法在一个小型数据集上进行检测 ,使用

更全面的评价指标 ,实现了 46.9% 的 AP ,92.4% 的 AP_{50} ,同时使用高效的轻量化模型 ,检测速度更具优势 ,在一秒内可以识别 53.7 张缺陷图片 ,实现了精度和速度的完美结合。而且参数量很小 ,可以实现在工业端的在线部署。

表 3 检测结果比较

Table 3 Comparison of detection results

序号	方法	缺陷类型	结果	图像尺寸
1	数据增强和轻量化的 YOLOV3(ours)	结节、裂缝、虫洞	精度 (IOU = 0.5: 0.95) : 46.9%	200 × 200
			精度(IOU = 0.5) : 92.4%	
			检测时间: 一副图像 18.7 ms	
2	灰度共生矩阵和前馈神经网络 ^[32]	结节	准确率: 78.26%	51 × 48
3	自动视觉检测和多分类的 支持向量机 ^[2]	鸟眼、污斑、开裂、蓝色斑点 ,	准确率: 91%	320 × 240
		木髓、结节、洞	检测时间: 一副图像 11.13 s	
4	主成分分析和压缩感知 ^[3]	活结、死结、裂缝	准确率: 92%	256 × 256
			检测时间: 一副图像 0.216 s	
5	创建特征字典、图像处理和 支持向量机分类 ^[33]	裂缝、结节	准确率: 91.5%	256 × 256
			检测时间: 一副图像 1.86 s	
6	混合的全卷积神经网络 ^[12]	死结、活结、蓝色斑点 ,	准确率: 99.14%	256 × 256
		灰色斑点、沥青条纹、裂缝	检测时间: 一个批次图像 0.368 s	
7	神经架构搜索和多通道的 Mask R-CNN ^[11]	死节、活节、裂缝	准确率: 98.70%	200 × 200
			检测时间: 100 张图片 2.5 s	
8	Faster R-CNN 和迁移学习 ^[13]	结节、裂缝、污斑	精度: 80.6%	800 × 600
			准确率: 96.1%	

7 结 论

木梁柱表面缺陷的高效检测是工业木材单板生产中不可缺少的环节,只有实时准确给出木材表面缺陷信息,才能进入下一步的排样和木材加工环节。本文提出一种结合数据增强和轻量化模型的木梁柱表面缺陷自动检测和定位方法,采用数据增强和轻量化模型改进 YOLOv3 网络,实现了木材缺陷高精度的实时检测。使用基于区域删除技术的数据增强方法 GridMask 和基于图像混合的方法 MixUp,来增强模型的泛化能力和鲁棒性,在一个较小的数据集上也能实现很好的检测精度。使用高效的 MobileNetv3 网络不仅实现了检测精度的进一步提高,而且大幅减少了模型的参数数量,提高了实时性能。最终结果表明: 实现了 53.5FPS 的检测效率,AP 为 46.9%,AP50 为 92.3%,参数数量相比原模型减少了 62.52%,完全满足工业实时木材缺陷检测的需要。此外,该方法还可以应用于数据较为稀缺的小缺陷表面检测领域。

文中的方法也为以后的工作提供了一些启发: 首先使用数据增广技术是完全可以实现深度神经网络在小数据集上的识别效果,我们会继续探索使用更高效的数据增强方法,以期达到更好的效果。第

二 轻量化模型是未来工业端使用的主流方向,探索实现更加高效的轻量化模型,是以后工作的重点。

第三,期望木材检测方法可以直接提供缺陷的轮廓和加工方案,方便后续的缺陷处理。

参 考 文 献:

[1] HASHIM U R , HASHIM S Z , MUDA A K. Auto-matedvision inspection of timber surface defect [J]. Jurnal Teknologi ,2015 ,77 (20) : 127.

[2] RUZ G A , ESTEVEZ P A , RAMIREZ P A. Auto-mated visual inspection system for wood defect classification using computational intelligence techniques [J]. International Journal of Systems Science ,2009 ,40(2) : 163.

[3] ZHANG Yizhuo , XU Chao , LI Chao , et al. Wood defect detection method with pca feature fusion and compressed sensing [J]. Journal of Forestry Research ,2015 ,26(3) : 745.

[4] XIE Yonghua , WANG Jincong. Study on the identification of the wood surface defects based on texture features [J]. Optik-International Journal for Light and Electron Optics , 2015 , 126 (19) : 2231.

[5] HE Kaiming , ZHANG Xiangyu , REN Shaoqing , et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition ,June 27 – 30 ,2016 , Las Vegas ,NV ,USA. IEEE Computer Society , 2016: 770 – 778.

[6] SIMONYAN K , ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. Computer Science ,

- 2014 2333.
- [7] REN Shaoqing, HE Kaiming, Girshick R, et al. Faster rcnn: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017 39(6): 1137.
- [8] HE Kaiming, GKIOXARI G, DOLLAR P, et al. Mask rcnn [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020 42(2): 386.
- [9] 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, June 27 – 30 2016, Seattle, WA. 2016: 779 – 788.
- [10] CHEN Liang-Chieh, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs [J]. Computer Science 2014 40(4): 834.
- [11] SHI Jiahao, LI Zhenye, ZHU Tingting, et al. Defect detection of industry wood veneer based on NAS and multi-channel mask rcnn [J]. Sensors 2020 20(16): 4398.
- [12] HE Ting, LIU Ying, XU Chengyi, et al. A fully convolutional neural network for wood defect location and identification [J]. IEEE Access 2019 7: 123453.
- [13] URBONAS A, RAUDONIS V, MASKELIUNAS R, et al. Automated identification of wood veneer surface defects using faster region based convolutional neural network with data augmentation and transfer learning [J]. Applied Sciences 2019 9(22): 4898.
- [14] TIAN Yunong, YANG Guodong, WANG Zhe, et al. Apple detection during different growth stages in orchards using the improved yolov3 model [J]. Computers and Electronics in Agriculture 2019 157: 417.
- [15] CHOI J, CHUN D, KIM H, et al. Gaussian yolov-3: an accurate and fast object detector using localization uncertainty for autonomous driving [C]//2019 IEEE International Conference on Computer Vision. 27 Oct. – 2 Nov. 2019, Seoul, Korea (South). 2019: 502 – 511.
- [16] PANG Shanchen, DING Tong, QIAO Sibo, et al. A novel YOLOv3-arch model for identifying cholelithiasis and classifying gallstones on CT images [J]. Plos one 2019 14(6): e0217647.
- [17] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//IEEE Conference on Computer Vision and Pattern Recognition. July 21 – 26, 2017, Honolulu, HI. 2017: 7263 – 7271.
- [18] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. arXiv eprints 2018 arXiv: 1804.02767.
- [19] CHEN Pengguang, LIU Shu, ZHAO Hengshuang, et al. Grid-mask data augmentation [J]. arXiv eprints, 2020, arXiv: 2001.04086.
- [20] ZHANG Hongyi, CISSE M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization [C]//6th International Conference on Learning Representations, April May 30, 2018, Vancouver, BC, Canada. 2018: 149806.
- [21] ZHONG Zhun, ZHENG Liang, KANG Guoliang, et al. Random erasing data augmentation [J]. arXiv e-prints, 2018, arXiv: 1708.04896.
- [22] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout [J]. arXiv e-prints, 2017, arXiv: 1708.04552.
- [23] INOUE H. Data augmentation by pairing samples for images classification [J]. arXiv eprints 2020 arXiv: 1801.02929.
- [24] HAN Song, MAOHUZI, DALLY W J. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding [J]. Fiber 2015 56(4): 3.
- [25] HINTON G, VINYAKS O, DEAN J. Distilling the knowledge in a neural network [J]. Computer Science 2015 14(7): 38.
- [26] HOWARD A G, ZHU Menglong, CHEN Bo, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications [J]. arXiv eprints 2017 arXiv: 1707.04861.
- [27] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18 – 23, 2018, Salt Lake City, UT. 2018: 4510 – 4520.
- [28] HOWARD A, SANDLER M, CHEN Bo, et al. Searching for MobileNetV3 [C]//2019 IEEE International Conference on Computer Vision (ICCV), 27 Oct. – 2 Nov. 2019, Seoul, Korea (South). IEEE 2019: 1314 – 1324.
- [29] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze and excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2019 42(8): 2011.
- [30] LIU Wei, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector [C]//European Conference on Computer Vision. Springer International Publishing, Oct 8 – 16, 2016, Amsterdam, Netherlands. 2016: 21 – 37.
- [31] LIN Tsungyi, DOLLAR P, Girshick R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21 – 26 2017, Honolulu, HI, USA. IEEE, 2017: 936 – 944.
- [32] PACKIANATHER M S. Feature selection method for neural network for the classification of wood veneer defects [C]//2008 World Automation Congress Waikoloa, 28 Sept – 2 Oct 2008, HI, USA. IEEE 2008.
- [33] HITTAWE M, MUDDAMSETTY S M, SIDIBED, et al. Multiple features extraction for timber defects detection and classification using SVM [C]//2015 IEEE International Conference on Image Processing (ICIP), September 27, 2015, Quebec City, QC, Canada. IEEE 2015: 427 – 431.

(编辑: 刘素菊)