

[文章编号] 1003-4684(2023)02-0011-06

改进的轻量级 YOLOv4 输电线路鸟巢检测方法

武明虎, 雷常鼎, 刘 聪

(湖北工业大学电气与电子工程学院, 湖北 武汉 430068)

[摘 要] 基于深度学习的输电线路缺陷检测方法往往需要昂贵的硬件支持来实现实时和高精度的技术指标要求, 难以大规模推广和普及。针对这一问题, 提出一种改进的轻量级 YOLOv4 神经网络模型方法。具体的, 为提高检测精度, 在特征提取网络中增加了轻量级注意力机制; 为提高检测速度, 选用轻量级 Mobilenet-v3 网络作为 YOLOv4 的骨干网, 同时采用了深度可分离卷积代替了传统 YOLOv4 模型中的多尺度特征金字塔结构普通卷积。实验结果表明, 改进的轻量级 YOLOv4 模型能有效检测出输电线路中的鸟巢, 平均检测精度(AP)值达到 97.56%, 相较于 YOLOv4 模型提高了 1.80%, 检测速度达到 YOLOv3 的 45.8 倍, 是 YOLOv4 的 56.6 倍, 具有较强的实际应用价值。

[关键词] 深度学习; 缺陷检测; YOLOv4; 注意力机制

[中图分类号] TP391.4 **[文献标识码]** A

高压输电线路多处于深山、林区、旷野等环境保护良好、水源丰富、鸟类活动频繁的区域, 鸟类经常会在输电线路的杆塔或绝缘子上方筑巢而形成鸟害, 严重威胁输电线路的安全运行, 定期清除高压输电线路的鸟巢非常重要。高压输电线路地域复杂, 人工检测非常不便, 随着电网建设的扩大, 新型的无人机巡检方式逐步取代人工巡检。

目前输电线路缺陷检测方法包括传统的图像处理和基于深度学习两类方法, 传统的图像处理方法^[1-3]缺乏鲁棒性、仅能识别和定位一种部件。随着深度学习运用在线路检测中, 基于深度学习输电线路检测领域也取得了一系列成果^[4-10], 然而目前基于深度学习的输电线路缺陷检测模型的参数量和计算量比较大, 检测速度较慢。

针对上述问题, 本文提出一种改进的轻量级卷积神经网络 YOLOv4 模型。针对 YOLOv4 在自建输电线路鸟巢数据集中检测硬件要求高和参数量过大的问题。在 YOLOv4 的基础上, 将主干特征网络由 CSPDarknet-53 替换成 Google 的轻量级 Mobilenetv3 网络, 大大降低网络参数量, 减少模型计算量, 使其可以无需昂贵的硬件用于输电线路鸟巢实时检测和通过远程控制无人机捕获的实时视频进行实时检测和本地化。其次在加强特征提取网络 PANet 中增加最新的轻量级 Coordinate Attention 注意模块, 数据集实验结果显示, 改进后的轻量级

YOLOv4 算法模型更小, 检测速度也有很大提升。

1 改进 YOLOv4 目标检测算法

改进 YOLOv4 模型融合了轻量级网络和双重注意力机制, 其总体结构如图 1 所示, 引入 MobileNetv3 轻量级网络模型, 用它替换 YOLOv4 模型的主干网络 CSPDarknet53, 可在不损失精度的前提下, 有效减少模型参数量, 提高模型检测速度。针对输电线路鸟巢目标被遮挡、多小目标的特点, 在 YOLOv4 特征融合网络的最后三个分支中引入 CA 模块, 以提高模型的检测精度。相较于通道注意力通过将输入进行二维全局池化转化为单个特征向量, CA 模块利用两个一维全局池化操作将沿垂直和水平方向的输入特征分别聚合为两个单独的方向感知特征图。然后将这两个方向感知特征图编码为两个注意力图, 这样做的优点是每个注意力图都可以沿着一个空间方向捕获长程依赖, 沿着另一个空间方向保留精确的位置信息。两个注意力图接着被乘到输入特征图上来增强特征图的表示能力。相比轻量级网络上的注意力方法, CA 存在以下优势: 首先, 它不仅能捕获跨通道的信息, 还能捕获方向感知和位置感知的信息, 这种注意力能帮助模型更加精准地定位和识别感兴趣的目标, 通过集成 CA 模块, 赋予特征图区分空间方向的能力, 使模型更加关注包含重要信息的目标区域, 抑制不相关信息, 提高目

标检测的整体精度。

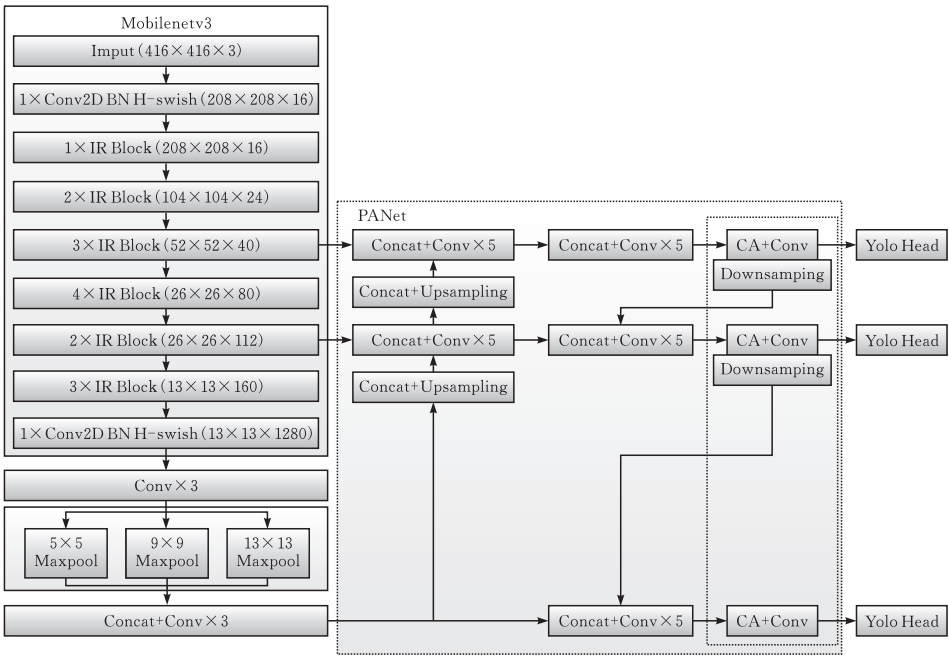


图 1 模型总体网络结构

1.1 轻量级骨干特征提取网络

图 2 中,卷积层由卷积层、批处理归一化层和 H-swish 激活函数组成。MobileNetv3 将普通卷积替换成深度可分离卷积,大大减少了参数,在不降低检测精度的同时减少了计算量,实现了速度与精度的完美结合。在训练过程中,图像被拉伸和缩放到 416×416 ,然后发送到卷积神经网络。经过 5 次 $3/2$ 卷积(卷积核大小为 3×3 ,步长为 2),缩减为 13×13 ,并选择 52×52 、 26×26 、 13×13 三个尺度输出 feature map。

1.2 基于 MobileNetv3 的 yolov4 网络结构改进

MobileNet 网络可分为 MobileNet-v1^[11]、MobileNet-v2^[12] 和 MobileNet-v3^[13] 三种类型。MobileNet-v3 分为大版本和小版本。在 ImageNet 分类任务中,大版本比小版本的准确率高 1.4%。但是大版本的检测速度与小版本相比降低了 10%。为保证实时检测和本地化需求,使用小版的 MobileNet-v3 取代 CSPDarkNet53 作为改进 YOLOv4 的骨干特征提取网络。

MobileNet-v3 在瓶颈层增加了挤压激励结构,将膨胀层通道修改为 $1/4$,在不增加计算时间的情况下提高检测精度。在 MobileNet-v2 中,平均池之前的 1×1 卷积层提高特征图的维数。然而,在 MobileNet-v3 中,特征图首先使用平均池化将其缩减为 1×1 。然后增加特征图的维数,将计算量减少 49 倍,提高了特征图的计算速度。图 1 显示了将 MobileNetv3 与 YOLOv4 融合的轻量级网络结构。为进一步降低网络的参数,将加强特征提取网络 PA-

Net 中的普通 3×3 卷积替换成深度可分离卷积。

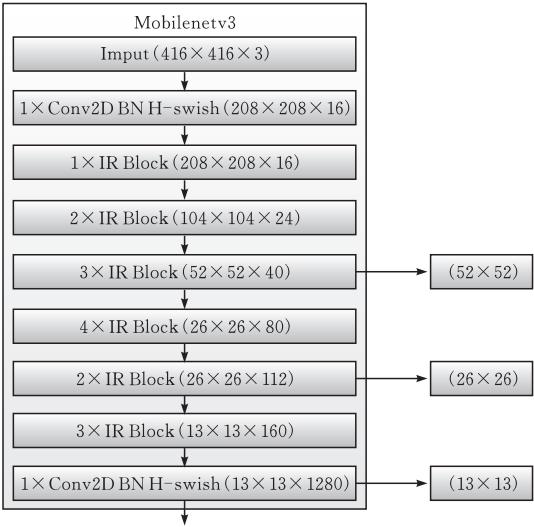


图 2 MobileNetv3 网络结构

1.3 特征融合网络

在 MobileNet-v3 特征提取网络后,使用特征融合网络 PANet 和 SPP 对提取的特征进行融合,以提高模型的检测能力。SPP 模块的作用是使卷积神经网络的输入不受固定大小的限制,在不降低网络运行速度的同时,可以增加接受域,有效分离重要的上下文特征。SPP 模块位于特征提取网络 MobileNetv3 之后。SPP 网络使用四种不同的最大池化比例来处理输入特征图。池化的核心大小为 1×1 、 5×5 、 9×9 、 13×13 ,而 1×1 相当于不经过处理,对四个 feature map 进行 concat 操作。最大池化采用填充操作,移动步长为 1,池化层后 feature map 的大小不变。

在 SPP 之后,使用 PANet 进行参数的聚合。与 YOLOv3 的特征金字塔结构不同的是,PANet 在自顶向下的特征金字塔之后增加了自底向上的路径增强结构,该结构包含两个 PAN 结构。并对 PAN 结构进行了改性。原 PAN 结构采用了一种快捷连接方式将下采样特征图与深度特征图融合,输出特征图的通道数量保持不变。修改后的 PAN 使用 concat 操作连接两个输入特征映射,并合并两个特征映射的通道号。自上而下的特征金字塔结构传达强烈的语义特征,自下而上的路径增强结构充分利用浅层特征传达强烈的定位特征。PANet 可以充分利用浅层特征,并针对不同的检测器水平,对不同的主干层进行特征融合,进一步提高特征提取能力,提高检测器性能。

2 基于 Coordinate Attention 的 yolov4 网络结构改进

Coordinate Attention^[14] 块可以看作是一个计算单元。CA 模块通过精确的位置信息对通道关系和长程依赖性进行编码,具体操作分为 Coordinate 信息嵌入和 Coordinate Attention 生成 2 个步骤。其模块结构见图 3。

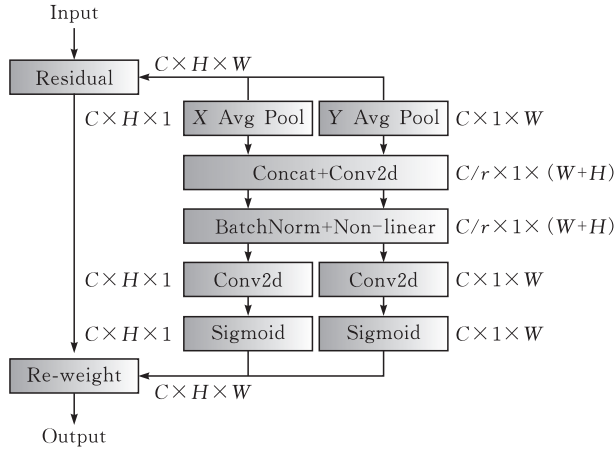


图 3 CA 模块结构

2.1 Coordinate 信息嵌入

为了促使注意力模块能够捕捉具有精确位置信息的远程空间交互,本文按照公式(1)全局池化分解为公式(2)和公式(3)的两个一维特征编码操作:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (1)$$

其中 z_c 是与第 c 个通道相关联的输出, H 和 W 为图像的高度和宽度, $x_c(i, j)$ 表示高度为 i 宽度为 j 的 c 通道输入,式(1)相当于将 $H \times W \times C$ 的输入转换成 $1 \times 1 \times C$ 的输出。式(2)和式(3)分别使用一个尺寸为 $(H, 1)$ 和 $(1, W)$ 的池化核分别沿着水平坐标方向和垂直坐标方向对每个通道进行编码。

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (2)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (3)$$

其中 $z_c^h(h)$ 表示 c 通道高度为 h 的输出, $z_c^w(w)$ 表示 c 通道宽度为 w 的输出, $x_c(h, i)$ 表示 c 通道高度为 h 的输入, $x_c(j, w)$ 表示 c 通道宽度为 w 的输入。公式(2)和公式(3)相当于将 $H \times W \times C$ 的输入转换成 $H \times 1 \times C$ 和 $1 \times W \times C$ 的输出。

式(2)和式(3)这 2 种特征变换利用两个一维全局池化操作,将沿垂直和水平方向的输入特征分别聚合为两个单独的方向感知特征图,然后将这两个方向感知特征图编码为两个注意力图。这种注意力机制与在通道注意力方法中只产生单一的特征向量的 SE Block^[15] 有本质不同。这样做的好处是每个注意力图都可以沿着一个空间方向捕获长程依赖,沿着另一个空间方向保留精确的位置信息。

2.2 Coordinate Attention 生成

该部分将上面的变换后的两个特征图进行级联操作,然后使用一个共享的 1×1 卷积进行变换,卷积变换函数表述如下:

$$f = \delta(F_1([z^h, z^w])) \quad (4)$$

式中 δ 为非线性激活函数,变换生成的 $f^h \in R^{C/r \times (H+W)}$ 是对空间信息在水平方向和竖直方向的中间特征图,这里 r 表示下采样比例,用来控制模块的大小。然后沿着空间维度将 f 分解为 2 个单独的张量 $f^h \in R^{C/r \times H}$ 和 $f^w \in R^{C/r \times W}$ 。利用另外 2 个 1×1 卷积变换 F_h 和 F_w 分别将特征图 f^h 和 f^w 变换为和输入 X 具有相同的通道数,得到:

$$g^h = \sigma(F_h(f^h)), g^w = \sigma(F_w(f^w)) \quad (5)$$

式中 σ 是 sigmoid 激活函数。为了降低模型的计算复杂性和计算开销,这里可以使用适当的通道缩减比 r (本文取 $r=16$) 来减少 f 的通道数。然后对输出 g^h 和 g^w 进行一次拓展,分别作为注意力权重。最后,CA 模块的最终输出可以表述为公式(6):

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

式中 $y_c(i, j)$ 表示 c 通道高度为 i 宽度为 j 的输出, $g_c^h(i)$ 表示 c 通道高度为 i 的注意力权重, $g_c^w(j)$ 表示 c 通道宽度为 j 的注意力权重。

在图 1 中,假设输入图像的大小是 $416 \times 416 \times 3$,以第一个 CA 模块为例,输入特征地图大小是 $52 \times 52 \times 256$,沿 X 和 Y 全局平均池化后,得到两个特征图大小 $52 \times 1 \times 256$ 和 $1 \times 52 \times 256$,然后经过一个拼接与共享的权重,维数变为 $1 \times 104 \times 16$,降维系数是 16,然后进行标准化和非线性激活,然后再分离并扩张维度到 $52 \times 1 \times 256$ 和 $1 \times 52 \times 256$,得到了两个激活后的特征图谱,将二者分别接一个

Sigmoid 激活函数,将输入的特征图分别与两个激活后的特征图谱相乘,得到尺寸为 $52 \times 52 \times 256$ 的带有位置信息的空间选择性注意力特征图。

3 实验分析

3.1 输电线路鸟巢数据集

本文数据采集自天津地区特高压与高压输电线路杆塔照片。这些图像来自不同电压等级、不同杆塔类型的输电线路。本文为了保证训练数据集的各种类型数据均衡,在标注数据之前从原始的数据中按比例抽取不同类型的杆塔。然后对数据集中不同线路的图片进行标注。选取了遮挡、多云、晴天、背光和强光条件下的 1500 幅图像作为数据集,保证了数据集的多样性。将标注的数据随机选取 72% 作为训练集、14% 作为验证集、14% 作为测试集。根据 VOC 数据集生成数据集的格式。部分输电线路鸟巢目标数据见图 4。



图 4 部分输电线路鸟巢数据集

3.2 实验环境

为了进一步加快网络训练速度,引入迁移学习技术,将预先训练好的模型加载到 voc 数据集上,对输电线路鸟巢数据集进行训练,模型的训练环境和测试环境是相同的。本文中使用的实验装置硬件由 windows10 操作系统,Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz, 32GB 的内存, NVIDIA TITAN RTX 32GB 显存的显卡。软件版本为 Pyc-harm2020, pytorch 框架, Python3.6.8, Opencv3.4.3, CUDA 10.0 和 CUDNN 7.5。

训练网络参数:批量大小设置为 16,初始学习率设置为 0.001,迭代次数 200,学习速率衰减 0.1,固定图像尺寸 416×416 。

3.3 输电线路鸟巢目标检测性能比较

为了验证改进的 YOLOv4 网络的有效性,对原始 YOLOv4 训练模型与改进的 YOLOv4 网络模型进行了对比实验。原来的 YOLOv4 训练参数与改进后的 YOLOv4 训练参数一致。采用常用的目标检测评估 mAP 对改进前后的模型进行比较。在目标检测任务中,根据交集 (IOU) 来判断目标是否被

成功检测,模型的预测框与地面真值框交集并的比例为 IOU。对于数据集中某一类型的目标,假设阈值为 α ,当预测框和 ground truth 框的 IOU 均大于 α 时,表示模型预测正确;当预测框和 ground truth 框的 IOU 小于 α 时,表示模型预测错误。 TP 是正确预测的阳性样本数, FP 是错误预测的阴性样本数, FN 是错误预测的阳性样本数, TN 是正确预测的阴性样本数。精准率和召回率的计算公式如下:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

AP 值通常用来作为目标检测模型性能的一个重要评价指标。AP 值为 P-R 曲线下的面积,以查全率为 x 轴,以精度为 y 轴。AP 表示一个模型在某一类别中的准确性。mAP 表示一个模型在所有不同类别的平均准确率,可以衡量网络模型在所有类别中的性能。其中 N 为检测到的类别个数。

$$\text{mAP} = \frac{1}{N} \sum_i \int_0^1 P(R) dR$$

本文算法的检测速度用 FPS (Frame Per Second) 来评估。它表示每秒可以处理的帧数。相同的模型在不同的硬件配置下有不同的处理速度。因此,本文在比较检测速度时使用相同的硬件环境。其中 t_{avg} 为处理一张图片所用的平均时间。

$$\text{FPS} = \frac{1}{t_{avg}}$$

PMs 是神经网络模型各层计算所用的参数。卷积层和全连通层的参数按照通用标准计算。对于 m 层的卷积层,假设输入通道数为 C_{in} ,卷积核大小为 $h \times w$,输出通道数为 C_{out} ,每个输出通道有 $C_{in} \times h \times w$ 参数,加上一个偏移参数。对于 k 层的全连接层,假设输入通道数为 F_{in} ,输出通道数为 F_{out} ,每个输出通道都有 F_{in} 参数,加上一个偏移参数。PMs (Parameters) 指总参数:

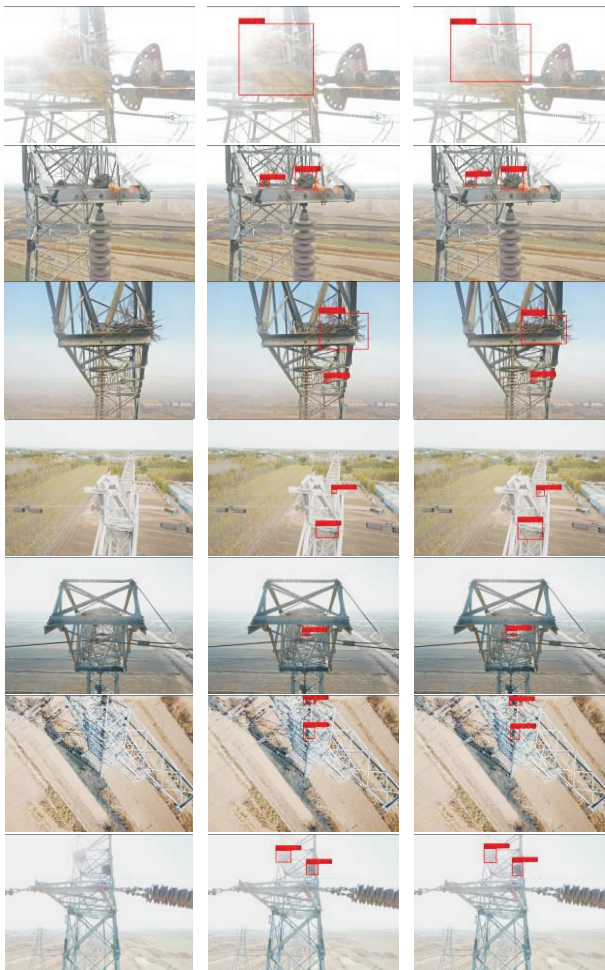
$$\text{PMs} = \sum_1^m (C_{in} \times h \times w + 1) \times C_{out} + \sum_1^k (F_{in} + 1) \times F_{out}$$

将测试集的数据发送到训练的目标检测模型中,选择不同的模型进行实验比较(表 1)。由表 1 可以看出,结合 MobileNetv3 和 CA 算法的 YOLOv4 的 mAP (IoU = 0.5) 得到改进,与原始 YOLOv4 算法相比,其 mAP 提高 1.80%。由于增加了三个 CA 模块,模型的体积变大,从 32.43 MB 增加到 34.80 MB,平均时间增加了 0.29 ms,模型权重增加了 2.37 MB。注意力机制增加了特征图的计算量,提高了检测精度,降低了检测速度,但速度仍然满足实时性要求。相较于原始 YOLOv4 算法,本

文漏检率下降了 13%,召回率也提升了 13.83%。

表 1 不同检测模型的性能比较

不同算法	平均准确率	平均时	权重文件	召回	漏检	FPS/
	AP/%	间/ms	大小/MB	率/%	率/%	(帧·s ⁻¹)
YOLOv3	91.25	136.13	246.5	36.57	48	7.30
YOLOv4	95.83	168.10	256.2	37.82	45	5.94
YOLOv4+ MobileNet-v2	95.27	5.92	48.62	52.38	42	168.91
YOLOv4+ MobileNet-v3	95.12	2.68	32.43	51.65	40	373.13
YOLOv4+ MobileNet-v3+CA	97.56	2.97	34.80	51.65	32	336.70



(a)原始输入 (b)YOLOv4 检测结果 (c)改进的YOLOv4结合 CA算法的检测结果

图 5 比较 YOLOv4 与改进的轻量级 YOLOv4 算法对输电线路鸟巢目标的检测结果

在图 5 中,第一列(a)是原始输入图片,第二列(b)是 YOLOv4 检测结果,第三列(c)是改进的 YOLOv4 结合 CA 算法的检测结果。在第一行图片中,YOLOv4 在高度强光照射环境下对鸟巢进行识别的置信度为 0.93,改进后的 YOLOv4 算法将鸟巢识别的置信度提升为 1;从第二、三、四、五行图片中可以明显看出,改进后的 YOLOv4 算法对小目标的检测比原算法更准确。其中,第二、四排鸟巢目标具有较多的背景环境干扰,改进后的 YOLOv4 具有更强的鲁棒性。由第六、七行图片可以看出,改进的 YOLOv4 算法成功检测到在逆光环境和背景干扰

条件下被遮挡鸟巢小目标;从第五和第六行也可以明显看出,当鸟巢目标小且被遮挡时,改进的 YOLOv4 算法对小目标的检测精度比原 YOLOv4 算法高。

4 结论

本文在 YOLOv4 的基础上,将主干特征网络由 CSPDarknet-53 替换成 Google 的轻量级 MobileNetv3 网络,大大降低了网络参数量,使其能应用在移动端和嵌入式平台上。同时在加强特征提取网络 PANet 中增加了最新的轻量级 Coordinate Attention 注意模块,使神经网络更加关注包含重要信息的目标区域,抑制无关信息,提高实时检测精度。实验结果表明,改进后的轻量级 YOLOv4 模型在目标检测任务中比原 YOLOv4 模型精度更高,权重文件大小为原来的 1/7,mAP50 提高了 1.80%,检测执行速度可以满足实时性检测要求,验证了这种改进算法的有效性。

[参 考 文 献]

[1] 焦红.直升机巡检输电线路图像中防振锤的识别定位[D].大连:大连海事大学,2011.

[2] 段旺旺,唐鹏,金炜东,等.基于关键区域 HOG 特征的铁路接触网鸟巢检测[J].中国铁路,2015(08): 73-77.

[3] LU J, XU X, XIN L, et al. Detection of bird's nest in high power lines in the vicinity of remote campus based on combination features and cascade classifier [J]. IEEE Access, 2018 (99):1-1.

[4] CHEN M Y, XU C. Birds nest detection method on electricity transmission line tower based on deeply convolutional neural networks[C].// 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2020.

[5] JIANG H, HUANG W, CHEN J, et al. Detection of bird nests on power line patrol using single shot detector[C].// 2019 Chinese Automation Congress (CAC). IEEE, 2020.

[6] BERG AC, FU CY, SZEGEDY C, et al. SSD:单发多

盒检测器:,10.1007/978-3-319-46448-0_2[P]. 2015.

[7] 王纪武,罗海保,鱼鹏飞,等.基于 Faster R-CNN 的多尺度高压塔鸟巢检测[J].北京交通大学学报,2019,43(05):37-43.

[8] 丁建,黄陆明,朱迪锋,曹浩楠.面向高空塔架鸟巢检测的双尺度 YOLOv3 网络学习[J].西安理工大学学报,2021,37(02):253-260.

[9] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv, 2018, 1804: 02767.

[10] W DONG, L WU, Q WANG, et al. An automatic detection method of bird's nest on electric tower based on attention full convolutional neural networks[C].//2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD), 2021:304-308, doi: 10.1109/ICAIBD51990.2021.9459088.

[11] HOWARD A.G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL].(2022-02-10) [2020-07-03]. <https://arxiv.org/abs/1704.04861>.

[12] SANDLER M, HOWARD A, ZHU M, et al. MobileNetv2: Inverted residuals and linear bottlenecks[C].// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Press,2018:4510-4520.

[13] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C].// In Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea: IEEE Press, 2019:1314-1324.

[14] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C].// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,2021: 13 713-13 722.

[15] JIE H, LI S, GANG S, et al. Squeeze and excitation networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 42(08): 2011-2023.

Improved YOLOv4 Transmission Lines Bird's Nest Detection Method

WU Minghu,LEI Changding, LIU Cong

School of Electrical and Electronic Engineering , Hubei Univ. of Tech., Wuhan 430068, China

Abstract: Transmission line defect detection methods based on deep learning often need expensive hardware support to achieve real time and high precision technical requirements, so it is difficult to be popularized on a large scale. To solve this problem, an improved lightweight YOLOv4 neural network model is proposed. Specifically, in order to improve the detection accuracy, a lightweight attention mechanism is added to the feature extraction network. In order to improve the detection speed, lightweight Mobilenet v3 network is selected as the backbone network of YOLOv4, and deep deprivable convolution is adopted to replace the traditional multi scale characteristic pyramid convolution in YOLOv4 model. The experimental results show that the improved lightweight YOLOv4 model can effectively detect nests in transmission lines, and the average detection accuracy (AP) value reaches 97.56%, which is 1.80% higher than that of YOLOv4 model. The detection speed is 45.8 times that of YOLOv3 and 56.6 times that of YOLOv4, which has strong practical application value.

Keywords: deep learning; defect detecting; YOLOv4; Attentional mechanism

[责任编辑: 张岩芳]