

文章编号: 2096-1618(2023)02-0129-07

基于视频识别的气象观测场设备监控技术研究

魏春梅, 马尚昌, 卢会国, 黄胃建
(成都信息工程大学电子工程学院, 四川 成都 610225)

摘要:针对观测场长期无人监管,存在设备被人破坏和偷盗,且目前还没有对观测场设备进行实时监控的有效算法和系统等问题,比较目前典型的目标检测算法,将其应用于观测场设备的检测和识别。根据研究目标搭建数据获取平台,研究数据预处理算法,比较经典目标检测算法 SSD 和 YOLOv5 两种模型在气象观测设备识别中的效果。两种模型实验对比结果表明,SSD 的识别准确率为92.09%,但训练模型速度较快;YOLOv5 的识别准确率为95.82%,收敛很快,且识别结果较佳。

关键词:目标检测;SSD;YOLOv5;数据增强;视频识别

中图分类号:TP391

文献标志码:A

doi:10.16836/j.cnki.jcuit.2023.02.001

0 引言

在《地面气象观测规范》^[1]中,对观测场的仪器安放、场内道路等有明确要求,其目的是为全面掌握环境现状,对观测环境进行客观评价,保证观测项目不因环境的影响而产生偏差。随着周边的植物生长和城市建设,对探测环境造成一定的影响^[2]。此外,气象站一般建在相对较偏远和人员稀少的地区,设备时常被偷盗和损坏。因此,识别与定位观测场设备,对观测场设备监控具有深远的研究意义。

国内外对图像的特征及其应用有大量的研究。图像特征提取^[3]是计算机视觉中最基础的问题,但同时也是最难的问题。目标检测就是判别图像中的目标物类别,并且用边框在图像中标记出位置。当前的目标检测方法主要有两种:一是传统的,二是基于深度的。传统的目标检测分为选择区域、特征提取和分类器分类3步。但是传统的目标检测算法有很大的局限性,如不能适应变化的光照强度,算法的时间复杂度高。

随着深度学习的不断发展,卷积网络可以提取更深、更好的特征层,相较于传统的特征提取,大大地提高了检测的准确性和实时性。基于卷积神经网络的目标检测算法,根据检测的思路不同和有无锚点可以将其划分为基于锚点的和无锚点的目标检测算法^[4]。基于锚点的目标检测算法,在特征图的每个特征点设置不同长宽比的先验框,再进行筛选和调整先验框获得最终的预测框。根据任务不同,目标检测算法可以分为基于区域的和基于回归的^[5]。基于回归的典型

算法分别有 YOLO (you only look once) 和 SSD (single shot multibox detector)。2016 年, Liu 等^[6]首次提出了 SSD 模型,在 VGG-16 之后去掉全连接层,采用多个卷积层来获得多个尺度的特征层,针对不同尺度的特征图设置不同的长宽比的先验框,使有重叠区域的目标具有较好的检测精度。2017 年, Jeong 等^[7]提出 RSSD (rainbow single shot detector) 算法,通过下采样和上采样相结合,对不同特征层进行信息融合,扩增特征图的数量,有效解决了 SSD 存在的问题。YOLO 由 Redmon 等^[8]在 2015 年提出,是首个引入深度学习后的目标检测算法。2016 年, Redmon 等^[9]提出 YOLOv2, 在 3×3 的卷积核中使用 1×1 卷积进行特征压缩,并使用全局平均池化来预测,在每一个卷积层加入批量归一化层 (batch normalization, BN), 从而解决了反向传播过程中的梯度消失和爆炸问题。2018 年, Redmon 等^[10]提出 YOLOv3, 采用更深的 DarkNet-53 提取更细粒的特征信息,采用 FPN 结构实现了多尺度预测,使用 1×1 卷积和 Logistic 激活函数替代 Softmax 分类层,更有效地进行数据拟合。2020 年, Bochkovskiy 等^[11]提出 YOLOv4, 采用先进的 CSPDarknet53 进行特征提取,用 SPP+PaNet 进一步增强特征的表达能力,同时将新提出的 Tricks 应用到网络的改进中,获得较高的检测精度和实时性。YOLOv5 是由 Ultralytics LLC 公司在 2020 年 5 月提出,每秒最多能处理 140 帧图像,满足视频图像实时检测的需求。

本文主要研究基于回归的两种典型目标检测算法,利用 SSD 和 YOLOv5 对于设备数据集进行训练,对比得到识别效果较好的设备识别模型,搭建目标检测平台观察设备识别的准确率,对模型做进一步的检验。

1 获取数据及预处理

在观测场内搭建了数据采集平台^[12],拍摄大量不同光线、角度、大小、位置的图片作为数据集。在观测场的4个角落各放置一个摄像头,离地高度为150 cm,以对角线的朝向放置。在室内和观测场中心处分别放置一个路由,使4个摄像头都能连上同一局域网。PC端通过局域网,访问实时流传输协议(real time streaming protocol, RTSP)获取视频数据,平台搭建如图1所示。

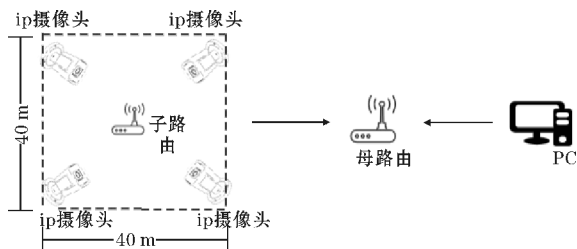


图1 平台搭建

百叶箱与周围物体以及环境有着难以分割的关系,尤其是背景中草的颜色越深,识别的效果越差,数据样本变化较少,特征不够突出,所以采用数据增强^[13]的方法来增强数据集。通过数据增强得到合适的训练样本,降低模型对参数的依赖,从而提高模型的鲁棒性和泛化能力。以百叶箱为例,通过加噪声、改变光线、翻转、切割、旋转、改变亮度来增强数据集。数据增强效果如图2所示。

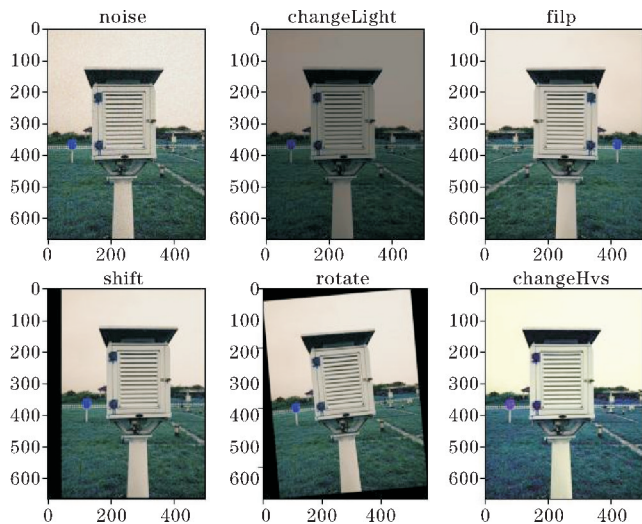


图2 数据增强效果图

此次训练数据的种类比较单一,并且在制作数据集的时候为了方便,所有的数据都是按每一类的顺序依次排列,会出现过拟合和过度收敛,导致训练效果不理想。因此,打乱数据集在一定程度上可以减轻这种问题。

2 方法及原理

当前主流的目标检测算法有 One-Stage 和 Two-Stage。One-Stage 只需一个 Stage 产生物体的类别概率和位置坐标值。Two-Stage 将检测问题分为两步:第一步产生候选区域和目标大概的位置信息;第二步对候选区域进行分类和定位精确。目标检测的主要性能指标是准确度和速度,Two-Stage 在准确率上有优势,One-Stage 在速度上有极大的优势。而本文是利用网络摄像头对目标进行检测,所以更看重速度。One-Stage 最典型的算法就是 SSD 和 YOLOv5,所以本次实验将对这两种算法进行对比。

2.1 SSD 原理

SSD 是典型 One-Stage 方法,此方法主要是对图片的不同位置进行抽样,抽样时使用不同的尺度和长宽比,利用卷积神经网络(CNN)提取特征后进行分类和回归。这个过程只有一步,所以速度比较快,在较大图片训练时,优势比较大。SSD 提取不同尺度大小的特征图,以便识别大小不一的物体。SSD^[14]算法的基本框架如图3所示。

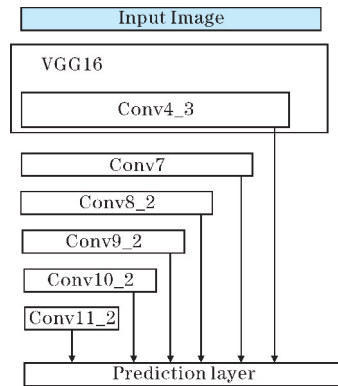


图3 SSD 算法框架图

SSD 模型主干网络采用的是 VGG-16,取代 VGG16 的全连接层为卷积层,并用一个卷积代替全连接层,再新增4个卷积层。输入 300×300 的图片,resize(改变大小不改变内容)到3通道的 shape;两次卷积网络和最大池化步长为2;三次卷积网络和最大池化步长为2;采用一次最大池化步长为1,用卷积取代全连接层,分别进行一次卷积核为3和1的卷积网络;此前的都是 VGG 的结构,在此基础上加上4个卷积层,前两层经过卷积核为1的卷积网络和一次步长为2的卷积网络,后两层经过卷积核为1的卷积网络和一次 padding 为 valid 的卷积网络。

特征提取过程如图4所示,共提取6个特征层来预测结果,每个特征层进行两次卷积操作。一次 num_

anchors×4 的卷积;用于预测该特征层上各单元的先验框的变化情况。一次 num_anchors×num_classes 的卷积;用于预测该特征层上各单元的每次预测对应的种类。num_anchors 指的是该特征层每个特征点所拥有的先验框数量,num_classes 指的是预测种类的数量。

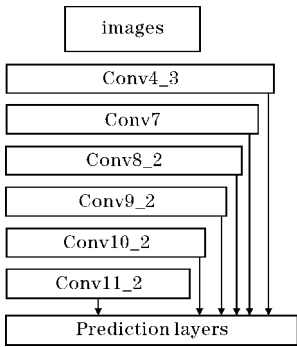


图4 特征提取过程

每个有效特征层都将图像分成长宽相对应的网络,网络参数如表 1 所示。Conv4-3 的特征层分为 38×38 个单元,各单元中心对应 4 个先验框,共有 5776 个先验框。各单元都有多个预测框,每个预测框都是针对这个方块(单元)本身,可以代表先验框的位置和大小信息,但是有一定的局限性。预测网络利用 num_anchors×4 的卷积结果,调整先验框,4 代表的是 c_x, c_y, h, w 的调整情况。其中 c_x, c_y 代表先验框与真实框中心的 x 和 y 轴偏移情况, h 和 w 代表真实框与先验框高宽的变化情况。预测框的中心就由单元的中心点加上 c_x 和 c_y ,高和宽是 h 和 w 调整先验框得到的,得到中心点和宽高就可以绘制预测框了。先验框示意图如图 5 所示。

表 1 网络参数

Name	Out-size	Prior_box_num	Total_num
Input_image	300×300		
Conv1	300×300		
Conv2	150×150		
Conv3	75×75		
Conv4_3	38×38	4	5776
Conv7	19×19	6	2166
Conv8	10×10	6	600
Conv9	5×5	6	150
Conv10	3×3	4	36
Conv11	1×1	4	4

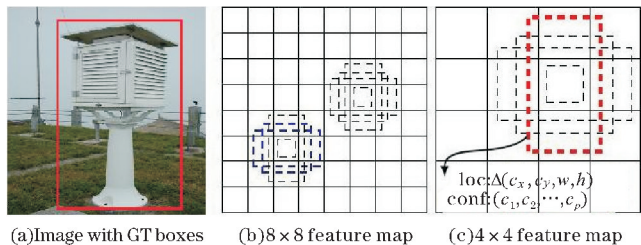


图5 先验框

针对每一个预测框进行得分排序和设置 NMS(非极大抑制筛选)筛选掉得分低的预测框,将得分高的预测框绘制在图片上。

2.2 Yolov5

Yolov5 主干网络采用的是 Focus 网络结构,其思想是在一张图像上每隔一个像素取一次值,获得 4 个独立的特征层,然后将 4 个独立特征层进行堆叠,则每个特征层的宽高信息就被堆叠起来,使输入通道扩大了 4 倍,由原来的 3 通道变成 12 通道。在 Yolo 中,采用的是 Mosaic(马赛克)数据增强,利用 4 张图片进行拼接实现图像增强,丰富目标检测的物体背景,在计算 BN 时同时计算 4 张图片的数据。Yolov5^[15] 的算法流程如图 6 所示。

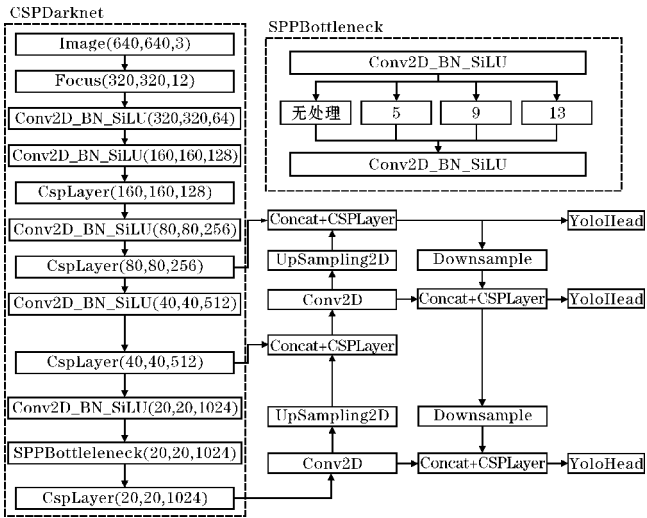


图6 Yolov5 算法流程图

主干网络采用的是 CSPDarknet,输入图像通过主干部分,得到 3 个有效特征层。CSPDarknet 网络由残差卷积构成,使用了残差网络 Residual。该网络中有两个残差卷积部分,一部分是主干网络,一次 (1,1) 的卷积和 (3,3) 的卷积;另一部分是残差边部分,不做任何处理,直接将主干部分的输入和输出结合。残差网络具有易于优化、能增大深度以提高精确度的优点,而其内部的残差块使用跳跃连接,有效地解决了由于深度神经网络中深度增大而导致的梯度消失问题。残差网络如图 7 所示。

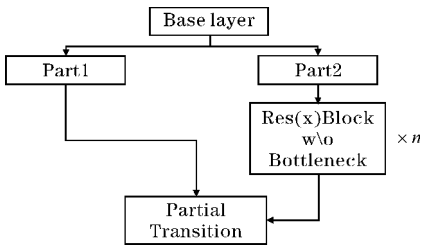


图7 残差网络

由图6可以看出,Yolo的激活函数是SiLU(Sigmoid和ReLU的改进型),具有无上界有下界、平滑、非单调函数的特性。与ReLU相比,该方法具有更好的深度神经网络性能,因此可以将其视为平滑的ReLU激活函数。

SPP(spatial pyramid pooling)结构,该结构最早是Yolov3的时候提出的,通过不同池化核大小的最大池化来进行特征提取,提高网络的感受野。感受野指的是在每个图层上,每个图层的像素点被映射到的面积。

通过构造FPN特征金字塔进行加强特征提取,如图8所示,Yolov5从3个有效特征层抽取用于识别的有效特征层,这些层分为位于中、下、底层的CspLayer,由图8可知,假设输入为(300,300),3个特征层的shape是 $feat1 = (80, 80, 256)$, $feat2 = (40, 40, 512)$, $feat3 = (20, 20, 1024)$ 。

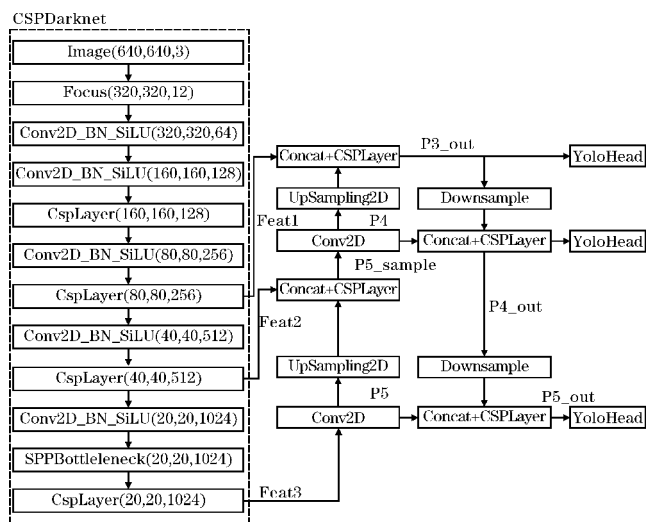


图8 FPN特征金字塔加强特征提取

FPN层的构建:

(1)将Feat3的特征层进行一次(1,1)的卷积,调整通道数获得P5,P5进行上采样,随后,将其与Feat2的特征层合并,再利用CSPLayer进行特征提取,获得P5_upsample,此时特征层为(40,40,512)。

(2)P5_upsample的特征层进行一次(1,1)的卷积,调整通道数获得P4,P4进行上采样,随后与Feat1的特征层合并,再利用CSPLayer进行特征提取,获得P3_out,此时特征层为(80,80,256)。

(3)P3_out的特征层执行一次(3,3)的卷积和下采样,下采样之后与P4叠加,再利用CSPLayer进行特征提取,获得P4_out,此时特征层为(40,40,512)。

(4)P4_out的特征层进行一次(3,3)的卷积和下采样,下采样后与P5叠加,再利用CSPLayer进行特征提取,获得P5_out,此时特征层为(20,20,1024)。

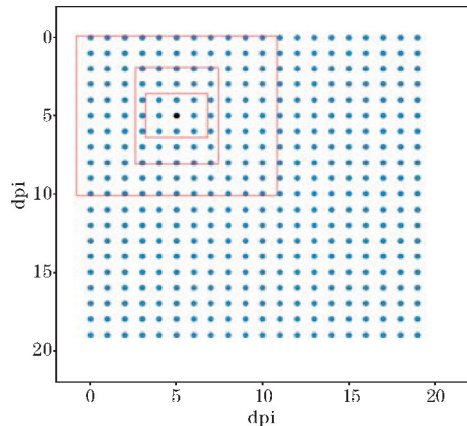
通过Yolo Head的计算,得到了预测的结果。将FPN特征金字塔获得3个加强特征层(20,20,1024),

(40,40,512),(80,80,256)传入YoloHead得到预测结果。对于每一个特征层,可以利用一个卷积调整通道数,最终输出的通道数与需要区分的种类个数相关,每个特征层上各特征点都有3个先验框。以本文数据集为例,种类为2种,最后的维度应该为 $21 = 3 \times 7$,这3个特征层的shape是(20,20,21),(40,40,21),(80,80,21)。最后的21拆分成3个7,对应3个先验框的7个参数,7个参数可以拆分成4+1+2。前4个参数用于判断每一个特征点的回归参数,回归参数调整后可以获得预测框;第5个参数用于判断每一个特征点是否包含物体;最后2个参数用于判断每一个特征点所包含的物体种类。以(20,20,3,7)这个特征层为例,相当于将图像划分为 20×20 个特征点,当某个特征点落在物体的对应框内,它就用于预测该物体。预测中心点如图9所示。图9中,蓝色的点代表 20×20 的特征点,对图9(a)黑点的3个先验框进行解码:

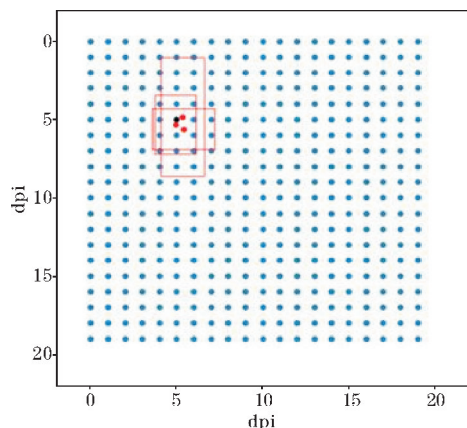
(1)预测中心点的计算,利用Regression预测结果前两个序号的内容对特征框的中心坐标进行偏移,偏移之后如图9(b)的3个红色点。

(2)预测框宽高的计算,利用Regression预测结果的后两个序号的内容求指数后获得预测框的宽高。

(3)此时就可以将获得的预测框绘制在图片上。



(a) 预测前



(b) 预测后

图9 预测中心点

为避免相同类型的预测框,既要用得筛选,选择满足 confidence 置信度的预测框,也要用非极大抑制操作,筛选出某一区域内同种类得分最大的框。在图像中找出比阈值更高的框,通过得分筛选来减少框的数量。对所有的种类进行循环,确保每个种类都能进行非极大抑制。按分数从大到小一次排序,去掉分数最大的框,然后根据其他的框进行比较,以排除重合程度太大的框。

3 实验结果

3.1 数据集的划分

数据集共有 10 类,共有 7429 张图片,分别用 SSD 和 YOLOv5 对数据集进行训练。数据集按照9:1的比例来划分,trainval_percent 被用于指定(训练集+验证集)与测试集的比例,train_percent 被用于指定(训练集+验证集)中训练集与验证集的比例,都按9:1来划分。

3.2 网络训练及性能分析

预训练权重在不同的数据集中具有普遍性,因为特征是通用的。SSD 模型输入的图片 resize 成 (300, 300) 的 shape,采用 ssd_weights 的预训练权重,先验框大小设置为 (30,60,111,162,216,264,315),迭代次数 200 次;YOLOv5 模型输入的图片 resize 成 (640,640) 的 shape,使用 YOLOv5_s 的预训练权重,先验框大小设置为 (10,13,16,30,33,23,30,61,62,45,59,119,116,90,156,198,373,326),迭代次数 200 次。训练种类为 10 种,包括:gc(自动观测站)、ACbox(电源箱)、FS(辐射计)、VA(vaisala)、SQ(水汽观测站)、ZHjc(综合集成器)、SY(酸雨观测仪)、JYgc(降雨自动观测仪)、BY(百叶箱)、person(行人)。

模型的评价指标有 4 个,Precision(准确度):在所有的预测都为正例的情况下,预测正确的比例;Recall(召回率):所有正例中被正确预测的比例;AP(平均精度):每一个类别的模型准确度。 F_1 : Precision 和 Recall 的加权调和平均。对于每一预测,分别统计 TP(预测正例是正确的),FP(错误地将其他类预测为本类),TN(预测反例是正确的),FN(该类别的标记是其他类别的标记),混淆矩阵见表 2。

表 2 混淆矩阵

预测结果	Positive(预测的正例)	Negative(预测的反例)
True(真)	TP	TN
False(假)	FP	FN

预测为百叶箱且确实是百叶箱的部分,占有所有预测为百叶箱的比例。计算公式为

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

预测为百叶箱且确实是百叶箱的部分,占有所有正例的比例。计算公式为

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

准确率和召回率的调和平均数。计算公式为

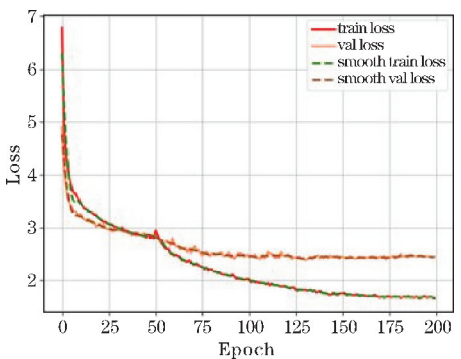
$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$
$$= \frac{2TP}{2TP+FN+FP}$$

平均准确度,将每一个阈值的准确率和召回率求得的乘积进行累积。计算公式为

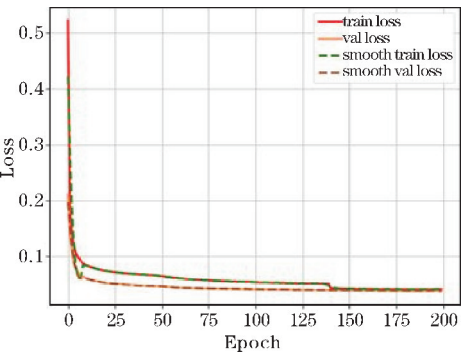
$$AP = \sum_{k=1}^N p(k) \Delta r(k)$$

其中: $p(k)$ 为不同阈值的准确率, $\Delta r(k)$ 为不同阈值的召回率。

通过分析模型的评价指标 Precision(准确度),AP(平均精度),Recall(召回率), F_1 来判断检测模型的性能。SSD 和 YOLOv5 训练集与验证集损失如图 10 所示。



(a) SSD



(b) YOLOv5

图 10 SSD 和 YOLOv5 训练集与验证集损失

3.3 设备识别结果

SSD 模型训练网络所产生的模型作为预测网络的

权值文件,依然是 VGG 为主干网络,每个单元都会有 4 个预测框,一张图片上会有很多个预测框,设置置信度为0.5,只有大于置信度的预测框才会被保留。根据非极大抑制框 NMS 来筛选掉低于0.45的预测框,anchors_size 不变,输出得分最高的预测框并绘制在图片上。Yolov5 预测是将 FPN 特征金字塔获得的 3 个加强特征金字塔传入 YoloHead。对先验框进行解码,计算预测中心点和预测框的宽高。经过得分筛选和非极大抑制,对每个种类进行循环,筛选出一定区域内同种类得分最大的框,与其他的预测框进行重合度计算,剔除重合度过大的框。选取摄像头拍摄的图片,用两种模型识别的效果如图 11、12 所示,可以看出 Yolov5 对于百叶箱的识别准确率要高于 SSD,并且识别较远百叶箱时,预测框的效果较佳。

10 类设备的 SSD 和 Yolov5 评价指标对比见表 3。



图 11 SSD 识别效果



图 12 Yolov5 识别结果

表 3 Yolov5 和 SSD 评价指标对比

设备简称	Precision/%		Recall/%		AP/%		F_1	
	Yolov5	SSD	Yolov5	SSD	Yolov5	SSD	Yolov5	SSD
gc	97.50	95.50	99.62	94.42	99.84	97.84	0.97	0.94
ACbox	90.86	84.76	98.92	86.02	98.92	93.42	0.99	0.98
FS	99.50	96.90	99.60	96.50	99.40	94.60	0.99	0.97
VA	96.22	91.72	92.44	88.04	99.42	93.92	0.95	0.92
SQ	96.19	92.19	98.67	87.47	98.36	94.76	0.96	0.93
ZHjc	97.33	90.53	98.65	92.65	99.93	97.93	0.98	0.97
SY	98.50	92.20	99.00	88.50	99.06	97.26	0.97	0.95
JYgc	98.90	95.90	97.83	89.73	99.91	98.41	0.99	0.97
BY	92.75	84.70	85.77	76.17	92.04	87.54	0.87	0.85
person	95.49	92.19	74.81	66.51	85.56	83.16	0.84	0.82

可以看出,Yolov5 的损失率更低,并且收敛速度更快。相较于 SSD,Yolov5 的 Precision 比 SSD 的高3.73, Recall 比 SSD 的高7.18,AP 比 SSD 的高3.46, F_1 比 SSD 的高2.02,但 Yolov5 的训练速度比 SSD 的慢几十个 epoch。

4 结束语

探讨了 SSD 和 Yolov5 两种目标检测算法,分析了模型结构,以十类设备为例,开展了实验对比,实验结果表明,使用 Yolov5 能得到较高的准确率和较低的损失函数,且收敛更快,相较于 SSD,Yolov5 的 Precision 比 SSD 的高3.73,准确度越高说明模型越好,Recall 比

SSD 的高7.18,AP 比 SSD 的高3.46, F_1 比 SSD 的高2.02,但 Yolov5 在模型训练时消耗的时间大于 SSD 模型的时间,因此,下一步将针对计算资源进行改进。

参考文献:

[1] GB 31221-2014,气象探测环境保护规范 地面气象观测站[S].
[2] 杨涛,张常亮,朱墨. 地面气象观测场实景监控系统设计[J]. 成都信息工程学院学报,2013,28(4):336-341.
[3] 索子恒. 图像特征检测与特征提取综述[J]. 产业创新研究,2022(4):33-35.

- [4] 张静,农昌瑞,杨智勇. 基于卷积神经网络的目标检测算法综述[J/OL]. 兵器装备工程学报: 1-12[2022-0602]. <http://kns.cnki.net/kcms/detail/50.1213.TJ.20220427.1656.035.html>.
- [5] 寇大磊,权冀,张仲伟. 基于深度学习的目标检测框架进展研究[J]. 计算机工程与应用,2019,55(11).
- [6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, et al. SSD: Single Shot MultiBox Detector[J]. CoRR, 2015, abs/1512.02325.
- [7] Jeong J, Park H, Kwak N. Enhancement of SSD by concatenating feature maps for object detection[J]. British Machine Vision Conference 2017, 1705:09587.
- [8] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. Computer Vision & Pattern Recognition, 2016.
- [9] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[J]. IEEE Conference on Computer Vision & Pattern Recognition, 2017:6517-6525.
- [10] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.
- [11] Bochkovskiy A, Wang C Y, Liao H. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020, 2004:10934.
- [12] 杨永辉,刘昌平,黄磊. 图像和视频分析在电力设备监控系统中的应用[J]. 计算机应用, 2010, 30(S1):281-284.
- [13] 朱晓慧,钱丽萍,傅伟. 图像数据增强技术研究综述[J]. 软件导刊, 2021, 20(5):230-236.
- [14] 谭红臣,李淑华,刘彬,等. 特征增强的 SSD 算法及其在目标检测中的应用[J]. 计算机辅助设计与图形学学报, 2019, 31(4):573-579.
- [15] 谈世磊,别雄波,卢功林,等. 基于 YOLOv5 网络模型的人员口罩佩戴实时检测[J]. 激光杂志, 2021, 42(2):147-150.

Research on Equipment Monitoring Technology of Meteorological Observation Field based on Video Recognition

WEI Chunmei, MA Shangchang, LU Huiguo, HUANG Weijian

(College of Electronic Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: In view of the long-term unsupervised observation field, the equipment is damaged and stolen, and there is no effective algorithm and system for real-time monitoring of observation field equipment, the current typical target detection algorithms are compared and applied to the detection and identification of observation field equipment. The data acquisition platform is built according to the research objectives, the data preprocessing algorithm is studied, and the effects of two classical target detection algorithms SSD and Yolov5 model in the recognition of meteorological observation equipment are compared. The experimental results show that the detection speed of SSD is faster, but the accuracy is only 92.09%, the accuracy of Yolov5 is 95.82% and the convergence is fast.

Keywords: target detection; SSD; Yolov5; data enhancement; video recognition