

doi:10.11835/j.issn.1000-582X.2020.11.007

基于 VGGNet 改进网络结构的多尺度大熊猫面部检测

何育欣^a, 郑伯川^{a,b}, 谭代伦^{a,b}, 刘丹^c, 蔡前舟^a

(西华师范大学 a.数学与信息学院; b.计算方法及应用软件研究所; c.计算机学院, 四川南充 637002)

摘要:大熊猫个体识别对研究大熊猫的种群数量非常重要,大熊猫面部检测是基于面部图像的大熊猫个体识别方法中的首要关键步骤。针对现有的大熊猫面部检测方法精确度不高的问题,提出基于 VGGNet-16 改进网络结构的多尺度大熊猫面部检测方法。首先,以 VGGNet-16 网络结构为基础,通过增加残差结构与 BN 层,降低卷积层通道数,并采用 LeakyRelu 激活函数等改进,构建一个新的特征提取主干网络。其次,将一个 3 尺度的特征金字塔网络结构与 SPP 结构结合用于目标检测。最后,使用深度分离卷积结构替代常规卷积结构。实验结果表明,提出的大熊猫面部检测方法在测试集上能够达到 99.48% 的 mAP,检测性能优于 YOLOv4。

关键词: VGGNet 网络结构; 大熊猫; 面部检测; 目标检测

中图分类号: TP183; TP391.4 **文献标志码:** A **文章编号:** 1000-582X(2020)11-063-09

Multi-scale giant panda face detection based on the improved VGGNet architecture

HE Yuxin^a, ZHENG Bochuan^{a,b}, TAN Dailun^{a,b}, LIU Dan^c, CAI Qianzhou^a

(a. School of Mathematics and Information; b. Institute of Computing Method and Application Software; c. School of Computer Science, China West Normal University, Nanchong, Sichuan 637002, P. R. China)

Abstract: Individual identification of giant pandas is very important for studying their population of them. Giant panda face detection is the first key step of giant panda individual identification method based on facial images. To solve the problem that the precision of the existing giant panda face detection methods are low, a multi-scale giant panda face detection method based on improved VGGNet-16 architecture was proposed in this paper. Firstly, based on the VGGNet-16 network architecture, a new feature extraction backbone network was constructed through certain improvements such as adding the residual block and BN (Batch Normalization) layer, reducing the channel dimensionality of convolution layer and adopting LeakyRelu active function as well. Secondly, a 3-scale feature pyramid network structure was combined with SPP (Spatial Pyramid Pooling) structure for object detection. Finally, the conventional convolution architecture was replaced with the depwise separation convolution architecture. Experimental results show that the

收稿日期: 2020-07-11

基金项目: 四川省科技计划资助项目(2019YFG0299); 四川省科技创新苗子工程(2019027); 西华师范大学基本科研项目(19B045)。

Supported by the Sichuan Science and Technology Program(2019YFG0299), the Sichuan Seedling Project of Science and Technology (2019027) and the Fundamental Research Funds of China West Normal University (19B045).

作者简介: 何育欣(1995—), 男, 硕士研究生, 主要从事深度学习研究, (E-mail)whmhyxhh@outlook.com。

通讯作者: 郑伯川, 男, 博士, 教授, 主要从事机器学习、深度学习、计算机视觉研究, (E-mail)zhengbc@vip.163.com。

proposed method can achieve 99.48% mAP(mean average recision) in the test dataset, and the detection performance is better than YOLOv4(You Only Look Once Version 4).

Keywords: VGGNet network structure; giant panda; face detection; object detection

大熊猫是中国的特有物种,也是珍贵的世界自然遗产。虽然,中国政府一直在采取各项措施保护大熊猫,但由于栖息地的丧失与破碎化等问题^[1],大熊猫的生存状况依然严峻。对野生大熊猫和野外放归大熊猫而言,对其进行个体监测掌握个体数量与分布情况对制定有效的大熊猫保护措施具有重要意义^[2]。通常跟踪和监测大熊猫个体的传统方法主要有:GPS 项圈法^[3]、基于经验的人工识别法^[4]、距离-咬节法与 DNA 法^[5-6]。虽然,这些方法能够识别大熊猫个体,但都存在一定的局限性,实施起来复杂且低效。近年来,研究人员主要采用红外相机等设备对大熊猫进行监测,获得了大量野生大熊猫图像。因此,迫切需要一种基于图像的大熊猫个体自动识别方法,实现对大熊猫个体的自动化监测。而大熊猫面部是较为理想的个体识别部位,一是面部图像数据易于采集,二是面部特征是天生的,具有稳定性与个体差异性。现有文献^[7]基于大熊猫面部特征设计大熊猫个体识别方法,但其使用的大熊猫面部图像是通过人工裁剪获得,并不是完整的大熊猫个体识别方法,缺少大熊猫面部自动检测的关键步骤。

有学者采用传统方法对大熊猫面部检测进行过相关研究^[8-10],由于是通过人工设置特征提取器,导致检测方法的泛化能力较弱,只适用于特定场景的检测。近年来,对深度卷积神经网络^[11-14]的研究取得了长足的进展,基于深度卷积神经网络的目标检测方法^[15-19]在精确度与召回率上都远远优于传统目标检测方法,成为了现阶段目标检测的主流方法。而采用基于深度卷积神经网络的目标检测方法对大熊猫面部进行检测的研究还鲜有报道。

相对于 COCO^[20](common object in context),ImageNet^[21]和 Pascal VOC^[22]数据集中对多类目标对象的检测,大熊猫面部检测任务中目标对象单一(只有 1 类对象),目标对象特征也相对明显。虽然直接采用现有的目标检测网络模型检测大熊猫面部能够达到较高的检测精确度和召回率,但是网络模型相对复杂,部分连接和参数显得多余。因此,根据大熊猫面部检测的特点,参照现有目标检测技术,有针对性地提出一种新的基于 VGGNet-16^[12]改进网络结构的多尺度大熊猫面部检测方法。该方法在光照不均、多姿态、遮挡等情况下都能较好地检测出大熊猫面部,其主要工作如下:

1)特征提取主干网络采用 VGGNet-16 改进网络结构。对 VGGNet-16 的改进包括引入残差结构增强前后层信息的传播,引入 BN 层^[23](batch normalization)保持训练的稳定性,以及采用 LeakyRelu 激活函数确保训练时梯度能够有效传递,并减少卷积层的通道数以获得更少参数的网络模型。

2)检测网络部分通过引入特征金字塔结构^[16, 17, 19]和空间金字塔结构^[24](spatial pyramid pooling, SPP)实现多尺度目标检测。特征金字塔结构通过不同尺度检测层可以检测不同大小的目标,并通过融合不同尺度的特征提高目标检测的精确度。SPP 结构则通过融合不同大小的感受野信息提高目标检测的精确度。

3)将特征提取主干网络和检测网络中的所有常规卷积结构替换成深度分离卷积结构^[25](depthwise separable convolution)。替换后的网络模型参数量大大减少,在不降低检测精确度与召回率的情况下,提高了检测速度,同时为今后将网络模型移植到边缘设备提供了可能。

1 大熊猫面部检测数据集

现阶段,由于没有公开的大熊猫面部检测数据集,所以需自行采集并制作大熊猫面部检测数据集。研究使用的大熊猫图像数据采集自成都大熊猫繁育研究基地,以及都江堰、雅安碧峰峡与卧龙 3 个中国大熊猫保护研究中心基地。数据集由 38 只大熊猫共 23 440 张图片组成,并使用 LabelImg 图像软件标记出大熊猫的面部区域。由于只对大熊猫面部进行检测,因此,所有标记信息只有 panda_face 这 1 类标签。图 1 展示了部分大熊猫面部标记图像。

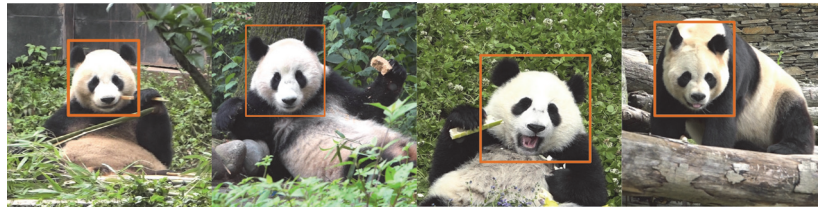


图 1 部分大熊猫面部标记图像
Fig. 1 Facial marker images of some giant pandas

2 大熊猫面部检测方案

基于 VGGNet 改进网络结构的多尺度大熊猫面部检测算法主要包括 2 个网络结构:特征提取主干网络和多尺度特征金字塔检测网络。特征提取主干网络采用 VGGNet-16^[12] 的改进网络结构,多尺度特征金字塔检测网络则借鉴了 YOLOv4^[17] 的多尺度检测结构。另外,为了使改进后的网络更加轻量化,将常规卷积结构替换为深度分离卷积结构^[25]。

2.1 VGGNet-16 改进网络结构

大熊猫面部检测模型使用 VGGNet-16 的改进网络结构作为特征提取主干网络,如图 2 所示。

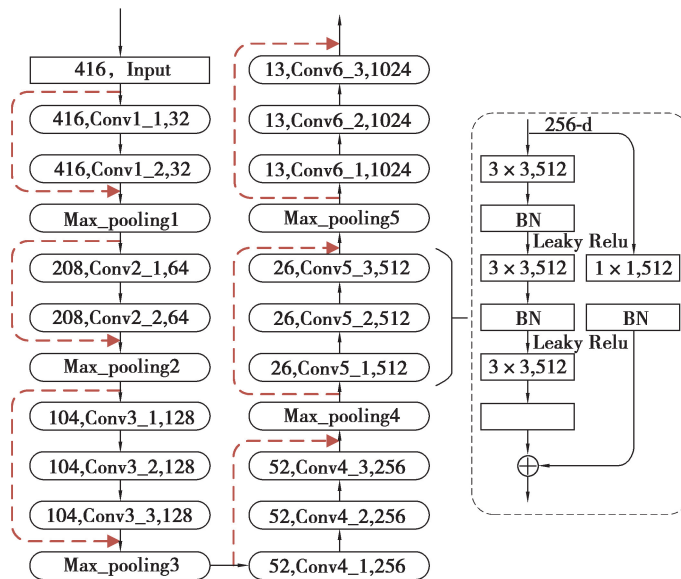


图 2 VGGNet-16 改进网络结构
Fig. 2 Improved network structure of VGGNet-16

原始 VGGNet-16 网络通过反复堆叠 3×3 大小的卷积核与 2×2 大小的最大池化层完成特征提取网络的搭建,具有泛化能力强的特性,在不同的图片数据集上都有良好的表现。原始 VGGNet-16 网络一共有 5 组卷积模块,为了提取出 3 种不同尺度的特征图用于目标检测,在第 5 个池化层后额外增加 1 组卷积模块 Conv_6,如图 2 所示。同时,由于大熊猫面部检测是单类别检测,类别数量少,造成原始网络中通道数冗余,所以,将网络浅层部分的 Conv_1、Conv_2、Conv_3、Conv_4 4 组卷积模块的通道数减小为原始网络的一半。改进后的网络共有 3 个用于目标检测的特征层,即:Conv4_3、Conv5_3、Conv6_3。当输入图像大小为 416×416 时,3 个特征层输出特征图的大小分别为 $52 \times 52 \times 256$ 、 $26 \times 26 \times 512$ 、 $13 \times 13 \times 1024$ 。

另外,借鉴 ResNet^[13] 的跨层连接结构,将网络结构中的特征图大小相同的相邻卷积层构成 1 个残差模块。改进后的网络结构一共包含 6 个残差模块,其中,第 5 个残差模块的具体结构如图 2 所示,其余 5 个残差模块结构与此类似。在残差模块中,每一个激活函数前都添加了 1 个 BN 层,用于提升网络收敛速度与训练稳定性。同时,采用 LeakyRelu 激活函数以减少静默神经元的出现,确保梯度可以有效通过整个网络架构。

2.2 多尺度特征金字塔检测网络结构

特征提取主干网络需要结合检测网络才能实现对目标的检测输出。检测网络部分则借鉴了 YOLOv4 的多尺度特征金字塔检测网络结构,如图 3 所示。

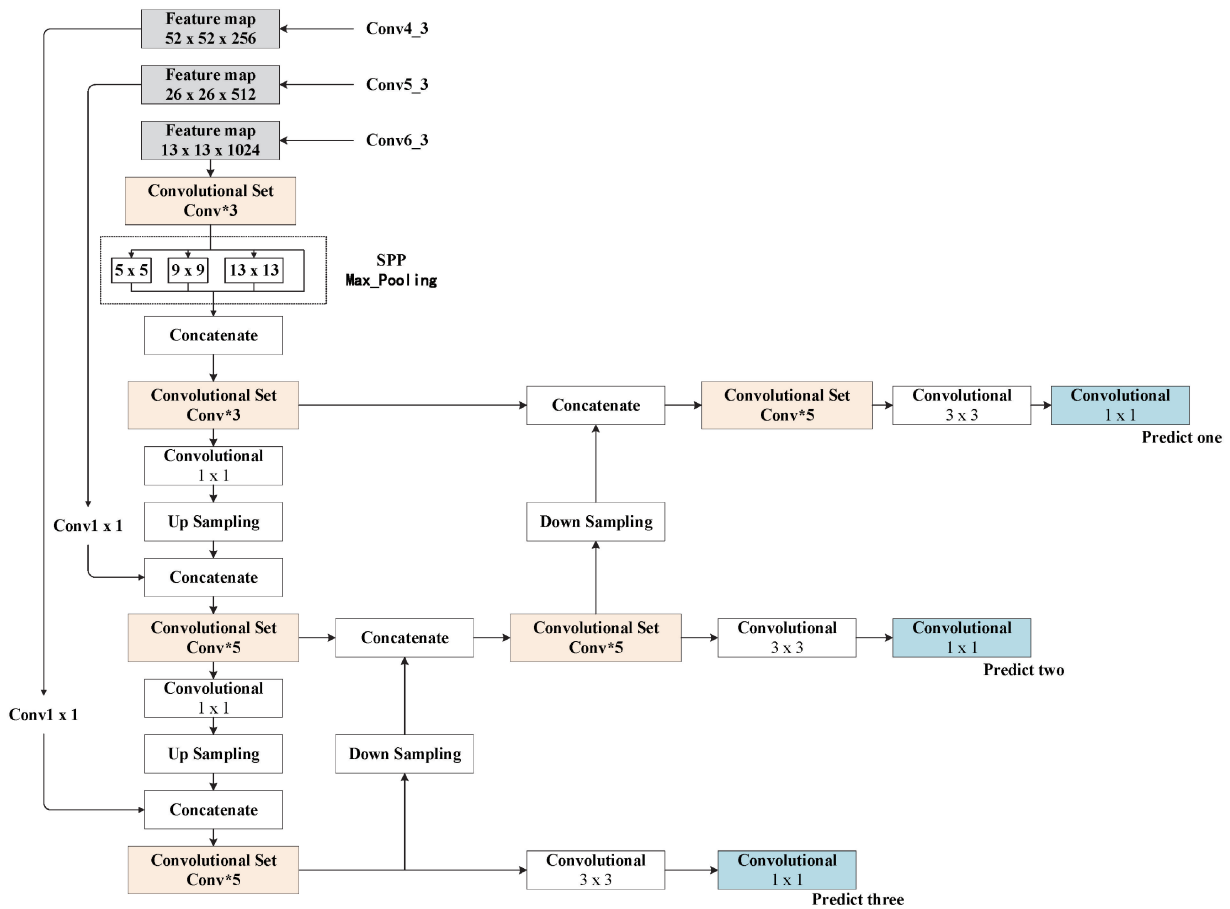


图 3 3 尺度检测网络结构

Fig. 3 3-scale detection network structure

首先,对特征提取主干网络最后一层 Conv6_3 输出的特征图进行一系列的卷积操作,特别是进行了空间金字塔池化。金字塔池化对特征图进行了 3 次步长为 1 的最大值池化操作,池化核大小分别为 13×13 , 9×9 , 5×5 ,池化后的 3 个特征图与输入的原特征图进行融合(concatenation),融合后的特征图不仅能够提升感受野,还能提取出不同尺度下的特征信息。

其次,进行两轮特征融合以实现细粒度的检测,第 1 轮通过上采样将 3 种不同尺度的特征图由深层到浅层进行融合,第 2 轮通过下采样将第一轮输出的 3 种不同尺度的特征图由浅层到深层再次进行融合,最终得到 3 种不同尺度的特征图。网络通过这样的结构实现了对特征的反复提取,使得网络对特征的利用更加充分。最后得到的不同尺度的特征图负责检测不同大小的目标,3 种不同尺度的特征图经过下方的若干个卷积层得到大小为 $13 \times 13 \times 18$, $26 \times 26 \times 18$, $52 \times 52 \times 18$ 的检测结果。

对于 13×13 大小的特征图,得到的张量大小是 $13 \times 13 \times [3 \times (4 + 1 + 1)]$,特征图的每个特征点设置 3 个检测框,每个检测框包含 4 个坐标偏移量、1 个目标预测置信度以及 1 个类别的预测概率。因此,使用多尺度特征图进行检测,将低分辨率与高分辨率特征图进行融合,使检测特征图拥有浅层细粒度特征和深层语义特征,能够获得更好的检测性能。

2.3 深度分离卷积结构

为了减少网络的参数量使得网络更加轻量化,将特征提取网络和检测网络中的所有常规卷积结构替换成深度分离卷积结构。与常规卷积操作不同,深度分离卷积的核心思想是将一个完整的卷积运算分解为两

步运行,分别为 Depthwise 卷积与 Pointwise 卷积。Depthwise 卷积的卷积核数量与输入特征图的通道数相同,且通道和卷积核一一对应。如图 4 中所示,卷积核大小为 $(3 \times 3 \times 1) \times 3$,1 个 3 通道的图像经过运算后生成了 3 个特征图。

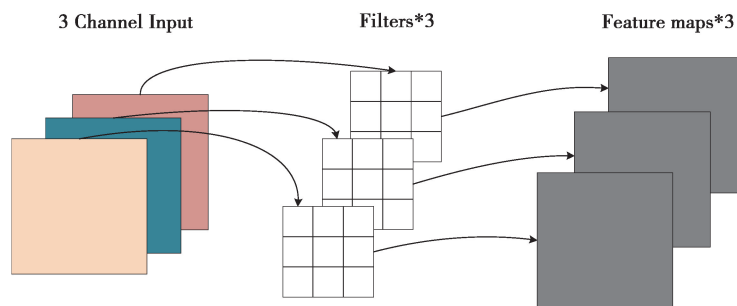


图 4 Depthwise 卷积

Fig. 4 Depthwise convolution

由于 Depthwise 卷积后不改变特征图的通道数,未有效地利用不同特征图在相同空间位置上的信息。因此,需要增加额外的操作来将这些特征图进行组合并实现特征图维度的扩展,即进一步进行 Pointwise 卷积。Pointwise 卷积结构,如图 5 所示,它的卷积核的尺寸为 $1 \times 1 \times M$, M 为输入特征图的通道数。通过 Pointwise 卷积运算,将上一层的特征图在深度方向上进行加权组合,生成新的特征图。

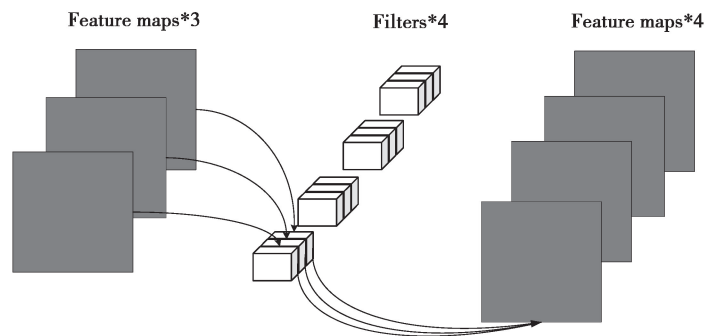


图 5 Pointwise 卷积

Fig. 5 Pointwise convolution

通过深度分离卷积可以大幅降低网络中的参数数量,当输入为 32 通道,输出为 64 通道时,常规卷积的参数数量为 $(3 \times 3 \times 32) \times 64 = 18\ 432$,而采用深度分离卷积的参数数量为 $(3 \times 3 \times 1) \times 32 + (1 \times 1 \times 32) \times 64 = 2\ 336$,参数的数量减少了 87.5%,网络参数数量的减少,将有利于加快检测速度。

3 实 验

实验环境为 Tensorflow 深度学习框架,Windows10 操作系统,Intel(R) Xeon(R) Silver 4114 CPU @ 2.20 GHz,32 GB 内存,单张 NVIDIA Quadro GP100 显卡,python3.5 编程语言。

在训练前,根据训练集的标记框大小,采用 K-means 方法^[16]聚类出 9 个不同大小的检测框(锚框),并按其大小分为 3 类,如表 1 所示。

表 1 9 个先验框聚类结果

Table 1 Clustering results of nine prior box

| 特征图 | 感受野 | 先验框 |
|----------------|-----|--|
| 13×13 | 大 | $(133 \times 131, 169 \times 188, 360 \times 320)$ |

续表 1

| 特征图 | 感受野 | 先验框 |
|----------------|-----|--|
| 26×26 | 中 | $(32 \times 55, 66 \times 48, 80 \times 93)$ |
| 52×52 | 小 | $(11 \times 20, 16 \times 28, 35 \times 22)$ |

实验部分将对比 4 种目标检测模型对大熊猫面部的检测性能,这 4 种模型如表 2 所示。

表 2 4 种目标检测模型

Table 2 Four object detection models

| VGG16 | VGG16_Res | VGG16_Res_Depwise | YOLOv4 |
|----------------------------------|---------------------------------------|----------------------------------|----------------------------------|
| 原始 VGGNet-16 网络 + 图 3 中 3 尺度检测网络 | 图 2 中 VGGNet-16 改进网络 + 图 3 中 3 尺度检测网络 | 用深度分离卷积替换 VGG16_Res 网络结构中所有的常规卷积 | CSPDarknet53 网络 + 图 3 中 3 尺度检测网络 |

3.1 模型训练

训练阶段采用 5 折交叉验证的方式。5 折交叉验证将整个数据集均分为 5 个子数据集,在训练过程中不重复的选择一个子数据集作为测试集,其它的 4 个子数据集作为训练集。整个训练过程重复 5 次,最终将 5 次的结果求平均作为最终的结果。该方法保证了每个样本都参与训练且被测试,从而降低了泛化误差。

训练时,采用 Adam 优化器,设置批次大小 $batch_size=8$,训练轮数 $epoch=30$,初始学习率 $learning_rate=0.001$ 。训练过程中采用缩放、平移变换、翻转、色彩抖动以及 mosaic 数据增强等操作增加训练样本以提升模型的泛化性能。训练过程中当验证集误差在 2 轮训练中都没有下降,则对学习率进行一次衰减,衰减权重 $decay=0.5$ 。当验证集误差在 5 轮训练中都没有下降,则进行“早停”操作^[26]。其中,验证集的损失函数变化情况如图 6 所示。

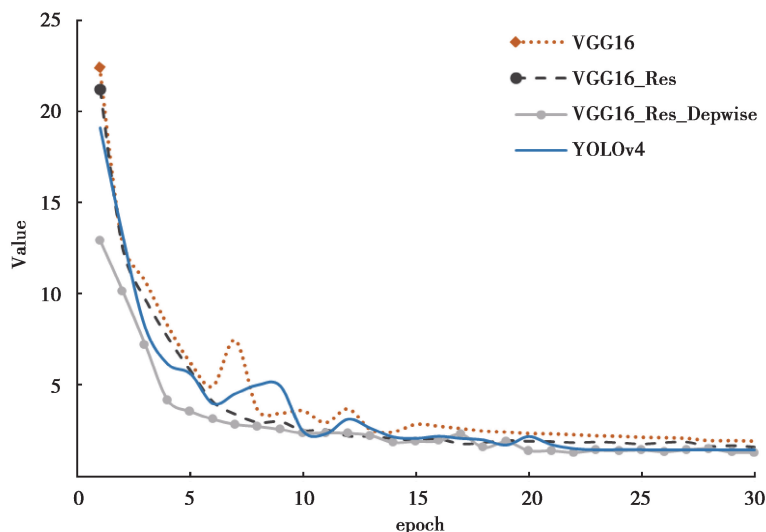


图 6 验证集损失函数曲线

Fig. 6 Loss function curves for validation dataset

由图 6 可知,最终经过 30 轮左右的迭代训练,模型基本上达到了收敛。采用 VGG16_Res_Depwise 模型的收敛速度最快并且稳定,VGG16 模型收敛速度最慢。

3.2 模型评价指标

3.2.1 平均精确度均值

平均精确度均值(mean Average Precision, mAP)是评估目标检测模型精度的重要指标,其值主要根据精确度和召回率来计算。精确度(P)与召回率(R)计算方式如下:

$$P_k = \frac{TP_k}{TP_k + FP_k},$$

$$R_k = \frac{TP_k}{TP_k + FN_k},$$

式中: k 代表当前是第几类;TP(True Positive)代表被正确检测为正样本的数量;FP(False Positive)代表被错误检测为正样本的数量;FN(False Negative)代表被错误检测为负样本的数量。 P - R 曲线的面积就是该类别的 AP(Average Precision)值。由于只对大熊猫面部进行检测,只有一个类,所以大熊猫面部类别的 AP 值也为模型的 mAP 值。

3.2.2 对数平均误检率

除了 mAP 评价指标外,还采用了文献[27]中的对数平均误检率(log-average miss rate, Lamr)指标来总结检测模型的性能,该指标越小代表模型的性能越好,是一个稳定且评价维度丰富的评价指标。

3.3 模型测试结果

图 7 是表 2 中的 4 种对比模型测试得到的大熊猫面部检测 P - R 曲线,可以看出,VGG16_Res_Depwise 模型的 P - R 曲线的积分面积最大,说明其检测性能最好。

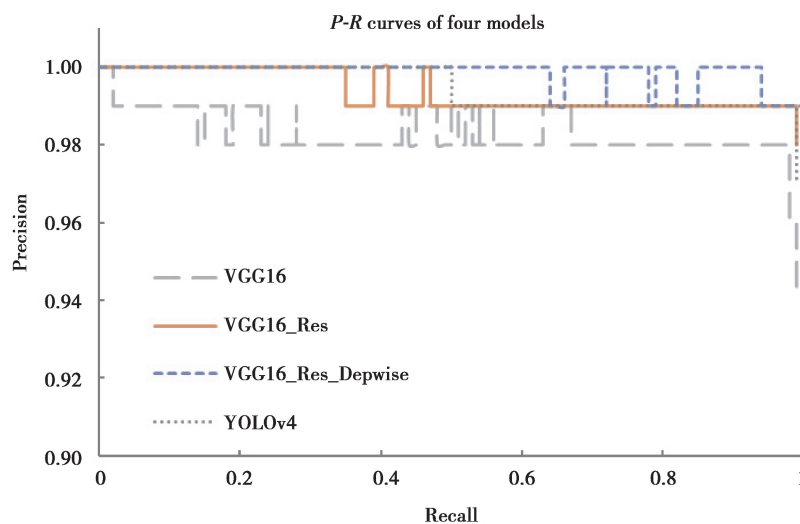


图 7 大熊猫面部检测 P - R 曲线

Fig. 7 P - R curves of giant panda face detection

表 3 中列出了 4 种模型的 mAP、Lamr 与单张图片检测时间以及模型参数大小。可以看出,4 种模型中,VGG16_Res_Depwise 不仅 mAP 更高,Lamr 更低,检测速度更快,同时网络参数也更少,性能上优于其他 3 种模型。同时也发现,改进的 VGGNet-16 特征提取网络只有 16 层,但其最终的性能优于 YOLOv4 的 53 层特征提取网络的性能。

表 3 4 种模型性能对比

Table 3 Performance comparison of four models

| 模型 | Lamr/% | mAP/% | 单张图检测时间/s | 模型保存参数大小/MB |
|-------------------|--------|-------|-----------|-------------|
| VGG16 | 3.32 | 96.39 | 0.071 | 289 |
| VGG16_Res | 0.91 | 98.72 | 0.075 | 265 |
| VGG16_Res_Depwise | 0.42 | 99.48 | 0.065 | 51 |
| YOLOv4 | 0.46 | 99.31 | 0.082 | 249 |

图 8 是 VGG16_Res_Depwise 模型对大熊猫面部进行检测得到的部分结果,可以看出,对于不同距离、面部有遮挡、夜间图像以及曝光状态下的大熊猫面部进行检测都能得到准确的检测结果。因此,VGG16_Res_Depwise 模型对复杂背景下的大熊猫面部有着不错的检测效果。

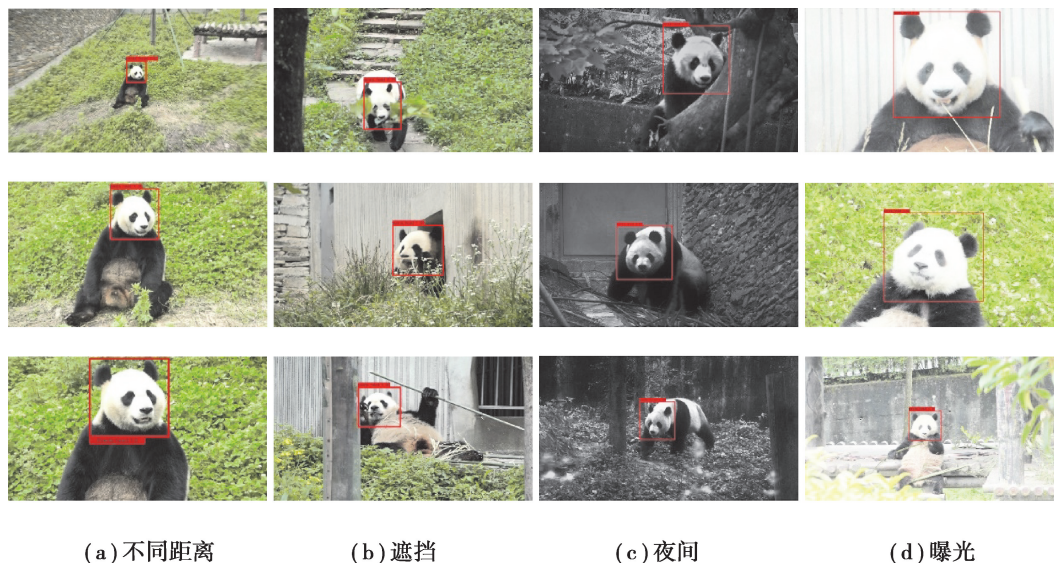


图 8 大熊猫面部检测结果

Fig. 8 Face detection results of giant pandas

4 结 语

随着红外相机技术在野生大熊猫监测工作中的广泛应用,研究基于图像的大熊猫个体识别技术有利于掌握大熊猫数量和种群分布,进而更好地保护大熊猫的繁衍生息。研究将 VGGNet-16 的改进网络结构作为特征提取主干网络,主干网络连接多尺度检测网络,同时用深度分类卷积替换常规卷积得到最终的大熊猫面部检测模型。改进后的大熊猫面部检测模型能准确地检测出大熊猫面部区域,为基于大熊猫面部图像的自动个体识别完成关键步骤。同时,提出的大熊猫面部检测模型对其他野生动物检测研究提供了重要的参考。

参考文献:

- [1] Li B V, Pimm S L. China's endemic vertebrates sheltering under the protective umbrella of the giant panda[J]. *Conservation Biology*, 2016, 30(2): 329-339.
- [2] Li B V, Alibhai S, Jewell Z, et al. Using footprints to identify and sex giant pandas[J]. *Biological Conservation*, 2018, 218: 83-90.
- [3] Zhang J D, Hull V, Huang J Y, et al. Activity patterns of the giant panda (*Ailuropoda melanoleuca*)[J]. *Journal of Mammalogy*, 2015, 96(6): 1116-1127.
- [4] Zheng X, Owen M A, Nie Y, et al. Individual identification of wild giant pandas from camera trap photos-a systematic and hierarchical approach[J]. *Journal of Zoology*, 2016, 300(4): 247-256.
- [5] 史雪威, 张晋东, 欧阳志云. 野生大熊猫种群数量调查方法研究进展[J]. *生态学报*, 2016, 36(23): 7528-7537. SHI Xuewei, ZHANG Jindong, OUYANG Zhiyun. Research progress on population investigation methods for wild giant panda[J]. *Acta Ecologica Sinica*, 2016, 36(23): 7528-7537. (in Chinese)
- [6] 唐小平, 贾建生, 王志臣, 等. 全国第四次大熊猫调查方案设计及其主要结果分析[J]. *林业资源管理*, 2015(1):11-16. TANG Xiaoping, JIA Jiansheng, WANG Zhichen, et al. Scheme design and main result analysis of the fourth national survey on giant pandas[J]. *Forest Resources Management*, 2015(1): 11-16. (in Chinese)
- [7] Hou J, He Y X, Yang H B, et al. Identification of animal individuals using deep learning: a case study of giant panda[J]. *Biological Conservation*, 2020, 242: 108414.
- [8] Chen J, Wen Q, Qu W M, et al. Panda facial region detection based on topology modelling[C]//2012 5th International

- Congress on Image and Signal Processing. Piscataway, NJ: IEEE, 2012: 911-915.
- [9] Chen J, Wen Q, Zhuo C L, et al. A novel approach towards head detection of giant pandas in the free-range environment [C]//2012 5th International Congress on Image and Signal Processing. Piscataway, NJ: IEEE, 2012: 814-818.
- [10] Zhang W W, Sun J, Tang X O. From tiger to panda: animal head detection[J]. IEEE Transactions on Image Processing, 2011, 20(6): 1696-1708.
- [11] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Proceedings of the International Conference on Neural Information Processing Systems. New York, USA: ACM, 2012: 1097-1105.
- [12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. Computer ence, 2014[2020-09-29]. <https://arxiv.org/abs/1409.1556>
- [13] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2016: 770-778.
- [14] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Los Alamitos: IEEE Computer Society Press, 2015: 1-9.
- [15] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [16] Redmon J, Farhadi A. Yolov3: An incremental improvement[J/OL]. Computer Vision and Pattern Recognition, 2018 [2020-09-29]. <https://arxiv.org/abs/1804.02767>
- [17] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J/OL]. Computer Vision and Pattern Recognition, 2020[2020-09-29]. <https://arxiv.org/abs/2004.10934>
- [18] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]. Cham: Springer International Publishing, 2016: 21-37.
- [19] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 2117-2125.
- [20] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]. Cham: Springer International Publishing, 2014: 740-755.
- [21] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [22] Everingham M, van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [23] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J/OL]. Machine Learning, 2015[2020-09-29]. <https://arxiv.org/abs/1502.03167>
- [24] He K M, Zhang X Y, Ren S Q, 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.
- [25] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 1251-1258.
- [26] Gencay R, Qi M. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging[J]. IEEE Transactions on Neural Networks, 2001, 12(4): 726-734.
- [27] Dollar P, Wojek C, Schiele B, et al. Pedestrian detection: an evaluation of the state of the art[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4): 743-761.