

残差网络下基于困难样本挖掘的目标检测

张超*, 陈莹

江南大学轻工过程先进控制教育部重点实验室, 江苏 无锡 214122

摘要 为了提高图像目标的检测精度,提出一种在残差网络下设计基于困难样本挖掘的目标检测算法。首先阐述基于深度学习的目标检测算法,即超快速区域卷积神经网络(Faster R-CNN)的工作原理,分析该算法存在的不足与改进方式。在Faster R-CNN的基础上,为了使模型能提取更有效的深度卷积特征,选取网络更深的残差网络替换原始的ZF或VGG网络。为了使学习到的网络模型有更强的泛化能力,在网络训练过程中,利用困难样本更新网络参数,使网络训练更充分。在Pascal VOC2007、Pascal VOC2007+Pascal VOC2012和BIT这三个数据集上进行训练和测试,实验结果显示,融合了两种方法的Faster R-CNN在这三个数据集上的检测精度分别提升了3.5%、7.1%、6.4%,提升效果明显。

关键词 图像处理; 目标检测; 超快速区域卷积神经网络; 残差网络; 困难样本挖掘

中图分类号 TP391 **文献标识码** A

doi: 10.3788/LOP55.101003

Object Detection Based on Hard Examples Mining Using Residual Network

Zhang Chao*, Chen Ying

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract In order to detect objects more accurately in images, an object detection algorithm based on hard example mining and residual network is proposed, which takes faster regional convolutional neural network (Faster R-CNN) as a benchmark. The working principle of Faster R-CNN is described based on deep learning, and the shortcomings and improvement methods of the algorithm are analyzed. Specifically, a deeper residual network is adopted to replace the original ZF or VGG network to extract more effective deep convolution features. In order to enhance the generalization ability of the learning network model, the network parameters are updated with hard examples during training. The experimental results on Pascal VOC2007, Pascal VOC2007+Pascal VOC2012 and BIT show that compared with Faster R-CNN, the proposed method improves detection accuracy by 3.5%, 7.1%, 6.4%, respectively, on the above three datasets.

Key words image processing; object detection; faster regional convolutional neural network; residual network; hard example mining

OCIS codes 100.4996; 100.3008; 100.5010

1 引言

目标检测是很多计算机视觉任务的必要前提,在智能视频监控^[1]、基于图像的内容检索^[2]、机器人导航^[3]和增强现实^[4]等领域都有广泛的应用。在国内,陈银等^[5]针对运动目标检测时传统高斯背景模型不能很好地自适应背景变化和检测不完整等

问题,提出应结合单高斯背景模型和mean shift原理对运动目标进行检测;刘德连等^[6]在针对遥感图像上的目标进行检测时,先对遥感图像进行高斯处理,将其作为近似理想背景,然后将原图像与高斯背景做差得到残差图,然后在残差图上进行目标检测;卢清华等^[7]针对目标检测中广泛应用的混合高斯模型的缺陷,通过采用新的权值和方差更新方法,加速

收稿日期: 2018-03-09; 修回日期: 2018-04-11; 录用日期: 2018-05-07

基金项目: 国家自然科学基金(61573168)

* E-mail: zcjndx@163.com

“鬼影”的消除,改善了针对车辆的目标检测性能。在国外:2012年前,最好的目标检测模型是基于可变形的组件模型(DPM)^[8]框架改进的各种算法,这些算法使用精心设计的人为特征来训练浅层分类器,从而实现目标分类与检测。2012年,Krizhevsky等^[9]训练出了一个含有5层卷积层的深度卷积神经网络用于图像分类算法,使得图像分类算法的准确率大幅度提升,证明了深度卷积特征在图像任务中的有效性。通过将目标检测中的传统特征替换为深度卷积特征,目标检测的准确率得到大幅度提升。

基于深度学习的目标检测主要有两类方法,分别是基于区域和不基于区域的方法。基于区域的方法将目标检测任务分为2个阶段:第一阶段,将候选区域生成的网络嫁接到深度卷积网络上生成高质量的候选框区域;第二阶段,设计一个后续子网络来分类改善候选框区域。基于区域的目标检测算法主要包括区域卷积神经网络(R-CNN)^[10]、快速区域卷积神经网络(Fast R-CNN)^[11]和超快速区域卷积神经网络(Faster R-CNN)^[12]模型,以及基于这些模型的改进算法^[13]。不基于区域的目标检测方法只有一个阶段,即对原图像使用滑动窗口分割,将分割后的各个窗口图像先后使用卷积神经网络提取特征并进行分

类回归,这类算法主要有YOLO(you only look once)^[14]和SSD(Single Shot MultiBox Detector)^[15]。随着基于深度学习的目标检测算法的快速发展,杨名等^[16]将基于深度学习的目标检测算法应用在船舶检测中,实验结果表明,所得到的准确率比传统的背景建模算法提高了25.75%。曹诗雨等^[17]将Fast R-CNN^[11]算法应用在车辆检测上,利用卷积特征替代传统手工特征,避免了传统检测的特征选择问题,验证了深层卷积特征具有更好的表达能力。

本文在Faster R-CNN的基础上进行改进。为了提取到更有效的卷积特征,使用101层的残差网络(ResNet)^[18]替换原先的塞勒-费格斯(ZF)^[19]或VGG(Visual Geometry Group)^[20]网络以提取更有效的卷积特征。在网络训练过程中,通过在线生成困难样本引导网络进行更充分的训练,以进一步提高目标检测精度。

2 基于Faster R-CNN的目标检测

Faster R-CNN经过R-CNN、Fast R-CNN的发展,在结构上,它已经将特征提取、候选区域提取、边界框回归及分类都整合到了一个网络中,这使得网络的综合性能有了极大的提高,在检测速度提升方面尤为明显。Faster R-CNN的算法框架如图1所示。

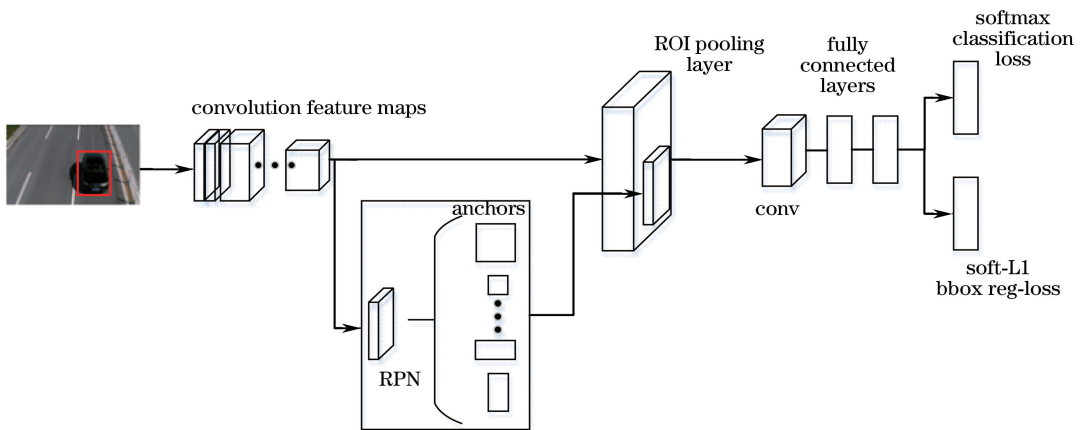


图1 算法流程图

Fig. 1 Flow chart of algorithm

将图像输入卷积神经网络提取卷积深度特征,接下来,卷积特征图输入区域生成网络(RPN)。RPN对于输入的卷积特征上的每一个特征点,选取该特征点对应原图像上的3种面积的窗口区域($128^2, 256^2, 512^2$),每一个窗口区域的长宽比又分为(1:1, 1:2, 2:1)。这样每一个特征点对应9个不同的窗口区域,这些窗口区域,称为anchors。RPN根据anchors产生相应的前景候选区域和背景候选

区域(ROI)。由于卷积神经网络中全连接层的输入特征维度是固定的,ROI池化层的作用就是对输入的不同大小的ROI在全图卷积特征上摘取固定尺寸的深度特征。最后,深度特征通过分类层损失和回归框层损失预测出目标的种类和回归框位置。

实践发现,在训练Faster R-CNN网络时存在如下问题。首先,Faster R-CNN使用的VGG网络的卷积层数比目前的残差网络要少,提取的卷积特征没有

残差网络提取的有效。与 VGG 网络相比,残差网络只在最后加入了一个池化层和一个全连接层,而 VGG 网络(以 VGG16 为例)有 5 个池化层、3 个全连接层,而且残差网络通过快捷链接的方式也进一步缩减了网络模型。因此,残差网络的计算复杂度要比 VGG 网络小。其次,由长尾效应^[21]可知,选出来的 ROIs 中大多数是易于训练的简单样本,少部分是较难训练的困难样本。在训练过程中,当训练到一个阶段后,困难样本对网络的优化往往起到至关重要的作用,但由于困难样本在一个批次中所占比例较低,对损失值的贡献小,使得网络得不到充分的训练。

为解决上述问题,在 Faster R-CNN 的基础上提出残差网络下融合困难样本挖掘的目标检测模型。

3 残差网络下基于困难样本挖掘的目标检测

残差网络下基于困难样本挖掘算法的目标检测模型可通过 101 层的残差网络提取更有效的卷积特征,并且由训练集中的困难样本来引导网络进行更有效的训练。

3.1 算法框架

改进后的 Faster R-CNN 算法框架如图 2 所

示。对于输入图片,首先通过 ResNet-101 网络提取比 VGG 和 ZF 网络更深的卷积特征,具体是将残差网络最后的池化层和全连接层去掉,将 ROI 池化层插入到最后一个卷积层之后。接着,卷积特征输入到 RPN 生成 ROIs,随后不同大小的 ROIs 映射到卷积特征图中,通过 ROI 池化层生成固定尺寸的 ROI 卷积特征并送入到后续网络。由于 ROI 池化层后面不再有卷积层或全连接层,每个 ROI 区域共享残差网络中的所有卷积运算,大大加快了训练检测速度。

在设计困难样本挖掘算法时,为了更高效地利用显存空间,并不只是将损失层输出的每个 ROI 损失进行排序后,简单保留损失值最大的前 n 个困难样本,并将简单样本损失值置为 0。因为这样做,那些损失值置为 0 的 ROIs 仍然分配显存占用空间^[22]。本文将 ROI 池化层后面的网络复制了两份:图 2 中绿色网络只进行前向计算,用来计算每个 ROI 的损失值,并对每个 ROI 损失值排序,选出前 n 个损失值最高的 ROI 作为困难样本;图 2 中红色网络可以进行前向计算和反向计算,并使用由绿色网络选出来的困难样本更新网络参数。

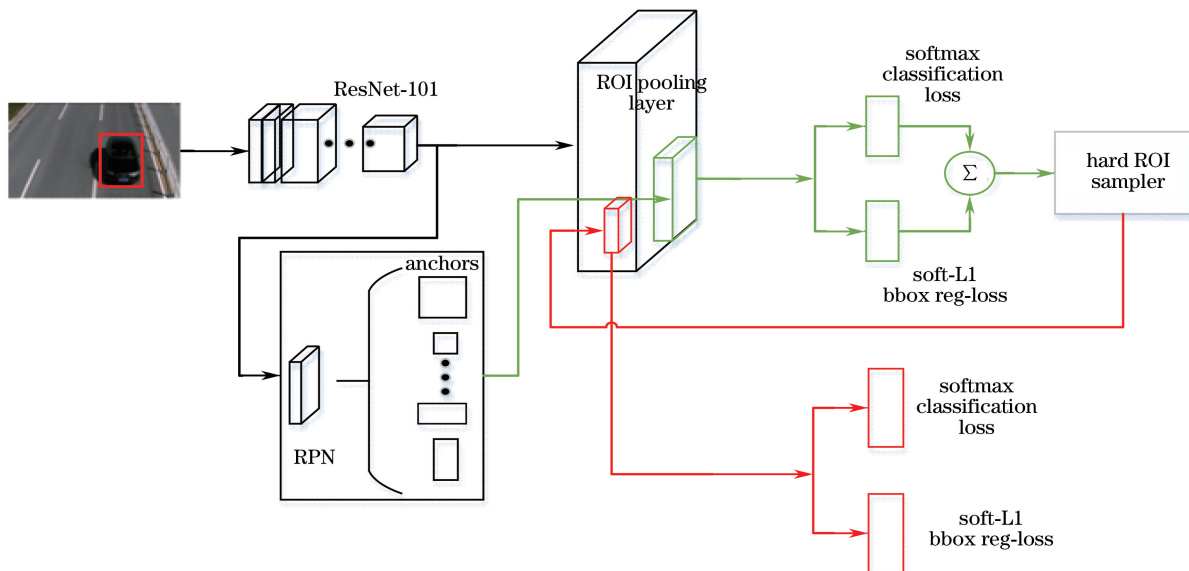


图 2 改进后的算法框架

Fig. 2 Improved framework of algorithm

3.2 融合残差网络的 Faster R-CNN

Faster R-CNN 模型主要使用 ZF 或 VGG 网络来提取卷积特征,但这两个网络存在的问题主要有: 1) 网络卷积层数较少,提取到的卷积特征没有卷积层数较多的网络提取到的特征有效; 2) 训练好的参数模型较大,训练和存贮时占用很大的资源。因此,

将 Faster R-CNN 中的浅层卷积神经网络替换为残差网络以获得更好的深度特征。

残差网络通过快捷链接的方式引入恒等映射,使得网络深度与网络的分类效果成正相关,它由多个构建块构成,如图 3 所示是一个由两个堆叠层组成的构建块,这个构建块在数学上可以表示为

$$y = f\{x, \{W_i\}\} + \bar{x}, \quad (1)$$

式中 x 是构建块的输入, y 是构建块的输出, $f\{x, \{W_i\}\}$ 表示残差映射, \bar{x} 表示 x 的恒等映射, 通过快捷链接到输出。由于图 3 的构建块由 2 个卷积层组成, 因此残差映射 $f\{x, \{W_i\}\}$ 可以进一步表示为

$$f = W_2 \sigma(W_1 x), \quad (2)$$

式中 σ 表示卷积神经网络的激活函数(ReLu), W_1 和 W_2 是卷积层中需要学习的参数。

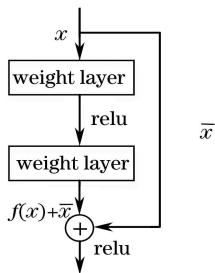


图 3 单个构建块

Fig. 3 A building block

3.3 困难样本挖掘

对 Faster R-CNN 作进一步改进, 通过融合困难样本挖掘来改善长尾效应给模型优化带来的问题。在改进后的模型中, 首先利用非极大值抑制算法(NMS)排除掉重叠度较高的 ROI, 然后计算区域生成网络产生的各个 ROI 的损失值, 挑选出困难样本。其中, 单个 ROI 损失值计算如下:

$$l(p_i) = \begin{cases} l_{\text{cls}}(p_i), & p_i \text{ is background} \\ l_{\text{cls}}(p_i) + \lambda l_{\text{loc}}(p_i), & p_i \text{ is foreground} \end{cases}, \quad (3)$$

式中 $l_{\text{cls}}(\cdot)$ 和 $l_{\text{loc}}(\cdot)$ 分别对应目标的分类损失值和回归框损失值。为了调和 $l_{\text{cls}}(\cdot)$ 和 $l_{\text{loc}}(\cdot)$ 之间的差距, 引入参数 λ 来平衡二者。目标分类损失 $l_{\text{cls}}(\cdot)$ 由 softmax 回归计算:

$$l_{\text{cls}}(p, i) = -\ln p_i. \quad (4)$$

回归框损失 $l_{\text{loc}}(\cdot)$ 通过 smooth 函数计算:

$$l_{\text{loc}}(t^u, v) = \sum_{k \in \{x, y, w, h\}} \text{smooth}(t_k^u - v_k), \quad (5)$$

$$\text{smooth}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| \geq 1 \end{cases}, \quad (6)$$

式中 $v = (v_x, v_y, v_w, v_h)$ 表示 ROI 的真实值坐标, $t^u = (t_x^u, t_y^u, t_w^u, t_h^u)$ 表示第 t 个 ROI 对应第 u 类的回归框坐标预测值。

在训练过程中使用批量随机梯度下降算法训练神经网络模型。假设每次训练时的批次为 n , 训练网络时, 对每个 ROI 计算损失并排序, 选取损失值最高的前 n 个样本组成迭代训练中的一个批次训练网

络, 在本文中 n 设置为 128。由此, 融合困难样本挖掘算法的 Faster R-CNN 可以有效缓解长尾效应在训练网络中带来的不利因素。

4 实验结果与分析

实验环境主要硬件配置如下: GPU 卡, TITAN XP; 处理器, Intel Xeon(R) E5-1630 V4 @ 4.0 GHz; 内存, 128 GB。实验软件工具为 Ubuntu14.04 下的 Caffe 框架。评价指标为平均准确率(AP), 也就是以查准率 P 为横坐标、查全率 R 为纵坐标所画的 P - R 曲线下的面积。mAP 则是对不同类别的 AP 再求平均值。

$$P = \frac{F_{\text{TP}}}{F_{\text{TP}} + F_{\text{FP}}}, \quad (7)$$

$$R = \frac{F_{\text{TP}}}{F_{\text{TP}} + F_{\text{FN}}}, \quad (8)$$

式中 F_{TP} 表示正例预测为正例的情况, F_{FP} 表示反例预测为正例的情况, F_{FN} 表示正例预测为反例的情况。

4.1 Pascal VOC

在近年来的计算机视觉竞赛中, Pascal VOC Challenge 具有举足轻重的地位。它包含了 20 种常见的目标类别, 竞赛内容由三个子任务组成, 分别是图像检测、图像分类、语义分割。本文使用与文献[12]一样的数据集 Pascal VOC2007 和 Pascal VOC2007+Pascal VOC2012 进行训练测试。两个数据集中训练图片、验证图片、测试图片的数量如表 1 所示。

表 1 数据集描述

Table 1 Dataset description

Dataset	Total	Training	Validation	Test No.
	No.	No.	No.	
VOC2007	9963	2505	2506	4952
VOC2007+ VOC2012	27088	11068	11068	4952

在上述两个数据集上分别对比了本文算法与近几年相关算法的准确率。从表 2 和表 3 可以看出, 在这两个数据集上, 本文算法较文献[12]方法在准确率上分别提升了 3.5% 和 7.1%。文献[11]和文献[22]都是基于 Fast R-CNN 算法, 其本身 ROI 区域的生成依靠的是传统方法, 没有实现网络的端到端训练, 而本文算法中 ROI 区域的生成统一到了网络中, 实现了网络的端到端训练, 相较于文献[11]和文献[22]中的方法, 本文算法在 VOC2007 数据集上准确率分别提升了 5.1% 和 2.1%, 在 VOC2007+VOC2012 数据集上准确率分别提升了 10.3% 和 5.7%。这些实验效果的提升主要是由于本文算法融合了残差网络强大的特征提取能力和困难样本挖掘算法。

表 2 不同方法在 Pascal VOC2007 上的准确率

Table 2 Accuracy of different methods on Pascal VOC2007

Image	Method in Ref. [11]	Method in Ref. [12]	Method in Ref. [22]	Method in Ref. [18]	Proposed algorithm
mAP	66.9	68.5	69.9	70.3	72.0
areo	74.5	74.1	71.2	70.6	74.4
bike	78.3	77.2	78.3	80.2	80.0
bird	69.2	67.7	69.2	72.7	72.7
boat	53.2	53.9	57.2	64.2	64.2
bottle	36.6	51.0	46.5	52.2	52.2
bus	77.3	75.1	81.8	78.4	78.4
car	78.2	79.2	79.1	79.9	79.7
cat	82.0	78.9	83.2	83.3	83.0
chair	40.7	50.7	47.9	48.9	53.9
cow	72.7	78.0	76.2	76.4	78.9
table	67.9	61.1	68.9	62.5	63.8
dog	79.6	79.1	83.2	83.9	84.3
horse	79.2	81.9	80.8	82.5	85.3
mbike	73.0	72.2	75.8	75.4	76.9
person	69.0	75.9	72.7	77.2	77.6
plant	30.1	37.2	39.9	39.9	41.2
sheep	65.4	71.4	67.5	74.3	74.1
sofa	70.2	62.5	66.2	71.5	73.4
train	75.8	77.4	75.6	76.3	79.2
tv	65.8	66.4	75.9	66.0	66.8

表 3 不同方法在 Pascal VOC2007+Pascal VOC2012 上的准确率

Table 3 Accuracy of different methods on Pascal VOC2007+Pascal VOC2012

Image	Method in Ref. [11]	Method in Ref. [12]	Method in Ref. [22]	Method in Ref. [18]	Proposed algorithm
mAP	70.0	73.2	74.6	76.4	80.3
areo	77.0	76.5	77.7	79.8	83.5
bike	78.1	79.0	81.2	80.7	82.3
bird	69.3	70.9	74.1	76.2	83.6
boat	59.4	65.5	64.2	68.3	78.0
bottle	38.3	52.1	50.2	55.9	75.2
bus	81.6	83.1	86.2	85.1	85.9
car	78.6	84.7	83.8	85.3	84.8
cat	86.7	86.4	88.1	89.8	86.1
chair	42.8	52.0	55.2	56.7	71.2
cow	78.8	81.9	80.9	87.8	85.0
table	68.9	65.7	73.8	69.4	70.5
dog	84.7	84.8	85.1	88.3	87.6
horse	82.0	84.6	82.6	88.9	84.6
mbike	76.6	77.5	77.8	80.9	82.0
person	69.9	76.7	74.9	78.4	80.3
plant	31.8	38.8	43.7	41.7	63.8
sheep	70.1	73.6	76.1	78.6	81.3
sofa	74.8	73.9	74.2	79.8	78.1
train	80.4	83.0	82.3	85.3	83.7
tv	70.4	72.6	79.6	72.0	79.6

4.2 BIT 数据集

车辆检测是目标检测的一个重要分支,为了验证本文方法的有效性,特别选用 BIT 数据集^[23]增加了对车辆检测的实验。BIT 最初是车型分类的数据集,但它的标注中含有车辆位置信息,因此本文采用该数据集用于车辆检测。BIT 数据集中一共有 9850 张车辆图片,每张图片的像素都是 1600 pixel×1200 pixel,是由现实场景中的高速公路交通卡口采集得到的。对 BIT 数据集中的车辆进行检测非常具有挑战性,它不仅包含了 5 种类型的车辆(公交车、中巴车、小型货车、轿车和卡车),而且考虑到了白天、黑夜、多云以及复杂背景等各种情况。

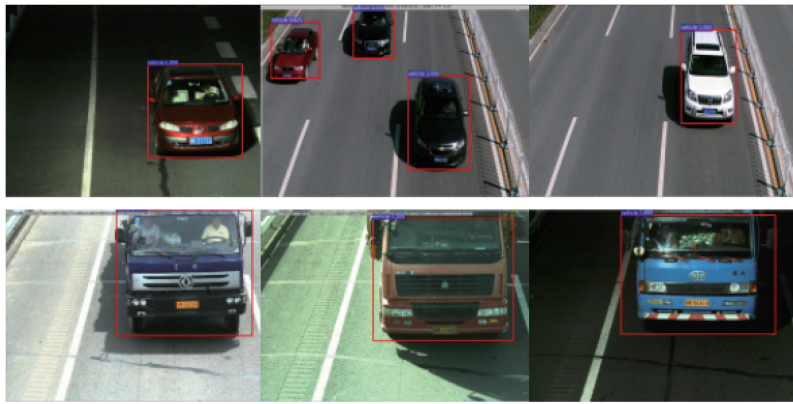


图 4 BIT 部分图片的检测结果

Fig. 4 Test results of part of BIT images

表 5 对比了不同算法在 BIT 数据集上的检测速度。检测速度是在大规模执行车辆检测时通过平均计算 BIT 数据集中 4925 张测试图片的检测时间得到的。由表 5 可知,本文方法在显著提高检测精度的情况下,相较于 Faster R-CNN 仅仅增加了微量的检测时间。表明改进后的算法在车辆检测中具有极大的优势,可以进一步用于交通领域的车辆识别。

表 5 算法检测速度对比

Table 5 Comparison of detection speed of algorithms

Method	Test time per image /s
Method in Ref. [12]	0.114
Method in Ref. [22]	0.282
Proposed algorithm	0.125

5 结 论

提出一种融合残差网络和困难样本挖掘算法的目标检测模型。首先将基于深度学习的目标检测模型 Faster R-CNN 中的特征提取网络用残差网络进行替换,由层数更多的残差网络提取到更有效的卷积特征。接着,将困难样本挖掘算法融合到网络中,

选取 BIT 数据集中 50% 的图片作为训练验证集,其余图片作为测试集,并且训练验证集中 50% 的图片作为训练样本,另外 50% 的图片作为验证样本。表 4 所示为不同方法在 BIT 上的实验结果,本文方法的准确率比文献[12]方法提升了 6.4%。图 4 为本文算法在部分 BIT 图像上的检测结果。

表 4 不同方法在 BIT 图像上的准确率

Table 4 Accuracy of different methods on BIT images

Method	AP
Method in Ref. [12]	63.0
Method in Ref. [22]	67.5
Proposed algorithm	69.4

由数据集中的困难样本引导网络进行更有效的训练。在数据集 Pascal VOC2007, Pascal VOC2007+Pascal VOC2012 上的对比实验显示,本文算法提高了目标检测精度。将本文算法应用在真实场景车辆数据集 BIT 上,结果显示,车辆检测效果获得明显提升,验证了本文算法的性能。在实际应用中,可以通过多线程计算来进一步提升算法的速度,或在检测精度可接受的前提下通过适当降低图片分辨率来加快图像中的目标检测。

参 考 文 献

- [1] Aggarwal J K, Ryoo M S. Human activity analysis: a review [J]. ACM Computing Surveys, 2011, 43 (3): 16.
- [2] Datta R, Joshi D, Li J, *et al.* Image retrieval: ideas, influences, and trends of the new age [J]. ACM Computing Surveys, 2008, 40(2): 5.
- [3] Krüger V, Kragic D, Ude A, *et al.* The meaning of action: a review on action recognition and mapping [J]. Advanced Robotics, 2007, 21(13): 1473-1501.

- [4] Palmese M, Trucco A. From 3D sonar images to augmented reality models for objects buried on the seafloor[J]. *IEEE Transactions on Instrumentation and Measurement*, 2008, 57(4): 820-828.
- [5] Chen Y, Ren K, Gu G H, *et al.* Moving object detection based on improved single gaussian background model[J]. *Chinese Journal of Lasers*, 2014, 41(11): 1109002.
陈银, 任侃, 顾国华, 等. 基于改进的单高斯背景模型运动目标检测算法[J]. *中国激光*, 2014, 41(11): 1109002.
- [6] Liu D L, Zhang J Q, He G J. Target detection for remote sensing image based on gaussian transformation of background [J]. *Acta Optica Sinica*, 2007, 27(4): 638-642.
刘德连, 张建奇, 何国经. 背景高斯化的遥感图像目标检测[J]. *光学学报*, 2007, 27(4): 638-642.
- [7] Lu Q H, Wu Z W, Fan Y B, *et al.* An improved mobile vehicle detection method based on Gaussian mixture model [J]. *Journal of Optoelectronics • Laser*, 2013, 24(4): 751-757.
卢清华, 吴志伟, 范彦斌, 等. 基于混合高斯模型的运动车辆检测方法[J]. *光电子 • 激光*, 2013, 24(4): 751-757.
- [8] Felzenszwalb P F, Girshick R B, McAllester D, *et al.* Object detection with discriminatively trained part-based models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627-1645.
- [9] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]. *Advances in Neural Information Processing Systems*, 2012: 1097-1105.
- [10] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580-587.
- [11] Girshick R. Fast R-CNN [C] // *Proceedings of the IEEE International Conference on Computer Vision*, 2015: 1440-1448.
- [12] Ren S Q, He K M, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [13] He K, Gkioxari G, Dollár P, *et al.* Mask R-CNN [C] // *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2980-2988.
- [14] 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*, 2016: 779-788.
- [15] Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector[C]. *European Conference on Computer Vision*, 2016: 21-37.
- [16] Yang M, Ruan Y D, Chen L K, *et al.* New video recognition algorithms for inland river ships based on faster R-CNN [J]. *Journal of Beijing University of Posts and Telecommunications*, 2017, 40(S1): 130-134.
杨名, 阮雅端, 陈林凯, 等. 甚高速区域卷积神经网络的船舶视频检测方法[J]. *北京邮电大学学报*, 2017, 40(S1): 130-134.
- [17] Cao S Y, Liu Y H, Li X Z. Vehicle detection method based on fast R-CNN [J]. *Journal of Image and Graphics*, 2017, 22(5): 671-677.
曹诗雨, 刘跃虎, 李辛昭. 基于 Fast R-CNN 的车辆目标检测[J]. *中国图象图形学报*, 2017, 22(5): 671-677.
- [18] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [19] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]. *European Conference on Computer Vision*, 2014: 818-833.
- [20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv: 1409.1556*, 2014.
- [21] Ouyang W, Wang X, Zhang C, *et al.* Factors in finetuning deep model for object detection with long-tail distribution [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 864-873.
- [22] Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 761-769.
- [23] Dong Z, Jia Y. Vehicle type classification using distributions of structural and appearance-based features[C] // *Proceedings of the IEEE International Conference on Image Processing*, 2013: 4321-4324.