改进的Faster RCNN煤矿井下行人检测算法

李伟山[1],卫 晨[2],王 琳[1]
1.西安邮电大学 通信与信息工程学院,西安 710121
2.西安邮电大学 经济与管理学院,西安 710121

摘 要:针对煤矿井下环境恶劣、光照差、背景混杂、行人模糊、行人多尺度等问题,提出了一种改进的Faster RCNN煤矿井下行人检测方法,使用深度卷积神经网络代替传统的手工设计特征方式自动地从图片中提取特征。利用深度学习通用目标检测框架Faster RCNN,以Faster RCNN算法为基础,对候选区域网络(Region Proposals Network,RPN)结构进行了改进,提出了一种"金字塔RPN"结构,来解决井下行人存在的多尺度问题;同时算法中加入了特征融合技术,将不同卷积层输出的特征图进行融合,增强煤矿井下模糊、遮挡和小目标行人的检测性能。实验结果表明:改进的Faster RCNN可以有效解决井下行人检测问题,在井下行人数据集上获得了90%的检测准确率,并在公测数据集VOC 07上对改进算法进行了验证。

关键词:深度学习;Faster RCNN;行人检测
文献标志码:A **中图分类号:**TP391 doi:10.3778/j.issn.1002-8331.1711-0282

李伟山,卫晨,王琳.改进的Faster RCNN煤矿井下行人检测算法.计算机工程与应用,2019,55(4):200-207.
LI Weishan, WEI Chen, WANG Lin. Improved Faster RCNN approach for pedestrian detection in underground coal mine. Computer Engineering and Applications, 2019, 55(4):200-207.

Improved Faster RCNN Approach for Pedestrian Detection in Underground Coal Mine

LI Weishan[1], WEI Chen[2], WANG Lin[1]
1.School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
2.School of Economics and Management, Xi'an University of Posts and Telecommunications, Xi'an 710121, China

Abstract: In order to solve the problems of harsh underground environment, poor lighting, mixed background and multi-scale pedestrian, this paper proposes a pedestrian detection method of underground coal mine based on improved Faster RCNN. Deep convolutional neural network can replace traditional manual design feature to extract features automatically from images. Based on the Faster RCNN algorithm, RPN(Region Proposals Network) structure is improved and a "pyramid RPN" structure is proposed to solve multi-scale detection problem of pedestrian underground. At the same time, by adding feature fusion technology, the feature maps of different convolution layers are merged to improve the detetion performance for under-mine blur, occlusion and tiny pedestrian. The experimental results indicate that the improved Faster RCNN can effectively solve the pedestrian detection problem of underground coal mine, which obtains 90% detection accurary on the under-mine pedestrian dataset. The improved Faster RCNN algorithm is validated in the VOC 07 benchmark.
Key words: deep learning; Faster RCNN; pedestrian detection

1 引言

随着信息化与工业化的深度融合,煤矿行业作为一个传统的重工业产业,正在逐步加快"两化融合"的脚步。煤矿作为一个高危产业,在入井口、出井口、井下的各个巷道等位置都安装有大量的监控摄像头,但是目前大量的视频资源没有得到有效的利用。矿井下的视频

基金项目:陕西省科技厅资源主导型产业关键技术(链)工业领域项目(No.2015KTCXSF-10-13)。
作者简介:李伟山(1991—),男,硕士研究生,研究领域为深度学习,E-mail:420702722@qq.com;卫晨(1983—),男,讲师,研究领域为图像理解与分析;王琳(1992—),女,硕士研究生,研究领域为深度学习。
收稿日期:2017-11-20 **修回日期:**2018-01-12 **文章编号:**1002-8331(2019)04-0200-08
CNKI**网络出版:**2018-05-24, http://kns.cnki.net/kcms/detail/11.2127.TP.20180522.0944.002.html

图像存在环境复杂、光线暗淡、噪声干扰大等问题，且矿井下摄像头安装位置在高处，监控视频中所监测到的行人存在尺寸偏小、分辨率低、尺度变化、行人重叠等问题。井下因其特殊的环境，井下图像中包含了目标检测和行人检测问题中常见的目标扭曲、多尺度、遮挡、光照等情况。因此，井下行人检测拥有较高的研究价值和意义，能够进一步提高工业视频的利用率，保障井下作业人员的安全。

传统的目标检测一般使用手提特征，然后采用一个分类器来实现目标的检测。如Dalal等人提出的HOG+SVM的行人检测方法[1]。这类方法一般使用滑动窗口的框架，大致分为三个步骤：(1)使用不同尺度大小的滑动窗口在图像中滑动，选取某一部分作为候选目标区域；(2)提取候选目标区域的视觉特征，如HOG(Histogram of Oriented Gradient)特征(常用于行人目标检测)、Harr特征(常用于人脸检测)[2]、LBP(Local Binary Pattern)特征[3]、积分通道特征[4]等；(3)应用分类器进行分类识别。这种传统的方法要求研究人员根据不同的检测任务，对相关领域深入研究设计出特定的适应性好的特征，泛化能力差。

近些年来，随着硬件设备的提升，深度学习技术得到了快速发展。卷积神经网络可以替代传统的手工设计特征且提取的特征拥有高级的语义表达能力，特征表达能力强，鲁棒性更好[5-8]，在图像分类、目标检测等计算机视觉领域取得了巨大的成果[9-10]，出现了大量的基于深度学习的检测算法[7-14]。

研究发现，目前大量的目标检测、行人检测的研究都是基于自然光场景的，其图像质量都较高，目标也比较清晰。如VOC数据集[15]、微软的COCO数据集[16]以及著名的大规模视觉挑战赛ImageNet[17]等都是基于自然场景下的研究。本文将基于Faster RCNN[18]的方法实现煤矿井下的行人检测。同时井下图片多来自于井下监控视频，将视频转换为图片会出现运动模糊，因此图片中会出现行人模糊不清晰的现象，本文的研究也可以归为模糊场景下的检测问题，并推广至相关领域。

2　相关研究

目前行人检测已经取得了大量的研究成果，2012年Dollar等人[19]对行人检测进行了综述，对比了近年来最优的行人检测方法；2014年Benenson等人[20]对近十年行人检测领域约40多种方法在Caltech数据集上进行了性能比较；2015年Hosang等人[21]对将卷积神经网络应用于行人检测进行了研究。

行人检测可以看作目标检测的一个子任务，通过对目标检测算法改进可以实现某一特定目标的检测。目前许多基于深度学习的特定目标的检测都是通过对通用目标检测算法的改进来实现的。Goirshick在2015年提出了基于深度学习的Faster RCNN通用目标检测算法，利用RPN网络生成候选区域，送入到Faster RCNN实现目标的检测，获得了非常高的准确率[18]。因其卓越的检测性能，被广泛应用到各类任务中。宋焕生等人[22]将Faster RCNN转换为二分类问题应用到复杂场景下的车辆检测中；Sun等人[23]通过特征融合、难例挖掘、多尺度训练等策略改进Faster RCNN，将其应用到人脸检测任务中。

目前主流的基于深度学习的目标检测算法分为两类：一类是以Faster RCNN为主的基于区域的目标检测算法，生成候选目标区域，对区域加以分类实现检测，如Faster RCNN、R-FCN[7]等。这类算法的优点是检测准确率较高，缺点是速度较慢。另一类是以YOLO(You Only Look Once)为代表的将目标检测转化为回归问题求解，输入原始图片直接输出物体的位置与类别，如YOLO[10]、SSD[11]等。该类方法的优点是检测速度快，每秒能实现几十帧的检测，但检测准确率低，针对小目标的检测不敏感。

本文将Faster RCNN通用目标检测算法引入到煤矿井下的行人检测这一复杂场景中。煤矿井下环境复杂，利用深度卷积神经网络可以很好地实现特征的提取；处于监控场景的行人，像素总体偏小，本文对RPN网络中的anchor大小做了进一步改进；原始的Faster RCNN在最后一层特征图上使用一个3×3的滑动窗来生成候选区域，本文进一步改进，提出了一种金字塔RPN方式来生成候选区域；针对图片中的行人运动模糊，将不同层级的特征进行融合来提高行人检测准确率。本文提出了一个改进的Faster RCNN井下行人检测算法，在井下数据集进行了实验分析，并在VOC 07公测数据集的行人类别上验证了本文算法的有效性。

3　井下行人检测方案的设计

本文设计的井下行人检测方案如图1所示，采用文献[18]中近似联合优化(Approximate Joint Optimization)机制实现模型的端到端的训练。选取井下行人数据集作为训练样本，将数据输入到网络中，图片缩放到600×1 000送入特征提取网络生成特征图，将输出的特征图送入RPN网络生成候选区域，再将提取的候选区域的特征经RoI Pooling层[18,24]处理为固定大小的特征向量，送入后面的全连接层实现分类与范围框的回归。整个

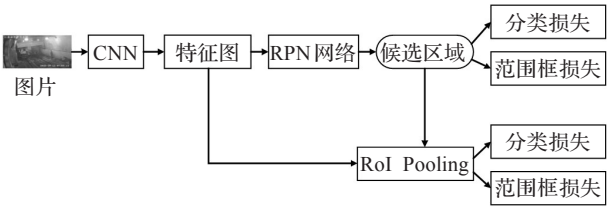


图1　井下行人检测方案

方案是一个端到端的结构,一个网络,四个loss。这样的方案设计是一种多任务学习策略,有助于提高模型的准确度[25]。

3.1　Faster RCNN简介

Faster RCNN目标检测网络分为两步,首先定位目标,然后对目标的具体类别进行分类。输入一张图片,首先利用特征提取网络经过一系列的卷积、池化操作提取图像的特征图(Feature Map);RPN网络在特征图上定位候选目标,使用softmax分类器来判别候选目标属于前景还是背景,同时利用范围框回归器修正候选目标的位置,最终生成候选目标区域。分类网络利用特征图和RPN网络生成的候选区域实现目标类别的检测,在本文中即实现行人的检测,判别候选区域属于行人还是背景。

3.2　特征提取网络

选择VGG16作为本文的特征提取网络,其是由牛津大学计算机视觉组和谷歌研究院一起研发的深度卷积神经网络,在2014年ILSVRC比赛中图像分类和目标定位分别获得第二和第一的成绩。整个网络通过堆叠相同尺寸的卷积核(3×3)和池化层(1×1)来实现。本文选取VGG16提取输入图片的特征,去掉了原网络中的全连接层和最后一个池化层,具体的网络参数如表1所示。

表1　VGG16网络结构参数表

类型/层数	卷积核数量	卷积核大小/步长	输出
Convl_x/2	64	3×3/1	600×1 000
Maxpool		2×2/2	300×500
Conv2_x/2	128	3×3/1	300×500
Maxpool		2×2/2	150×250
Conv3_x/3	256	3×3/1	150×250
Maxpool		2×2/2	75×125
Conv4_x/3	512	3×3/1	75×125
Maxpool		2×2/2	38×63
Conv5_x/3	512	3×3/1	38×63

本文将VGG16网络中输出大小相同的卷积层归为一部分,如表1中第1列所示,整个网络分为5组卷积层,每一组分别包含x层,如Conv5_x/3表示第5部分共包含3层卷积。从表中1可以发现,整个网络卷积核大小均为3×3。3×3的卷积核是最小的能够提取特征的尺寸。同时这样反复地堆叠小尺寸的卷积核,能够提升CNN对特征的学习能力。因此,选择VGG16作为本文井下行人检测的特征提取网络。表中第2、3、4列分别表示卷积核数量、卷积核大小/步长、每一层对应的特征图输出大小。

3.3　金字塔RPN结构

RPN网络输入特征提取网络生成的特征图,输出目标候选区域矩形框集合。原始的RPN网络结构通过在输出的特征图上利用滑动窗口通过3×3的卷积直接实现候选区域的提取,送入网络后续部分进一步实现前景背景的分类和候选区域位置框的回归。在特征提取网络输出的最后一层特征图上经过3×3卷积之后每一个像素点映射回原始图片对应的坐标点,以该点为中心生成3种比例1∶1/1∶2/2∶1,3种尺度128/256/512,共9种不同大小的粗粒度的候选区域,即“anchor”。如图2、图3所示。

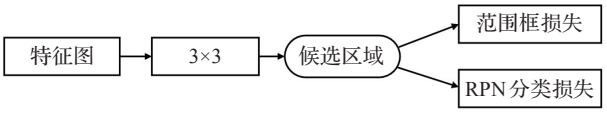


图2　原始RPN网络结构

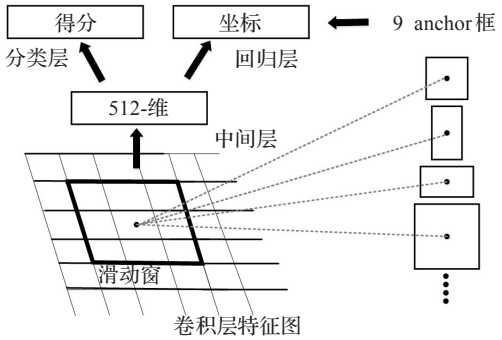


图3　anchor生成示意图

井下行人的数据来源于井下监控摄像头,监控视频中行人通常离摄像头位置较远,因此图片中行人的尺寸普遍偏小。为了使模型对小目标更加敏感,将anchor的scale修改为64/128/256,比例保持不变,生成9种不同的候选区域。

井下行人在经过摄像头过程中,由远到近会呈现出不同尺度变化,为了提高网络对多尺度目标的检测能力,本文提出了一种金字塔RPN结构。原始的RPN网络利用VGG16卷积层Conv5_3输出的最后一层特征图生成候选区域,经过3×3滑窗后每一个像素点的感受野是228×228。不能仅通过一种感受野来生成候选区域,不同尺度的目标可以使用不同大小的感受野来获得更好的候选区域。本文提出了在最后一层特征图上使用3种不同大小的滑动窗来生成候选区域,分别通过1×1、3×3、5×5卷积实现,如图3所示,将这种RPN结构,命名为“金字塔RPN”。

感受野[26]是卷积神经网络的每一层输出的特征图上像素点在原图像上映射的区域大小。Fisher和Valdlen[26]使用扩张卷积(Dilated Convolution)聚合多尺度的上下文信息提高了图像分割的准确率,其中扩张卷积的作用主要是在不损失图片信息的情况下增大感受野。卷积神经网络从2012年的7层LeNet[5]到发展到2015年152层的残差网络[27],使图像分类和检测性能大幅提升,一方面得益于网络结构的设计以及深度网络提

取到更加鲁棒的特征，另一方面网络越深导致感受野也越大。受上述相关工作的启发以及检测问题中存在多尺度问题，RPN网络可以使用感受野不同的特征图来定位不同尺度的目标（即生成目标的候选区域），因此提出了一种利用三种不同大小的卷积核的金字塔RPN结构，如图4所示。这样的结构设计对目标的多尺度可以更加鲁棒，从而提高整个模型的检测能力。

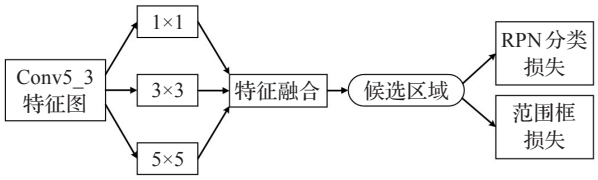


图4　“金字塔RPN”网络结构

3.4　特征融合

原始的Faster RCNN中，RPN网络生成的候选区域通过坐标映射到特征提取网络的最后一层特征图上，得到候选区域的特征图，经RoI Pooling层生成固定大小的特征图，送入网络后续部分实现目标的分类和区域框的回归，只利用了网络最深层的特征。卷积神经网络中深层的特征具有较强的语义特征，感受野也较大，是一种全局信息，粗粒度的特征；浅层的特征拥有更强的细节信息，是一种局部信息，细粒度的特征。

图像语义分割任务是一种基于像素点的预测，需要对每一个目标进行分类并从背景图片中分离出来。FCN[28]使用跨层连接（Skip Architecture）结构将全局信息和局部信息、粗粒度和细粒度特征相结合来改善预测的准确率，具体的操作就是将卷积神经网络不同层输出的特征图对应的像素相加。PSPnet[29]利用来自四个不同池化层的特征图，将全局和局部、不同层级的信息相结合，提高了区别不同目标类别的能力。这类结合特征信息的方式统称为“特征融合”，将不同卷积神经网络中不同层输出的特征图以不同的方式进行了组合。

井下数据集中行人较模糊，且行人之间会出现遮挡。受图像分割[25-26,28-29]中将不同层输出的特征图进行融合来提高模型的检测能力以及曹洁等人[23,30]利用特征融合实现人脸识别和人脸检测的启发，在煤矿井下行人检测任务中，可以将不同卷积层输出的特征图进行融合，来提高行人的检测性能。这里的特征融合使用的是特征图的拼接（Feature Concatenation），具体的操作方式在下文中会论述。井下行人数据集图片中行人较模糊，容易与背景相混叠，浅层的特征图包含一些局部信息，可以帮助行人的准确定位，仅使用深度的特征会使遮挡严重的行人漏检或者检测的位置不准。具体的实现过程如下：将候选区域映射到特征提取网络Conv5_3层和Conv4_3层生成的特征图中，得到候选区域位于这两层上的特征图，经RoI Pooling和L2正则化得到固定大小的特征向量，送入后续的全连接层实现行人的检测与范围框的回归。值得注意的一点是，如果去掉L2正则化，在实验过程中会导致网络过拟合。具体的细节如图5所示。

如图5中所示，这里特征融合使用的方法是拼接，将输入的特征图在指定维度进行堆叠。例如：输入两组大小为(N, C, H, W)的数据，输出的数据为$(N, 2C, H, W)$。其中N表示图片数量，C表示通道数，H和W分别表示特征图或者图片的高和宽。不足的是，特征融合会增加整个网络的运算量。特征融合将不同卷积层输出的特征图堆叠送入全连接层，因此网络的运算量主要增加在了与全连接层相连接的部分。RoI池化层在卷积层输出的特征图上进行池化操作，输出大小为7×7的特征图。由图5可知，RoI Pooling 4和5分别输出512张7×7的特征图，经特征融合后为1 024张7×7特征图送入全连接层，全连接层包含4 096个神经元，因此特征融合后全连接层共包含4 096×1 024×7×7=205 520 896，约2 000万个参数。原始的Faster RCNN中RoI Pooling仅在最后一层卷积层输出的特征图上进行池化操作并送入全连接层，此时全连接层包含4 096×512×7×7=102 760 448，约1 000万个参数。特征融合后的结构较原始结构增加了约1 000万个参数，因此在训练过程中会更加耗时，测试时检测速度会有所降低。

4　实验分析与结果

本文在井下行人检测数据集上评估了上述结构，以验证本文算法性能。并在VOC 07公测数据集上对改进的Faster RCNN煤矿井下行人检测算法进行了评估。

4.1　数据集

本文煤矿井下数据集来自于某煤业井下监控视频，整个数据集共包含23 210张图片，图片大小均为1 280×720，选择11 605张图片作为训练集，11 605张图片作为

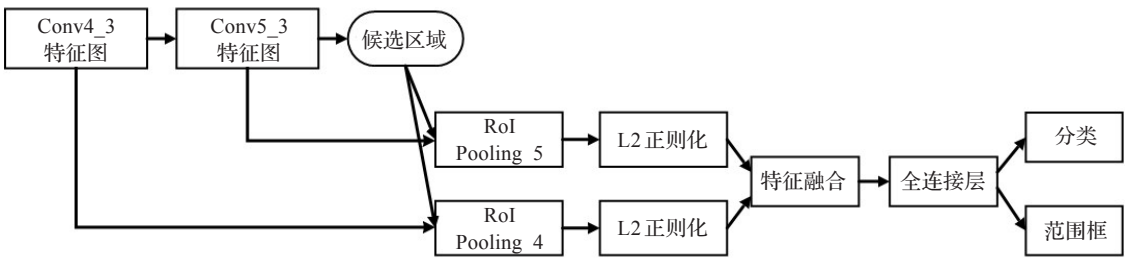


图5　不同层级特征融合结构

测试集。每张图片中行人个数从1~20人不等,包含各种尺度。

4.2 模型训练

网络实现部分本文选择主流的深度学习框架[31] caffe(https://github.com/BVLC/caffe)作为实验平台。按照目前基于深度学习的目标检测方法的标准惯用策略[7-8,10-11,30],选择在ImageNet分类[17,32]任务上预训练好的模型初始化训练网络。将ImageNet分类预训练所得到的VGG16卷积神经网络来初始化特征提取网络卷积层的权重。整个网络的训练过程使用SGD反向传播优化整个网络模型。学习率为0.001,momentum为0.9,weight_decay为0.000 5,每5万次迭代衰减一次学习率,衰减因子为0.1,共进行8万次的迭代。实验所用设备为ubuntu15.04,GeForce GTX 1070。

4.3 井下行人检测数据集结果及分析

最终本文提出的改进的Faster RCNN模型在井下行人检测数据集上取得了90%的平均检测准确率AP。训练的loss曲线如图6所示,网络的P-R曲线如图7所示。

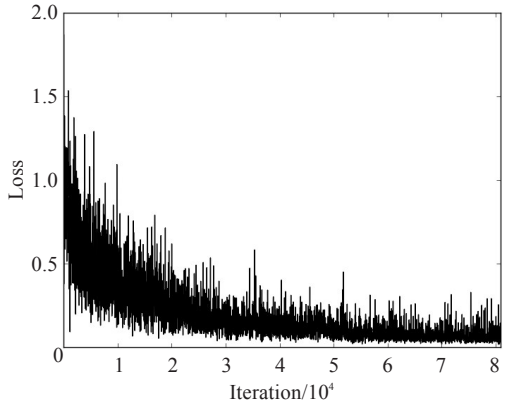


图6　Loss曲线

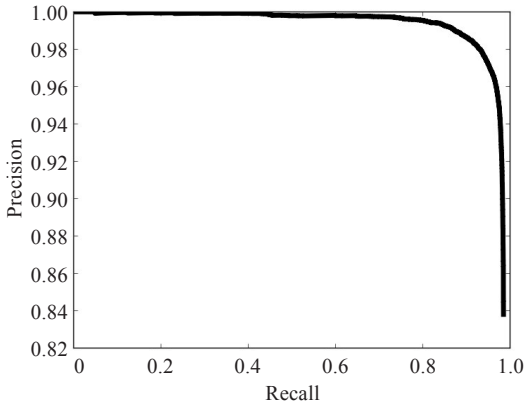


图7　P-R曲线

由图6中可以看出,经过8万次迭代本文的模型达到了收敛。本文算法在井下行人测试数据集获得了90%的平均检测准确率AP,P-R曲线如图7所示。通过分析,改进的Faster RCNN算法在煤矿井下数据集能获得90%的检测准确率,主要原因:(1)大量的训练数据集,训练样本包含1万多张图片,整个训练样本中包含约3万个标记的行人;(2)Faster RCNN利用PRN生成行人候选区域,后续网络再对后续区域进行分类的这种基于区域检测的优良结构;(3)针对不同尺度的矿井下行人,本文提出来的“金字塔”RPN网络结构,对不同尺度的行人生成候选区域;(4)针对图片中行人存在的模糊问题,使用了特征融合的方式作为候选区域的特征进行分类与范围框的回归。最终本文的改进算法取得了上述的检测结果。

4.4 检测结果展示

图8中左侧图片是改进的Faster RCNN煤矿井下行人检测算法在测试数据集上的检测效果。从图中可以观察到改进算法能够很好地检测出行人,并非常准确地标记出具体的位置。为了验证改进算法的先进性,将井下数据集在YOLO目标检测算法[10]上训练得到了一个基于YOLO的井下行人检测模型。YOLO也是基于深度学习的通用目标检测框架,使用回归的方式直接进行目标的预测。在测试集上选取了4张具有代表性的图片,分别利用改进的Faster RCNN和YOLO算法进行检测,检测结果如图8所示。左侧图片是利用改进的算法获得的检测结果;右侧图片是利用YOLO获得的检测结果。由图8可得,在前两张图片中,两种算法的检测性能相当,都准确地检测出图片中所有的行人;第3张图片当行人出现遮挡情况下,改进Faster RCNN仍能准确地检测出行人,而YOLO出现了漏检;第4张图片上包含了两个小尺度的行人,改进的Faster RCNN准确地检测出两个行人,YOLO出现了漏检且位置不够精确。从图8可以直观地看出,在行人被遮挡、小目标的行人数据上改进的Faster RCNN算法性能优于YOLO。

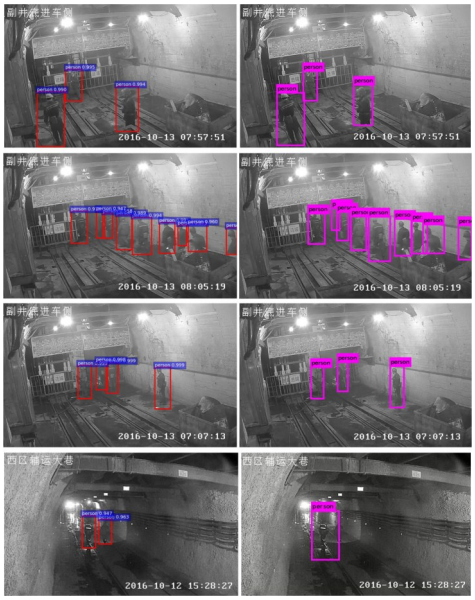


图8　改进的Faster RCNN与YOLO在井下行人测试数据集上的检测结果

为了证明本文模型的鲁棒性，从百度图片中找了一张煤矿井下的图片，该图片与井下行人数据集分布不同且是彩色图片。利用改进的Faster RCNN和YOLO井下行人检测模型分别对该图片进行检测，实验结果如图9所示。YOLO检测结果出现了错判（图9右侧），改进的Faster RCNN准确地检测出图中的行人，这充分说明改进的Faster RCNN算法鲁棒性更好。

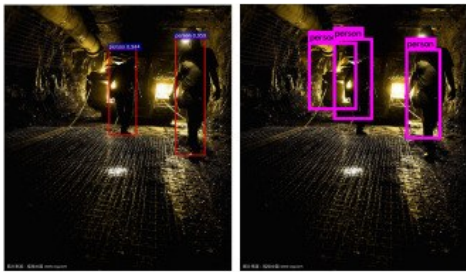


图9　与井下数据集不同分布图片上检测结果

4.5　可视化特征图

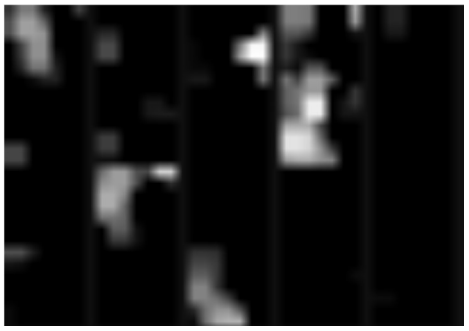
可视化理解卷积神经网络已经超出了本文的研究范畴，相关的研究工作可以参考Zeiler和Fergus可视化卷积神经网络的研究工作[33]，其分析了卷积神经网络的每一层学习到了什么样的特征。这里对本文中使用的算法网络中间层输出的特征图进行了可视化，结果如图10所示。

图10是可视化了网络中不同层输出的特征图，每一层都会输出大量的特征图，这里只挑选了部分图片来展示。输入的图片如图(a)所示，图(b)是Conv1_1层输出的特征图，图(c)是Conv5_3层输出的特征图，图(d)是RoI Pooling层输出的特征图，图(e)是RPN分类层输出的特征图。由表1可知Conv1_1层共有64个卷积核，每一个卷积核输出一张大小为600×1 000的特征图，这里只可视化了其中的25张特征图，如图(b)所示。同样由表1可知Conv5_3层输出512张大小为38×63的特征图，这里同样只可视化了其中的25张，如图(c)所示。观察图(b)、(c)、(d)可以发现每一张特征图都不相同，这是因为卷积神经网络不同的卷积核会提取到不同的特征。观察图(b)中的特征图，用肉眼能够看出所提取的特征，可以理解这些特征的意义；但是观察图(c)中的特征图，已经无法用肉眼解释这些特征的意思。这是因为卷积神经网络中浅层会提取一些低级特征，而深层提取的是图像中的高级语义特征。更多该方面的研究可以参考前文提到的Zelier和Fergus的工作。由图(a)可知输入的图片中共有3人，观察图(e)可以发现每一幅特征图中白色激活部分映射到图(a)中恰好是图片每一个行人的位置，由此可以发现RPN网络可以非常准确地定位行人的位置，图片中的其他信息都被判为了背景。

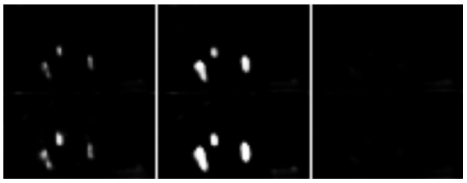
(a)输入的原始图片

(b)Conv1_1层输出的特征图

(c)Conv5_3层输出的特征图



(d)RoI Pooling层输出的特征图



(e)RPN分类层输出的特征图

图10　卷积神经网络中不同层输出的特征图

4.6　VOC 07数据集上的实验

为了验证本文改进算法的有效性，进一步在公测数据集VOC 07上进行了验证。VOC 07数据集共包括9 963张图片，20个类别，共标注了24 640个目标，尺寸基本为375×500的自然场景下的图片。模型训练的参数设置、实验平台与4.2节相同，本文复现了原始的Faster RCNN结构；分别实现了3.3节所述的改进的“金字塔RPN”结构和3.4节所述的特征融合。不同模型在VOC 07上的检测结果如表2所示。

观察表2中第2、3列数据，“金字塔RPN网络”较原

始算法在VOC 07数据集上总的平均检测性能仅提高了约0.7%,性能提升不明显。值得注意的是,瓶子的检测性能降低了5%,而椅子的性能竟然提高了15%,其他类别的目标性能基本保持不变或小幅提升。针对这一现象,对VOC 07数据集中瓶子和椅子两个类别的数据进行了分析。VOC 07数据集中包含瓶子的图片共502张,其中244张图片被用来训练,标注的数据中有大量的瓶子尺寸较小;VOC 07椅子类别的数据共包含1 117张,其中445张图片共包含了798个目标被用来训练,尺寸普遍较大且通常有不同尺度遮挡,因为数据集中椅子上会坐有人或者摆放着其他物品等。经过分析认为:“金字塔RPN结构”使用三种不同尺度的卷积核来定位目标,5×5的卷积核在原始图片上感受野太大,瓶子的尺寸较小,可以划分为小目标检测,太大的感受野会损害小目标的检测性能;数据集椅子尺度较大且尺度不一,“金字塔RPN结构”中5×5卷积核可以帮助定位尺寸较大的目标,三种尺度的卷积核对不同尺度的目标更加鲁棒,因此“金字塔RPN”检测椅子的性能较原始算法大幅提高。分析表2中第4列,特征融合利用了图片的浅层特征和深层高级语义特征,浅层特征可以利用一些细节信息帮助目标的定位,不同幅度地提高了VOC 07数据集上各个类别的检测性能。表2中第5列数据是将“金字塔RPN”和特征融合结合后在VOC 07数据集上的检测结果。

表2　不同模型VOC 07数据上的检测结果　%

模型	Faster RCNN	Faster RCNN+金字塔RPN	Faster RCNN+特征融合	Faster RCNN+金字塔RPN+特征融合
飞机	66.17	68.83	69.43	70.88
自行车	78.13	78.41	79.86	79.69
鸟	67.83	68.64	69.97	69.02
轮船	56.25	56.75	58.37	57.43
瓶子	**50.54**	**45.71**	**59.79**	55.38
公交车	75.94	78.15	76.63	79.51
小轿车	79.65	79.90	80.04	81.27
猫	86.96	83.74	87.74	86.30
椅子	**49.30**	**64.16**	52.07	66.72
奶牛	75.77	74.16	77.49	77.07
桌子	64.36	64.16	65.32	66.74
狗	81.05	79.26	82.57	81.37
马	80.18	80.58	81.09	82.26
山地车	72.59	74.80	76.55	74.96
人	**76.84**	**77.85**	**78.68**	**79.76**
植物	41.76	41.43	41.63	41.85
绵羊	67.14	67.98	69.25	70.02
沙发	64.83	64.49	65.03	65.96
火车	74.85	74.86	76.98	76.71
电视机	71.31	71.68	73.65	73.03
平均值	69.10	69.77	71.16	71.80

本文的研究主要针对井下行人检测,因此特别关注改进的Faster RCNN算法在VOC 07人类别数据上的检测结果。本文绘制了改进算法在VOC 07人类别上的P-R曲线,如图11所示。同时VOC 07人类别的数据与井下的数据集分布不同,都是全高清彩色的图片,涵盖了各类姿态的人,包括大尺度人脸、行人等,而井下的数据只包含行人。

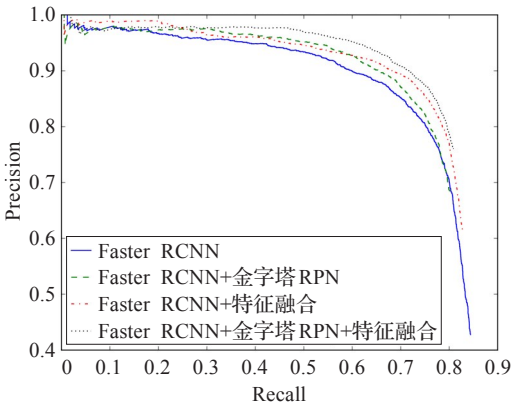


图11　不同模型在VOC 07人类别上P-R曲线图

整个VOC 07数据集中有4 192张图片包含人,其中2 096张用于训练,2 096张用于测试。由表2和图11可以得到,本文的改进算法是有效的,在自然场景下改进算法将原始的Faster RCNN性能提升了约3%,性能优于原始算法。同时也证明了改进算法也适用于自然场景下的行人检测。

5　结论与展望

本文以Faster RCNN为基础,提出了一种改进的Faster RCNN煤矿井下行人检测算法。井下行人尺寸偏小,首先对anchor的大小进行了调整;进一步对RPN网络结构进行了改进,提出了一种“金字塔RPN”网络结构;最后利用特征融合技术,将底层特征和高层语义特征进行融合来共同实现目标的分类。最终在井下行人检测数据集上获得了90%的检测准确率,并在VOC 07行人类别数据上对本文的算法进行了验证。下一步研究方向包括井下行人姿态检测,工业视频的智能化应用,这将有助于煤矿安全管理和智慧煤矿的发展。

参考文献:

[1] Dalal N,Triggs B.Histograms of oriented gradients for human detection[C]//Proc IEEE Conf Comput Vis Pattern Recognit,2005:886-893.

[2] Mita T,Kaneko T,Hori O.Joint Haar-like features for face detection[C]//10th IEEE International Conference on Computer Vision,2005:1619-1626.

[3] Ahonen T,Hadid A,Pietikäinen M.Face recognition with local binary patterns[C]//European Conference on Com-

puter Vision.Berlin，Heidelberg：Springer，2004：469-481.
[4] Dollár P，Tu Z，Perona P，et al.Integral channel features[C]// British Machine Vision Conference，London，Sep 7-10，2009.
[5] Krizhevsky A，Sutskever I，Hinton G E.Imagenet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems，2012：1097-1105.
[6] Lecun Y，Boser B，Denker J S，et al.Backpropagation applied to handwritten zip code recognition[J].Neural Computation，2014，1(4)：541-551.
[7] Li Y，He K，Sun J.R-FCN：object detection via region-based fully convolutional networks[C]//Advances in Neural Information Processing Systems，2016：379-387.
[8] 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.
[9] 闫喜亮，王黎明.卷积深度神经网络的手写汉字识别系统[J].计算机工程与应用，2017，53(10)：246-250.
[10] 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.
[11] Liu W，Anguelov D，Erhan D，et al.SSD：single shot multibox detector[C]//European Conference on Computer Vision.Cham：Springer，2016：21-37.
[12] 熊丽婷，张青苗，沈克永.基于搜索区域条件概率CNN的精确目标探测方法[J].计算机工程与应用，2017，53(20)：134-140.
[13] 杜玉龙，李建增，张岩，等.基于深度交叉CNN和免交互GrabCut的显著性检测[J].计算机工程与应用，2017，53(3)：32-40.
[14] Li J，Liang X，Shen S M，et al.Scale-aware Fast R-CNN for pedestrian detection[J].IEEE Transactions on Multimedia，2018，20(4)：985-996.
[15] Everingham M，Gool L，Williams C K，et al.The Pascal visual object classes（VOC） challenge[J].International Journal of Computer Vision，2010，88(2)：303-338.
[16] Lin T Y，Maire M，Belongie S，et al.Microsoft COCO：common objects in context[C]//European Conference on Computer Vision.Cham：Springer，2014：740-755.
[17] 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.
[18] Ren S，He K，Girshick R，et al.Faster R-CNN：towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems，2015：91-99.
[19] Dollár P，Wojek C，Schiele B，et al.Pedestrian detection：an evaluation of the state of the art[J].IEEE Transactions on Pattern Analysis & Machine Intelligence，2012，34(4)：743-761.
[20] Benenson R，Omran M，Hosang J，et al.Ten years of pedestrian detection，what have we learned?[C]//European Conference on Computer Vision.Cham：Springer，2014：613-627.
[21] Hosang J，Omran M，Benenson R，et al.Taking a deeper look at pedestrians[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2015：4073-4082.
[22] 宋焕生，张向清，郑宝峰，严腾.基于深度学习方法的复杂场景下车辆目标检测[J/OL].[2017-03-31].http://www.arocmag.com/article/02-2018-04-004.html.
[23] Sun X，Wu P，Hoi S C H.Face detection using deep learning：an improved Faster RCNN approach[J].arXiv：1701.08289，2017.
[24] He K，Zhang X，Ren S，et al.Spatial pyramid pooling in deep convolutional networks for visual recognition[J].IEEE Transactions on Pattern Analysis & Machine Intelligence，2015，37(9)：1904.
[25] He K，Gkioxari G，Dollár P，et al.Mask RCNN[J].arXiv：1703.06870，2017.
[26] Yu F，Koltun V.Multi-scale context aggregation by dilated convolutions[J].arXiv：1511.07122，2015.
[27] He K，Zhang X，Ren S，et al.Identity mappings in deep residual networks[C]//European Conference on Computer Vision.Springer International Publishing，2016：630-645.
[28] Long J，Shelhamer E，Darrell T.Fully convolutional networks for semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition，Boston，2015：3431-3440.
[29] Zhao H，Shi J，Qi X，et al.Pyramid scene parsing network[C]//IEEE Conference on Computer Vision and Pattern Recognition，Hawaii，2017：2881-2890.
[30] 王娇娇，刘政怡，李辉.特征融合与objectness加强的显著目标检测[J].计算机工程与应用，2017，53(2)：195-200.
[31] Jia Y，Shelhamer E，Donahue J，et al.CAFFE：convolutional architecture for fast feature embedding[C]//ACM International Conference on Multi-Media，2014：675-678.
[32] 桑军，郭沛，项志立，等.Faster-RCNN的车型识别分析[J].重庆大学学报(自然科学版)，2017，40(7)：32-36.
[33] Zeiler M D，Fergus R.Visualizing and understanding convolutional networks[C]//European Conference on Computer Vision.Springer International Publishing，2014：818-833.