

基于 SE Detection Net 的安全帽检测方法

刘霞

(中国石油大学(华东)计算机与通信工程学院, 山东 青岛 266580)

摘要:在日常的生产中,安全帽对工人的安全提供了保障。为了减少因未佩戴安全帽而引起的安全事故的发生,安全帽的识别在工人安全生产方面具有极高的应用价值。针对利用传统的机器学习方法对安全帽的检测效果不理想的问题,提出一种全卷积深度神经网络:分辨率增强检测网络,对工人是否佩戴安全帽进行检测。该方法利用 VGG16 网络中的三个不同层次的特征图,采用提出的分辨率增强模块,使三个特征图的分辨率达到一致;将此三个特征图根据通道数相连的方式进行融合;利用融合后的特征图生成特征金字塔,用于多尺度安全帽的检测。实验表明,相比于常见的神经网络检测模型,此方法不仅实现了实时检测安全帽的速度要求,同时具有较高的检测准确率,提高了小尺度安全帽的召回率和大尺度安全帽的检测精度。

关键词:多尺度安全帽检测;特征融合;特征金字塔;分辨率增强模块;卷积神经网络

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2020)06-0156-04

doi:10.3969/j.issn.1673-629X.2020.06.030

Safety Helmet Detection Method Based on SE Detection Net

LIU Xia

(School of Computer and Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: In the daily production, the safety of the workers is guaranteed by the safety helmet. In order to reduce the occurrence of safety accidents caused by not wearing safety helmet, the identification of safety helmet has a high application value in the safety production of workers. Aiming at the problem that the traditional machine learning method is not ideal for safety helmet detection, a full convolution depth neural network, resolution enhanced detection network, is proposed to detect whether workers wear safety helmet or not. This method uses three different levels of feature maps in VGG16 network and the proposed resolution enhancement module to achieve the same resolution of the three feature maps. The three feature maps are fused by connecting the number of channels. The fused feature maps are used to generate feature pyramids for multi-scale safety helmet detection. Experiments show that compared with common neural network detection model, this method not only achieves the speed requirement of real-time detection of safety helmet, but also has high detection accuracy, and improves the recall rate of small-scale safety helmet and the detection accuracy of large-scale safety helmet.

Key words: multi-scale safety helmet detection; feature fusion; feature pyramid; resolution enhancement module; convolutional neural network

0 引言

在日常的生产过程中,如建筑工地、电厂等,由于场景环境较为复杂,存在着威胁人身安全的各种因素,由于头部是人体中最易受到伤害的部位,在上述场合中工作的人员均要求佩戴安全帽^[1]。智能化检测工人是否佩戴安全帽,能有效提高监管人员对安全帽佩戴情况的监管效率,大幅度减少了人工参与,一定程度上解放了劳动力;对工人本身而言,提供了安全保障。

现有的安全帽识别技术主要有冯国臣^[2]、胡恬^[3]、刘晓慧^[4]等基于传统方法对安全帽进行检测。

以上方法都是基于人为主观设计的特征,需要研究者具备丰富的实践经验和牢固的专业知识。人为设计特征,不但费时费力,而且泛化能力不好,不具备鲁棒性。

随着近年来深度学习方面相关技术的蓬勃发展,越来越多的研究者将深度学习技术应用到图像分类^[5]、目标识别^[6]、图像分割^[7]等复杂任务上,并且都取得了相当可观的效果。但是对于这些算法而言,目标的尺度变化仍然是一个挑战。现有的用以解决多尺度目标检测的方法,如:SSD^[8]、DSSD^[9]、FPN^[10]、ION^[5]、HyperNet^[11]、SNIP^[12]、SNIPER^[13]和 Trident-

收稿日期:2019-07-25

修回日期:2019-11-25

基金项目:国家自然科学基金(61402533);山东省自然科学基金资助项目(ZR2019MF049)

作者简介:刘霞(1994-),女,硕士研究生,研究方向为计算机视觉。

Net^[14]等存在着网络结构复杂,计算量大的问题。虽然精度提升了,但检测效率低下,增加了检测时间。

针对上述问题,文中在 VGG16 神经网络的基础上,选用 Conv4_3、Conv6_2 和 Conv7_2 卷积层;利用分辨率增强模块,使后两个卷积层的特征图的大小和第一个卷积层中的特征图的大小保持一致;然后对转换后的特征图进行融合,这样融合后的特征图既包含高层语义信息,又包含低层细节信息;最后再根据融合后的特征图生成特征金字塔对安全帽进行检测,实现对小尺度安全帽的高召回,对大尺度安全帽的精确定位。

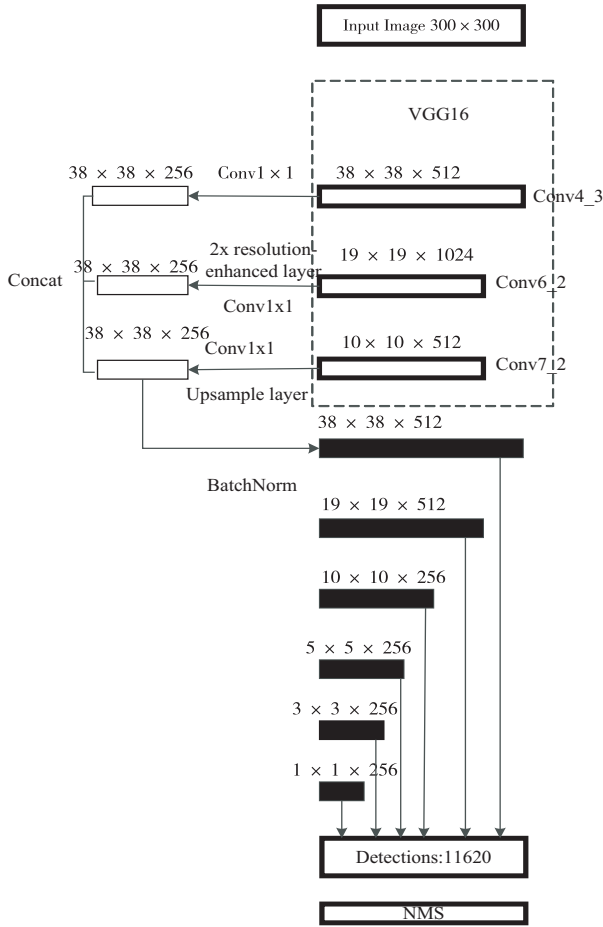


图1 SE Detection Net 网络结构

1 SE Detection Net 网络结构

SE Detection Net 网络结构如图1所示。它以 VGG16 为特征提取网络,利用其中的 Conv4_3、添加的 Conv6_2、Conv7_2 特征图进行特征融合,利用融合后的特征图生成特征金字塔,用于不同尺度的安全帽的检测,最后利用 NMS (non—maximum suppression) 算法,筛选重复的候选框,得到最后的检测结果。分辨率增强模块包括分辨率增强层和上采样层,分辨率增强层在不增加计算量的前提下,扩大了 Conv6_2 卷积层中特征图的大小,并减少了通道数,上采样层通过插

值的方法对 Conv7_2 层实现了从低分辨率到高分辨率的转化。

1.1 VGG16 网络

文中采用的特征提取网络是 VGG16^[15]。根据组成网络的层数,16 层的 VGG 命名为 VGG16。VGG16 网络结构比较规整,是由多个卷积层加上一个池化层叠加组成的结构。VGG16 使用 3×3 小卷积核和 2×2 小池化核。它总共包含 5 组卷积,每组卷积由 2 个或者 3 个卷积层组成,后面的最大池化层用于减少参数。另外,在每组卷积中,卷积核的数量是固定的。从前往后,每组的卷积核的数量依次是 64、128、256、512、512。为了提高网络提取特征的能力,删除了 VGG16 网络中的最后一个池化层和三个全连接层,增加了两个卷积块,每个卷积块都有两个卷积层组成,最终网络生成的特征图的大小为 10×10×512。

1.2 分辨率增强模块

特征图的分辨率问题是目标检测中的核心问题。结合不同分辨率的特征图,有利于多尺度目标的识别。在进行特征融合前,特征图的尺寸需保持一致。分辨率增强模块包含两个部分:一部分是分辨率增强层,另一个是上采样层。对于上采样层,先利用 1×1 的卷积核把通道数从 1 024 减少到 256,有利于训练参数的减少,降低模型复杂度,然后利用插值方法,把特征图从 10×10 扩大到 38×38。

为了得到高分辨率的特征图,还不增加额外的计算量,设计了分辨率增强层,该层对于不同分辨率的图像的转换非常高效,可以直接插入到 VGG16 模型中。假设输入的特征图的维度为: $h \times w \times c \times g^2$,其中 g 是上采样因子。如图2所示,分辨率增强层通过减少特征图通道数对特征图的宽度和高度进行扩大。用数学形式表示如下:

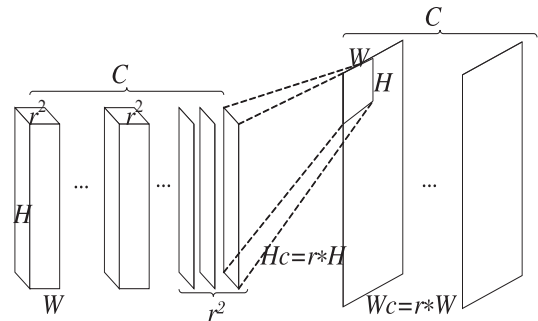


图2 分辨率增强模块

$$f(I^{LR}) = O^{SR} \quad (1)$$

每个具体的转换函数如下:

$$f(I_w^{LR}) = O_w^{SR} \quad (2)$$

$$f(I_h^{LR}) = O_h^{SR} \quad (3)$$

$$f(I_c^{LR}) = O_c^{SR} \quad (4)$$

其中, O^{SR} 为高分辨率特征图, I^{LR} 为低分辨率特征图。 w, h, c 为输入的低分辨率图像的宽、高、通道数, $w * g, h * g, c / g^2$ 分别为转化后的高分辨率图像的宽、高、通道数。分辨率增强模块有效地减少了通道数, 增大了图像的尺寸。这种方式没有引入额外的参数, 也没有增加计算量。

在分辨率增强层中, Conv7_2 的特征图的大小为 $19 \times 19 \times 1024$, 取上采样因子 $g = 2$, 因此特征图的大小从 $19 \times 19 \times 1024$ 转换成 $38 \times 38 \times 256$, 然后再通过大小为 1×1 的卷积核, 对变化后的特征图进行特征的整合, 生成 $38 \times 38 \times 256$ 的特征图。经过转化后的三处特征图在通道数维度上进行融合, 随后通过 BatchNorm^[16] 层正则化融合后的特征图, 最终生成的特征图大小为: $38 \times 38 \times 768$ 。

1.3 特征金字塔

对融合后大小为 $38 \times 38 \times 768$ 的特征图进行卷积操作后, 生成特征金字塔。然后 SE Detection Net 使用两个卷积核大小为 3×3 的卷积层对特征金字塔中每一层进行卷积。其中一个卷积层输出每个区域候选框的种类作为类别的预测结果, 另一个则输出每个区域候选框的位置偏移量, 再根据每个区域候选框的置信度排序, 利用 NMS 算法去除多余的框, 留下最有可能存在目标的框, 根据预测出的偏移量调整候选框的形状和位置, 最后生成最终的检测结果。

2 实验

这个数据集总共有 3 个种类: 蓝色安全帽、红色安全帽和黄色安全帽。训练集总共有 20 500 张, 测试集有 1 000 张。其中蓝色安全帽数据集共 6 500 张, 红色 6 500 张, 黄色 6 500 张。

选用在 ImageNet^[17] 数据集训练得到的 VGG16 为初始模型, 然后开始训练此网络。模型输入的图片尺寸为: 300×300 , 由于 GPU 显存的限制, 训练时的 batch_size: 28, 总共的训练次数为 100 个 epoch, 在前 30 个 epoch 的学习率为 10^{-3} , 在以后的迭代中, 学习率依照余弦函数进行衰减^[18], 在第 100 次 epoch 时, 学习率下降为 0。

2.1 检测效果分析

在收集的安全帽数据集上评价此网络, 采用 mAP 分数^[19] 作为评价检测性能的指标。在整个测试集中, 红色安全帽大部分是大尺度, 蓝色安全帽各个尺度都有, 黄色安全帽小尺度居多。

表 1 展示了在安全帽测试集上的检测结果。在这些算法中, SSD 没有融合高低层的特征, 仅利用不同层次的单一层进行检测。从表中可以看出, SE Detection Net 相较于 SSD 对这三种类别的安全帽都有

提升, 表明 SE Detection Net 采用融合后的特征图生成特征金字塔, 促使了低层的特征图含有语义信息, 高层的特征图包含了细节信息。其中, 黄色安全帽提升最大, 提高了 7.6%, 说明 SE Detection Net 对小目标的检测效果提升明显。另外, Faster RCNN 虽然使用 RPN 网络 (region proposal network) 进行候选框的选择, 但只使用顶端的一个特征图进行检测, 而 SE Detection Net 使用多个特征图进行检测, 提高了各个尺度的安全帽的检测效果, 从表中可以看出, SE Detection Net 相比于 Faster RCNN 在类别上的提升分别为: 1.5%、0.7%、3.9%。从表中可以得出, 在这三种方法中 SE Detection Net 的检测效果最好, 提高了安全帽检测的精度。

表 1 各个模型的检测结果

方法	mAP	bluehat	redhat	yellowhat
Faster RCNN	83.9	85.6	89.1	77.9
SSD	81.7	84.4	86.5	74.2
SE Detection Net	85.8	87.1	88.4	81.8

为了更加直观地观察 SE Detection Net 的检测效果, 文中在黄色安全帽和红色安全帽测试集上做了 SSD 和 SE Detection Net 的对比实验, 检测效果如图 3 所示。SSD 由于只使用单一特征图做预测, 低层特征图缺乏语义信息, 高层特征图缺少细节信息, 对于小尺度的黄色安全帽没有检测出来, 存在漏检现象; 对于大尺度的红色安全帽虽然检测出来, 但是定位没有 SE Detection Net 精确。SE Detection Net 由于有效的特征融合, 使低层的特征图含有语义信息, 高层的特征图包含细节信息, 提高了小尺度的安全帽的召回率和大尺度安全帽的置信度, 并且定位更加精准。

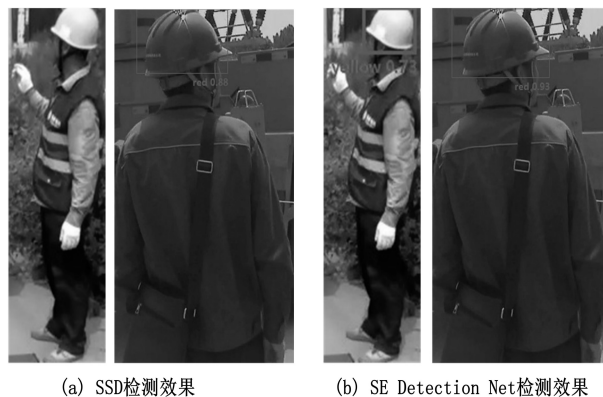


图 3 检测效果对比

2.2 检测速度分析

为了测试 SE Detection Net 的速度, 测试了整个测试集的检测时间, 然后计算了每秒的帧数 (fps)。把 batch_size 设置为 1, 并在 GPU 为 1080Ti 的机器上测试速度。

表2显示了基于安全帽测试集 SE Detection Net 和其他算法关于速度和精度的比较。从表中可以看出,SE Detection Net 实现了精度和速度的均衡。它融合了三个特征图的信息,增加了时间的开销,虽然没有 SSD 的检测速度快,但精度提高了 4.1%。Faster RCNN 是二阶段检测器的代表,它先由 RPN 网络生成区域候选框,再进行分类和回归,因此它的检测时间较长。SE Detection Net 利用卷积操作直接进行分类和回归,实现了比 Faster RCNN 更理想的检测效果,且识别速度增加了一倍。

表2 基于测试集速度与精度的比较

方法	mAP	Speed/fps
Faster RCNN	83.9	9.7
SE Detection Net	85.8	20.9
SSD	81.7	22.7

3 结束语

为了解决安全帽检测问题,提出了分辨率增强检测网络(SE Detection Net)用于多尺度安全帽的识别。SE Detection Net 先由 VGG16 提取安全帽的特征,然后利用分辨率增强模块把需要融合的低分辨率特征图转化到高分辨特征图,再通过融合后的特征图生成特征金字塔,最后利用金字塔中的特征图实现安全帽的检测。实验结果表明 SE Detection Net 的效果优于 SSD 和 Faster RCNN,对各个尺度的安全帽的检测效果都有提升,其中,小尺度安全帽的提升效果较为明显。检测效率也能达到实时性的要求,实现了检测精度和速度的平衡。

参考文献:

- [1] 贾峻苏,鲍庆洁,唐慧明. 基于可变形部件模型的安全头盔佩戴检测[J]. 计算机应用研究,2016,33(3):953-956.
- [2] 冯国臣,陈艳艳,陈宁,等. 基于机器视觉的安全帽自动识别技术研究[J]. 机械设计与制造工程,2015,44(10):39-42.
- [3] 胡恬. 利用几何分析法和 BP 神经网络进行人脸识别的研究[J]. 计算机工程与设计,2012,23(9):18-21.
- [4] 刘晓慧,叶西宁. 肤色检测和 Hu 矩在安全帽识别中的应用[J]. 华东理工大学学报:自然科学版,2014,40(3):365-370.
- [5] BELL S, ZITNICK C L, BALA K, et al. Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016:2874-2883.
- [6] 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. Columbus: IEEE, 2014:580-587.
- [7] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Hawaii: IEEE, 2017:3431-3440.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European conference on computer vision. Berlin: Springer, 2016:21-37.
- [9] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector[C]//Computer vision and pattern recognition. [s. l.]: [s. n.], 2017:1701.06659.
- [10] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Computer vision and pattern recognition. Hawaii: IEEE, 2017:1612.03144.
- [11] KONG T, YAO A, CHEN Y, et al. Hypernet: towards accurate region proposal generation and joint object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2017:845-853.
- [12] SINGH B, DAVIS L S. An analysis of scale invariance in object detection snip[C]//Computer vision and pattern recognition. Salt Lake City: IEEE, 2018:1711.08189.
- [13] SINGH B, NAJIBI M, DAVIS L S. SNIPER: efficient multi-scale training[C]//Advances in neural information processing systems. [s. l.]: [s. n.], 2018:9310-9320.
- [14] LI Y, CHEN Y, WANG N, et al. Scale-aware trident networks for object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2019, 16(4):4694-4702.
- [15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International conference on learning representations. San Diego, USA: ICLR, 2015:1-5.
- [16] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 16(4):4694-4702.
- [17] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. [s. l.]: NIPS, 2012:1097-1105.
- [18] LI H, LIU Y, OUYANG W, et al. Zoom out-and-in network with recursive training for object proposal[C]//Computer vision and pattern recognition. [s. l.]: [s. n.], 2017:1702.05711.
- [19] CHEN X, FANG H, LIN T Y, et al. Microsoft COCO captions: data collection and evaluation server[C]//Computer vision and pattern recognition. [s. l.]: [s. n.], 2015:1504.00325.