

文章编号 1004-924X(2019)05-1196-10

## 基于改进 YOLOv2 的快速安全帽佩戴情况检测

方 明\*, 孙腾腾, 邵 桢

(长春理工大学 计算机科学技术学院, 吉林 长春 130022)

**摘要:**施工现场光照多变、背景复杂、施工人员形态多样,给安全帽佩戴情况检测带来很大的困难。针对传统检测方法准确率低、鲁棒性差的问题,本文提出了一种基于深度学习的安全帽佩戴情况检测方法。该方法以 YOLOv2 目标检测方法为基础,对其网络结构进行了改进。首先借鉴了密集连接网络思想,在原网络中加入了密集块,实现了多层特征的融合以及浅层低语义信息与深层高语义信息的兼顾,提高了网络对于小目标检测的敏感性;然后,利用 MobileNet 中的轻量化网络结构对网络进行压缩,使模型的大小缩减为原来的十分之一,增加了模型的可用性。采用自制的 HelmetWear 数据集对改进后的网络模型进行训练和测试,并将该模型与原 YOLOv2 和最新的 YOLOv3 进行了对比,结果显示:该模型的检测准确率为 87.42%,稍逊色于 YOLOv3,但是其检测速度提升显著,比 YOLOv2 和 YOLOv3 分别提高了 37% 和 215%,可达 148 frame/s。实验表明,改进后的网络模型能在保证检测准确率的同时,有效减小参数量,显著提升检测速度。

**关键词:**深度学习;目标检测;安全帽检测;密集连接网络;MobileNet

**中图分类号:**TP391.4 **文献标识码:**A **doi:**10.3788/OPE.20192705.1196

## Fast helmet-wearing-condition detection based on improved YOLOv2

FANG Ming\*, SUN Teng-teng, SHAO Zhen

(School of Computer Science and Technology, Changchun University of  
Science and Technology, Changchun 130022, China)

\* Corresponding author, E-mail: fangming@cust.edu.cn

**Abstract:** Construction sites comprise changeable illumination, complicated backgrounds, and various types of construction personnel, which makes the detection of helmet wearing challenging. To address the problems of low accuracy and poor robustness of traditional detection methods, this paper proposed a method of helmet wearing detection based on deep learning. The proposed method was based on the YOLOv2 target detection, and its network structure was improved. First, utilizing the notion of densely connected networks, dense blocks are added to the original network, which aids in the realization of the fusion of multi-layer features and the combination of shallow low semantic information and deep high semantic information, thereby improving the network sensitivity to enable the detection of small targets. Subsequently, the lightweight network structure in MobileNet was used for network compression, thus reducing the size of the model to one tenth of its original and also increasing model availability. The improved network model was trained and tested on the self-made HelmetWear data-

收稿日期:2018-10-16;修订日期:2018-12-14.

基金项目:吉林省重点科技成果转化项目资助(No. 20170307002GX)

set and compared with the original as well as the latest YOLOv3. The obtained results show that the detection accuracy of the model is 87.42%, which is slightly lower than that of YOLOv3, but its detection speed is significantly improved, i. e., 37% and 215% higher than that of YOLOv2 and YOLOv3, respectively, while reaching 148 frames/s. Experiments confirm that the proposed model can effectively reduce parameter quantity and significantly enhance detection speed while ensuring detection accuracy.

**Key words:** deep learning; target detection; helmets detection; densely connected networks; MobileNet

## 1 引言

安全帽是一种常见的头部防护工具,在施工现场,佩戴安全帽对保护施工人员的头部安全起着至关重要的作用。但是由于一些施工人员缺乏安全意识,不按要求佩戴安全帽,导致很多安全事故的发生。因此,对施工现场安全帽佩戴情况进行实时检测意义重大。然而,目前主要的检测方法仍然依赖于人工监视,由于现场情况复杂、监视人员易疲劳等原因,很容易导致漏检的发生。

利用计算机视觉技术对安全帽佩戴情况进行检测是当前研究的热点。国内学者已对此做了大量工作:刘晓慧等<sup>[1]</sup>采用肤色检测的方法定位到人脸区域,然后提取脸部以上的Hu矩特征向量,最后用SVM完成对安全帽的识别。但是该方法受视角影响较大,应用场景单一,主要限制在工地的出入口处。Park等<sup>[2]</sup>利用背景差法提取前景,基于HOG特征和SVM进行人体检测,同样利用这两类特征对安全帽进行检测,检测出人体和安全帽以后,根据他们的空间关系和几何关系进行匹配,从而确定工作人员的安全帽佩戴情况。但对于工作人员非站立、被遮挡、静止不动等情况,该方法有一定的局限性。周艳青等<sup>[3]</sup>先对工地进出口的视频中的人头区域进行截取,并提取该区域的统计特征、局部二进制模式特征、快速主成分分析特征,再利用分类器和BP人工神经网络进行分类预测。但是该方法对于背景比较复杂的情况,识别率还有待提高。以上基于传统图像处理和机器学习的方法,存在手工设计特征困难、泛化能力差等问题。

近年来,基于深度学习的目标检测方法取得了巨大突破<sup>[4]</sup>,该方法可以在训练样本的驱动下自动地学习有价值的特征来帮助完成目标检测,

泛化能力比较强。基于深度学习的目标检测方法主要包括两类,一类是基于区域提名的,首先生成候选框,然后对每个候选框进行分类,主要研究有R-CNN(Region-based Convolutional Neural Networks)<sup>[5-7]</sup>、SPP-net(Spatial Pyramid Pooling networks)<sup>[8]</sup>和R-FCN(Region-based Fully Convolutional Network)<sup>[9]</sup>等。另一类是基于端到端的,对整张图像进行预测,一次性完成对目标的定位与分类,该类方法的速度要快于前一类,但是精度要差一些,主要研究有YOLO(You Only Look Once)<sup>[10-12]</sup>、SSD(Single Shot multi-box Detector)<sup>[13]</sup>。

实际上,安全帽佩戴情况检测问题可以看成典型的目標检测问题,只需要将佩戴安全帽与不佩戴安全帽的人作为两类不同目标进行检测即可。为了实现安全帽佩戴情况的准确、快速检测,本文选择基于端到端的在检测速度和准确率方面都有不错效果的YOLOv2目标检测算法作为基础,通过在其网络中引入密集网络结构和轻量化网络结构构建出Mobile-Dense-YOLOv2(我们命名为:MD-YOLOv2)网络,最后对MD-YOLOv2网络进行训练得到相应的安全帽佩戴情况检测模型。实验表明,本文提出的模型能在保证检测准确率的前提下,有效减小参数量,显著提升检测速度。

## 2 YOLOv2 目标检测

YOLOv2检测方法使用的是基于GoogLeNet的定制网络,结构如图1所示。其基础网络是Darknet-19网络模型(包含19个卷积层和5个最大值池化层),采用了常用的 $3 \times 3$ 卷积核,并在每一次池化操作后把通道数翻倍。借鉴了network in network的思想,网络使用了全局平均池化

(global average pooling)做预测,把  $1 \times 1$  的卷积核置于  $3 \times 3$  的卷积核之间,用来压缩特征。YOLOv2 检测网络采用了带有 anchor 预测框的卷积层,并利用聚类算法自动选择先验预测框最佳的大小和数量。YOLOv2 还添加了一个转移

层(passthrough layer),把  $26 \times 26 \times 512$  的浅层特征图转化为  $13 \times 13 \times 2\,048$  的特征图,并结合原有的深层特征图组成  $13 \times 13 \times 3\,072$  的特征图。YOLOv2 在训练时,采用了多尺度训练的策略,提高了对不同分辨率图像的鲁棒性。

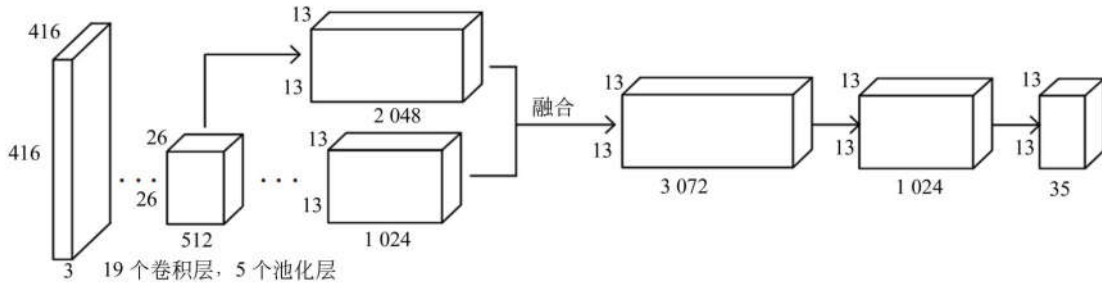


图 1 YOLOv2 网络结构

Fig.1 Network structure of YOLOv2

YOLOv2 在目标检测方面取得了很好的效果,对于 VOC2007 数据集中常见物体的检测,在检测速度为 67 frame/s 的情况下,其平均准确率达到 76.8%<sup>[11]</sup>。本文将 YOLOv2 用于安全帽佩戴情况检测任务,采用自制的 HelmetWear 数据集<sup>[14]</sup>重新训练了 YOLOv2,试验结果表明,检测效果并不理想。一方面该模型不能很好地适应景深较大的小目标情况;另一方面其网络参数量和计算量比较大,检测耗时长。

人与前面所有层的输出都相连,这种连接方式可以实现多层特征的融合,充分利用各层的特征。在密集块中满足:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), (l = 1, 2, 3, 4), \quad (1)$$

其中:  $x_0$  代表密集块输入特征图,  $x_l$  代表第  $l$  层的输出,  $[x_0, x_1, \dots, x_{l-1}]$  表示对第 0 到  $l-1$  层输出特征图的拼接,拼接后的特征图作为第  $l$  层的输入。  $H_l(\cdot)$  表示对第  $l$  层进行非线性变换,一般包括批量归一化(batch normalization),ReLU (Rectified Linear Units)以及卷积操作。整个密集块的输出是密集块内所有特征图的拼接,输出维度一般都比较大。一个密集连接网络包含多个密集块,密集块之间通过过渡层(transition layer)进行相连。过渡层由一个  $1 \times 1$  卷积层和一个步长为 2 的平均池化层组成,主要作用是对前一个密集块的输出进行降维和融合。

### 3 Mobile-Dense-YOLOv2 (MD-YOLOv2)

#### 3.1 小目标检测性能提升策略

密集连接网络(Densely connected convolutional Networks, DenseNet)<sup>[15]</sup>的基本组成单元是密集块(dense block),一个密集块的结构如图 2 实线框内所示,其最大的特点就是每一层的输

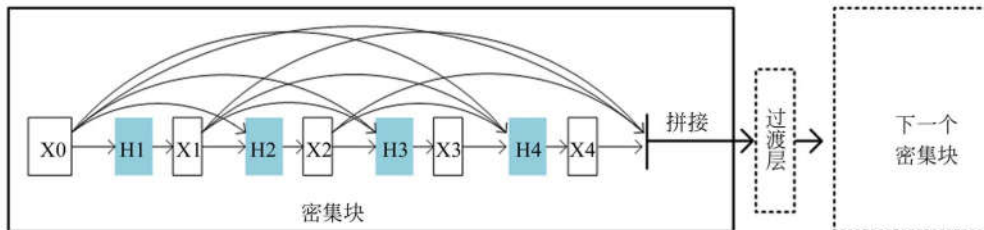


图 2 密集块

Fig.2 Dense block

小目标的检测既需要高分辨率的特征图来提供小目标的位置信息,也需要丰富的语义信息来对小目标进行有效地分类<sup>[16]</sup>。密集连接网络的过渡层对前一个密集块内的高分辨率低语义信息特征图进行融合,提取出有用的特征信息(包括对小目标检测有用的信息),再将这些特征信息传递给下一个密集块与其内的低分辨率高语义信息特征图进行融合,这种特征提取和传播方式使得密集连接网络的深层特征图在具备高语义信息的同时,还具备了来自浅层特征图的低语义信息,进而具有了一定的小目标检测能力。

### 3.2 模型压缩策略

MobileNet<sup>[17]</sup>是 Google 针对移动视觉应用提出的一种轻量化的卷积神经网络。该网络模型基于深度可分离卷积(depthwise separable convolutions),它将标准卷积分解为深度卷积(depthwise convolution)和 $1\times 1$ 卷积。图 3 表示了将标准卷积 3(a)分解为深度卷积 3(b)和 $1\times 1$ 卷积 3(c)的过程。其中卷积的输入通道数为  $M$ ,

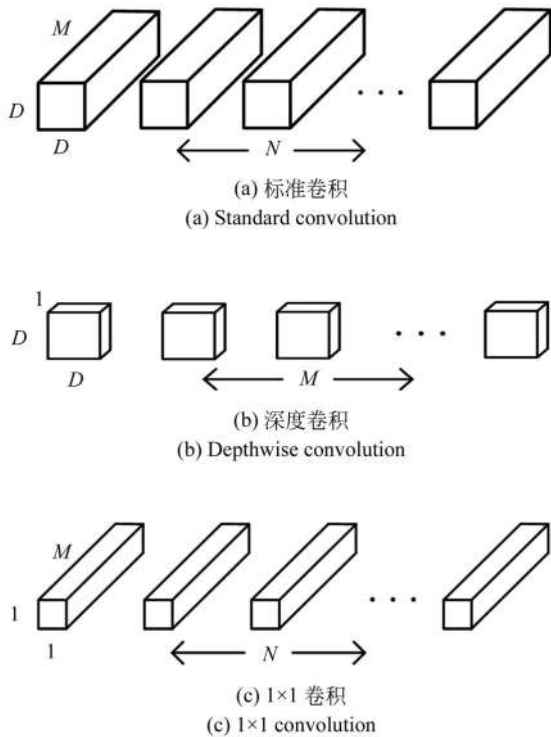


图 3 将标准卷积分解为深度卷积和 $1\times 1$ 卷积的过程  
Fig. 3 Process of factoring a standard convolution into a depthwise convolution and a  $1\times 1$  convolution

卷积的输出通道数为  $N$ ,卷积核的大小为  $D\times D$ 。深度卷积一个卷积核负责处理一个输入通道, $1\times 1$ 卷积将深度卷积的输出进行融合,这种处理方式具有显著减小计算量和模型大小的效果。

### 3.3 重构 YOLOv2 网络结构

结合前面提到的密集连接网络和 MobileNet 对 YOLOv2 网络进行重构,构建了如图 6 所示的 Mobile-Dense-YOLOv2 (MD-YOLOv2) 网络结构。

YOLOv2 的基础网络包含 19 个卷积层和 5 个最大值池化层,特征信息在传播过程中逐层丢失,由于前 8 个卷积层分辨率相对较高,可以保留足够的特征信息,所以 MD-YOLOv2 网络的前 8 个卷积层和前 4 个池化层与 YOLOv2 的保持一致。在第 4 个池化层之后,特征图分辨率相对较低,MD-YOLOv2 网络使用了两个密集块结构(两密集块之间由过渡层进行连接)对特征进行提取,实现了多层特征的融合,传播了更多有用的特征信息。本文对密集块之间的过渡层也进行了调整,将其中步长为 2 的池化层用步长为 2 的深度卷积进行了替换,如图 4 所示,这样可以减轻池化带来的信息损失,将前一个密集块的特征信息尽可能多地传到下一个密集块与其内的高语义信息特征图进行融合,提升网络对小目标检测的敏感性。

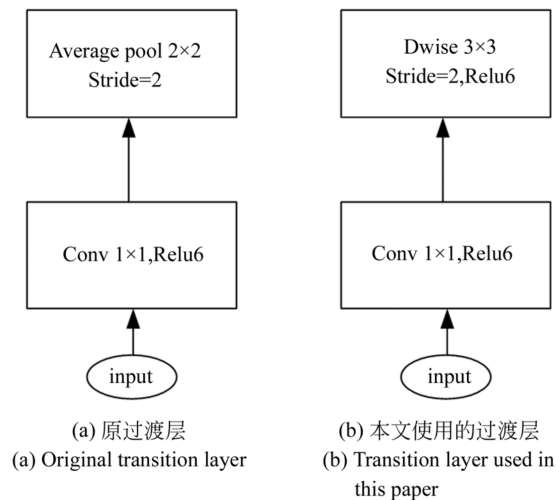


图 4 过渡层  
Fig. 4 Transition layer

为了减小网络的参数量和计算量,提升检测速度,本文使用轻量化网络结构作为密集块中的  $H_l(\cdot)$  变换。为了适应密集块的内部结构,这里采用的轻量化网络结构与 MobileNet 基本结构有所不同,如图 5(b)所示,首先使用的是  $1 \times 1$  卷积,由于密集块中  $H_l(\cdot)$  变换的输入来自前面的各层,后面的  $H_l(\cdot)$  输入通道数会比较大,所以先使用  $1 \times 1$  卷积,一方面是实现各输入通道的融合,另一方面可以对通道数较大的输入进行降维,减小后续的计算量;然后进行深度卷积操作(Dwise);最后采用线性的  $1 \times 1$  卷积进行降维操作,这里进行降维是为了避免网络宽度过大,使用线性卷积可以防止非线性变换对特征的破坏<sup>[18]</sup>。MD-YOLOv2 网络结构的具体参数如表 1 所示,  $s$  代表步长, M-block 代表本文使用的轻量化网络结构。

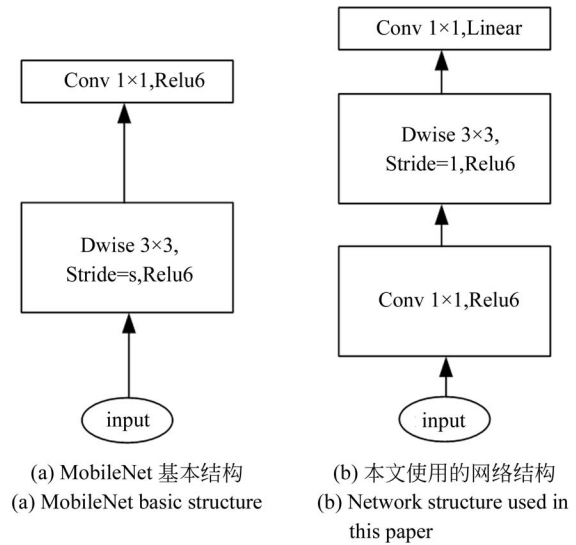


图 5 轻量化网络结构  
Fig. 5 Light-weight network structure

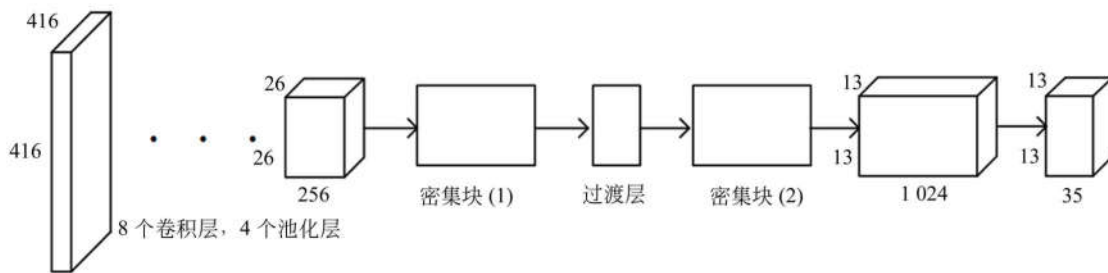


图 6 MD-YOLOv2 网络结构  
Fig. 6 Network structure of MD-YOLOv2

表 1 MD-YOLOv2 网络参数  
Tab. 1 Parameters of MD-YOLOv2 network

网络层	参数	输出通道数	输出尺寸
Convolution	$3 \times 3, s=1$	32	$416 \times 416$
Maxpool	$2 \times 2, s=2$	32	$208 \times 208$
Convolution	$3 \times 3, s=1$	64	$208 \times 208$
Maxpool	$2 \times 2, s=2$	64	$104 \times 104$
Convolution	$3 \times 3, s=1$	128	$104 \times 104$
Convolution	$1 \times 1, s=1$	64	$104 \times 104$
Convolution	$3 \times 3, s=1$	128	$104 \times 104$
Maxpool	$2 \times 2, s=2$	128	$52 \times 52$
Convolution	$3 \times 3, s=1$	256	$52 \times 52$
Convolution	$1 \times 1, s=1$	128	$52 \times 52$
Convolution	$3 \times 3, s=1$	256	$52 \times 52$
Maxpool	$2 \times 2, s=2$	256	$26 \times 26$

续表 1 MD-YOLOv2 网络参数  
Tab. 1 Parameters of MD-YOLOv2 network

网络层	参数	输出通道数	输出尺寸
Dense block(1)	M-block(1) $\begin{bmatrix} 1 \times 1 \times 256, \text{relu} \\ 3 \times 3 \text{ Dwise } s=1 \\ 1 \times 1 \times 64, \text{linear} \end{bmatrix}$	512	26×26
	M-block(2~4) $\begin{bmatrix} 1 \times 1 \times 384, \text{relu} \\ 3 \times 3 \text{ Dwise } s=1 \\ 1 \times 1 \times 64, \text{linear} \end{bmatrix} \times 3$		
	Transition layer $\begin{bmatrix} 1 \times 1 \times 512, \text{relu} \\ 3 \times 3 \text{ Dwise } s=2 \end{bmatrix}$	512	13×13
Dense block(2)	M-block(1) $\begin{bmatrix} 1 \times 1 \times 512, \text{relu} \\ 3 \times 3 \text{ Dwise } s=1 \\ 1 \times 1 \times 96, \text{linear} \end{bmatrix}$	896	13×13
	M-block(2~4) $\begin{bmatrix} 1 \times 1 \times 576, \text{relu} \\ 3 \times 3 \text{ Dwise } s=1 \\ 1 \times 1 \times 96, \text{linear} \end{bmatrix} \times 3$		
	Convolution	1 024	13×13
	Convolution	35	13×13

## 4 实 验

### 4.1 数据集

样本获取:本文的测试样本通过网络收集得到。为了提高在施工现场下安全帽佩戴情况检测的适应性,所获取的样本基本都是施工现场人员

目标的图片。

样本标注:需要在样本中标注两类目标,分别是佩戴安全帽的人和不佩戴安全帽的人。标注时记录下目标的类别以及左上角和右下角的坐标信息。样本标注如图 7 所示,实线框标注的是戴安全帽的人,虚线框标注的是未戴安全帽的人。



图 7 样本标注示例

Fig. 7 Sample labeling examples

最后将收集并标注的样本制作成 HelmetWear 数据集<sup>[4]</sup>,该数据集下共有 2 721 张图片,其中 2 272 张是训练集,449 张是测试集。在

整个数据集中,标注出的目标总数为 8 259 个,其中戴安全帽的目标有 4 081 个,不带安全帽的目标有 4 178 个。数据集具体分配如表 2 所示。

表 2 数据集分配表

Tab. 2 Dataset allocation table

目标类	标注	训练集	测试集
	目标数	目标数	目标数
戴安全帽目标	4 081	3 511	570
不带安全帽目标	4 178	3 646	532

#### 4.2 网络训练

本文的实验环境为: Intel Core i7-7740X CPU @4.30 GHz, 32 G 运行内存, Nvidia Geforce GTX1080, ubuntu16.04, 64 位操作系统, darknet 深度学习框架。

实验采用 HelmetWear 数据集从零开始训练了 MD-YOLOv2 网络, 最终得到该网络对应的安全帽佩戴情况检测模型。动量(momentum)设为 0.9, 衰减系数(decay)设为 0.000 5, 初始学习率(learning rate)为 0.000 1, 学习率调整策略为 steps, 最大迭代次数为 140 000 次, 每迭代一次训练的样本数为 64, 对自制数据集中的目标框进行维度聚类, 确定预选框; 学习率在迭代次数为 100, 50 000, 80 000, 115 000 次时, 分别再乘以 10, 0.1, 0.1, 0.1。

在训练的过程中, 不断地对图像的曝光度、饱和度和色调进行随机地调整, 以及对图像进行旋转和随机翻转, 达到数据集扩增的目的; 采用多尺度训练策略, 每训练十轮就重新调整输入尺寸的大小, 增强对不同分辨率图像的鲁棒性。图 8 为 MD-YOLOv2 网络在训练时的平均损失随迭代次数的变化曲线图, 在训练初期, 损失下降比较快, 随着迭代次数增加, 损失下降越来越慢, 迭代 100 000 次以后损失基本趋于平稳。

本文还在同样的实验环境下采用 HelmetWear 数据集重新训练了 YOLOv2 和 YOLOv3, 得到相应的检测模型, 以便于在测试集上与 MD-YOLOv2 的检测结果进行对比。

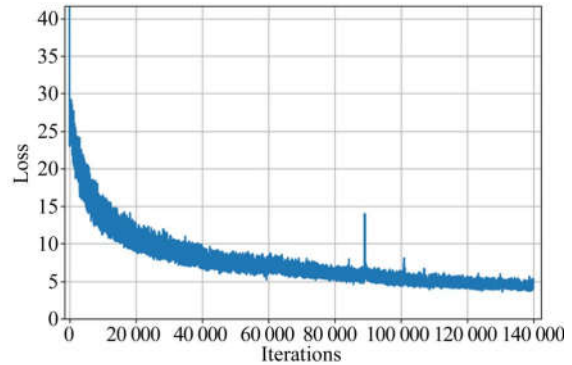


图 8 损失曲线

Fig. 8 Loss curve

#### 4.3 实验结果分析

首先选出一些样本分别输入到 YOLOv2, YOLOv3 以及本文提出的 MD-YOLOv2 检测模型中进行检测, 并将检测结果进行对比。图 9(a) 是选出的 4 张测试样本, 从左到右依次为样本(1), 样本(2), 样本(3), 样本(4)。图 9(b), 图 9(c) 和图 9(d) 分别是测试样本在 YOLOv2, YOLOv3 和 MD-YOLOv2 下对应的检测结果, 其中, 黑色框表示检测出的戴安全帽的人, 白色框表示检测出的未戴安全帽的人(彩图见期刊电子版)。样本(1)中的场景比较简单, 3 种模型的检测结果都比较理想; 样本(2)场景复杂, YOLOv2 出现了将背景误检为目标的情况, 而 MD-YOLOv2 和 YOLOv3 的检测效果较为理想, 说明本文对 YOLOv2 网络的改进在一定程度上提高了网络对复杂场景的适应能力; 样本(3)的场景中存在一些小目标, YOLOv2 没有检测出来, 而 MD-YOLOv2 和 YOLOv3 模型都检测出了其中的一些, 这表明密集连接网络结构的引入在提升小目标检测能力方面起到了一定的作用; 由于样本(4)中存在与背景颜色比较相近的施工人员, 3 种模型在对该样本进行检测时都出现了一定的漏检。

(a) 原始样本  
(a) Original samples



图 9 部分样本检测结果  
Fig.9 Detection results of some sample

然后利用 HelmetWear 数据集中的测试集对 YOLOV2, YOLOV3 和 MD-YOLOv2 模型分别进行测试,结果如表 3 所示。MD-YOLOv2 的检测准确率为 87.42%,稍逊色于 YOLOv3,但是其检测速度提升显著,比 YOLOv2 和 YOLOv3 分别提高了 37% 和 215%,可达 148 frame/s。另外, YOLOv2 和 YOLOv3 模型的参数量都要 200 M 以

上,而本文模型只有 17.5 M,可以节省 90% 以上的存储空间。分析其原因,一方面,本文模型采用了轻量化网络结构,使得参数量和计算量上都显著减小,提升了检测速度。另一方面本文模型采用了密集连接网络结构,实现了多层特征的融合以及浅层低语义信息与深层高语义信息的兼顾,在参数量很少的情况下,仍能保证较好的检测效果。

表 3 测试结果  
Tab.3 Test results

检测模型	测试样本数量	平均准确率(AP)/%		平均准确率 均值(mAP)/%	参数量 /MB	检测速度 /(frame · s <sup>-1</sup> )
		不带安全帽目标	戴安全帽目标			
YOLOV2		87.14	85.89	86.52	202.3	108
YOLOV3	449	87.42	89.05	88.24	234	47
MD-YOLOv2		87.79	87.05	87.42	17.5	148

## 5 结 论

针对施工现场安全帽佩戴情况检测问题,本

文提出了基于深度学习的解决方法,克服了传统方法对手工设计特征依赖的问题。该方法以 YOLOv2 目标检测方法为基础,在其基础网络上引入了密集连接网络结构和轻量化网络结构,提高

了网络对于小目标检测的敏感性,有效减小了网络的参数量和计算量。从实验结果来看,本文提出的基于改进 YOLOv2 的安全帽佩戴情况检测方法,检测的平均准确率均值为 87.42%,检测速度为

148 frame/s,能在保证较高检测准确率的同时,显著提升检测速度。但是当施工人员与背景颜色比较相近的时候,本文方法也会存在漏检的情况,如何有针对性地解决该问题是下一步研究的重点。

#### 参考文献:

- [1] 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.
- LIU X H, YE X N. Skin color detection and hu moments in helmet recognition research[J]. *Journal of East China University of Science and Technology(Natural Science Edition)*, 2014,40(3):365-370. (in Chinese)
- [2] PARK M W, ELSAFTY N, ZHU Z. Hardhat-wearing detection for enhancing on-site safety of construction workers[J]. *Journal of Construction Engineering & Management*, 2015, 141(9).
- [3] 周艳青,薛河儒,姜新华,等. 基于 LBP 统计特征的低分辨率安全帽识别[J]. 计算机系统应用, 2015, 24(7):211-215.
- ZHOU Y Q, XUE H R, JIANG X H, et al.. Low-resolution safety helmet image recognition combining local binary pattern with statistical features [J]. *Computer Systems & Applications*, 2015, 24(7):211-215. (in Chinese)
- [4] 潘蓉,孙伟. 基于预分割和回归的深度学习目标检测[J]. 光学精密工程, 2017, 25(10s):221-227.
- PAN R, SUN W. Deep learning target detection based on pre-segmentation and regression[J]. *Opt. Precision Eng.*, 2017, 25(10s):221-227. (in Chinese)
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al.. Region-based convolutional networks for accurate object detection and segmentation[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, 38(1):142-158.
- [6] GIRSHICK R. Fast R-CNN[C]. *IEEE International Conference on Computer Vision*. 2015:1440-1448.
- [7] REN S, HE K, GIRSHICK R, et al.. Faster R-CNN: towards real-time object detection with region proposal networks[C]. *International Conference on Neural Information Processing Systems*. MIT Press, 2015:91-99.
- [8] 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*, 2014, 37(9): 1904-16.
- [9] DAI J, LI Y, HE K, et al.. R-FCN: Object detection via region-based fully convolutional networks [J]. *arXiv:1605.06409v2*, 2016.
- [10] REDMON J, DIVVALA S, GIRSHICK R, et al.. You only look once: unified, real-time object detection[C]. *Computer Vision and Pattern Recognition*. IEEE, 2016:779-788.
- [11] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017:6517-6525.
- [12] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement [J]. *arXiv: 1804. 02767v1*, 2018.
- [13] LIU W, ANGUELOV D, ERHAN D, et al.. SSD: Single Shot MultiBox Detector[C]. *European Conference on Computer Vision*. Springer, Cham, 2016:21-37.
- [14] 孙腾腾. HelmetWear Dataset[EB/OL](2018-10-03) [2018-12-15]. <https://pan.baidu.com/s/1PYGcfD9nA5pRANzL4UJJFg>
- [15] HUANG G, LIU Z, WEINBERGER K Q, et al.. Densely connected convolutional networks [C]. *IEEE conference on Computer Vision and Pattern Recognition*, 2016: 4700-4708.
- [16] 吴天舒,张志佳,刘云鹏,等. 基于改进 SSD 的轻量化小目标检测算法[J]. 红外与激光工程, 2018, 47(7):47-53.
- WU T SH, ZHANG ZH J, LIU Y P, et al.. A lightweight small object detection algorithm based on improved SSD[J]. *Infrared and Laser Engineering*, 2018, 47(7):47-53. (in Chinese)
- [17] HOWARD A G, ZHU M, CHEN B, et al.. Mo-

MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. *arXiv*:1704.04861v1, 2017.

[18] SANDLER M, HOWARD A, ZHU M, *et al.*. MobileNetV2: Inverted residuals and linear bottlenecks[J]. *arXiv*:1801.04381v3, 2018.

#### 作者简介:



方 明(1977—),男,吉林梨树人,副教授。2000 年于长春光学精密机械学院获学士学位,2006 于长春理工大学获硕士学位,2011 年于日本北海道大学获工学博士学位,主要从事鲁棒图像处理,计算机视觉技术研究。E-mail: fangming@cust.edu.cn



孙腾腾(1992—),男,江苏徐州人,硕士研究生,2015 年于河北工业大学城市学院获学士学位,主要从事计算机视觉、深度学习等方面的研究。E-mail: suntengteng@foxmail.com