

基于图像分析的施工场景安全帽检测

王帅, 项学智

(哈尔滨工程大学信息与通信工程学院, 哈尔滨, 150000)

摘要: 本文主要针对工地场景下的施工工人是否正确佩戴安全帽的问题, 利用图像分析的方式对施工场景下的安全帽进行检测。传统的人工检测手段耗时长、效率低下且浪费资源, 利用图像的方式可以实现自动化检测的目的。主要检测的目标为安全帽与头, 主要采用的方法为 YOLOv3 目标检测方法, 利用该方法进行施工场景的自制数据集微调, 通过调整超参数不断训练, 得到一款能够稳定检测安全帽的模型, 以监控施工场景下工人是否遵规守纪, 切实保障施工人员的生命安全。经过实验验证, 最终模型对安全帽的检测率在 95% 以上, 能够应用于实际的施工场景。

关键词: 图像分析; 施工场景; YOLOv3; 安全帽;

中图分类号: TN911.73

Helmet Detection under the Scene of Construction Based on Image Analysis

Wang Shuai, Xiang Xuezhi

(Harbin Engineering University, College of Information and Communication Engineering, Harbin, 150000)

Abstract: This paper mainly focuses on the problem of whether the construction workers wear the helmets correctly under the construction site scene, and uses the image analysis method to detect the helmets under the construction scene. Traditional manual detection methods are time-consuming, inefficient, and waste resources. The use of images can achieve the purpose of automated detection. The main target of detection is the helmet and the head. The main method used is the YOLOv3 target detection method. This method is used to fine-tune the self-made data set of the construction scene. By adjusting the super-parameters and continuously training, a model capable of stably detecting the helmet is obtained. In order to monitor the compliance of the workers in the construction scene, the construction personnel are guaranteed to ensure the personal safety of the construction personnel. After experimental verification, the final model has a detection rate of more than 95% for the helmet, which can be applied to the actual construction scenario.

Key words: Image analysis; Construction background; YOLOv3; Helmet;

0 引言

近年来, 施工场景中安全事故频发, 安全帽是一种常见的保护头部安全的有效安全用具, 它的坚硬程度足以抵挡高空坠物以及其他坚硬物体撞击带来的伤害^[1], 减小施工人员的受伤几率。因此, 施工场景中工人是否佩戴安全帽关乎其生命安全, 能够稳定可靠的检测施工工人是否正确佩戴安全帽对于安全施工至关重要。

施工场景配备考勤制度, 通过人工排查的手段对施工人员是否佩戴安全帽进行检查, 但是这种方式存在弊端。第一, 人工方式无法实时对施工人员进行监控, 无法保证所有施工人员在不同时段均按规定佩戴安全帽; 第二, 人工手段进行排查工作效率低, 耗时长, 浪费大量的人力物力资源^[2]。因此利用人工方式已经无法满足工业需要。而基于图像分析的目标检

作者简介: 王帅 (1995), 男, 研究方向: 计算机视觉与模式识别

通信联系人: 项学智 (1979), 男, 硕士, 主要研究方向: 计算机视觉与模式识别. E-mail: xiangxuezhi@hrbeu.edu.cn

测方法,是在施工现场安装摄像机装置,利用摄像机采集图像,通过前端传输回的图像进行图像分析,从而将图像中存在的目标检测出来。利用图像分析的方式可以实现从人工到机器的自动化转变,大大提升了工作效率。在神经网络兴起以前,目标检测方法大多是利用特征和分类器相结合,完场对相应场景内目标的检测。常见的特征有 SIFT^[3]、HOG^[4]、SURF^[5]、颜色、纹理特征等;常见的分类器有 Adaboost^[6]分类器、SVM^[7]分类器等。目前,在目标检测领域,算法分为一阶段和两阶段方法。一阶段的方法有 SSD^[8]、YOLO^[9]、YOLOv2^[10]、YOLOv3^[11]等,两阶段的方法有 RCNN^[12]、SPP-Net^[13]、Fast RCNN^[14]、Faster RCNN^[15]、MR-CNN^[16]等。一阶段方法不生成候选框,而是直接采用回归的方式解决目标边界框的定位问题;两阶段的方法要通过算法生成一系列的样本候选框,再通过卷积神经网络进行分类实现目标检测。两阶段的方法在检测精度上相对较高,在检测速度上相对不足;一阶段方法速度快,但是精度上与两阶段相比稍稍逊色。

本文面向施工场景对施工人员的安全帽佩戴情况进行检测,采用的是基于图像分析的目标检测算法,利用一阶段的 YOLOv3 网络进行微调,实现对施工现场的稳定性检测,达到以自动化的方式代替传统人工方法。

1 YOLOv3 目标检测算法

YOLOv3 目标检测算法是在 YOLO 和 YOLOv2 的基础上进行改进,该算法以 DarkNet-53^[17]为基础网络架构,利用 DarkNet-53 的特征提取部分进行特征提取,结合 FPN 思想,将多个尺度特征图进行融合,得到一个对多尺度目标均具有范化性的鲁棒模型。如图 1 所示,为 YOLOv3 检测算法的框架图。

(1) 基础网络

YOLOv3 算法以 Darknet-53 为基础网络,利用其 53 个卷基层作为体征提取结构,如图 1 所示,为 YOLOv3 算法的具体结构。Darknet-53 结构中结合了全卷积的思想,同时也将 residual 结构引入其中,对于深层网络中的梯度问题效果显著。引入残差模块 residual 的原因是保证在网络很深的情况下,训练的结果依然能够收敛,同时深层次的网络使得特征表达的更好,最终分类和检测的效果均会有相应提升。同时,残差模块使用 1*1 卷积,降低了参数的数量以及训练过程中的计算量。

将网络的层数加深,同时在 32、16、8 倍下采样处进行目标检测。整个网络共有 5 次下采样,3 次降采样,将 32 倍降采样与第 5 次下采样的特征图进行拼接,同理,将 16 倍降采样与第 4 次下采样、8 倍降采样与第 3 次下采样的特征图进行拼接,使网络同时学习深层次与浅层次的特征,表达效果更好。在 3 次降采样的特征拼接中,32 倍降采样的感受野最深,通常对较大目标的检测的效果较好;16 倍降采样适用于一般大小的物体;8 倍降采样的感受野最小,适用于小目标的检测。

(2) 结果输出

在回归预测过程,YOLOv3 不使用 softmax 进行回归运算,而采用 logistic 回归的方式进行逻辑回归,避免同一目标具有两个类别无法同时输出,每个边界框与真值之间的匹配为 1:1 匹配,同时 YOLOv3 算法只为每个真实的对象分配一个边界框。边界框的预测方式延续了参考了 SSD 中的思路,采用了 anchor 的方式进行回归框的预测,通过 Kmeans 聚类的方式进行回归框的选取。

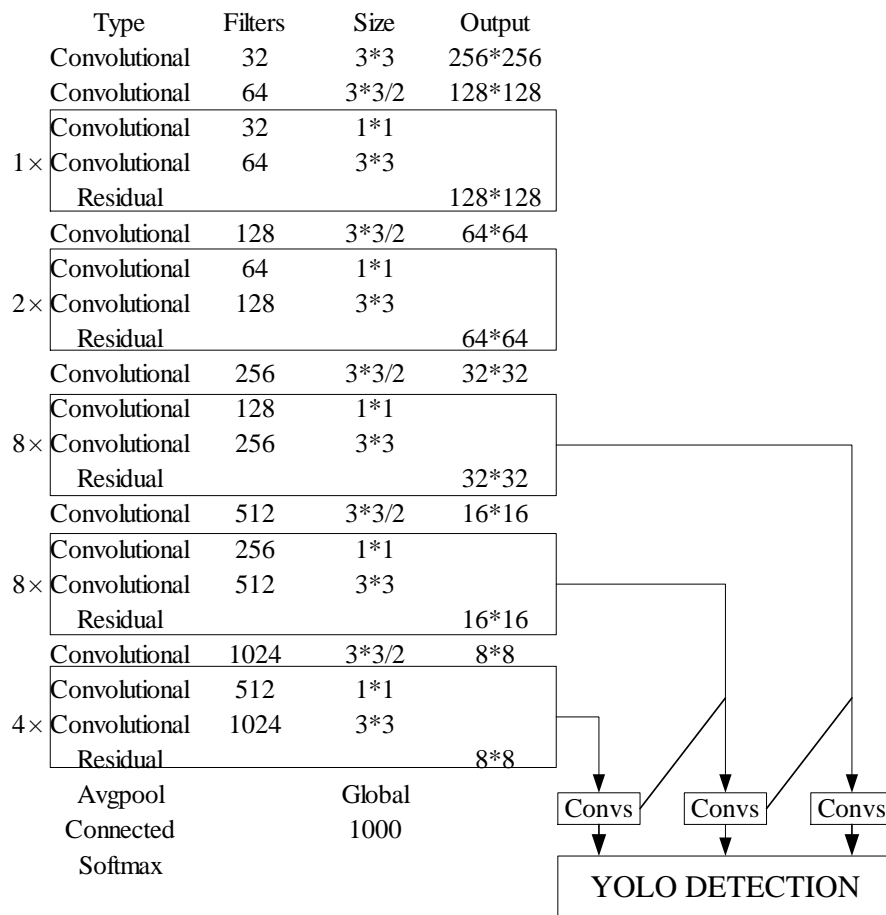


图 1 YOLOv3 目标检测算法

Fig. 1 objective detection by YOLOv3

损失函数的选择沿用了YOLO中的方式，对四类信息预测框长宽(w,h)、坐标(x,y)、类别class以及置信度confidence进行损失计算。式（1）为长宽的损失函数，利用的函数为和方差函数。坐标、类别与置信度的损失函数为式（2），以交叉熵损失函数为算法的损失函数。由公式可以看出，预测值与真值间的差别越大，最终的损失越大。

$$SSE = \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2 \quad (1)$$

$$loss = - \sum_{i=1}^n y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \quad (2)$$

$$\frac{\partial loss}{\partial y} = - \sum_{i=1}^n \frac{\hat{y}_i}{y_i} - \frac{1 - \hat{y}_i}{1 - y_i}$$

YOLOv3 的速度远远超过其他算法，在 320*320 输入分辨率上的检测时间仅有 22ms，map值也达到了 51.5。如表 1 中对比了YOLOv3 与目标检测领域中较为先进的方法，在速度与精度上综合考量，YOLOv3 均有良好表现，因此本文确定了基础网络为YOLOv3 目标检测算法。

表 1 YOLOv3 与其他方法对比

Tab. 1 Comparative Experiment of YOLOv3 and other methods on VOC datasets

	mAp	time (ms)
[B]SSD321	45.4	61
[C]DSSD321	46.1	85
[D]R-FCN	51.9	85
[E]SSD513	50.4	125
[F]DSSD513	53.3	156
FPN FRCN	59.1	172
RetinaNet-101-800	57.5	198
YOLOv3-320	51.5	22
YOLOv3-416	55.3	29
YOLOv3-608	57.9	51

2 实验方案及结果

综合速度与精度的考量，本文以 YOLOv3 为基础网络，利用图像分析的方案，设计了目标检测模型训练方案。如图 2，通过自制的数据集进行 YOLOv3 算法微调，得到一个具有鲁棒性的模型检测器，检测器可以实现检测施工场景下安全帽的实时检测。

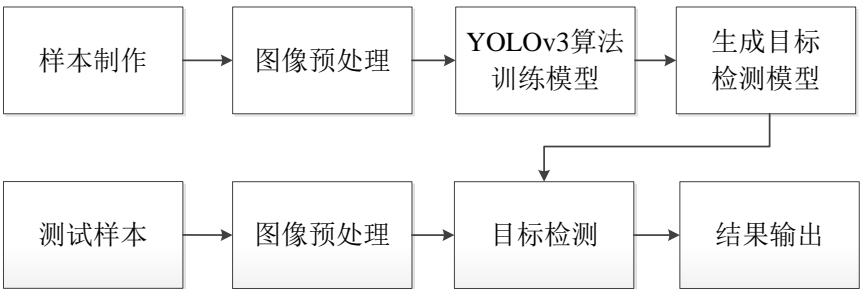


图 2 目标检测方案
Fig. 2 method of objective detection

2.1 自制数据集

自制数据集以施工的工地为工程项目背景，主要包括安全帽和头两类基本目标，共 1555 张图像，以VOC的格式进行标注。其中安全帽的个数为 2253 个，头的个数为 2346 个，样本处于均衡状态。如图 2 所示，均为工地施工场景下的目标。图像的标注方式为矩形框标注，记录目标的类别信息、位置信息、坐标信息，最终存入xml文件中。



(a) 施工场景下工人



(b) 建筑场景下工人

图 3 样本数据

Fig.3 our dataset

2.2 数据扩增

网络本身的训练过程需要大量的数据进行支撑，否则在训练时易出现过拟合现象，与 ImageNet 数据集不同，自制数据集中没有足够多的图像去参加模型的训练，因此进行数据扩增是训练模型必要的手段。

本文利用的图像数据扩充包含图像的旋转、饱和度、曝光以及色调调整的方式进行数据扩充。图像的数据样本为 1555 张，通过数据增强的方式进行扩增，增加样本的数量，有利于模型的微调。适当的数据扩增手段不仅有利于模型的优化，更加能够防止因为数据过少而使模型出现的过拟合现象。

2.3 模型微调

实验使用了自制数据集进行微调，根据自制数据所做的图像标注，利用 kmeans 算法聚类，得到施工场景安全帽数据集的 9 个 anchor 的值。在原作者给出的实验中，在公共数据集上将图像的输入分辨率设置为 416*416，使用的数据集为 COCO 数据集。在数据集上通过聚类得到 9 种结果分别为：(10*13); (16*30); (33*23); (30*61); (62*45); (59*119); (116*90); (156*198); (373*326)。由于实际数据中存在目标较小的情况，因此将输入大小改为 608*608，根据实际数据集得到的 anchor 值为(31*38); (49*52); (60*71); (69*112); (87*119); (88*115); (115*113); (140*162); (204*296)。

调整网络结构中的类别参数，划分训练数据集，利用程序实现训练集、验证集、测试集的比重为 3: 1: 1，迭代次数为 30000 次，batchsize 为 64，学习率为 0.0001，进行微调得到最终的检测模型。

3 实验结果分析

3.1 实验结果展示

在经过数次调整超参数后，得到一版稳定的模型，利用该模型对数据进行测试，得到了最终的实验结果。在测试程序中，可设置的阈值为目标检测的置信度与 NMS 阈值，其中置信度设置为 0.3，NMS 阈值设置为 0.4。



(a) 遮挡情况检测



(b) 小目标检测

图 4 实验结果展示

Fig.4 Experimental result

由图 4 中实验结果可以看出,场景中的工人所佩戴的安全帽均可以正确检测,算法不仅能够对多目标进行正确检测,同时具有一定的抗遮挡特性,如图 4 (a) 所示,检测器的检测效果具有一定的鲁棒性,可以对小目标进行检测,如图 4 (b) 所示,并且在不同光照的影响下也可以稳定的对目标进行检测。

3.2 实验结果统计

根据最终版的检测模型,利用测试数据进行测试。将得到的最终实验结果进行数学统计,从结果中分析最终的检测效果。式 (3) 为模型统计的基本公式, *precision* 为精确率, *FalseAlarm* 为虚警率, *MissingAlarm* 为漏警率, 其中 *TP* 为在实验统计中能够被正确检测出来的正样本, *FP* 为不能正确检测的正样本, *FN* 为错误检测的负样本。

$$\begin{aligned} precision &= \frac{TP}{TP + FP} \\ FalseAlarm &= \frac{FP}{TP + FP} \\ MissingAlarm &= \frac{FN}{TP + FN} \end{aligned} \quad (3)$$

如表 2 所示,为该模型最终的精确率、漏警率以及虚警率。测试数据共有 1326 张图像,其中安全帽的目标数为 2423 个,头的数目为 2541 个。

表 2 统计结果

Tab. 2 statistic results

	精确率	虚警率	漏警率
安全帽	96.58%	0.89%	3.42%
头	97.12%	0.69%	2.88%

根据表格中统计结果,可以看出在工地场景的自制数据下微调的YOLOv3 模型的准确率可以达到 95% 以上,能够应用于实际的施工场景中进行工人安全帽的佩戴检测。

4 结论

本文利用深度学习中的目标检测算法 YOLOv3,通过自制施工场景安全帽数据集,利用该数据集进行微调,得到一个稳定可靠的安全帽检测模型。通过实验结果可知,该检测器可以应用于工地施工现场进行现场检测,对安全帽的检测率可以在 95% 以上,满足现实生活

需求。

165 [参考文献] (References)

- [1] 刘云波, 黄华. 施工现场安全帽佩戴情况监控技术研究[J]. 电子科技, 2015(4):69-72.
- [2] 李太华, 王迪. 基于安全帽颜色识别的人员身份认证算法在变电站的应用[J]. 电脑知识与技术, 2014(5):1018-1020.
- 170 [3] Histograms of Oriented Gradients for Human Detection[C]// null. IEEE Computer Society, 2005.
- [4] Viola P, Jones M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2):137-154.
- [5] Lowe D G. Object Recognition from Local Scale-Invariant Features[C]// iccv. IEEE Computer Society, 1999:1150.
- 175 [6] Freund, Yoav, Schapire, Robert E. A decision-theoretic generalization of on-line learning and an application to boosting[C]// European Conference on Computational Learning Theory. Springer, Berlin, Heidelberg, 1995:23-37.
- [7] Vapnik V N. The Nature of Statistical Learning Theory[J]. IEEE Transactions on Neural Networks, 1997, 8(6):1564.
- 180 [8] W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]. European Conference on Computer Vision. Springer, Cham, 2016:21-37.
- [9] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:779-788.
- [10] Redmon J, Farhadi A. [IEEE 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Honolulu, HI (2017.7.21-2017.7.26)] 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - YOLO9000: Better, Faster, Stronger[J]. 2017:6517-6525.
- 185 [11] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. 2018.
- [12] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// Computer Vision and Pattern Recognition. IEEE, 2013:580-587.
- 190 [13] 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.
- [14] Girshick R. Fast R-CNN[J]. Computer Science, 2015.
- [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, 2015:1-1.
- 195 [16] Gidaris S, Komodakis N. [IEEE 2015 IEEE International Conference on Computer Vision (ICCV) - Santiago, Chile (2015.12.7-2015.12.13)] 2015 IEEE International Conference on Computer Vision (ICCV) - Object Detection via a Multi-region and Semantic Segmentation-Aware CNN Model[J]. 2015:1134-1142.