

# 基于 Mask R-CNN 的高空作业安全带检测<sup>①</sup>



冯志珍, 张卫山, 郑宗超

(中国石油大学(华东) 计算机科学与技术学院, 青岛 266580)

通讯作者: 冯志珍, E-mail: fengzhizhenabc@163.com

**摘要:** 随着计算机视觉近几年的发展, 相关工作者越来越侧重人工智能算法在电力安全管控系统的实际应用. 本文针对电力检修工作人员安全带规范问题, 基于 Mask R-CNN 算法提出了一种新型高空作业安全带低挂高用违规检测算法, 实时高效率完成作业者安全带违规检测问题. 针对安全带挂环违规现象的复杂性和场景多变性等问题, 本文提出适用于安全带检测和人体关键点信息相结合检测的 Mask-Keypoints R-CNN 新型高空作业安全带违规挂法的检测方法, 该算法基于人体关键点定位检测模块进行裁剪人体关键部位有用安全带数据集, 结合安全带检测模块进行判断作业人员违规情况, 算法本身具有很强的实用性和高效性, 并取得了较高的精确率.

**关键词:** 卷积神经网络; 目标检测; 实例分割; 关键点定位; 深度学习

引用格式: 冯志珍, 张卫山, 郑宗超. 基于 Mask R-CNN 的高空作业安全带检测. 计算机系统应用, 2021, 30(3):202-207. <http://www.c-s-a.org.cn/1003-3254/7812.html>

## Safety Belt Detection Algorithm for Aerial Work Based on Mask R-CNN

FENG Zhi-Zhen, ZHANG Wei-Shan, ZHENG Zong-Chao

(College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

**Abstract:** With the development of computer vision in recent years, more and more attention is paid to the practical application of artificial intelligence algorithms in power security systems. In this paper, aiming at the safety belt specification of power maintenance workers, based on the Mask R-CNN algorithm, we propose a new detection algorithm of safety belts hanging lower than the operator position during aerial work, which can complete the detection of safety belt violation in real-time and efficiently. Furthermore, we propose a new detection method of safety belt violation for aerial work, i.e., Mask-Keypoints R-CNN, which is applicable to the combination of safety belt detection and human key point information. The algorithm cuts the useful safety belt data set from the key parts of human bodies based on the positioning and detection module of the key points of human bodies and judges the violation of operators by combining with the safety belt detection module. In conclusion, the proposed algorithm has strong practicability and high efficiency and has achieved high accuracy.

**Key words:** Convolutional Neural Network (CNN); target detection; instance segmentation; key point positioning; deep learning

在电力检修高处作业施工现场, 重叠交叉的高空作业繁杂多样. 为增强保障工作人员的人身安全, 以防作业者高空作业坠落事故的发生, 企业管理部门按照

相关制度, 规定高空作业人员施工检修时必须佩戴安全带. 高空作业是指作业人员在位于坠落基准面一定高度以上作业. 目前高空作业电力检修安全带挂位有

① 基金项目: 山东省自然科学基金 (ZR2014FM038, ZR2019MF049)

Foundation item: Natural Science Foundation of Shandong Province (ZR2014FM038, ZR2019MF049)

收稿时间: 2020-07-10; 修改时间: 2020-08-11; 采用时间: 2020-08-17; csa 在线出版时间: 2021-03-03

低挂高用和高挂低用两种,把安全带挂环部分挂到工作人员腰部以上的位置叫安全带高挂低用,有助于减少意外坠落事故发生时的冲击距离,这种挂法是安全正规科学的系挂方法.另一种系挂方法是低挂高用,把安全带挂环部分挂到工作人员腰部以下的位置,人员坠落时安全带起不到安全缓冲的作用.所以安全带必须高挂低用,杜绝低挂高用.

在电力检修高空作业时,高空作业施工人员对企业佩戴安全带规定的法规遵守,仅依赖于其自身自觉和现场安全员的管理提醒,只有少数公司会增加一些违章作业的经济处罚规定.针对公司安排的临时活,部分工作人员觉得系安全带左右移动不便,直接怕麻烦不佩戴安全带或者佩戴安全带不规范,发生坠落趋势时安全带没能有效发挥安全缓冲作用,因安全带佩戴问题发生安全事故,给作业者造成难以挽回的人身安全伤害.针对安全管理制度执行存在盲点的问题,设计出一套智能化的安全带违规检测算法是至关重要的,既可以大大提升对于高处作业人员安全带佩戴情况的监管效率,又可以实现违章人员智能化的安全管理,保障工作人员的安全.

## 1 相关工作

在深度学习兴起并逐渐成为计算机视觉的核心方法之后,基于深度学习的计算机视觉目标检测算法大致可以分为 one-stage 和 two-stage 两大流派.普遍应用的目标检测 one-stage 和 two-stage 方法的根本区别在于候选区域框的差异.与 one-stage 直接对输入图像应用算法并输出类别和相应的定位不同的是 two-stage 设计了候选区域目标推荐的过程,即通过锚点机制去找到不同的目标推荐区域 proposals,锚点机制属于 RPN 网络的核心组件,由此可以看来,one-stage 和 two-stage 最典型特征在于是否存在 RPN 网络.本论文使用 two-stage 目标检测的 Mask-Keypoints R-CNN 方法进行安全带检测,即先产生候选区域,然后再进行卷积神经网络分类.

### 1.1 One-stage 目标检测算法

主流的 one-stage 直接通过预训练的主干网络给出类别和位置信息,利用检测目标的数据集进行微调达到快速收敛的效果,精确度相对 two-stage 目标检测网络略低,速度相比较快<sup>[1]</sup>.目前常用的典型 one-

stage 目标检测网络:YOLOv1、YOLOv2、YOLOv3、YOLOv4 和 SSD、DSSD 等<sup>[2-7]</sup>,如图 1 所示.

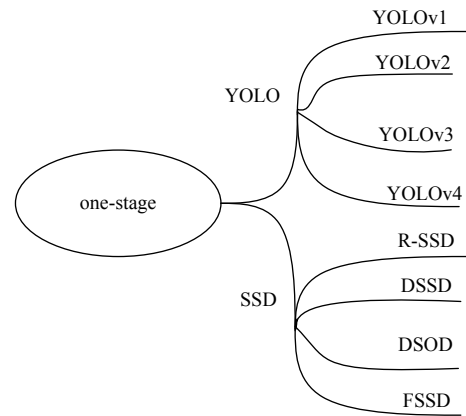


图 1 One-stage 相关算法

### 1.2 Two-stage 目标检测算法

Two-stage 的目标检测算法不同于端到端的目标检测方法,算法首先生成一系列作为输入样本的略微精准的区域候选框,再通过卷积神经网络进行样本数据分类和回归,因此目标检测精确率相对较高,但速度相对于 one-stage 慢. Girshick 等在 2014 年提出的 R-CNN 到 Faster R-CNN 网络<sup>[8-11]</sup>.主要通过一个卷积神经网络来完成目标检测过程,其提取的是 CNN 卷积特征,在训练网络时,其主要训练两个部分,第一步是训练 RPN 网络,第二步是训练目标区域检测的网络.目前常用的典型 one-stage 物体检测算法: R-CNN<sup>[8]</sup>、SPP-Net<sup>[9]</sup>、Fast R-CNN<sup>[10]</sup>、Faster R-CNN<sup>[11]</sup> 和 Mask R-CNN<sup>[12]</sup> 等,如图 2 所示.

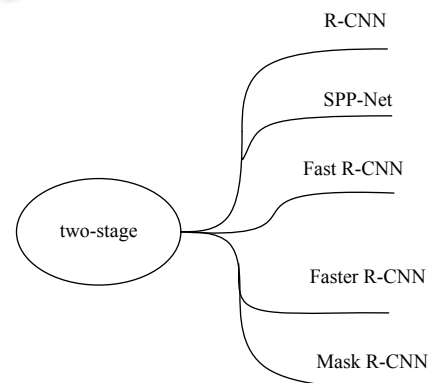


图 2 Two-stage 相关算法

### 1.3 Mask R-CNN 发展史

深度学习 (Deep Learning, DL) 发展火热,尤其在目标检测方面表现突出.2012 年在 ILSVRC 竞赛上基

于深度学习 CNN 网络的 AlexNet 取得突破性成果<sup>[13]</sup>, 2014 年 Girshick 在卷积神经网络的基础上设计出了 R-CNN 模型<sup>[8]</sup>, 在此之后, R-CNN 模型不断发展改进, 出现了具有代表性的 R-CNN 系列算法: SPP-Net、Fast R-CNN、Faster R-CNN、Mask R-CNN 等<sup>[9-12]</sup>. 2017 年何恺明团队提出了由新的 backbone 和 FPN<sup>[14]</sup> 组成, 可应用于实例分割、目标检测、人体关键点检测的 Mask R-CNN<sup>[12]</sup>. Fast R-CNN 汲取 SPP-Net 算法的优点, 采用两个 loss 结合训练的方法, 大大提高了模型预测的速度<sup>[9-10]</sup>. Mask R-CNN 在基础上 Faster R-CNN 基础上再加一个输出图像 Mask 信息的分支, 同时采用双线性插值的 ROI Align 代替原本的 ROI pooling, 算法不仅可以进行目标检测<sup>[15]</sup>, 而且可以做语义分割<sup>[16]</sup>.

随着人工智能算法的发展, 人体骨骼关键点定位检测算法的高效提升, 已经开始广泛应用于计算机视觉的姿态估计及动作分类等诸多相关领域. 人体关键点很多是直接通过卷积神经网络 (CNN) 来直接获取人体关键点的三维信息<sup>[17-20]</sup>, 其中文献 [20] 首次证明了深度神经网络可以从单视图中获得合理精度的三维人体关键点预测值. 本论文结合人体骨骼关键点定位技术进行人体安全带佩戴规范检测. 目前深度学习在高空作业安全带领域研究应用相对较少, 设计电力作业安全带系挂装置的传统方法比较耗时耗资源, 人工智能新方法值得进一步深度挖掘研究.

本课题基于深度学习的电力高空作业安全带佩戴规范检测和识别, 将人工智能和安全带检测工作相结合, 不仅可以降低人工监控的程度以减少人力物力, 大大提升对于高空作业人员安全带佩戴情况的监管效率, 还可以避免因人力因素导致的重要监控信息的遗漏, 有效提高工作效率, 实现了智能化的安全管理, 捕捉工人的安全带佩戴情况, 及时发现没有正确使用安全带的工人并发出警报, 对安防高空作业重大事故能够进行很好的预警及监控作用, 从而避免事故的发生, 对该产业安全智能管控系统发展具有重大指导意义.

## 2 目标检测算法

对于各种安全带的有效检测是保证高空作业人员安全的首要前提. 但可能存在的问题是由于现实场景的复杂性, 安全带区域检测不完全, 即存在部分区域未检测出来的状况, 对于低挂高用违规使用判断难度大, 因此要求目标检测算法足够鲁棒, 能够适应高空作业

复杂的场景. 基于高空作业现实场景中工作人员位置与安全带挂绳的相对位置特点, 以及安全带高挂低用规范挂位需要高于腰部使用的特点, 通过膝盖以下区域重要骨骼关键点位置进行判断安全带低挂高用违规方法是可行有效的. 在人体关键点定位算法设计和选取时, 充分考虑到安全带佩戴规范所接触的人体关键特征, 从而设计出 Mask-Keypoints R-CNN 的高效算法. 整体流程图如图 3.

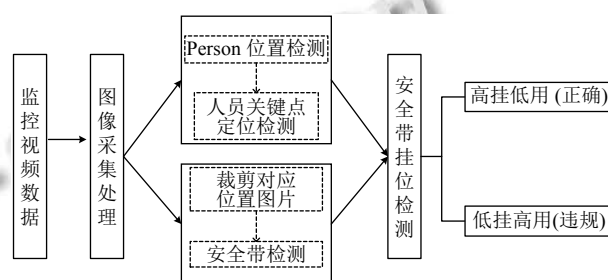


图3 整体流程图

### 2.1 整体思路

针对高空作业安全带场景复杂挂位难辨等问题, 无法进行直接的判断挂位是否规范. 本论文采用关键点定位技术与安全带的检测相结合的 Mask-Keypoints R-CNN 的高效算法, 通过膝盖以下区域重要骨骼关键点的位置是否存在安全带, 来判断安全带挂位是否规范. 第一阶段进行关键点截取膝盖以下部位的人员信息检测边界框, 第二阶段处理第一阶段保存的数据集, 使用标注工具标注安全带区域, 采用 ResNet 神经网络提取方法, 结合 Mask R-CNN 算法进行腿部以下安全带目标检测. 回归原数据查看是否为低挂高用或者是高挂低用, 判断高空作业人员安全带挂位检测的规范性.

### 2.2 Mask-Keypoints R-CNN 算法

为了解决从 R-CNN 就遗留下来的候选框生成问题, Faster R-CNN 的关键在于设计了 RPN (Region Proposal Network) 区域候选网络, 将候选框的选择和判断交给 RPN 进行处理, 将 RPN 处理之后的候选区域进行基于多任务损失的分类定位. 由于 RPN 网络能够在固定尺寸的特征图中生成多尺寸的区域候选框, 导致出现多目标尺寸和固定感受野不一致的情况. Faster R-CNN 的优点在于 CNN 提取的特征信息能够做到全网络的权值共享, 从而解决了因生成候选框数量多导致的速度慢的问题.

Mask R-CNN 算法改善了 R-CNN 系列算法, 类标签、边界框偏移量和输出对象掩码是每个目标候选对象的 3 个输出. Mask R-CNN 算法在输出对象掩码 mask 时, 实现了像素到像素的对齐, 从而提取出目标对象的更精细的空间布局. Mask R-CNN 为 two-stage 目标检测算法, 第一阶段流程主要是通过区域提议网络 (RPN) 输出区域候选框. RPN 网络结构主要是用来生成锚点 anchors, 然后经过分类器 Softmax 提取含有目标的 positive anchors, 边界框回归定位 positive anchors, 最后经过 proposal layer 生成目标区域建议 proposals. 第二阶段中, 在进行预测分类和边界框回归的同时, Mask R-CNN 算法中还为每个感兴趣区域 ROI 输出一个二进制的掩码, 依靠专用的卷积网络分类分支来预测用于选择掩码的类标签, 这使得掩码和类预测解耦<sup>[12]</sup>.

本论文中的 Mask-Keypoints R-CNN 算法由关键点检测和 Mask R-CNN 检测两部分组成, 首先经过人员检测输出人员位置及关键点信息, 裁剪所需的部分数据图片, 本论文主要采用左膝盖关键点作为裁剪指标. 不同尺寸的数据图片经过 FPN 网络生成多个不同尺寸的特征图, 采用 FPN 网络将底层特征和高层特征进行融合, 强调语义信息和位置信息, 如图 4 为 FPN 网络结构图, 左侧为 ResNet 骨架, 从低向上进行 0.5 倍的下采样, 输出结果分别为 C2、C3、C4、C5, 中间为高层特征层进行双线性插值 2 倍的上采样, 同时与低层特征进行横向连接, 相当于特征图相加操作, 特征图尺寸呈倍数逐渐增大输出, 输出结果分别为 M5, M4, M3, M2. 此处为消除混叠效应进行 3×3 的卷积得到 FPN 的最后结果 P2, P3, P4, P5.

RPN 网络用来从不同特征图中产生目标区域候选

框, 再根据共享特征图的特征输出某个尺寸的特征图上的目标区域建议, ROI 区域特征框的选择相当于裁剪的过程, 不同尺寸的目标区域建议经过 ROI Align 模块调整尺寸大小, 然后输出相同尺寸的特征图, 最后经过全连接层进行分类和目标回归定位, 同时经过设定的卷积层输出目标掩码, 在原数据图片上展示目标类别、边界框和目标掩码信息. 安全带目标检测算法基本流程如图 5 所示.

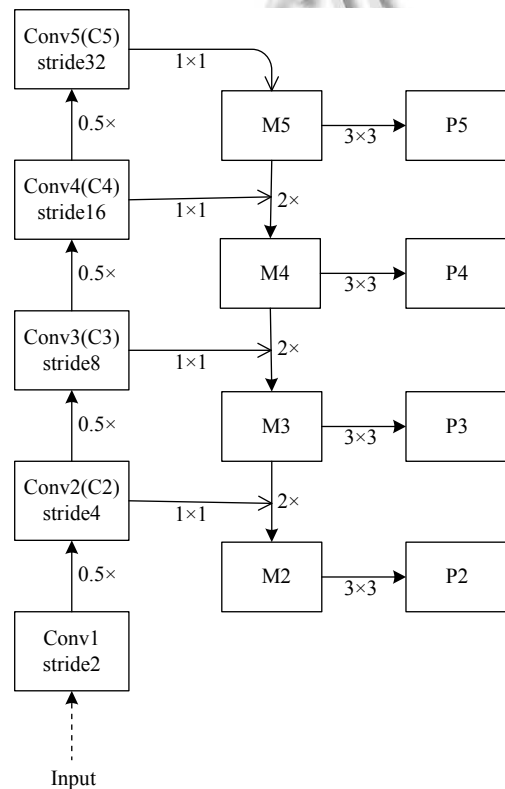


图 4 FPN 网络结构图

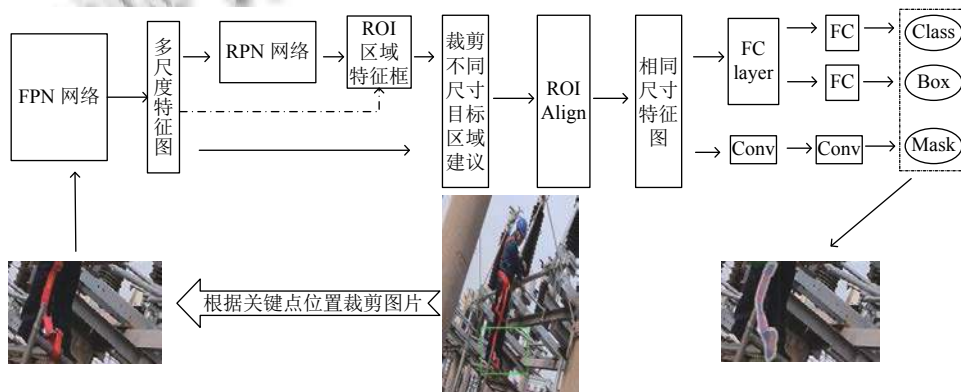


图 5 Mask-Keypoints R-CNN 算法

### 3 实验过程与分析

实验数据总共 1200 张, 其中 700 张对应不同的光照强度和不同的场景低挂高用违规数据和 500 张高挂低用正规使用数据图片。

#### 3.1 实验环境与数据集处理

本实验使用的实验设备为 64 位 Linux 系统计算机, 计算机的 GPU 显卡类型 GeForce RTX 2080 Ti, GPU 显存 11 GB. CUDA Version 为 10.2. 同时配备 CPU 类型为 Intel(R) Core(TM)i9-9900K, CPU 的频率为 3.60 GHz. 实验训练及测试数据均在设备 GeForce RTX 2080 Ti 上进行, 其他机型可能稍有偏差。

我们采用 Mask R-CNN 算法用数据图片迭代 6000 次训练人员身上的安全带和裁剪膝盖以下部分安全带检测, 基于数据的特点, 使用 UPCLable 工具进行标注标签, 数据训练采用 JSON 格式读取。

#### 3.2 实验过程及结果

对于安全带佩戴标准检测问题在实验中进一步尝试, 实现关键点定位检测和安全带检测模块的高效结合方法, 人体的全部骨骼关键点共分为 17 个, 人体骨骼关键点详细介绍见表 1, 此处使用 left\_knee 关键点的纵坐标结合人员边界框进行裁剪所需数据。

表 1 人体关键点信息

序号	名称	序号	名称
1	nose	10	left_wrist
2	left_eye	11	right_wrist
3	right_eye	12	left_hip
4	left_ear	13	right_hip
5	right_ear	14	left_knee
6	left_shoulder	15	right_knee
7	right_shoulder	16	left_ankle
8	left_elbow	17	right_ankle
9	right_elbow		

在 Mask R-CNN 算法的基础上, 使用 Mask-Keypoints R-CNN 算法的关键点检测算法进行关键点的定位。在关键点定位检测中结果如图 6 所示, 相对于整张数据图片而言, ["person", 0.994871 14, 183, 46, 466, 425] 分别代表: 检测类别为 person, 检测准确率为 99.487114%, 183、46 和 466、425 分别为检测出来的人员位置边界框 bbox 左上角像素坐标值和右下角像素坐标值。["nose", 373, 186] 则代表关键点 nose 的横纵坐标像素坐标值, 其他骨骼关键点同样代表像素值。

我们采用以图像为中心的训练方法。根据数据集

设计 e2e\_mask\_rcnn\_R-101-FPN\_2x.yaml 配置文件, 采用 GPU 训练, 分为背景类和 belt 两类, 配置文件相关参数值如表 2 所示。

```
[[{"person": 0.99487114, 183, 46, 466, 425}]
[["nose", 373, 186], ["left_eye", 361, 186], ["right_eye", 376, 95], ["left_ear", 349, 81], ["right_ear", 349, 82], ["left_shoulder", 322, 103], ["right_shoulder", 333, 112], ["left_elbow", 348, 145], ["right_elbow", 364, 155], ["left_wrist", 372, 153], ["right_wrist", 372, 153], ["left_hip", 295, 222], ["right_hip", 284, 228], ["left_knee", 281, 315], ["right_knee", 232, 307], ["left_ankle", 266, 403], ["right_ankle", 215, 393]]]
##### 2
```

图 6 人体关键点输出信息

表 2 配置文件信息

配置文件相关参数	配置文件参数值
NUM_CLASSES	2
WEIGHT_DECAY	0.0001
BASE_LR	0.001
MAX_ITER	6000
STEPS	[0, 3000, 5000]
MAX_SIZE	1500

本文以深度学习中的实例分割技术和关键点检测算法为基础, 提出新的检测方法, 以深度学习中的卷积神经网络技术为基础, 通过构建的安全带检测神经网络对图片中的高处作业人员进行精确定位, 得到高空作业人员在图片中的位置后, 裁剪膝盖以下边界框, 如图 7(b), 裁剪部分进行训练, 如图 7(c), 最后回归原图判断安全带低挂高用和高挂低用规范佩戴, 如图 7(d), 相关检测过程结果图 7. 本实验对安全带检测训练模型和关键点模型分别用 100 张数据进行了测试, 安全带和关键点识别率分别为 98% 和 95%, 两者结合使用处理低挂高用违规图片的检测识别准确率达 94%, 识别每张图片的平均速度 4 秒左右, 满足电力检修实际场景中的实时应用要求。

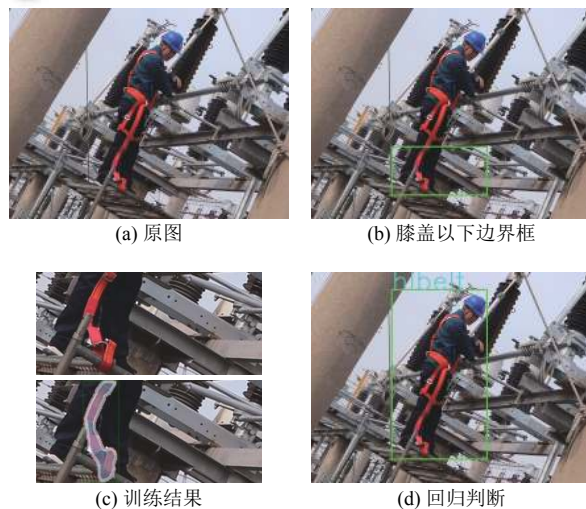


图 7 安全带违规检测过程图

## 4 结论与展望

本文数据集将更进一步增加和更新,为以后人工智能计算机视觉研究在国家电力工业方面的研究奠定基础。该课题虽然还存在问题,针对此问题本作者将继续跟进探索研究关键点检测和人员检测相结合的算法。安全带检测模型和关键点检测模型相结合,准确检测出安全带低挂高用的违规操作,及时发现问题并解决问题。本课题基于上述背景和思想,提出了基于深度学习的安全带规范检测算法,利用深度学习的方式减少了传统人工选取特征的种种麻烦,通过计算机视觉检测的方法,可以较精准的判断高处作业施工人员安全带的佩戴情况,有效的预防意外安全事故的发生,为当前工业检测和识别领域提供新的解决办法,场景安全带的检测与识别也具有较强的现实意义。

### 参考文献

- 1 Wang TC, Anwer RM, Cholakkal H, *et al.* Learning rich features at high-speed for single-shot object detection. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 1971–1980.
- 2 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 779–788.
- 3 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 7263–7271.
- 4 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv preprint arXiv: 1804.02767, 2018.
- 5 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv: 2004.10934, 2020.
- 6 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 21–37.
- 7 Fu CY, Liu W, Ranga A, *et al.* DSSD: Deconvolutional single shot detector. arXiv preprint arXiv: 1701.06659, 2017.
- 8 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587.
- 9 He KM, Zhang XY, Ren SQ, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916. [doi: 10.1109/TPAMI.2015.2389824]
- 10 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1440–1448.
- 11 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]
- 12 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386–397. [doi: 10.1109/TPAMI.2018.2844175]
- 13 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84–90. [doi: 10.1145/3065386]
- 14 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 936–944.
- 15 蔡念, 周杨, 刘根, 等. 鲁棒主成分分析的运动目标检测综述. 中国图象图形学报, 2016, 21(10): 1265–1275. [doi: 10.11834/jig.20161001]
- 16 魏云超, 赵耀. 基于 DCNN 的图像语义分割综述. 北京交通大学学报, 2016, 40(4): 82–91. [doi: 10.11860/j.issn.1673-0291.2016.04.013]
- 17 Tekin B, Katircioglu I, Salzmann M, *et al.* Structured prediction of 3D human pose with deep neural networks. Proceedings of the British Machine Vision Conference. York, UK. 2016. 130.
- 18 Tekin B, Márquez-Neila P, Salzmann M, *et al.* Learning to fuse 2D and 3D image cues for monocular body pose estimation. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 3961–3970.
- 19 Tome D, Russell C, Agapito L. Lifting from the Deep: Convolutional 3D pose estimation from a single image. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 5689–5698.
- 20 Li SJ, Chan AB. 3D human pose estimation from monocular images with deep convolutional neural network. Proceedings of 12th Asian Conference on Computer Vision. Singapore. 2014. 332–347.