

桥式起重机部件多任务学习 Mask R-CNN 分割与关键点识别方法*

杨帆¹ 梁敏健² 杨宁祥² 彭晓军²

(1.华南理工大学机械与汽车工程学院, 广东 广州 510640

2.广东省特种设备检测研究院珠海检测院, 广东 珠海 519002)

摘要: 起重机运行状态实时检测是工业安全生产的重要保障之一。针对起重机部件提出基于多任务学习 Mask R-CNN 的分割与关键点网络结构, 该结构由 Mask R-CNN 定位框与区域语义分割网络、DeepLabCut 关键点检测网络构成; 在吊钩桥式起重机中采集数据进行测试, 利用贪婪线性搜索算法与贝叶斯优化算法, 搜索得到此模型的最优超参数组合为: 学习率 0.005, 批数 2, 学习率策略为余弦衰减。该模型测试误差为 2.46 个像素点, 测试 AP 可达 95%, 像素点误差反映到实际误差在 5 cm 以内, 满足实际检测需求, 可拓展应用于无人化、自动化起重机运行状态监测。

关键词: 桥式起重机; 关键点识别; 深度学习; 卷积神经网络; 多任务学习; 语义分割

中图分类号: TN911.73

文献标识码: A

文章编号: 1674-2605(2021)02-0003-05

DOI: 10.3969/j.issn.1674-2605.2021.02.003

0 引言

起重机是现代工业生产不可或缺的设备, 其作业范围广、作业环境复杂、吊运对象多样, 需要多机构同时操作、多工种协同作业。起重机在作业过程中危险因素较多, 需实时监测其关键部件的运行状态, 及时发现故障并发出预警信息。起重机关键部件监测方法主要有基于多物理传感器直接测量^[1-3]、基于机器视觉传感^[4-6]两大类, 前者装置拆装不便, 适应性较差; 后者拆装方便, 以高鲁棒性图像算法为基础, 在满足精度误差范围内具备不同场景测量的高适应性。基于机器视觉传感的监测方法是基于图像传感器及图像处理算法, 识别图像中起重机关键部件位置, 并结合图像深度信息进行定量测量。文献[4]采用归一化互相关图匹配算法对起重机吊绳图像进行目标对象定位框识别, 并作为 Mean Shift 跟踪的目标区域进行跟踪。文献[5]利用灰度化、二值化、边缘检测、霍夫变换等图像处理算法识别吊绳边缘, 再根据相机小孔成像原理预测吊绳角度。上述研究均利用经典图像处理算法进行起重机部件识别, 无法抽取高层图片信息, 对场景变化及环境变化适应性差。相比于经典图像处理算

法, 深度学习模型具有图像高级语义特征理解力, 对不同场景适应能力更强。为此, 文献[6]利用深度学习检测算法 YOLO^[7]实现不同角度起重机图片吊钩定位, 适应性较好, 但算法仅适用于图像定位框识别任务, 没有对起重机运行部件的关键点坐标信息及部件区域分割信息进行预测。

为此, 本文应用 Mask R-CNN 结合多任务学习^[8-9], 接入多预测分支, 构建多任务识别深度学习网络, 实现桥式起重机关键部件定位框、区域语义分割与关键点的同时识别。

1 网络设计

桥式起重机关键部件定位与关键点识别方法框架包括 Mask R-CNN 定位框与区域语义分割网络、DeepLabCut 关键点检测网络, 如图 1 所示。

1.1 Mask R-CNN 定位框与区域语义分割网络

Mask R-CNN 架构图如图 2 所示。输入图像经过预训练的 ResNet50-FPN 结构, 融合 ResNet50 各层不同尺度的特征张量, 输出特征图 $A_i (i=1, \dots, 4)$, 即图像语义特征。

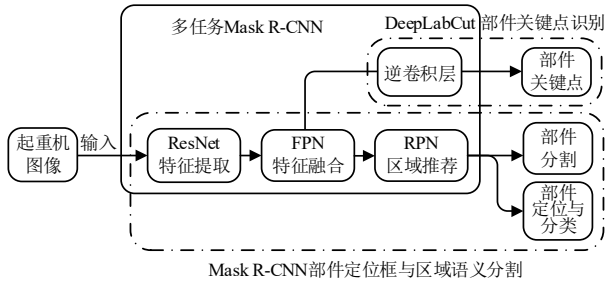


图1 桥式起重机关键部件定位与关键点检测方法框架图

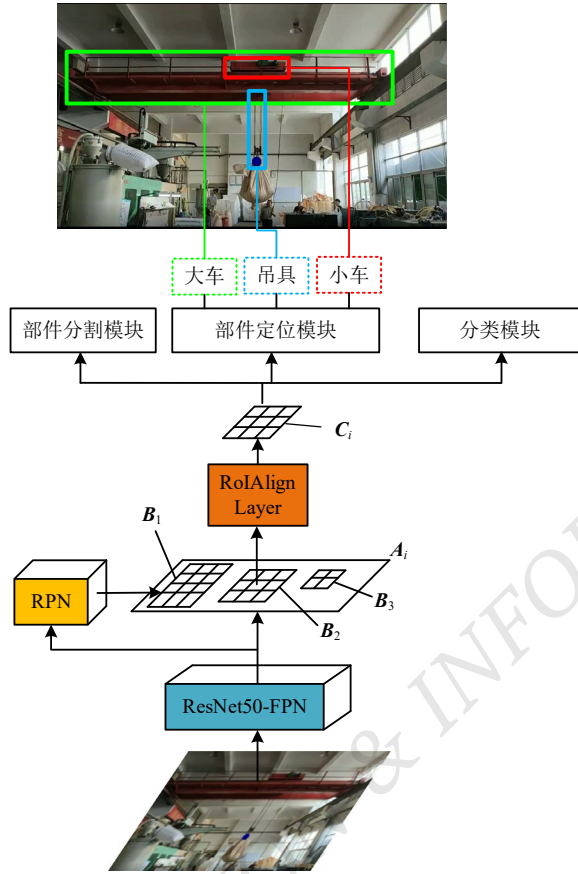


图2 Mask R-CNN 整体架构图

特征图 A_i 输入感兴趣目标区域候选框生成网络 (region proposal network, RPN) 后, 获得语义特征张量对应候选框区域的局部特征张量 B_i ($i=1, \dots, n$)。设输入图像的宽和高分别为 W, H , 若 RPN 输出某一 RoI 的宽与高分别为 w_j, h_j , 左上顶点的图像像素坐标为 (x_j, y_j) , 且该 RoI 为特征张量 A_i 通过 RPN 生成, 则此图像上的 RoI 对应的 A_i 局部特征张量 B_i 为

$$k_j = \left\lfloor i + \frac{1}{2} \log_2 \left(\frac{w_j h_j}{WH} \right) \right\rfloor \quad (1)$$

$$x_j' = x_j / k_j; y_j' = y_j / k_j; w_j' = w_j / k_j; h_j' = h_j / k_j \quad (2)$$

$$x_j'' = x_j' + w_j'; y_j'' = y_j' + h_j' \quad (3)$$

$$B_i = A_i[x_j': x_j'', y_j': y_j''] \quad (4)$$

每个局部特征张量 B_i 输入感兴趣目标区域对齐层, 即 RoIAlign 层, 首先, 该层均匀地将 B_i 分割成 $7 \times 7 = 49$ 个子区域; 然后, 利用双线性插值法获得每个区域中心位置的插值, 得到插值后张量; 接着, 对插值后张量进行最大值池化操作; 最后, 得到 RoIAlign 层的输出, 区域对齐后的张量 C_i ($i=1, \dots, N$)。

张量 C_i 分别输入部件分割模块、部件定位模块和分类模块, 以完成每个已分类的部件定位区域内的分割任务。

各预测模块的内部结构如图3所示。首先, 部件定位模块与分类模块共用多个全连接 (fully connected, FC) 层以提取张量 C_i 的语义信息矢量; 然后, 部件定位分支将语义信息矢量输入到4个神经元的 FC 层, 获得第二次精炼的部件定位框回归预测值 x, y, w, h ; 分类分支将矢量输入到3个神经元的 FC 层; 接着, 输入 Softmax 层获得小车、大车与吊具的预测概率 p_c, p_b 和 p_s ; 最后, 部件分割分支利用 $3 \times 3 \times 256$ 卷积层与 $1 \times 1 \times 3$ 卷积层对 C_i 提取语义张量信息, 再输入 Sigmoid 层, 得到 $14 \times 14 \times 3$ 掩码张量, 张量的每一维度代表一个类别掩码区域的预测, 张量的任意元素值表示该位置属于目标区域概率。

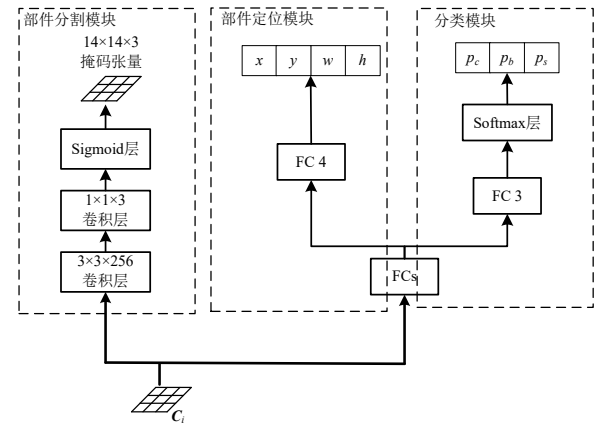


图3 预测模块内部结构图

分类分支、部件分割分支的输出分别输入交叉熵损失函数得到分类损失 L_{cls} 与部件分割损失函数 L_{mask} ；部件定位分支输出输入最小二乘回归损失函数得到定位框损失 L_{box} ，各损失之和即为 Mask R-CNN 总损失函数 L

$$L = L_{box} + L_{cls} + L_{mask} \quad (5)$$

1.2 DeepLabCut 关键点检测网络

DeepLabCut 关键点检测网络结构如图 4 所示。设图像中起重机关键点个数为 k ，起重机图像输入 ResNet 网络获得特征图后，分别输入关键点空间概率分布预测分支、偏移量预测分支。

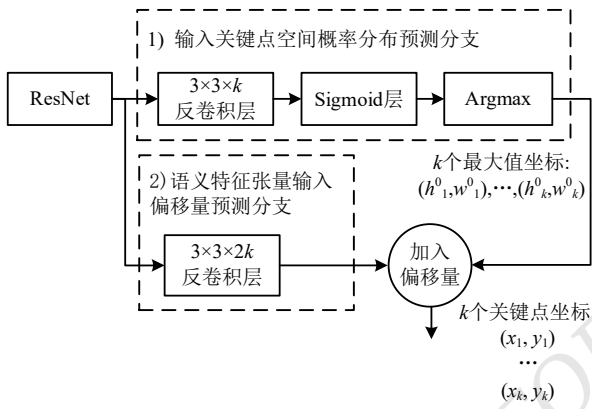


图 4 DeepLabCut 关键点检测网络结构

1) 输入关键点空间概率分布预测分支，ResNet 输出经尺度为 $3 \times 3 \times k$ 的反卷积层，再通过 Sigmoid 层获得特征图 $M_{H \times W \times k}$ ，特征图中第 i 维第 (m, n) 个元素的值表示原图第 m 行，第 n 列对应区域出现第 i 个关键点的概率。故由特征图 M 可初步得到每个关键点大致出现的区域：

$$(h^0_i, w^0_i) = \text{Arg max}_{h_i, w_i} M[h_i, w_i, i] (i = 1, \dots, k) \quad (6)$$

2) 语义特征张量输入偏移量预测分支，ResNet 输出经 $3 \times 3 \times 2k$ 的反卷积层，输出偏移量回归张量 $F_{H \times W \times 2k}$ ，则第 i 个关键点 (x_i, y_i) 的预测输出为

$$\begin{aligned} x_i &= h^0_i + F[h^0_i, w^0_i, 2i - 1]; \\ y_i &= w^0_i + F[h^0_i, w^0_i, 2i]; i = 1, \dots, k \end{aligned} \quad (7)$$

DeepLabCut 的损失函数 L_{kpt} 为交叉熵损失函数，设 p_{out_n} 表示关键点分支通过 Softmax 层后的输出； p_n 表示标签真值； N_{ap} 指一个批次的所有样本量，则 L_{kpt} 表示为

$$L_{kpt} = \frac{\sum_{n=1}^{N_{ap}+1} [p_n \log p_{out_n} + (1 - p_n) \log(1 - p_{out_n})]}{N_{ap} + 1} \quad (8)$$

1.3 多任务深度网络损失函数设计

边框预测、语义分割预测与关键点预测 3 类子任务应用于起重机实例对象的预测、识别。子任务之间具有同质性，可提高特征提取器的泛化性，最终提升各子任务预测准确性，并且通过公用骨干网络，减少算法时间复杂度与空间复杂度。

Mask R-CNN 总损失函数 L 添加 DeepLab 关键点分支损失函数项 L_{kpt} 得到多任务深度网络损失函数 $L_f = L + L_{kpt}$ 。最后利用 Adam 算法^[11]，最小化总损失函数 L_f ，训练获得最优网络参数。

2 实验与结果分析

本实验对象为广东省珠海市某厂房的通用式吊钩桥式起重机，选用 MV-CA060-10GC、4 K 分辨率工业相机为图像采集传感器；深度学习硬件设备为 Nvidia 1080ti 显卡，CPU 为 i7 8700，32 GB 内存。

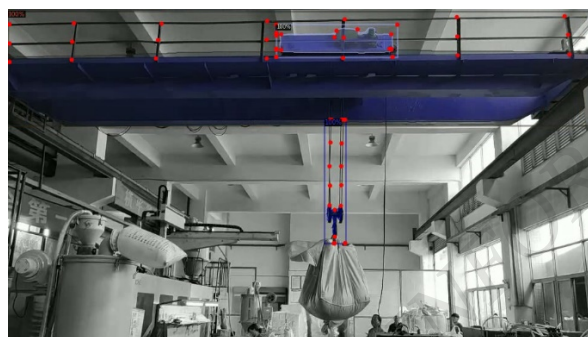
采集全天候起重机运行视频，并抽帧 500 张图像作为训练集，利用贪婪线性搜索算法，获得近似超参数最优组合。基本超参数组合为：学习率 0.00003，批数 1，学习率策略为多步衰减。在此基本超参数组合上，依次改变单个超参数的取值，不同组合下的误差结果如表 1 所示。

将单步搜索最优值对应超参数取值进行组合，得到近似最优组合：学习率为 0.00003，批数为 3，学习率策略为多步衰减，对应训练误差为 2.33，测试误差为 2.46。

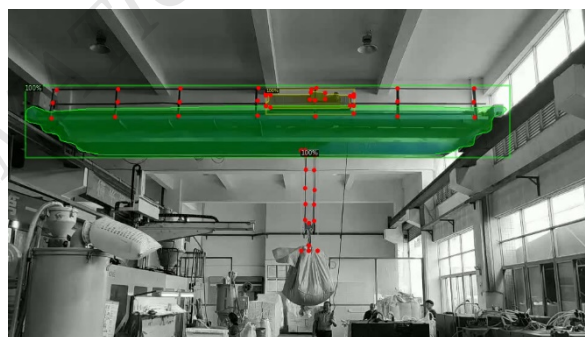
表 1 超参数优化效果

超参数取值	训练误差	测试误差	测试 AP
0.00003	6.33	6.61	88%
0.0001	3.82	4.06	90%
0.001	2.39	2.88	91%
0.002	2.49	2.92	91%
0.005	2.57	<u>2.85</u>	92%
1	6.61	6.61	87%
2	6.24	6.38	87%
3	2.44	<u>2.56</u>	93%
余弦衰减	8.76	8.66	72%
多步衰减	6.33	<u>6.61</u>	88%
学习率: 0.00003 批数: 3 学习策略: 多步衰减	2.33	<u>2.46</u>	95%

利用树状结构 Parzen 估计方法^[14]获取超参数最优组合,这是一种贝叶斯优化算法,与贪婪线性搜索算法不同,它依据历史搜索结果,得出此次实验的超参数组合,从而提升搜索效率。搜索结果表明:最优组中的最优超参数组合为:学习率 0.005,批数 2,学习率策略为余弦衰减,对应的测试误差与训练误差分别为 2.29 与 2.10。综合 2 种优化算法,最终得到最优超参数组合为:学习率 0.005,批数 2,学习率策略为余弦衰减,测试误差 2.46 个像素点,测试 AP 可达 95%。最优训练效果下多任务深度网络语义分割与关键点识别效果如图 5 所示。实际测试结果表明基于多任务学习 Mask R-CNN 分割与关键点识别网络特征点检测实际误差在 5 cm 以内,满足实际检测需求。



(a) 5 m 距离检测效果



(b) 10 m 距离检测效果



(c) 15 m 距离检测效果



(d) 20 m 距离检测效果

图 5 多任务深度网络语义分割与关键点识别效果

3 结语

1) 依据起重机日常运行状态检测信息,提出面

向起重机部件的多任务学习 Mask R-CNN 分割与关键点识别网络,包括 Mask R-CNN 关键部件定位框与区域语义分割网络结构、DeepLabCut 关键点检测网

络结构;

2) 利用贪婪线性搜索算法与贝叶斯优化算法,搜索此模型的最优超参数组合,训练收敛后,得到用于起重机大车、小车与吊具的定位框识别、区域实例分割与关键点识别最佳深度学习检测模型,测试误差2.46个像素点,测试AP可达95%,特征点实际检测误差在5cm以内。

下一步可依据边缘计算理论,将加速算法模型部署在桥式起重机现场边缘端设备,构成云边端检测系统,实现起重机运行状态量实时测量。

参考文献

- [1] 张昊.基于物联网的门式起重机结构健康监测管理系统[D].南京:南京理工大学,2018
- [2] 姜长城.门式起重机远程监测及诊断分析系统研发[D].南京:东南大学,2018.
- [3] 罗伟.门式起重机管理中安全监控管理系统的应用[J].中国设备工程,2018(19):32-33.
- [4] 陈洪志.大型起重机臂架旁弯图像测量算法研究[D].广州:

广东工业大学,2015.

- [5] 李明聪.基于图像的嵌入式吊绳状态监测系统[D].大连:大连海事大学,2016.
- [6] 张锐.基于机器视觉的塔式起重机控制策略研究[D].合肥:合肥工业大学,2019.
- [7] Redmon Joseph, Divvala Santosh, Girshick Ross, et al. You only look once: unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, America: Computer Vision Foundation, 2016. 779-788.
- [8] He Kaiming, Gkioxari Georgia, Dollar Piotr, et al. Mask R-CNN[C]// IEEE international conference on computer vision. Venice, Italy: IEEE, 2017. 2961-2969.
- [9] CARUANA R. Multitask Learning[J]. Machine Learning, 1997, 28(1):41-75.
- [10] KINGMA D P, BA J. Adam: a method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [11] MOCKUS J, TIESIS V, ZILINSKAS A. The application of Bayesian methods for seeking the extremum[J]. In L.C.W. Dixon and G.P. Szego, editors, Towards Global Optimization, North Holland, New York, 1978, 2:117-129.

Multi-task Learning Mask R-CNN Segmentation and Key Point Recognition Method for Bridge Crane Components

Yang Fan¹ Liang Minjian² Yang Ningxiang² Peng Xiaojun²

(1.School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China 2.Guangdong Institute of Special Equipment Inspection and Research Zhuhai Branch, Zhuhai 519002, China)

Abstract: Real time detection of crane operation status is one of the important guarantees for industrial safety production. A segmentation and key point network structure based on multi task learning mask R-CNN is proposed R-CNN crane key parts positioning frame and regional semantic segmentation network, crane key point detection network based on DeepLabCut; in a hook crane to collect data for testing, using greedy linear search algorithm and Bayesian optimization algorithm, the optimal super parameter combination of this model is: learning rate 0.005, batch size 2, learning rate strategy is cosine decay. The test error of the model is 2.46 pixels, and the test AP is up to 95%. The pixel error can be reflected within 5 cm of the actual error, which can meet the actual detection needs, and can be extended to the unmanned and automatic crane operation state monitoring.

Key words: bridge crane; key point recognition; deep learning; convolutional neural networks; multi-task learning; semantic segmentation

作者简介:

杨帆,男,1996年生,硕士研究生,主要研究方向:人工智能与精密检测技术研究。E-mail: 201820100566@mail.scut.edu.cn