计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2019,28(9):258−263 [doi: 10.15888/j.cnki.csa.007074]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于改进 Faster RCNN 的工业机器人分拣系统①

孙雄峰[1,2], 林　浒[2], 王诗宇[1,2], 郑飂默[2]

[1](中国科学院大学, 北京 100049)
[2](中国科学院 沈阳计算技术研究所 高档数控国家工程研究中心, 沈阳 110168)
通讯作者: 孙雄峰, E-mail: sunxiongfeng16@mails.ucas.edu.cn

摘　要: 传统的分拣作业无法伴随工作环境的变化进行相应的调整, 针对此种不足, 出现了基于机器视觉的分拣机器人的相关研究, 通过将图像处理和特征工程技术引入视觉模块, 使得分拣系统能适时的调整. 不同于这些方法, 本研究基于实验室的工业分拣系统, 将深度学习方法应用其中. 通过将 Faster RCNN 检测算法引入视觉模块并对区域提取网络 RPN 进行相关改进, 加快 Faster RCNN 模型的检测过程, 使得该系统满足工业的实时性要求. Faster RCNN 作为一种端到端的方法, 能自动对输入图像生成更具表达力的特征, 对相应目标提取相应特征, 这避免了人工设计特征, 它的特征自动生成能力使其能适用于各种场景, 这提升了工业分拣机器人的环境适应能力.

关键词: 物体检测; Faster RCNN; 区域提取网络; 分拣系统

引用格式: 孙雄峰,林浒,王诗宇,郑飂默.基于改进 Faster RCNN 的工业机器人分拣系统.计算机系统应用,2019,28(9):258–263. http://www.c-s-a.org.cn/1003-3254/7074.html

Industrial Robots Sorting System Based on Improved Faster RCNN

SUN Xiong-Feng[1,2], LIN Hu[2], WANG Shi-Yu[1,2], ZHENG Liao-Mo[2]

[1](University of Chinese Academy of Sciences, Beijing 100049, China)
[2](National Engineering Research Center for High-end CNC, Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

Abstract: Traditional sorting operation can not be adjusted with the change of working environment. In view of this shortcoming, a sorting robot is researched based on machine vision. By introducing image processing and feature engineering technology into the visual module, the sorting system can be adjusted in time. Unlike these methods, this research is based on the industrial sorting system in the laboratory and applies the deep learning method to it. By introducing faster RCNN detection algorithm into visual module and improving of Region Proposal Network (RPN), the detection process of faster RCNN model is accelerated, so that the system meets the real-time requirements of industry. faster RCNN, as an end-to-end method, can automatically generate more expressive features for input images and extract corresponding features for corresponding targets. This avoids the manual design features. Its automatic feature generation ability makes it suitable for various scenarios, which improves the environmental adaptability of industrial sorting robots.

Key words: object detection; Faster RCNN; region proposal network; sorting system

分拣机器人作为一种专用机器人的形式, 通常只能完成特定的工件分拣任务. 在目前实际应用中, 很多机器人是通过示教或是离线编程方式完成一些固定的操作[1]. 近年来, 随着机器视觉的发展, 在机器人中引入视觉模块, 在机器人视觉系统中利用图像处理技术[2]对工件图像进行预处理, 利用特征工程技术[3]抽象出特征

① 基金项目: 国家重大科技专项 (2017ZX04018001-003)
Foundation item: National Science and Technology Major Program (2017ZX04018001-003)
收稿时间: 2019-03-06; 修改时间: 2019-04-02; 采用时间: 2019-04-08; csa 在线出版时间: 2019-09-05

© 中国科学院软件研究所 http://www.c-s-a.org.cn

向量, 进而由具体的像素坐标框定出目标位置并对目标进行分类. 这种方法虽快速然精度不足, 难以满足更加智能化场景的需求.

随着深度学习的发展, 物体检测的相关研究呈现出从具体目标到泛化目标, 从特征工程到深度学习的发展趋势. 检测的目标从特定的几种物体到任意物体, 检测算法采用的特征从简单的像素, 到统计型描述子, 再到卷积神经网络中间层, 并且特征从稀疏到稠密、从主观设计的人工特征到可以抽象高层语义信息的自动生成特征. 在物体检测中, 特征是核心. 更具表达力的特征往往带来更优良的结果. 而在工业生产中, 并无成熟的深度学习算法应用, 究其原因, 就是深度学习算法尚且无法满足工业的实时性要求.

面向实时的 Faster RCNN[4]算法在检测速度上的提升给工业机器人带来了新的变化, 它基本满足了工业实时性要求. 同时, Faster RCNN 作为一种通用算法框架, 不同于特征工程技术主观设计的人工特征, 只要数据满足, 经训练后的模型能自动生成更具表达力的特征, 使得可以检测更为复杂、更为广泛的目标, 就能做到万事万物皆可检测.

1 分拣机器人的硬件构成

机器人分拣系统包括机器人模块、检测模块、通信模块、传送带装置和待分拣对象[1]. 如图 1 所示, 为分拣机器人硬件构成, 其中不包括计算机.

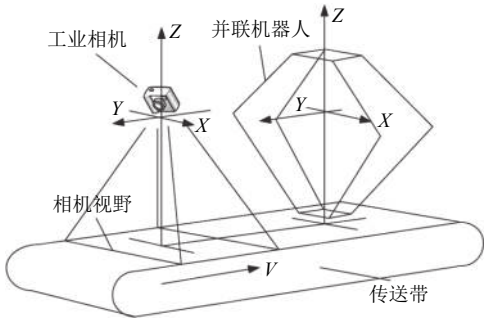

图 1 机器人分拣系统硬件构成

机器人模块包括阿童木并联机械手机器人, 机器人控制柜、示教编程器和驱动各关节的伺服交流电机组成. 机械臂末端为气动吸盘, 用于吸取传送带上物体, 并放入特定位置.

检测模块分为硬件和软件两部分. 康耐视 In-Sight7000 型工业相机作为硬件部分捕获图像, 相机将固定在传送带上方. 软件部分为 Faster RCNN 检测算法, 处理输入图像数据并得到目标类别和位置.

通信模块为一小型局域网, 用于工业相机图像和检测模块数据间传输.

传送带为长为一米的皮带输送机, 用于输送目标对象. 本研究中以工厂工件作为分拣对象.

2 分拣机器人工作原理

本系统中研究的重点在于检测算法的实现, 分拣系统分为训练和测试两部分. 通过已标定的数据训练 Faster RCNN 算法模型, 将训练好的模型部署到系统的检测模块中. 测试时, 将芯片放在传送板上以模拟工业传送环境. 相机以等时间间隔方式拍摄图像, 保证芯片在通过相机视野区域时, 相机将捕获到此芯片图像, 考虑到经由通信模块将传入的图像作为输入, 在计算机中检测模块运行 Faster RCNN 算法, 将处理后结果作为输出. 输出结果为此图像中目标物体的类别和矢量位置坐标, 再由通信模块传入并联机器人控制器. 计算机会记录当前图像的拍摄时间、数据传输时间和算法检测消耗时间, 由于传送带传送速度恒定, 机器人控制系统将获取到此目标的运动后位置和类别. 考虑到机械臂的运动速度, 机械臂随之执行抓取动作, 将目标物体放入其所属类别指定位置.

3 Faster RCNN

Faster RCNN 作为一种通用的物体检测框架, 它是在已有的算法框架上进行的改进. RCNN[5](Regions with CNN features) 方法将 CNN(卷积神经网络) 与区域候选框相结合, 在此基础上又出现了 Fast RCNN[6] 和 Faster RCNN 算法. 它们的基本思想都是将原始图片划分成不同的候选框, 然后将卷积神经网络 CNN 作为一种特征提取器, 将候选框提取出一个特征向量, 并训练一个分类器对特征向量进行分类. 如图 2 所示, 目标检测由这 3 部分构成.

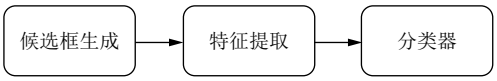

图 2 目标检测构成

R-CNN 采用了迁移学习的方法, 训练时基于 softmax 分类器在 ImageNet[7]上预训练好的模型使用领域特定的 PASCAL VOC[8]数据集进行网络微调.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

在 RCNN 基础之上, Fast RCNN 借助选择性搜索方法独立生成候选框, 使用经典分类网络 AlexNet[9]或 VGG-16[10]对候选框进行特征提取得到 ROI 层, 并使用多任务损失训练分类器和回归器. 其中的 RoI (Region-of-Interest) 池化层, 有效解决了网络低层无法训练的问题, 从而提高了检测精度.

然而 Fast RCNN 依旧无法进行实时检测, 其候选框生成阶段是独立于模型训练和测试过程的, 在测试时必须先进行候选框生成, 因而候选框生成阶段反而成了实时检测的瓶颈[11].

不同于 Fast RCNN 中借用选择性搜索方法生成候选框, Faster RCNN 基于卷积神经网络网络提出了区域提取网络 (Region Proposal Network, RPN) 生成候选框, 依然使用 Fast R-CNN 作为检测子. RPN 与 Fast RCNN 其实是共享了提取特征的卷积层, 从而 RPN 与 Fast RCNN 结合成了单个统一的 Faster RCNN 网络. 如图 3 所示, 可以将其大体分为卷积主干网络、RPN 微型网络、Fast RCNN 检测子和多任务损失四部分.

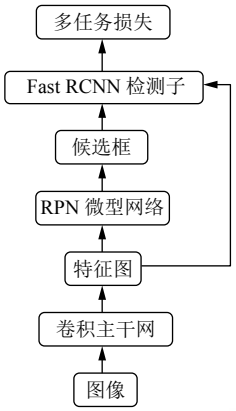

图 3 Faster RCNN 算法框架

本研究基于工业实时性要求, 对 Faster RCNN 的 RPN 微型网络进行改进, 对卷积主干网和多任务损失函数进行了调整, 更改了 RPN 微信网络的 anchor 尺度并降低了特征提取维度, 加快模型的检测过程.

3.1 卷积主干网

Faster R-CNN 中的卷积层借用的是经典分类模型的卷积层架构及其预训练好的权重, 将分类任务预训练好的模型用于相似的检测任务上, 直接对分类模型进行权值调整, 这大大减少了模型训练量.

如图 4 所示, 本研究中的卷积主干网借用 VGG-16 分类模型的卷积层部分, 特征图输出前没有进行池化, 这一点稍有变化. 所有的卷积操作的步长为 1, 边界填充为 1, 卷积核宽、高为 3×3, 这保证了卷积前后图像宽高不变, 池化层 (pooling) 采用 2×2 且步长为 2 的最大池化, 池化层不影响图像的通道数目, 但每次池化过后图像的宽高都将减半. 卷积的通道数有 64、128、256、512 等情况, 通道数量表示图像经卷积提取特征后的特征图数量. 每个卷积层之后 ReLu 激活函数进行非线性变换, 此操作不影响特征的宽高及通道数目. 因而输入图像经过 13 层卷积和 4 层池化后得到的输出特征图的宽高变为原图像的 1/16, 通道数目由 RGB 三通道变为 512.

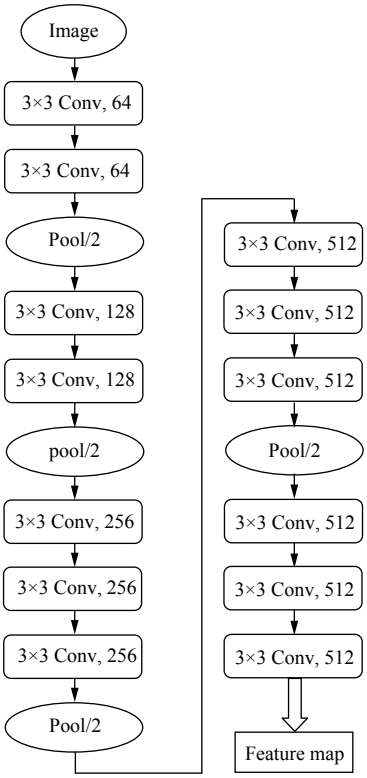

图 4 Faster RCNN 的卷积主干网

3.2 改进的 RPN 微型网络

RPN 微型网络采用滑动窗口方式对特征图上的每个点生成输入图像上的 9 个 anchor. 如图 5 所示, 为特征图中心点对应在输入图片上的 anchor. 外黑框为 800×600 像素点的原始图像, 内、中、外三种粗细方框分别代表 128、256、512 三种大小的尺度, 每个尺度下又有纵横比为 1∶2、1∶1、2∶1 三种情况, 因而每个滑动窗口对应 9 个 anchor.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

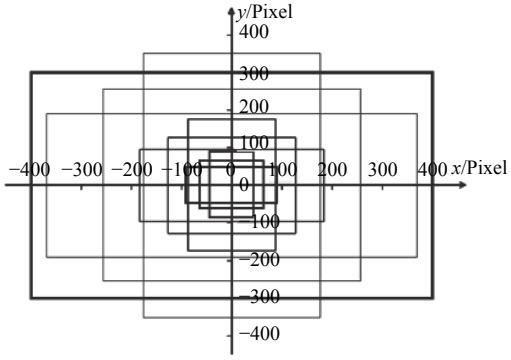

图 5 特征图中心点对应的 anchor

在原始 anchor 中, 为了保证能自适应各种尺度的目标物体, 设置了 128、256、512 这三种尺度, 由于在实验工业相机固定不动, 实际得到的目标图片不会由于距离产生大尺度的变化, 因而在本研究中, 进行尺度上的缩小和集中, 改进为 32、64、128 三种尺度, 纵横比依旧不变.

而三种尺度三种比例的 anchor 设置, 这相当于特征图中的一个点能够对应到原始图片感受野中的 9 个区域, 每个区域对应了一个 anchor. 而通过有监督学习的参数训练, 模型能够将参数调整到使计算得到的特征图能够对应到原始图片中的物体. 较小的尺度能够捕获到物体间的细小差异, 这使得不同类别物体能够得到区分, 而较大的尺度能够保证覆盖原始图片即全部感受野, 这使得原始图片不会遗漏未检测到的物体.

由于摄像机的视野固定, 因而传送带上的待分拣物体呈现在摄像机中的大小不会由于距离的远近产生变化, 因而其尺度变化具有一致性, 也就是说, 相同类别的物体在图像中占据的像素大小差异不会产生大的差异. 而原始 anchor 中设置的尺度变化过大, 并不适宜物体尺度变化不大的情况. 因而将原始尺度进行缩小以适应尺度一致的情况是有必要的.

如图 6 所示, 为 RPN 网络结构. 对于给定的输入图像, 经由卷积层产生卷积后特征图. RPN 微型网络在卷积后特征图上滑动一个 3x3 的小窗口, 每个窗口映射到一个 256 维的特征向量, 接着将特征向量送入两个分支网络：cls 分类网络和 reg 回归网络. 在这里, 将原始的 512 维特征向量改进为 256 维特征向量, 加快了检测的速度.

Cls 分类器对窗口映射的特征向量进行分类, 对每个 anchor 预测一个前景概率和后景概率, 因而会有 2×9=18 个概率值, 用 18 个神经元表示. Reg 回归器对窗口映射的特征向量进行回归, 对每个 anchor 预测中心点坐标及宽高的偏移量, 用(t_x,t_y,t_w,t_h)表示, 因而会有 4×9=36 个偏移量, 用 36 个神经元表示. 注意到对特征图的处理是以滑动窗口方式进行的, 因而这些过程可以通过卷积操作实现.

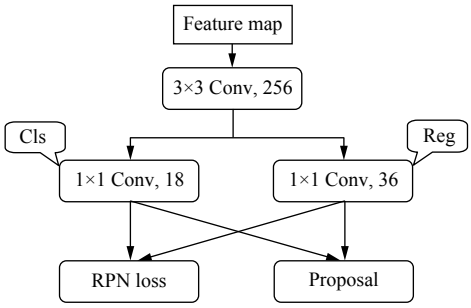

图 6 RPN 微型网络

3.3 多任务损失

对于 RPN 的训练, 本研究采用多任务损失, 将分类器的交叉熵损失与回归器的 SmoothL1 损失相结合. 为了多任务损失$L(\{p_i\},\{t_i\})$, 其分类损失为$L_{cls}(p_i,p_i^*)$, 回归损失为$L_{reg}(t_i,t_i^*)$. 对于所有样本的多任务损失$L(\{p_i\},\{t_i\})$为

$$\frac{1}{N_{cls}}\sum_i L_{cls}+\lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg} \tag{1}$$

其中, N_{cls}和N_{reg}为标准化项, λ为权衡系数.

考虑单个样本的分类损失

$$L_{cls}(p,p^*)=-\log p_{p^*} \tag{2}$$

其中, p^*为 anchor 对应的类别标记, p为其对应类别标记的预测概率.

多任务损失中只计算标记为正的 anchor 的回归损失, 单独考虑回归器损失函数

$$L_{reg}(t_i,t_i^*)=Smooth_{L1}(t_i-t_i^*) \tag{3}$$

其中,

$$Smooth_{L_1}(x)=\begin{cases}0.5x^2, & \|x\|<1\\ \|x\|-0.5, & others\end{cases} \tag{4}$$

t_i和t_i^*由四元组表示, 为了表示上的简便, 去掉下标i,只考虑单个样本下的回归器对 anchor 预测偏移t_i及 ground truth 对 anchor 的真实偏移t_i^*.

$$t=(t_x,t_y,t_w,t_h),\ t^*=(t_x^*,t_y^*,t_w^*,t_h^*) \tag{5}$$

其中,

© 中国科学院软件研究所 http://www.c-s-a.org.cn

$$t_x = (x - x_a)/w_a, t_x^* = (x^* - x_a)/w_a \tag{6}$$

$$t_y = (y - y_a)/h_a, t_y^* = (y^* - y_a)/h_a \tag{7}$$

$$t_w = \log(w/w_a), t_w^* = \log(w^*/w_a) \tag{8}$$

$$t_h = \log(h/h_a), t_h^* = \log(h^*/h_a) \tag{9}$$

RPN 网络采用随机梯度下降 SGD 方法优化多任务损失函数$L(\{p_i\},\{t_i\})$, 使得损失函数最小, 模型在优化过程中完成参数的调整, 找到一个局部最优解. 在测试时, 使用 RPN 对每个 anchor 预测出类别概率及标记为正的 anchor 的回归偏移量. 将 RPN 微型网络的输出采用非极大抑制方式得到回归偏移校正的候选框.

3.4 Fast RCNN 检测子

在 Fast RCNN 检测子中需要注意两点, 其一是分层采样加快随机梯度下降的训练速度, 其二是感兴趣 ROI 池化层反向传播特征图到池化层的映射可能会有重叠, 需要对重叠部分的梯度残差进行累加计算.

如图 7 所示, 将候选框 proposal 在特征图 feature 上对应部分进行 ROI 池化, 得到 ROI 特征图, 接着使用全连接 FC 进行特定类别的权值计算. 同样的, 检测部分还是采取 Fast R-CNN 中的多任务损失作为优化目标从而调整权值. 测试时, 检测网络在 FC 层之后只计算虚线部分得到预测的类别及位置, 并经过非极大抑制后得出最终的预测结果.

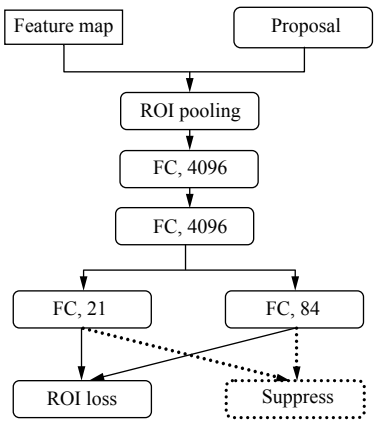

图 7 Fast RCNN 检测子

4 实验结果及分析

实验采用自标定数据集对模型进行训练, 将训练好的模型部署在机器人视觉模块中. 实验过程为在传送带上放置不同类型的待分拣对象, 随着传送带的恒定运动到达工业相机视野, 相机等时间间隔拍摄照片同时通过通信模块传输到视觉系统进行处理. 此时 Faster RCNN 算法的卷积主干网络将逐层提取图片特征, 注意到此时模型的参数已经固定, 在一次计算完毕后得到此张图片的卷积后特征图. 一方面改进后 RPN 网络将对此特征图生成候选框, 另一方面此特征图直接送入 Fast RCNN 检测子, 同候选框一起生成感兴趣区域得到 ROI 池化层, 通过分类损失和回归损失对候选框的类别和位置参数进行调整, 得到更加精细的结果. 若判断出相机视野中存在待分拣目标, 视觉系统会将此结果以一定的数据通信格式发送给机器人控制系统, 随后在经过时间计算后, 机械臂将摆动到相应位置吸取此目标.

如表 1 所示, 为不同方法的训练时间和测试时间及准确率比较结果. 可以看出这些模型训练时间耗时较长, 这是由于数据集较大, 数据的存储与传输也比较耗时. 而基于 Canny 算子的方法训练时间较短, 是因为相比卷积神经网络方法, 它只需要训练一个 SVM 分类器, 其特征向量的提取采用图像处理和 Canny 边缘检测算子等操作, 无需训练. 从测试时间可以看出, Fast RCNN 无法满足工业实时性需求, 而 Faster RCNN 也只是勉强实时, 并不能做到真正实时.

表 1 不同方法对比

方法	训练时间 (h)	测试时间 (s)	准确率 (%)
Canny	3.5	0.032	73.6
Fast RCNN	8.75	0.32	85.1
Faster RCNN	17.2	0.185	91.4
Faster RCNN 改	15.8	0.049	89.7

分析改进的 Faster RCNN 模型检测时间的效果, 可以发现测试时间显著缩减. 这是由于将原始维特征向量改进为 256 维特征向量, 将特征维度缩减了一半, 这减少了模型的参数数量, 并减少了每张图片候选框生成的个数, 因而通过特征图生成候选框需要的时间更短, 而实验结果也证实了这一点.

虽然模型的训练时间相比而言变化不大, 但其检测速度大大加快, 这使其真正做到了工业实时, 能够应用在工业生产中. 这种改进是基于相机固定不变的情况下, 因而其准确率并没有太大的降低. 相比于 Canny 算子方法而言, 虽然没有其快速, 但是已经能达到工业实时性标准, 并且相比而言, 其准确率也有显著提升.

5 总结与展望

相比传统方法而言, 深度学习方法准确率高, 然而

© 中国科学院软件研究所 http://www.c-s-a.org.cn

其在工业中实时性要求大多算法难以满足. 针对相机固定的工业分拣机器人, 通过对勉强实时的 Faster RCNN 算法进行改进, 本研究将其应用到机器人视觉系统中, 做到了真正实时, 在分拣正确率上有了显著提升. Faster RCNN 的改进, 使其能自由应用于各种场景, 这提升了机器人的环境适应能力, 并提升了其智能程度和技术水平. 然而此种方法在分拣成功率上尚有提升空间, 接下来的研究计划是分析工业场景的特点, 研究单阶段算法框架如 YOLO[12]和 SSD[13]算法, 选取更针对性的算法框架, 在满足工业实时性的基础上进一步提升分拣效果和智能化程度.

参考文献

1 王诗宇, 林浒, 孙一兰, 等. 基于机器视觉的机器人分拣系统的设计与实现. 组合机床与自动化加工技术, 2017, (3): 125–129, 133.

2 王诗宇, 林浒, 孙一兰, 等. 一种改进的 Canny 算子在机器人视觉系统中的应用. 计算机系统应用, 2017, 26(3): 144–149. [doi: 10.15888/j.cnki.csa.005635]

3 朱良, 林浒, 吴文江. 基于机器视觉的工业机器人工件定位. 小型微型计算机系统, 2016, 37(8): 1873–1877. [doi: 10.3969/j.issn.1000-1220.2016.08.048]

4 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada. 2015. 91–99.

5 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587.

6 Girshick R. Fast R-CNN. arXiv: 1504.08083, 2015.

7 Deng J, Dong W, Socher R, *et al*. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA. 2009. 248–255.

8 Everingham M, Van Gool L, Williams CKI, *et al*. The PASCAL Visual object classes (VOC) challenge. International Journal of Computer Vision, 2010, 88(2): 303–338. [doi: 10.1007/s11263-009-0275-4]

9 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.

10 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.

11 Szegedy C, Liu W, Jia YQ, *et al*. Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. 2015. 1–9.

12 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.[doi:10.1007/978-3-319-46448-0_2]

13 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. arXiv: 1506.02640, 2015.

© 中国科学院软件研究所 http://www.c-s-a.org.cn