

基于难负样本挖掘的改进 Faster RCNN 训练方法

艾 拓 梁亚玲 杜明辉

(华南理工大学电子与信息学院 广州 510641)

摘 要 目标检测方法甚高速卷积神经网络(Faster Region-based Convolutional Neural Network, Faster RCNN)在训练过程中存在负样本远多于正样本的问题,即数据集不平衡问题。针对该问题,提出了一个综合定位误差和分类误差的判别函数用于判别难正样本,基于该函数和难负样本挖掘提出了改进的自助采样法,并提出了基于该自助采样的“五步训练法”用于训练 Faster RCNN。与传统的 Faster RCNN 训练方法相比,五步法加强了对难样本的学习,提高了网络泛化能力,减少了误判;训练出的模型在 Pascal VOC 2007 数据集上测试的平均正确率均值(mean Average Precision, mAP)提高了 2.4%,在 Fddb(Face Detection Data Set and Benchmark)相同检出率下误检率降低了 3.2%,且边框拟合度更高。

关键词 甚高速区域卷积网络,目标检测,难负样本挖掘,自助采样

中图分类号 TP183 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.05.043

Improved Faster RCNN Training Method Based on Hard Negative Mining

AI Tuo LIANG Ya-ling DU Ming-hui

(School of Electronics and Information, South China University of Technology, Guangzhou 510641, China)

Abstract In the training process of object detection method named Faster RCNN(Faster Region-based Convolutional Neural Network), there is a data imbalance problem which means that training data contains an overwhelming number of negative examples. Aiming at this problem, a discriminant function was proposed to distinguish hard positive examples, which combines location loss and classification loss. Based on this function and hard negative mining, an improved bootstrap sampling method was proposed. Five-step training method was proposed by introducing the bootstrap sampling into traditional Faster RCNN training. Comparing with the traditional training, this method improves network's generalization ability, reduces false positive rate, and can learn hard example better. The experimental results show that the model trained by five step attains 2.4% higher mAP(mean Average Precision) on Pascal VOC 2007 dataset, reduces false positive by 3.2% on Fddb(Face Detection Data Set and Benchmark) with the same true positive rate, and gets higher fitting degree of boundary box.

Keywords Faster RCNN, Object detection, Hard negative mining, Bootstrap sampling

Faster RCNN^[1]是目标检测中非常具有代表性的算法,拥有强大的检测能力和实时性,在船舶检测^[2]、合成孔径雷达图像检测^[3]、电力小部件检测^[4]等领域有着深入的研究和广泛的应用。但在训练目标检测器其存在负样本大幅多于正样本的问题,即数据集不平衡的问题。Faster RCNN 以不同的比率通过对正负样本进行下采样来解决该问题。该方法容易实施,但是对难样本的学习且鲁棒性较差。文献[5-9]在 VJ, DPM+SVM 和 HOG+SVM 等目标检测算法中引入了难负样本挖掘(Hard Negative Mining, HNM)来解决数据集不平衡的问题,减少了检测器误判。但是这些方法都缺少对难正样本的学习,正样本的检测精度依然有提升的空间。

因此,本文基于难负样本挖掘提出了一种改进的自助采

样方法,该方法在难负样本挖掘的基础上加入了对难正样本的挑选和训练,并且设计了一种“五步训练法”将该自助采样法引入到 Faster RCNN 训练中。

本文第 1 节和第 2 节分别介绍难负样本挖掘和 Faster RCNN 的原理;第 3 节详细介绍改进的自助采样法;第 4 节给出实验结果与分析。

1 难负样本挖掘介绍

难负样本挖掘的思想源于文献[10],Sung 和 Poggio 针对训练目标检测器存在的数据不平衡问题提出了自助采样方法,即在大量的背景区域中挑选出易被误判为目标的难负本来训练目标检测器,使得检测器的鲁棒性更强,其过程如下。

到稿日期:2017-03-31 返修日期:2017-06-05 本文受广州市科技计划项目(201707010070)资助。

艾 拓(1995-),男,硕士,主要研究方向为计算机视觉、深度学习;梁亚玲(1977-),女,博士,讲师,主要研究方向为模式识别、图像处理, E-mail:ylliang@scut.edu.cn(通信作者);杜明辉(1964-),男,教授,博士生导师,主要研究方向为信号处理、图像处理。

- 1) 将所有目标区域作为正样本,随机挑选一部分背景区域作为负样本;
- 2) 训练分类器;
- 3) 对所有不含目标的图像进行检测,将所有被误判成目标的区域加入到负样本中,并将这些负样本称为难负样本;
- 4) 重复步骤 2) 和步骤 3)。

该方法多次训练难负样本,减少了检测器的误判,被广泛用于目标检测器训练中,并被称为难负样本挖掘。

2 Faster RCNN 原理

Faster RCNN 是基于深度卷积神经网络的目标检测算法,其结构如图 1 所示。Faster RCNN 分为 RPN (Region Proposal Network) 和 Fast RCNN^[11] 两个部分。RPN 用于预测输入图像中可能包含目标的候选区域;Fast RCNN 用于分类候选区域,并修正候选区域的边界框。这两个网络使用共享的卷积神经网络来提取特征。

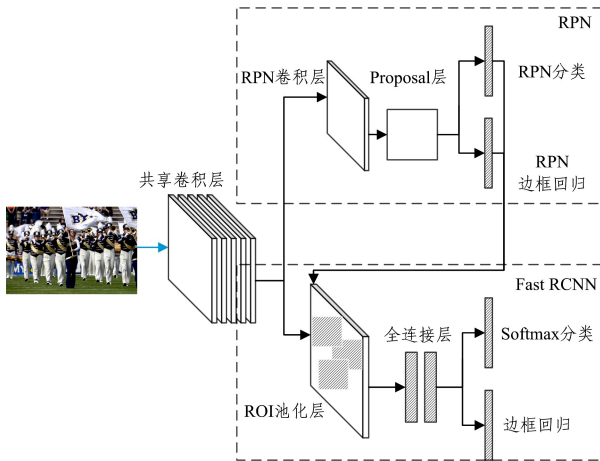


图 1 Faster RCNN 结构

Fig. 1 Structure of Faster RCNN

RPN 网络的工作原理为:使用共享的卷积网络对图像进行特征提取,然后将特征输入到 RPN 卷积层进行压缩。Proposal 层在大小为 $w * h$ 的特征图上的每个像素位置采集 k 个初始区域,共得到 $w * h * k$ 个被称为锚点的初始区域。使用分类器挑选出可能包含目标的锚点,再使用边框回归^[6]修正锚点边界框以得到目标候选区域。

在 RPN 的训练阶段,Proposal 层先对生成的锚点做标注,满足下列两个条件的锚点被标注为正样本,否则为负样本。

- 1) 锚点与目标相交率(IoU)最大;
- 2) 锚点与至少一个目标相交率不低于 0.7。

标注锚点后,将式(1)作为 RPN 的代价函数来对 RPN 网络进行训练:

$$L_{RPN}(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

其中, i 代表锚点的索引值; N_{cls} 为锚点数量; N_{reg} 为正的锚点数量; p_i 为预测置信度,其取值范围为 $0 \sim 1$; p_i^* 为标注值, p_i^* 取 1 表示正样本, p_i^* 取 0 表示负样本; t_i 为预测的锚点边

界框修正值; t_i^* 为实际的锚点边界框的修正值(边界框的修正值为一个四维向量,其中包含了边界框的中心点位置修正值以及边界框的宽高修正值); L_{cls} 是分类代价,其使用交叉熵代价函数,表示预测置信度与目标值的误差; L_{reg} 是边框回归代价,使用文献[11]中定义的 smooth L1 代价表示预测的修正值与实际值的误差。式(1)通过平衡权重 λ 将分类代价和边框回归代价结合在一起,利用该代价函数训练好的 RPN 能够预测目标锚点,修正这些锚点边界框即可得到目标候选区域(Region of Interest, ROI)。这些候选区域被输入到 Fast RCNN 部分做进一步检测。

Fast RCNN 中的 ROI 池化层使用的是空间金字塔池化^[12],其输入为目标候选区域对应的特征图,输出为固定长度的特征向量。特征向量通过全连接层分别输入到 Softmax 分类器和边框回归中以预测目标类别和边框修正值。最后,使用非极大值抑制^[13]消除重复的边界框以得到检测结果。训练 Fast RCNN 的代价函数如下:

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda [u \geq 1] L_{reg}(t^u, v) \quad (2)$$

其中, L_{cls} 为分类代价; L_{reg} 为边框回归代价; p 为分类预测结果; u 表示第 u 类; t^u 为第 u 类边框回归预测的修正值; v 为实际修正值。

训练 Fast RCNN 的数据来自于 RPN 预测的候选区域。由于 RPN 是对目标的粗略预测和定位,这些候选区域中的背景区域远多于目标区域,背景与前景的比率高达 $100:1$ ^[14]。对于这种正负样本不平衡的情况,原 Faster RCNN 通过随机采样抽取一定比率的正负样本,得到一个固定大小的 minibatch,再使用这个 minibatch 训练 Fast RCNN。该方法简单且容易实现,但是缺少对难负样本和难正样本的学习,其分类精度和边框精度仍然有提升的空间。为了减少误判并提高边框精度,本文提出了一种自助采样法来提升 Faster RCNN 的训练效果。

3 本文训练方法

本文的自助采样法的实施方式为:在 Faster RCNN 训练前对每个样本测试一次,挑选出结果中误差较大的正负样本加入训练,以加强对难样本的训练。不同于难负样本挖掘,该方法增加了对难正样本的挑选。采集难正样本需要综合分类误差和边框误差两个因素。如果采样策略偏向于分类误差较大的正样本,训练结果中预测概率就会偏高,漏检率减小,但边框精度可能会降低,反之亦然。置信度小容易造成漏检,边框与目标相交率小也易造成误检。因此,本文挑选难正样本时将优先挑选与目标相交率小的样本,以减少训练结果的误检;当样本与目标相交率较大时,则偏向于挑选分类精度低的样本,从而减少训练结果中的漏检。对此,本文设计了一个判别函数,如式(3)所示:

$$L(o, p) = L_{IoU}(o) + L_{score}(p) \quad (3)$$

$$L_{IoU}(o) = \frac{1}{1 + e^{k * (o - 0.5)}} \quad (4)$$

$$L_{score}(p) = (1 - p) \quad (5)$$

其中, L_{IoU} 为边框误差; L_{score} 为分类误差, 其定义分别如式(4)和式(5)所示。 L_{IoU} 中 o 为样本与目标的相交率, k 为 L_{IoU} 对阈值 0.5 的敏感系数, 通常将相交率 0.5 作为判断边框是否正确的阈值, 将该值引入公式。 L_{score} 中 p 为分类置信度, L_{score} 与 p 呈线性负相关。 每个样本与目标相交率 o 和预测置信度 p 的取值范围为 $0 \sim 1$ 。 使用判别函数计算每个样本对应的判别值, 当 $L(o, p)$ 大于某个阈值时将其判定为难正样本。

图 2 是当 k 取不同值时, L_{IoU} 和 L_{score} 的函数曲线。 图中, 当 k 取 0 时, L_{IoU} 为一个常数, 此时判别函数 L 的大小只与预测置信度 p 相关, 与相交率 o 无关; 当 k 取 10 时, 判别函数 L 对 o 小于 0.5 的样本比较敏感, 这些较多样本被挑选; 当 k 为无穷大时, L_{IoU} 为 0.5 处的一个阶跃函数, o 小于 0.5 的样本完全被优先判定为难样本。 因此, k 越大, 相交率 o 对判别函数的影响力越大, 挑选出的难正样本中边框误差大的样本比重更大。

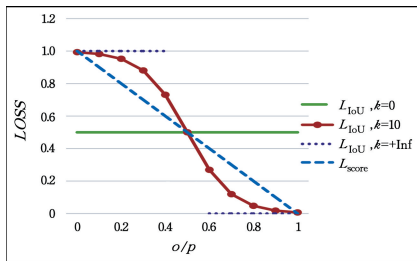


图 2 分类误差函数和 k 取不同值时的边框误差函数

Fig. 2 Bounding-box error function of classification error function and different k values

负样本只影响分类误差而不影响边框误差, 因此只需要挑选分类误差大的样本作为难负样本。 由于检测时背景的预测置信度不会被使用, 而是通过目标类别的置信度拒绝背景区域, 因此负样本在目标类别中的预测置信度越大, 误差就越大。 本文通过设定阈值的方法来挑选出预测置信度不小于该阈值的难负样本。

最后将得到的所有正负难样本组合起来, 并从中随机挑选一定数量样本组成 minibatch, 挑选时并不维持前景与背景

的固定比率, 因为数据的分布反映了类别的难易程度, 难样本多的类应该更多地被学习^[15]。 当难样本数量不够时, 将负样本按照之前计算的预测置信度从大到小排序, 挑选排序靠前且误差大的样本补齐 minibatch。

四步法和近似联合训练是训练 Faster RCNN 网络的两种方法。 由于四步法中 RPN 和 Fast RCNN 是被分开训练的, 自助采样难以实施。 因此, 本文结合这两种方法设计了一种“五步法”将自助采样法加入到 Faster RCNN 训练中。 “五步法”的步骤如下:

- 1) 训练 RPN 网络和共享的卷积层, 并对每张图片预测 2000 个候选区域;
- 2) 使用第一步生成的区域训练 Fast RCNN 和共享的卷积层;
- 3) 固定卷积层, 训练 RPN 并预测 2000 个候选区域;
- 4) 固定卷积层, 训练 Fast RCNN;
- 5) 每次迭代载入一张图片, 使用 RPN 预测 2000 个候选区域并使用本文设计的自助采样法挑选出难样本, 使用这些难样本通过近似联合来训练整个网络。

4 实验结果及分析

为了验证本文所提自助采样法的有效性, 本文将所提方法分别在目标检测和人脸检测上与传统方法进行对比。 本文实验在 Caffe^[16] 上进行, 操作系统为 Ubuntu 16.04, 硬件信息为: Intel Xeon 24 核 CPU 和 GTX Titan Black 显卡, 64GB 内存。

训练 Faster RCNN 网络时每次迭代载入一张图片, 训练前保持图像宽高比不变并将其大小进行缩放, 缩放策略为在保证最大边不超过 1000 个像素的前提下按最小边放缩至 600, 否则按最大边缩放至 1000。 网络结构和模型参数采用文献[1]中定义的 ZF^[17] + RPN 模型, 并使用在 ImageNet^[18] 数据集上预训练好的 ZF 模型来初始化网络, 然后使用“五步法”训练这个网络, 式(3)中判别函数的敏感度参数 k 取 10, 难正样本采样阈值取 1.0, 难负样本采样阈值取 0.05, minibatch 的大小为 128, 每个阶段的迭代次数和学习率如表 1 所列。

表 1 各阶段训练的迭代次数和学习率

Table 1 Iteration number and learning rate of each training step

	Step 1 RPN		Step 2 Fast RCNN		Step 3 RPN		Step 4 Fast RCNN		Step 5 bootstrap	
Iteration/k	60	20	30	10	60	20	30	10	15	10
Learning rate	0.001	0.0001	0.001	0.0001	0.001	0.0001	0.001	0.0001	0.0001	0.00001

目标检测部分实验使用 Pascal VOC 2007^[19] 数据集, 其中包含 20 种物体类别。 训练数据包含了各种场景的图片, 每张图片对应一个标签文件, 标签文件拥有每个目标所属类别以及边界框左上角和右下角两个坐标。 数据集一共有 9963 张图片、24640 个被标记的目标。 本文使用其中 50% 的图片用于训练, 使用另外 50% 的图片用于测试。

表 2 是近似联合训练、四步法和五步法在 Pascal VOC 2007 测试集上的测试结果对比。 相对于四步法和近似联合训练, 五步法提升了大部分类的正确率均值 (AP), 从表中可以看出, 其尤其对单车、桌子、盆栽、沙发等类别有非常明显的

提升效果, 对只有鸟、摩托、船和狗等类别的正确率均值略有下降, 平均正确率均值 (mAP) 提高了 2.4%。 各个类别提升幅度不一致是由类别之间数据数量的不同和难易程度的不同造成的。 鸟、船、狗和摩托这 4 个类别的正确率均值略微下降, 这是由于自助采样带来了数据样本扰动, 3 种不同的训练方式带来了算法的参数扰动。 通常, 训练时加入数据样本扰动和算法参数扰动是集成学习中增强基学习器之间的差异性的办法。 因此, 在少数类别上五步法训练的模型检测性能下降正是这种差异性的体现, 但其总体性能依然得到了提高。

表 2 近似联合训练、四步法与五步法在 Pascal VOC 2007 上的测试结果
Table 2 Detection results of approximate joint training, Four-step and Five-step

方法	mAP	飞机	单车	鸟	船	瓶子	大巴	汽车	猫	椅子	牛	桌子	狗	马	摩托	人	盆栽	羊	沙发	火车	电视
近似联合训练	59.1	63.6	71.7	52.0	46.2	31.9	65.8	74.0	69.6	37.8	61.3	60.5	67.2	75.1	67.6	65.0	33.9	58.3	51.8	70.7	57.9
四步法	59.2	63.9	69.9	56.7	44.5	30.3	66.8	73.5	71.0	35.7	63.3	59.7	65.7	76.7	70.8	64.4	30.4	58.3	53.6	72.3	56.7
五步法	61.6	64.8	74.4	54.2	45.6	35.0	69.4	75.8	73.9	38.3	66.8	63.1	66.7	77.2	70.6	66.4	34.8	61.3	58.3	72.9	62.6

人脸检测部分实验的训练数据集为 WIDER FACE^[20], 其中包含 32203 张图像, 一共标记了 393703 张人脸。本文使用 FDDB 数据集^[21] 测试训练结果, FDDB 一共有 2845 张图像, 标记了 5171 张人脸, 测试结果用 ROC 曲线表示。ROC 曲线表示受试者工作特征曲线, 用于分析目标检测结果。通过设定一系列检测阈值, 得到一系列真正类率(True Positive Rate)和假正类率(False Positive Rate)。图 3 中, 横轴代表假正类率, 纵轴代表真正类率。真正类率为检出率, 假正类率为误检率。曲线上升得越快, 代表鲁棒性越强, 最大值越大代表检测器检出能力越好, 因此当 ROC 总体处于上方时, 代表检测器的综合性能更强。

将 WIDER FACE 数据集作为训练集, 训练后的结果如图 3 所示。从图中可以看出, 当检出率相同时, 五步法的误检率比四步法低 3.2%, 这说明五步法的综合性能优于四步法。

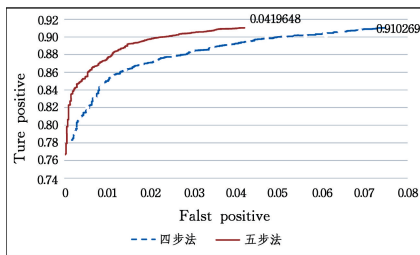


图 3 FDDB 测试结果 ROC 曲线
Fig. 3 ROC curves of FDDB test results

为了直观表示本文方法在目标检测中的效果, 图 4 和图 5 分别给出四步法和五步法在 Pascal VOC 2007 和 WIDER FACE 数据集上的检测结果。在图 4 和图 5 中, 图(a)和图(c)对应的是四步法的结果, 图(b)和图(d)为五步法的结果。图 4 中四步法的检测阈值为 0.7, 五步法的检测阈值为 0.5; 图 5 中四步法的检测阈值为 0.8, 五步法的检测阈值为 0.4。从图中可以看出, 即使使用了更低的阈值, 五步法的误检仍少于四步法, 并且边框精度更高。

本次研究的实验结果与预期一致, 由测试结果的数据可知, Faster RCNN 训练中加入基于 HNM 改进的自助采样后, 其精度和鲁棒性都得到加强。本文设计了一个判别函数, 使得采样更加偏向于挑选边框误差大的正样本, 因此五步法的训练结果中边界框更加精确。例如, 图 4(a)和图 4(c)是四步法的训练结果, 图 4(b)和图 4(d)是五步法的训练结果。图 4(b)中小女孩和成年男人的边界框和图 4(d)中黑衣男人的边界框刚好将人物包围, 而四步法的预测结果只包含了一部分。由于五步法加强了对难负样本的学习, 因此其鲁棒性更强。图 5(a)和图 5(c)是四步法的训练结果, 其中玻璃位置和手的位置分别存在误检。图 5(a)和图 5(c)是五步法的训练结果, 其中没有出现误检。图中五步法的预测置信度相对于四步法都有所降低, 但是其强鲁棒性弥补了这一缺点。五步法可以使用更低的阈值来提高检出率并且可保证较低的误判率。本

文方法同样检测出了所有目标, 且没有出现误判, 而四步法的结果出现了误判, 且边界框不够精确, 因此五步法总体提升了 Faster RCNN 的检测性能。

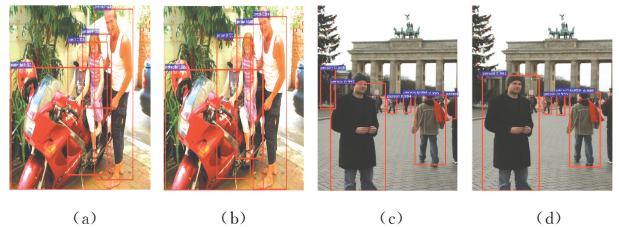


图 4 四步法和五步法目标检测对比结果
Fig. 4 Comparison of object detection result of four-step and five-step



图 5 四步法和五步法人脸检测对比结果
Fig. 5 Comparison of face detection result of four-step and five-step

结束语 本文基于难负样本挖掘提出了一种改进的自助采样法, 并为该方法设计了判别函数以用于挑选难正样本。判别函数使用较为灵活, 可以根据实际需要, 通过调整敏感系数 k 在检出和误检之间做取舍。本文使用该方法改进了 Faster RCNN 的训练方式, 相较于原来的四步法, 本文方法能够减少检测器的误判, 在 FDDB 相同的检出率下, 误检率降低了 3.2%, 主观上提高了目标边界框的精确度。通过图 4 和图 5 可以看出, 五步法边框更加准确, 总体上提升了检测器的性能, 在 Pascal VOC 2007 数据集上测试的 mAP 提高了 2.4%。

参 考 文 献

[1] 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, 2017, 39(6): 1137-1149.

[2] YANG M, RUAN Y D, CHEN L K, et al. New Video Recognition Algorithm for Inland River Ships Based on Faster RCNN [J]. Journal of Beijing University of Posts and Telecommunications, 2017, 40(S1): 130-134. (in Chinese)
杨名, 阮雅端, 陈林凯, 等. 甚高速区域卷积神经网络的船舶视频目标识别算法[J]. 北京邮电大学学报, 2017, 40(S1): 130-134.

[3] LI J B, YANG W H, XU J Q, et al. Deep Convolutional Network Based SAR Image Object Detection and Recognition [J]. Navigation Position and Timing, 2017, 4(1): 60-66. (in Chinese)
李君宝, 杨文慧, 许剑清, 等. 基于深度卷积神经网络的 SAR 图像目标检测识别[J]. 导航定位与授时, 2017, 4(1): 60-66.

[4] WANG W G, TIAN B, LIU Y, et al. Study on the electrical

- devices detection in UAV images based on region based convolutional neural networks[J]. *Journal of Geo-information Science*, 2017, 19(2): 256-263. (in Chinese)
- 王万国, 田兵, 刘越, 等. 基于 RCNN 的无人机巡检图像电力小部件识别研究[J]. *地球信息科学学报*, 2017, 19(2): 256-263.
- [5] VIOLA P, JONES M J. Robust Real-Time Object Detection[C]// *International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and Sampling*. Vancouver: IEEE Press, 2001.
- [6] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2010, 32(9): 1627-1645.
- [7] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]// *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*. IEEE Computer Society, 2005: 886-893.
- [8] CHEN L Y. Object Detection Based on Ensemble of Exemplars [D]. Shanghai: Shanghai Jiao Tong University, 2015. (in Chinese)
- 陈璐艳. 基于范例集成的目标检测模型研究[D]. 上海: 上海交通大学, 2015.
- [9] ZHANG X S. Research on Traffic Sign Detection in Cluttered Outdoor Scene[J]. *Computer Applications*, 2015, 34(10): 39-42. (in Chinese)
- 张雪松. 复杂室外场景中交通标志检测研究[J]. *自动化技术与应用*, 2015, 34(10): 39-42.
- [10] SUNG K K. Learning and example selection for object and pattern detection[M]. Cambridge: Massachusetts Institute of Technology, 1996.
- [11] GIRSHICK R. Fast R-CNN[C]// *IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015: 1440-1448.
- [12] 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*, 2015, 37(9): 1904-1916.
- [13] NEUBECK A, VAN GOOL L. Efficient Non-Maximum Suppression[C]// *International Conference on Pattern Recognition*. IEEE Computer Society, 2006: 850-855.
- [14] WAN S, CHEN Z, ZHANG T, et al. Bootstrapping Face Detection with Hard Negative Examples[OL]. <https://arxiv.org/abs/1608.02236>.
- [15] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training Region-based Object Detectors with Online Hard Example Mining [OL]. <https://arxiv.org/abs/1604.03540>.
- [16] JIA Y Q, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding [J]. *Eprint Arxiv*, 1408.5093.
- [17] ZEILER M D, FERGUS R. Visualizing and Understanding Convolutional Networks [M]. *Computer Vision – ECCV*. 2014: Springer International Publishing, 2014: 818-833.
- [18] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [19] EVERINGHAM M, GOOL L, WILLIAMS C K, et al. The Pascal Visual Object Classes (VOC) Challenge [J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [20] YANG S, LUO P, CHEN C L, et al. WIDER FACE: A Face Detection Benchmark[C]// *IEEE Conference on Computer Vision & Pattern Recognition*. 2015: 5525-5533.
- [21] JAIN V, LEARNED-MILLER E. FDDB: A Benchmark for Face Detection in Unconstrained Settings [M] // *UMass Amherst Technical Report*. University of Massachusetts, 2010.
- (上接第 231 页)
- [7] MYRONENKO A, SONG X. Point-set registration: coherent point drift[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(12): 2262-2275.
- [8] AGUILAR W, FRAUEL Y, ESCOLANO F, et al. A robust graph transformation matching for non-rigid registration [J]. *Image and Vision Computing*, 2009, 27(7): 897-910.
- [9] XIONG Z, ZHANG Y. A novel interest-point-matching algorithm for high-resolution satellite images[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2009, 47(12): 4189-4200.
- [10] CAETANO T S, MCAULEY J J, CHENG L, et al. Learning graph matching[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(6): 1048-1058.
- [11] ZHANG K, LI X, ZHANG J. A robust point-matching algorithm for remote sensing image registration [J]. *IEEE Geoscience & Remote Sensing Letters*, 2014, 11(2): 469-473.
- [12] MA J, ZHOU H, ZHAO J, et al. Robust feature matching for remote sensing image registration via locally linear transforming [J]. *IEEE Transactions Geoscience & Remote Sensing*, 2015, 53(12): 6469-6481.
- [13] ZHAO M, AN B, WU Y, et al. A robust delaunay triangulation matching for multispectral/multidate remote sensing image registration[J]. *IEEE Geoscience & Remote Sensing Letters*, 2015, 12(4): 711-715.
- [14] LIU Z X, SHAO F, JING Y, et al. Application of graph structure in aerial image feature matching[J]. *Computer Engineering and Application*, 2018, 54(1): 19-24. (in Chinese)
- 刘朝霞, 邵峰, 景雨, 等. 图结构在航空遥感图像特征点匹配中的应用[J]. *计算机工程与应用*, 2018, 54(1): 19-24.
- [15] PAUL S, PATI U C. Remote Sensing Optical Image Registration Using Modified Uniform Robust SIFT[J]. *IEEE Geoscience & Remote Sensing Letters*, 2016, 13(9): 1300-1304.
- [16] MA W, WEN Z, WU Y, et al. Remote Sensing Image Registration with Modified SIFT and Enhanced Feature Matching[J]. *IEEE Geoscience & Remote Sensing Letters*, 2017, 14(1): 3-7.
- [17] AL-KHAFAJI S L, ZHOU J, ZIA A, et al. Spectral-Spatial Scale Invariant Feature Transform for Hyperspectral Images[J]. *IEEE Transactions on Image Processing*, 2018, 27(2): 837-850.
- [18] HARSANYI J C. Detection and classification of subpixel spectral signatures in hyperspectral image sequences[D]. Baltimore: University of Maryland Baltimore County, 1993.
- [19] DU Q, REN H, CHANG C. A comparative study for orthogonal subspace projection and constrained energy minimization [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2003, 41(6): 1525-1529.