# 行人领域综述

Author：ydwu

Create\_Date：2019.11.19

Update\_Date：

人体活动理解大致可以分为identity， communication和state/action三个大方向进行。其中，行人重识别ReID是identity的一个代表分支。

## 行人检索

行人检索Person Search = Pedestrian Detection + Person Re-ID

目前大家的焦点大都集中在 Person Re-ID 之上，比如 CVPR'18 的 Re-ID 相关文章数目就达到了惊人的三十多篇。Person Re-ID 处理的对象：probe image 和 gallery image 都是已经切好块了的，即图像中基本只包含了我们所感兴趣的这个人，其他可能的干扰信息（比如背景、遮挡、另外的行人等）已经尽可能降到了最低。

当然了，这是为了创建一个统一的、标准化的研究课题而做的一个实验室环境下的约束。随着 Person Re-ID 的性能被逐渐刷高，是时候考虑一下与实际应用更接近一些的任务了，比如 Person Search。

## 行人跟踪

在视频中进行Keypoint检测怎么加入tracking？能否用keypoint检测做行人检测和行人多目标跟踪？

能否用keypoint检测做行人检测和行人多目标跟踪这个的话，其实在做行人检测上，可以用keypoint的检测结果做post filter滤掉行人检测的FP。Bottom Up的方法可以用紧框去做行人检测，最好与现有检测框架共同去检测行人，行人多目标跟踪的话，现在有很多方法是用人体关节点作为特征来比对跟踪目标，keypoint的作用主要在用于提供特征。

## 行人重识别

即人员重新识别，跨镜追踪，这是一个2018年开始热门的研究课题，旨在自动匹配多摄像机网络中的特定人。特征表示和度量学习是人员重新识别的两个重要问题。

### 行人重识别的算法

reid的主流研究方向基本是随着人脸识别来的；最开始是facenet，后来用度量学习；再后来用GAN。

query image是所指定的图片，而gallery images是指检索的集合。

在行人识别领域有两类流型的模型（都是基于 CNN），一类是 verification model，一类是 identification model，由于损失函数不同，两类模型各有优缺点。

1. verification model 以图片对作为输入，经过一个非线性函数（CNN），得到特征后进行相似度判断，这样只利用到了弱的 Re-id 标签。
2. identification model 训练时是以一张图片作为输入，经过一个非线性函数（CNN），得到特征，然后进行多类判断。在测试阶段，把两张图片经过全连接的网络，得到特征，然后再做相似度判断。identification model 虽然利用了更多的标注信息，但是训练的目标并不直接是行人重识别。

其他思路：借鉴DPM(Deformable Part Model)是一种思路，人在判断时会观察其中很多的细节，例如：两个人的胳膊像不像，衣服穿的是否一样，腿部是不是一样等，也就是通过将人的组成部件解析出来，从而达到精细识别的目的。

行人检测需要改变输入尺寸，一般行人的高度是宽度的3倍，可以从这点出发进行实验。

当前的行人重识别方法，如PCB， MGN，为了使网络学习到更精密有区分度的特征，一般将最后得到的feature map进行水平切分然后分别求loss，简单做水平切分可能会造成一块完整的feature，如背包，被丢失，失去信息。所以他们都做了相应的改进，如PCB加上了RPP，所以PCB之后加上了RPP把没有切分好的feature重新划分，需要进行两阶段的训练。而MGN的则是用了一个比较复杂包含多种切分的方法的网路，切两次，切三次，最后把这些feature融合起来。

切分feature map的本质是为了不让网络太关注于那些太显而易见的全局特征。因为行人重识别是一个zero shot learning的问题，就是说训练中的人和测试中的人是不同的人，而且主流的行人数据集都较小，容易过拟合， 导致这些在训练的时候表现很好的全局特征在测试的时候反而达不到最好的效果。

### 挑战和注意事项

目前存在着诸多如视角方向，人体移动，遮挡等主要挑战，提出了利用GAN将不同的投影学习出典范视角类型进行识别，提高了识别性能。

目前已发表的大部分研究还是在标准数据库上比性能，但其实这些研究离实际应用还很遥远。ReID中有很多其他问题很值得研究，比如夜晚用红外相机成像的问题或者人群混杂的情况下的严重遮挡问题等。在研究这些重要问题方面，英国伦敦玛丽女王学院的龚邵刚老师和国内中山大学的郑伟诗老师组的很多工作都很值得参考。另外可以进行跨数据集的迁移学习模型尝试等。

问题5：某个数据集上训练得到的行人重识别模型在其他数据集上表现很差，您觉得背后的原因是什么呢？

回答：跨数据库的方法现在做的是不太好的。目前原因还是模型过拟合于某些数据库了，而且数据库也比较小。不过最近已经有不少工作在关注这个问题了，比如中科院自动化所组织的ECCV2018上的一个有关ReID的Tutorial，最后就有总结这方面的最新进展。其中有的做得好的跨库模型甚至可以做出比在原数据库上直接训练更好的结果。

### 3.3 实际落地算法

澎思科技的行人重识别很厉害，刷新了三大数据集的世界纪录

云从科技的行人跨镜追踪技术通过多支路网络结构的设计，利用三粒度图片分割方法整合图片的全局信息以及局部显著特征信息，结合独SoftmaxLoss与TripletLoss联合训练结构与方法，实现对行人衣着、姿态、配饰等语意信息的提取，克服缺少人脸、图片模糊、光线变化、姿态多变等行人识别中的客观困难，该技术可以广泛应用于视频监控、智能安保、智能商业等领域。

2018年各大会议都有行人重识别ReID的论文，特别是CVPR 2018，2018年ReID特别的火。

## 行人检测/人体检测

### 4.1 数据集

现在用的最多的数据集是\*\*CityPersons、Caltech-USA和KITTI\*\*，二者都是比较大且具有挑战性的。Caltech-USA有大量的方法使用因而比较起来比较方便，而KITTI的数据集更加丰富一些但是用的。

1）Caltech dataset，其评价方法FPPW>FPPI

2）CityPersons是目前的最佳选择。行人检测任务存在一个领域迁移能力差的问题，例如在 Caltech dataset 上训练的模型在其上的性能很好，但是其在 KITTI dataset上的性能却比较差。之所以出现这样的问题是因为现有的数据集的多样性不够，CVPR 2017上将会公布一个新的行人检测数据集：CityPersons。

CityPersons数据集是脱胎于语义分割任务的Cityscapes数据集，对这个数据集中的所有行人提供 bounding box 级别的对齐性好的标签。CITYPERSON数据集是CITYSCAPES数据集的子集，CITYPERSON的标注文件只标注了其中HUMAN的类别。

由于CityPersons数据集中的数据是在3个不同国家中的18个不同城市以及3个季节中采集的，其中单独行人的数量明显高于Caltech 和 KITTI 两个数据集。实验结果也表明，CityPersons 数据集上训练的模型在 Caltech 和 KITTI 数据集上的测试漏检率更低。也就是说，CityPersons数据集的多样性更强，因而提高了模型的泛化能力。

3）KITTI数据集

4）INRIA数据集：该数据库是“HOG+SVM”的作者Dalal创建的，该数据库是目前使用最多的静态行人检测数据库，提供原始图片及相应的标注文件。

训练集有正样本614张（包含2416个行人），负样本1218张；测试集有正样本288张（包含1126个行人），负样本453张。图片中人体大部分为站立姿势且高度大于100个象素，部分标注可能不正确。图片主要来源于GRAZ-01、个人照片及google，因此图片的清晰度较高。在XP操作系统下部分训练或者测试图片无法看清楚，但可用OpenCV正常读取和显示。

5）TUD行人数据库

6）CrowdHuman数据集，比较新的数据集，是face++于2018年发布的

### 4.2 性能指标

1）FPPW (False Positive per Window)

基本含义：给定一定数目N的负样本图像，分类器将负样本判定为“正”的次数FP，其比率FP/N即为FPPW。意义与ROC中的假阳率相同。FPPW中，一张图就是一个样本。

评价方法一般是FPPW>FPPI。FPPI 相比于FPPW来说，更接近于分类器的实际应用情况。

2）FPPI (False Positive per Image)

基本含义：给定一定数目N的样本集，内含N张图像，每张图像内包含或不包含检测目标。FPPI = （False Positive 次数）/N。(即平均每张图中能正确检索到的数目)。

每张图像均需要标定：包含目标的个数；目标的准确位置L。而后在每张图像上运行分类器，检测目标并得到位置p。然后，检查每张图像内的检测结果是否“击中”标定的目标：

a. 若图像内无目标，而分类器给出了n个“目标”检测结果，那么False Positive 次数 +n；

b. 若图像内有目标，则判断p是否击中L，判断标准参看上述文章（主要看p与L的重叠率）。若判断未击中，则False Positive 次数 +1。

3）miss rate，漏检率

miss rate = sum(fn) / (sum(tp) + sum(fn)) = (1-recall rate)，即miss rate = 测试集正例判别为负例的数目 / 全部groud truth的数量。

同时，recall rate + miss rate = 1

4）行人检测评估表的理解

行人检测评估表即：miss rate VS fppi。

评估表的miss rate 是log-average miss rate。若每幅图像miss rate为x。则log-average miss rate为：2的[log(x1)+log(x2)+log(x3)+...log(xn)]/n 次幂

实例：测试的n幅Full Image 图像下出现了多少个误捡的窗口，假设是100幅中有1个误捡窗口，即fppi为10-2，那么此时的100幅图像中每一幅图像都有自己的miss rate，然后对这些miss rate 求log平均，从而得到log-average miss rate。

### 4.3 传统算法

2016年以前，主流方法是多特征融合的方法加上级联分类器，即手工特征 + SVM + adaboost。

常用的特征有Harry-like、Haar特征、Hog特征、改进的HOG、LBP特征、Edgelet特征、CSS特征、COV特征、积分通道特征以及CENTRIST特征。

一般的离线训练过程：

1）输入为标准图像库中的图像，对于正样本，直接提取特征；对于负样本，多尺度遍历整幅图像，进行特征提取。提取Harr，LBP，HOG三种特征并保存为积分图形式。

2）对于cascade的每一级进行如下操作：用Adaboost进行特征选择，选出最好的特征。在前面的几级用Harr和LBP这两种计算较快的特征，后几级用HOG。

3）进行级数调整，直到达到要求的检测率和错误率。

一般的检测过程(influence过程)：

1）输入图像，多尺度遍历整幅图像，进行特征提取：提取Harr，LBP，HOG三种特征并保存为积分图形式。

2）将所有子窗口输入训练得到的模型中，经cascade分类器后，根据检测结果即可标定行人区域，通过一些后处理（例如，合并临近的窗口），就可以对行人进行定位，达到检测的目的。

**特别介绍一下HOG+SVM：**

方向梯度直方图（Histogram of Oriented Gradient, HOG）特征是一种在计算机视觉和图像处理中用来进行物体检测的特征描述子。

出于速度上的实用性考虑，至今虽然有很多行人检测算法，但基本都是以HOG+SVM的思路为主（如果要进一步提升精度的话，还会加入adaboost，这样就需要相当的优化技巧了）。

HOG+SVM作为经典算法直接就集成到opencv里面去了，可以直接调用实现行人检测。例子：https://github.com/icsfy/Pedestrian\_Detection

**对于传统机器学习的改进：**

对于传统机器学习的改进，做的最多的就是增加或者多样化输入图像的特征。通过更多的和更高维度的特征，分类的任务似乎是变简单了，结果也有了改进。很多种类的特征已经被发现：边缘信息，颜色信息，纹理信息，局部形状信息，协方差特征，还有其他等等。越来越多的特征已被证明可以系统性的改善性能。

还有hand-crafted feature，part and occlusion建模。

**深度学习兴起前的基于NN的传统算法：**

在2015年，最好的行人检测方法是通过Adaboost学习的决策树，包括SquaresChnFtrs，InformedHaar，SpatialPooling，LDCF和Regionlets，这些方法都是ICF结构的变体。

之前的基于cnn的方法使用的是人工设计的特征，最早的方法ConvNet的输入是YUV图像，DBN-Isol和DBN-Mut的输入是HOG，MultiSDP使用HOG+CSS作为输入，JointDeep和SDN使用YUV+梯度作为输入。

### 4.4 NN算法

用cnn的有太多了，但没有一个特别主流的cnn算法，而目前主流的行人检测算法大多基于两阶段的检测器框架，即rcnn相关方法，如 fast faster等等。

svm加hog这类方法更多用于做分类classification ，而不是检测detection。即拿到一个对象后对其进行判断，而不是在一副图像里找一个对象。如果遍历来做计算量就上去了。

较早时候，基于CNN的行人检测方法中比较好的是SDN（在这之前的方法没有直接使用LeNet的）。

后来，将CNN结合DPM，提出了新的形变池化层。

https://github.com/samylee/Multi\_Scale\_PersonDet这一个是用mtcnn的，实时但效果也没说。

RepLoss是face++的，效果很好。

### 4.5 难点

1）一个是低分辨率情况的背景和小尺寸的人很难区分。小尺度的弊端：a）大多数小尺寸实例出现带有边框模糊，外观模糊。将它们与杂乱的背景和其他重叠实例区分开很难。b）大尺寸的行人实例通常表现出与小尺寸实例显着不同的视觉特征。

解决：对与小目标检测，用结合FPN+在线hard samples数据增强，再调整合适大小的anchors，就很不错了。

2）另一个是遮挡问题。对于行人检测而言，实际情况中几乎都是遮挡的，所以根本性的难题就是遮挡问题。

也有两种：a）多个人在一起的时候基于CNN的单个行人的检测定位误差相对较大，即行人个体之间的相互遮挡，这往往会引入大量的干扰信息，进而导致虚检；b）其他物体对行人的遮挡，这往往会带来目标信息的缺失，进而导致漏检。

对于遮挡情况下的行人检测，主要有两种解决思路：一是对目标候选框分成不同part逐一处理、区别对待再加以特征融合；二是从loss的角度使得目标候选框对相互遮挡的情况更具判别性，如CVPR2018的RepLoss，ECCV2018的OR-CNN。

尽管近年来有大量工作聚焦于行人检测中的遮挡问题，但现有算法距离该问题的解决依然还有很长的路要走，从论文中展现的结果来看，遮挡情况下的检测结果要远远劣于非遮挡情况下的结果。一个很关键的原因来自于人群密集场景下行人自遮挡的情况普遍存在，在训练数据有限的情况下，基于有监督训练得到的检测器难以学习到无法穷尽的各种遮挡情况，因此如何利用现有的无监督或者半监督方式来改善这个问题，应该是非常值得探索的一个方向，期待你的精彩发现。

对于“不同物体之间的距离非常接近”的问题：

a.旷世的repulsion loss:让proposal尽量远离和他overlap第二大的GT，同时也让不同GT上的proposals之间相互远离。我们目前测试这个方法确实很有效

b.softnms来适当的解决nms阈值原因造成的问题。

c.改进的nms。

d.有人提到分part来做，其实这个比较适合人体的，并不通用。可以看看我对这个paper的一点理解。

e.做crop数据增强，这样能对遮挡识别有利（参考SSD），标注时要尽量将重叠遮挡区域都标注出来。

f.cascade-rcnn的做法，调整IOU，获得高质量的样本，有效提升检测精度。

g.增加所谓的上下文信息。

### 4.6 对行人检测问题的分析

对行人检测问题的分析，来自《How Far are We from Solving Pedestrian Detection》中，caltech行人测试集上的failure case进行分析，failure case主要有两类，一类是false positive，一类是false negative。

1）false positive（不是人，却预测成了人）来源的原因分析如下图，主要有9个来源，9个来源又被归为三类原因：定位不准、背景、标注错误。定位不准的问题文章认为来自于训练集标注的框不准，于是文章重新标注了一遍数据集，最后实验下来发现提高训练集标注框的精度的确有助于提高检测器的效果。而背景问题也是行人检测里边的老大难问题，比如路边的指示牌、车后面的方向灯、树叶等等都可能被识别为行人，现在也没啥好的解决方法。

2）false negative来源的原因分析，主要有6个来源。其中尺寸小和遮挡是众所周知的难题。而side view和cyclists，文章认为是由于训练集里这两类情况表达不足，也就是训练集里边这类情况比较少，解决这种问题的最好方法是在训练集中增加对应图片。对于尺寸小的问题，文章又进一步分析，尺寸小的行人包含三种问题：模糊不清、像素点少、对比度差，经过实验，文章发现像素点少才是尺寸小的行人容易出现false negative的原因。

### 4.7 2018年行人检测回顾

在过去的一年中，行人检测领域的工作大致可被归为以下三类：

卷积神经网络为通用物体检测任务带来了全面的性能提升。而行人检测技术也使用了Faster R-CNN 这样的通用检测框架，因而性能也得到了很大的提升。在 CVPR 2015中，当时在竞赛中的最好方法使用的还是传统的ACF 检测器，其漏检率是18.5%；而在 ECCV 2016中，在使用了 CNN 后，行人检测漏检率降低到了9.6%。

第一类是将传统的检测方法Boosting trees 和 CNN 结合起来。张姗姗等人在CVPR 2016的工作是使用 ICF 提取proposal，然后使用 CNN 进行重新打分来提高检测的性能；在 ECCV 2016上，中山大学林倞教授课题组使用RPN 提取 proposal，同时提取卷积特征，然后使用 Boosting trees进行二次分类，性能得到了很大的提升。

第二类是解决多尺度问题，例如在视频数据中人的尺度变化问题。颜水成教授课题组提供了一种解决方法：训练两个网络，一个网络关注大尺度的人，另一个网络关注小尺度的人，在检测时将两个网络进行加权融合得到最终的结果，这样能使性能得到很大的提升；UCSD 在 ECCV 2016上有一个类似的工作，提出在高层提取大尺度人的特征，在低层提取小尺度人的特征，这样能保留尽量多的信息量，使得对小尺度的行人也有较好的检测效果。

第三类是使用语义分割信息来辅助行人检测。首先对整个图像进行语义分割，然后将分割的结果作为先验信息输入到检测网络中（包括传统的 ICF 网络，以及现在常用的CNN），这样可以通过对整体环境的感知来提高检测的效果。

### 4.8 瞎逼逼(个人)

人脸检测的算法也可以用于行人检测，它们都是目标检测，只是针对的应用场合不同（所以可能会有先验条件上的不同，这就导致了算法上的不同）

只要detection够强，这算法就效果很好，但detect并不能每一帧都检测到，所以后来又有了deep sort，多目标一般都是tracking by detection。

看知网时，发现还有很多论文是基于红外的行人检测

用上GPU，稍微快一点的都是实时行人检测

旷视科技俞刚的历经：基于深度学习，Detection 组开始做的第一个任务就是行人检测，最开始使用 fast rcnn，后来慢慢切换到 faster rcnn。同时 Detection 组开始基于行人检测做人体骨架算法。另外，为了更精细的定位人体框，Detection 组又做了一些分割算法。

### 4.9 实际代码测试（更详细可见“行人检测算法实测”）

见“行人检测算法性能实测”的excel。

行人检测有三个可以尝试的(有代码和model的)：ALFNet、HOG和faster-rcnn的变种：

1）ALFNet是2018年ECCV的，优先尝试，性能好，速度差不多是ssd/mtcnn的速度（https://github.com/liuwei16/ALFNet）

2）HOG基本上是2015年前最好的了，就算是现在精度也不差，而且直接用opencv就行。（https://github.com/icsfy/Pedestrian\_Detection）

3）faster-rcnn的变种有很多，基本上是刷榜跑分的论文，可以去试一下 （https://github.com/CharlesShang/TFFRCNN）和（https://github.com/longcw/faster\_rcnn\_pytorch）

# 参考资料

行人检测：https://blog.csdn.net/xiaofei0801/article/details/72800977