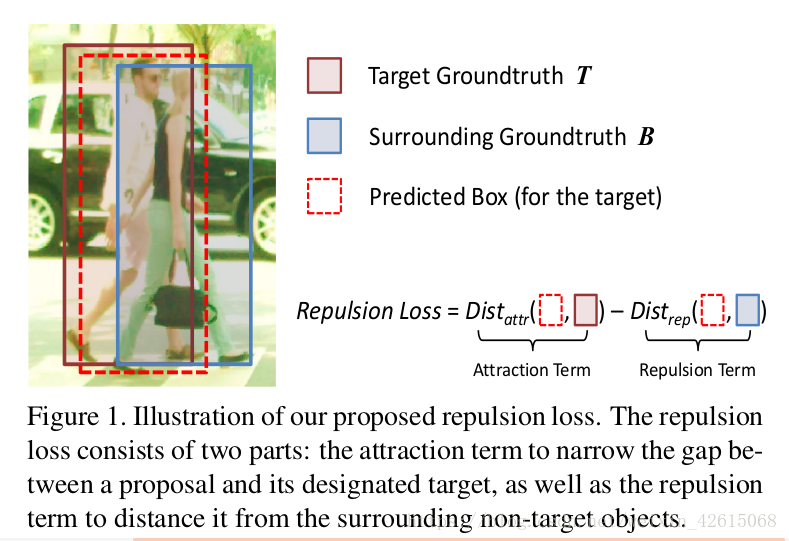
**解决密集遮挡问题**

**1. Repulsion Loss: 解决行人检测中密集遮挡问题**

**motivation：**

目标检测的目的是使产生的proposal能更加接近真是目标框GT，而行人检测中经常会遇到行人密集拥挤的情况，文中称作crowd occlusion，使得训练输出的预测框会由于周围其他目标的影响而干扰最终的输出结果（从下图可以看出）。那么可不可以通过一种方式，阻止proposal由于crowd occlusion而偏移到周围目标框的情况 ？论文中写道“the attraction by target, and the repulsion by other surrounding objects”，即使得预测框更加接近所负责的真实目标框，而远离周围的目标。



**innovation point：**   
通过设置损失函数的方式，即*Repulsion Loss*,使预测框和所负责的真实目标框的距离缩小，而使得其与周围非负责目标框（包含真实目标框和预测框）的距离加大 。如下式，如果与周围目标的距离越大，损失值会越小。



**detailed information in the paper：**

1. **crowd occlusion 的影响**

occlusion的定义https://img-blog.csdn.net/20180904200517983?watermark/2/text/aHR0cHM6Ly9ibG9nLmNzZG4ubmV0L3dlaXhpbl80MjYxNTA2OA==/font/5a6L5L2T/fontsize/400/fill/I0JBQkFCMA==/dissolve/70BBox-visible在[CityPersons](https://arxiv.org/abs/1702.05693" \t "_blank)数据集中会单独标记出来，即没有被遮挡的部分。

*occlusion case*：真实目标的*occ>=0.1*   
*crowd occlusion case*: 真实目标的*occ>=0.1*，并且与其他任何真实目标的*IoU>=0.1*

从上图可以看出，预测框偏移相邻目标而造成更多的False Positives, 而且这些预测框往往会有更高的confidence score（这使得会更加容易被归类为False Positives）.

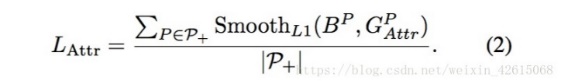
1. **Repulsion Loss**



损失函数共包含三部分，第一部分为**预测框与真实目标框所产生的损失值**（attraction term）;第二部分为**预测框与相邻真实目标框所产生的损失值**（repulsion term(RepGT)）;第三部分为**预测框与相邻不是预测同一真实目标的预测框所产生的损失值**（repulsion Box(RepBox))。通过两个相关系数alpha和beta来平衡两部分repulsion损失值。论文中试验结果显示两个相关系数都设置为0.5取得了较好的效果。

**Attraction term**

目的使预测框与真实目标框更加接近，沿用 Smooth\_L1 构造吸引项。给定一个 proposal P ∈ P\_+（所有的正样本），为每个proposal匹配一个与之有最大IoU值的真实目标框：G^P\_Attr = arg max\_G∈G IoU(G,P)。B^P 是回归 proposal P 的预测框。Attraction Term的定义为：

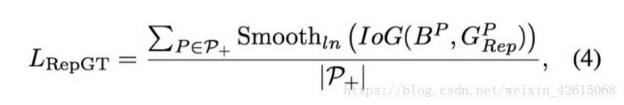


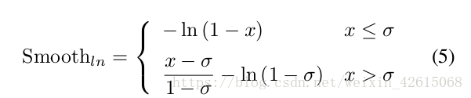
**Repulsion Term(RepGT)**

目的使预测框远离与之相邻的真实目标框。给定一个 proposal P ∈ P\_+，它的排斥 ground truth 物体被定义为除了其指定目标之外带有最大 IoU 区域的 ground truth 物体https://img-blog.csdn.net/20180904212431972?watermark/2/text/aHR0cHM6Ly9ibG9nLmNzZG4ubmV0L3dlaXhpbl80MjYxNTA2OA==/font/5a6L5L2T/fontsize/400/fill/I0JBQkFCMA==/dissolve/70

RepGT 损失被计算以惩罚 B^P 和 G^P\_Rep 之间的重叠https://img-blog.csdn.net/20180904212617815?watermark/2/text/aHR0cHM6Ly9ibG9nLmNzZG4ubmV0L3dlaXhpbl80MjYxNTA2OA==/font/5a6L5L2T/fontsize/400/fill/I0JBQkFCMA==/dissolve/70

IoG(B, G) ∈ [0, 1] ，从而 RepGT 损失可写为：

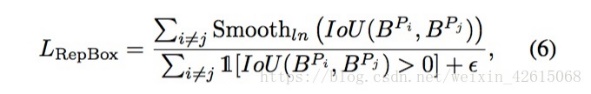




**Repulsion Term(RepBox)**

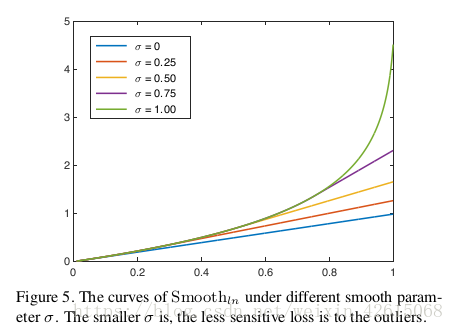
目的使预测框远离相邻不是预测同一真实目标的预测框。首先根据真实目标框GT将P\_+分为不同的子集，这里写图片描述，然后使得来自与不同子集的proposal的overlap尽可能的小.

RepBox 损失可写为：



分母中的示性函数表示：必须是有交集的预测框才计入损失值，如果两个proposal完全不相邻，则不计入。 **RepBox 损失可以降低 NMS 之后不同回归目标的边界框合并为一的概率，使得检测器在密集场景中更鲁棒。**

**注：关于Smooth\_ln**



通过σ，来调整repulsion loss对outliers（这些有较大overlaps的boxes pair）的敏感性。RepBox相对于RepGT有更多的outliers，所以应该弱化其对σ的敏感性，论文实验中RepGT和RepBox分别在σ=1和σ=0取得更好的效果。

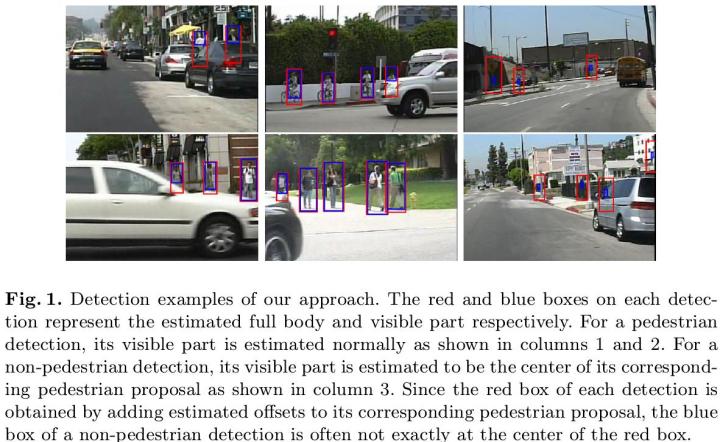
**2. Bi-box行人检测（Bi-box Regression for Pedestrian Detection and Occlusion Estimation）**

对于遮挡情况下的行人检测，总结已有的相关工作，主要有两种思路：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[1]，ECCV2018的OR-CNN[2]。

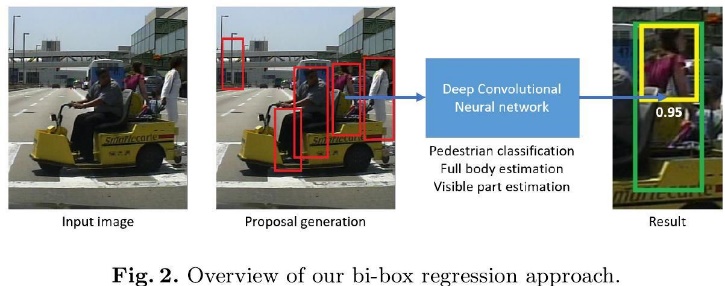
目前主流的行人检测算法大多基于两阶段的检测器框架，本文也不例外，但另辟蹊径，**通过让网络同时输出目标候选框的可见部分**，指导网络在学习的过程中对遮挡情况下的行人目标更具有判别力，实验表明该思路对严重遮挡情况下的行人检测十分有效。

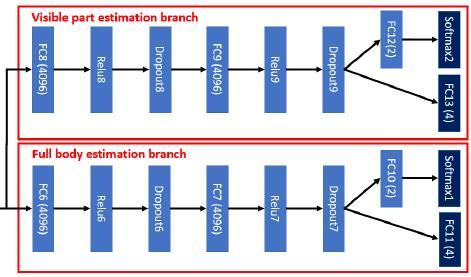
**本文方法**

本文方法遵循Fast R-CNN检测框架。为了处理遮挡问题，本文提出利用一个网络同时进行行人检测和遮挡估计，具体而言就是让网络并行两个分支，分别输出两个bounding box，一个是**完整的行人框**，另一个是**行人的可见部分**，如下图1所示，前者为红色框后者为蓝色框，对于没有行人目标的候选框而言，则可见部分分支的输出要尽量逼近候选框的中心。



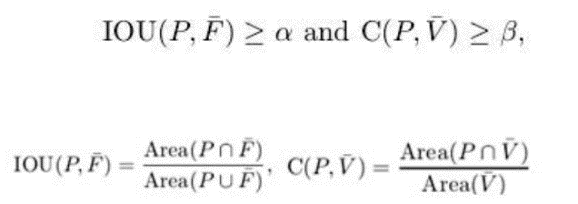
文中图2给出了本文算法的具体流程，基于Fast R-CNN检测框架，首先产生目标候选框，将图像和目标候选框输入卷积神经网络，然后通过RoI pooling提取每个目标候选框的特征，**对这些特征进行分类和回归，分别得到两个box**，图2中的绿色框代表行人目标整体，黄色框代表可见部分。文中图3给出了具体的网络架构示意图，其主要包括两个分支：**可见部分估计分支和行人整体估计分支**。



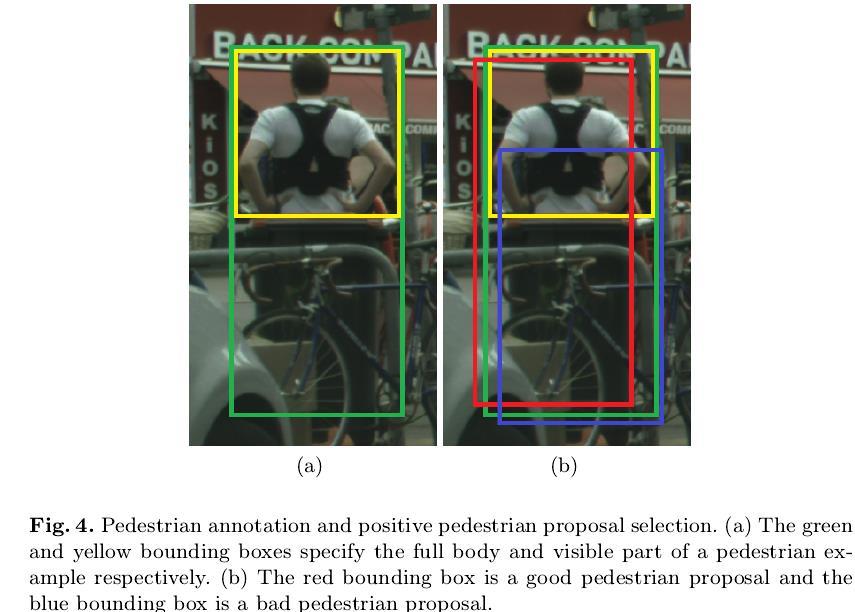


对于行人整体估计分支的具体处理流程和通用目标检测一致，不再赘述，这里重点论述下**可见部分估计分支**，也是本文的贡献所在。

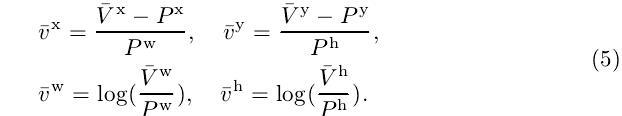
对于网络训练，一个**目标候选框P被当成正样本**需要满足如下两个条件：



也即目标候选框**P不仅要和整个标注框的重叠程度大于一定阈值**，还要和标注框内行人**可见部分的重叠程度大于一定阈值**，这种设定正样本的方式显然是合理的，能够更有效地指导网络学习进而获得更强的特征判别力。以图4为例，如果和已有方法一样，只用式（1）中的第一个条件时，图（b）中的蓝色框也会被当成正样本，然而蓝色框中包含的目标信息非常有限，显然是会干扰网络的正常学习的，而如果加上式（1）中的第二个条件，则只有红色框才会被当成正样本。



对于可见部分估计分支的正样本而言，其回归目标和整体估计分支一致，定义如下：



但对于负样本也应当有回归目标，因为有些负样本和真实目标没有重叠，而**有些负样本和真实目标有部分重叠**，为了更好地区分正负样本，作者将**负样本的回归目标定义为目标候选框的中心区域**，实验表明这一点非常重要。如果可见部分估计分支只为正样本分配回归目标的话，这个分支的训练将被没有遮挡的目标所主导（因为这部分目标的可见部分和整体部分基本是一致的），结果会导致预测的负样本的可见部分和整体部分是一致的，这样两个分支就不具有互补性了，也就达不到本文利用两个分支的目的。

**本文贡献：**

（1）本文重点关注行人检测中的遮挡问题，并基于Fast RCNN的检测框架提出了一套解决方案。

（2）本文实验给出两个结论：1、预测行人框的可见部分是可行的，并可以带来整体行人检测的性能提升；2、训练时，结合遮挡程度来定义高质量的正样本，可以为训练带来帮助，不失为一种非常经济有效的提升性能的trick。

**1) 文章出发点**

针对行人遮挡问题（occlusions），提出一种新颖的行人检测和遮挡估计的方法，通过回归两个bounding boxes分别定位**全身**和行人的**可见部分**。

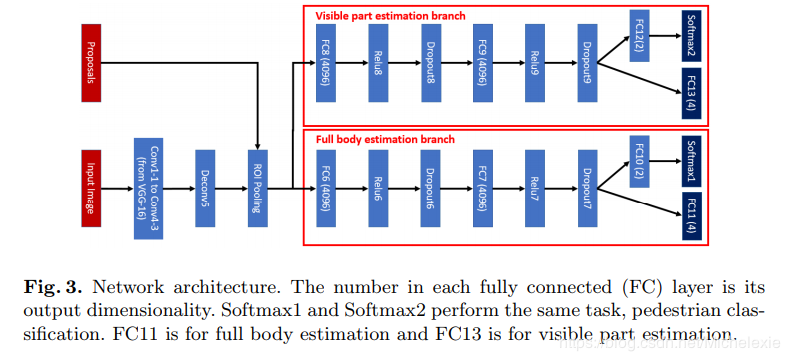
1. **创新点**

* 一个CNN包含两个分支：全身和可见部位估计
* 一种新的训练策略，**对两个分支的分数有效融合**，改进行人检测的表现
* 提出一种新的选择positive pedestrian proposal准则，有效提升遮挡行人检测

1. **方法介绍**

整个网络结构分成两个任务分支：行人全身及可见部位估计。

用于特征提取的基础网络为：**VGG16**



文章在proposed approach部分中提到，本文采用**Fast RCNN**框架（Fast RCNN产生proposals是通过selective search）

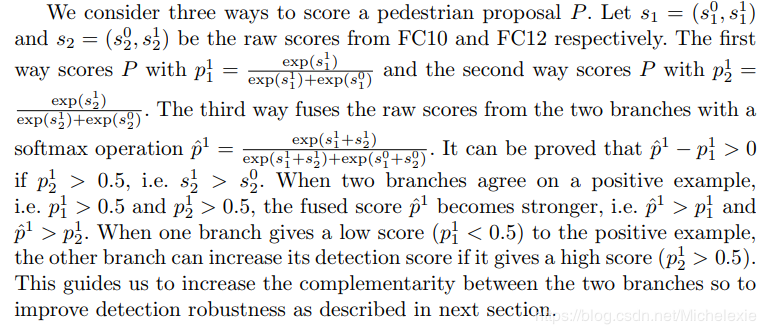
1. **网络输入部分**

特别注意输入部分：input image和 proposals。此处proposals是什么呢？

原文：A set of region proposals which possibly contain pedestrians are generated for an input image by a proposal generation approach (e.g. [38, 4]). 参考论文38和4分别是：Is faster rcnn doing well for pedestrian detection?和Illuminating Pedestrians via Simultaneous Detection & Segmentation . 这两篇论文产生region proposals的方法都是Faster RCNN中用到的--RPN！也就是说，本文的proposals是由RPN产生的，而网络框架是Fast RCNN, 其实Faster RCNN = RPN+Fast RCNN. 所以本文可以看做是使用Faster RCNN来做的。（个人理解）

对于两个分支，每个分支都由分类和bounding box 回归组成，和Fast RCNN检测部分一致。

1. **两个分支的分数融合**



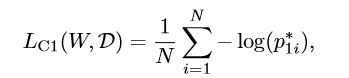
这种融合方式与Illuminating Pedestrians via Simultaneous Detection & Segmentation相同,即针对每个proposal，分别计算出属于全身和可见部分的softmax分数，再通过一个softmax操作融合，这种做法可以使两个分支互补，提高行人检测正确率。

1. **损失函数：**

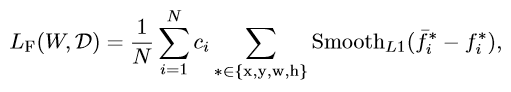
C:\Users\12544\AppData\Roaming\Tencent\Users\1254468141\QQ\WinTemp\RichOle\OUGJ3S7OW6{ZBFYMIF)C1)O.png

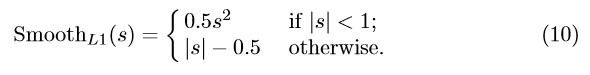
以上损失分别为full部分的分类和回归损失以及可见部分的分类和回归损失。

其中分类损失为多项选择损失：



回归损失为平滑L1损失：





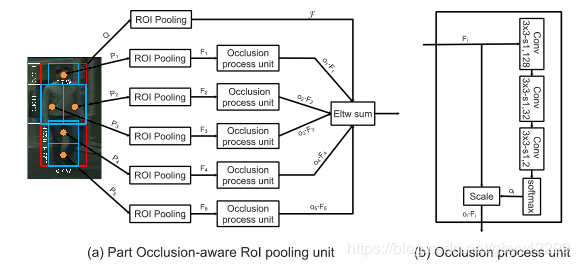
其中λF = λC2 = λV = 1.

LF和LV之间的差异是在LF中没有考虑负面的例子，因为对于这些例子中的等式，ci = 0。 （9），虽然在LV中考虑了正面和负面的例子。在训练期间，负面例子的可见部分区域被迫缩小到它们的中心。以这种方式，学习可见部分估计分支和全身估计分支以产生互补输出，其可以被融合以提高检测性能。如果可见部分估计分支被训练为仅对正行人建议的可见部分进行回归，则该分支的训练将由非遮挡或略微遮挡的行人示例支配。对于这些行人提议，他们的地面实况可见部分和全身区域在很大程度上重叠。结果，负面行人建议的估计可见部分区域通常接近其估计的全身区域，并且训练后两个分支之间的差异不会像负面例子的可见部分区域的情况那样大。被迫缩小到他们的中心。如我们的实验所示，当两个分支融合时，**迫使负例的可见部分区域缩小到它们的中心**实现了比不这样做时更大的性能增益。

**3. 遮挡下的行人检测(Occlusion-aware R-CNN: Detecting Pedestrians in a Crowd)**

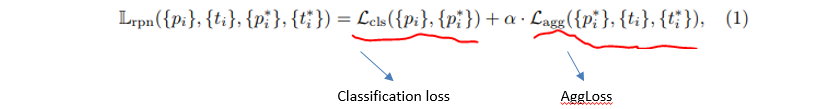
**1) 概述**

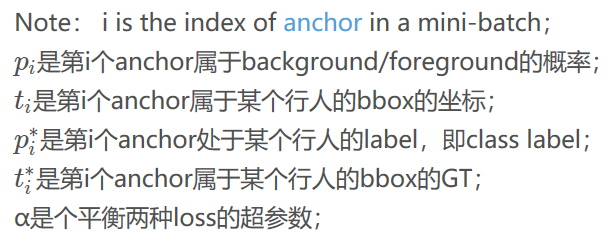
依然是解决在遮挡的情况下对人的检测的文章，作者分别从loss和two stage detector中核心的ROI Pooling操作这两个角度出发改善遮挡物体的检测问题。在之后我会根据这两个方面对文章进行一个简单的总结。先简单的放个在2.2中我将会提到的网络结构图：



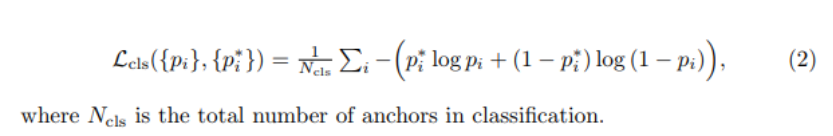
1. **Occlusion-aware R-CNN**

为了让RPN module能够更好地、更精确地提取到proposal，作者设计了一种AggLoss来限制proposals能够更接近GT并且同一个object的proposals之间能够尽可能的靠的紧致。所以整体的RPN的Loss函数如下：





其中：分类loss采用log loss：

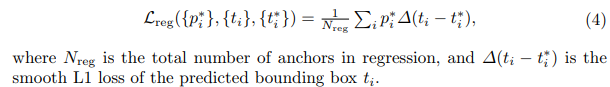


1. **Aggregation Loss（loss角度改善遮挡的问题）**

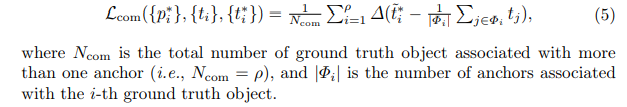
为了减少对于紧邻行人的错检，作者强制使proposals能够靠近真正的GT，并且能够proposals之间能够紧密分布。所以作者提出了AggLoss在RPN和Fast-RCNN层。这个loss能够让不同的proposals能够接近各自的GT，并且让属于同一GT的proposals之间的距离能够最小.

https://img-blog.csdnimg.cn/20181112091406494.png

(3）式中第一项是regression loss，目的是让每个proposal能够更接近其真实的GT，采用smooth L1损失，如下（4）式所示；第二项是compactness loss，目的是让属于同一GT的proposals能够分布紧致。



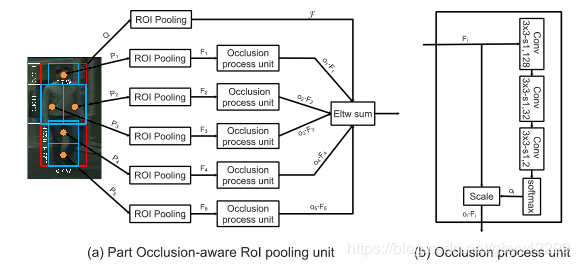
对于compactness term Lcom目的是让属于同一GT的proposals之间能够尽可能的分布紧致，使得可以减少对于临近的两个人之间的错检问题。如下：



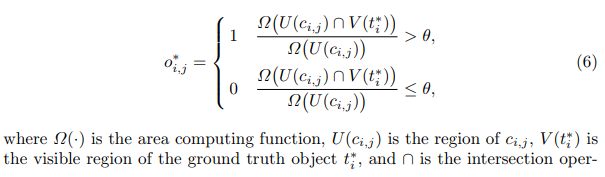
compactness term Lcom也即对多个匹配上真实目标的anchor的回归目标取平均值再与真实目标计算误差，误差度量依然采用的是smooth l1 loss。值得注意的是，Aggregation Loss既可以用在第一阶段也可以用在第二阶段。

1. **Part Occlusion-aware RoI Pooling Unit（ROI Pooling操作改善遮挡的问题）**

Faster R-CNN的第二阶段需要提取每个目标候选框的特征，如果对目标候选框分成不同的Part分别提取特征，就可以减少遮挡位置对全局特征的影响，这也是处理遮挡问题的一种常用方式。与已有方式不同的是，本文并非随机选取part，而是根据先验将行人目标分成5个part（如下图1所示），分别提取这5个part的特征，这种根据人体布局的先验能够获得互补的局部特征，再结合**整个目标的全局特征做加权求和**（在各部分基础上加权），得到融合特征再做进一步的分类和回归。显然，这里的加权方式就显得尤为重要，为此作者提出了一种遮挡处理单元（图1（b）所示），该单元的输出是一个sigmoid值，意味着该part没被遮挡的程度，因此**遮挡处理单元就是一个二分类问题**，可以和检测网络联合训练了。



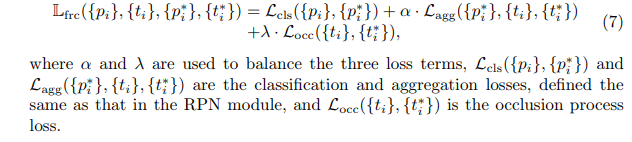
正如上图Fig.1中所示，作者将人分为5个区域，用RoI pooling层把feature map固定在HxW的尺寸。



https://img-blog.csdnimg.cn/20181112092736887.png

作者分别对着5个局部区域分别进行了预测一个0到1的遮挡的score，用来表示着5个区域是否被遮挡，可以把这个部分认为是一个简单的mask信号，最后作者使用这5个score和其对应区域的feature相乘后相加，得到最终的RoI feature。

综上，第二阶段的损失函数包括三个部分：分类loss、定位loss和遮挡处理单元loss，如下式所示：



**注意：处理遮挡的loss是和t即坐标有关的，所以可以改善有遮挡的回归。**