**Ten Years of Pedestrian Detection,**

**What Have We Learned?**

**摘要：**

越来越多的研究成果令我们很容易忽视这些树木所在是森林。我们通过讨论最近在Caltech 行人检测基准上40多个检测器的核心思想来分析近十年来那些重大的进展。我们发现主要存在三种类型的方法，都在最近达到了相似的检测质量。根据我们的分析，我们通过结合多个已发表的策略来研究最有前景的一些思想之间的互补性。这个决策森林检测器在富有挑战的Caltech-USA数据集上达到了目前最好的表现。

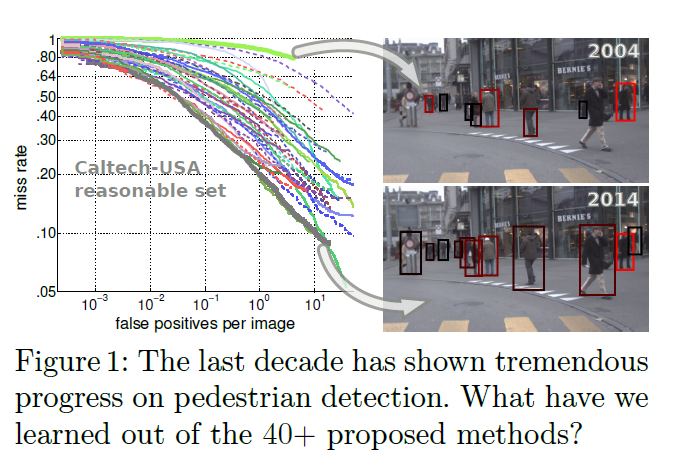
【关于这篇文章的思考】: 行人检测领域经过十几年的发展，已经出现了不少标志性的成果。虽然不断的发展是必然趋势，但一味的专注于某个方向的发展而忘记整个领域的发展情况是不可取的。正如文中所说的一样“Paper-by-paper results make it easy to miss the forest for trees”。因此本文旨在对过去十年中行人检测领域的发展做一个综述，通过分析多个检测器的核心思想来总结那些重大进展。

我认为综述性文章对于初入领域的人来说十分重要，它可以帮助我们了解整个领域的概括，发展方向以及发展进度等。通过阅读该篇论文虽然并不能了解某个特定的方法，但可以看出人们在解决行人检测问题时不断的提出问题，解决问题，从而形成了一个螺旋上升的发展路线。了解过去，才能更好的看清未来。在本文中也提出了该领域下一步科学的发展，值得人们深思。

**1．简介**

行人检测是物体检测的一个典型实例。由于其可直接应用于车辆安全，监控以及机器人学，在过去几年中吸引了不少关注。重要的是，基于已建立的标准和评价指标，这是一个已被很好定义的问题。同样的，他也作为探索目标检测不同思想的一个平台。作为目标检测主要的范例“Viola&Jones 变种”，HOG+SVM刚性模板，可变形部件检测器（DPM）和卷积神经网络都被研究过了。

本文的目的在于回顾过去十年中行人检测的发展（40+种方法），确定所探索的主要思想并且量化哪个思想对于最终的检测质量有着最深的影响。在接下来几节种，我们将回顾已有的数据集（第二节），提供一个对不同方法的讨论（第三节），以及对近年进展再造/量化的实验（第四节，使用20个新训练的检测模型来进行试验）。尽管我们并不是要介绍一种新奇的技术，但我们会通过把现有的方法放在一起比较来报告在富有挑战的Caltech-USA数据集上最好的检测结果。



【本文的工作】 ：

[1]:讨论行人检测领域主要所使用的数据集以及各自的特点。

本文中简要介绍了如INRIA,ETH,TUD-Brussels，Daimler，Caltech-USA和KITTI等数据集。

* INRIA：图片相对较少，标注质量高。
* ETH、TUD-Brussels：中等大小视频数据集。
* Daimler：缺少颜色通道。
* KITTI：测试集更多变。

【思考】: 数据集显然对于行人检测问题来说十分重要。到底需要怎样的数据集以及需要多少才能足够检测器使用一直是一个问题。训练集与测试集的比例，选取方式也十分重要。值得思考的是训练集中的行人数据是否足够于模型的学习，而模型在测试集上的泛化程度是否就确实代表了模型本身的性能。如果训练集多样性较小，而测试集的多样性较大，那么如此所测试的结果就不能很好的表示我们模型的能力。于是，训练集与测试集的设置就成了至关重要的问题，而对于此似乎并没有一个统一的说法。我们的目的是得到更具鲁棒性的模型，这样就需要我们从数据中学习到更具代表性的特征以应对各种姿势、各种角度、各种状态的行人。

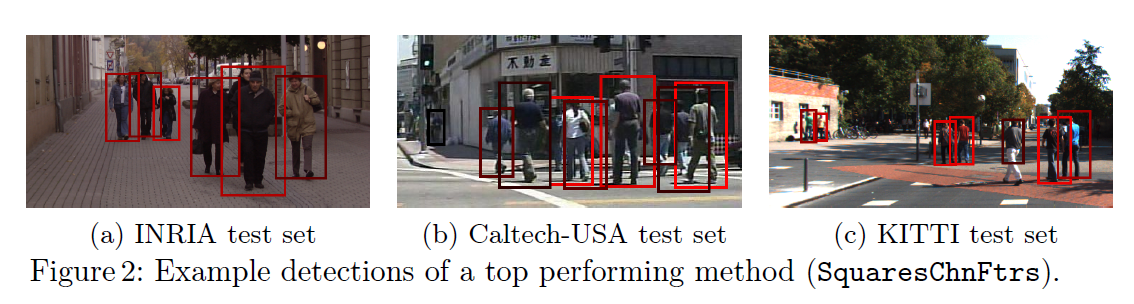
**2. 数据集**

几年来很多公共行人数据集都被收集起来；INRIA，ETH，TUD-Brussels，Daimler，Caltech-USA和KITTI是使用最普遍的数据集。他们有不同的特点，劣势和优势。

INRIA是这之中最古老并且相对图片最少的数据集。然而，在不同的环境下，行人的高质量标注也带来了不少好处，这也是为什么该数据集会被普遍选择作为训练集。ETH和TUD-Brussels是中等大小的视频数据集。Daimler由于缺少颜色通道而没有被使用。Daimler ste-reo,ETH 和KITTI提供了立体的信息。所有的数据集除了INRIA都是从视频中获取，因此可以使用光流作为附加的线索。

目前，Caltech-USA和KITTI是行人检测主要的基准。两者都是相对较大并富有挑战性的数据集。Caltech-USA较为杰出是因为很多方法在其上被评估。KITTI较为杰出是因为它的测试集更多变，但它也并没有被频繁使用。更多数据集的详细讨论可以查阅[8,7]。INRIA，ETH，TUD-Brussels,Daimler和Caltech-USA都可以使用一个统一的评估工具箱；KITTI在一个未被公开的测试集上使用不同的评估工具箱。这两个工具箱都维护了一个在线排名，在其上发布的方法可以进行并列的比较。

本文主要使用Caltech-USA来比较各个方法，其次会使用INRIA和KITTI。如图2。INRIA上的结果使用log-average miss-rate（越低越好）来评估，而KITTI使用precision-recall curve（越高越好）下面积来评估。



**基准的价值** 个别的论文通常只展示了最好的方法有限的一面。有一个从各种方法收集了检测结果的官方基准可以减轻作者把曲线加入上下文中的负担并使得评论者可以更简单的看到最好方法的结果。对结果的收集使回顾性的分析变的可行，正如下一节所提到的。

【本文的工作】:

[2]:对不同检测方法的讨论。

本文以对提高行人检测的方法为切入点对不同的检测方法做了一个分析比较。

【思考】:

1. 训练数据的不同会导致检测器性能的不同，一般认为更具多样性的数据会产生更鲁棒的检测器。如文中提到的表现较好的方法使用了Caltech的拓展版本。

2. DPM变体，深度网络和决策树虽然都在行人检测领域达到了最好的结果，但文中提到单提升的决策树结构似乎更适合行人检测。但到底该结构的优势在什么地方依然有待考察。

3. 本中虽然说什么结构的分类器是最好的问题是未知的，但可以发现在不同评估标准下一些的方法的表现不尽人意。评估标准的发展同样体现了人民发现问题的过程。一个好的评估标准应该是合理的，而无法使某种结构的模型可以“钻空子”，针对该标准做出相应的优化。

4. 附加数据的使用从理论上来讲应该具有提升检测性能的功效，即增加了模型学习的可用信息，同理于上下文信息，但事实证明一些模型不使用附加数据也可以达到同样的效果。因为模型的学习过程以及学习结果对于我们来说是透明的，所以理解模型到底学习到了什么内容也许对提升其性能有帮助。

5. 可变型部件和多尺度模型都是针对于特定问题的解决方案，但整体来讲对检测器的提升不大。

6. 使用深度结构确实会产生不错的结果，但由于深度网络的黑箱性质，我们很难理解该模型，从而也阻碍了我们对其进行改进。深度网络的可解释性一直是一些学者致力于解决的问题，相信在该方面如果有了突破性进展，深度学习的浪潮又会更上一个台阶。

7. 大量的特征类型已经被研究过：边缘信息，颜色信息，纹理信息，局部形状信息，协方差特征等。怎样的特征称为更好的特征以及为什么该特征好成为了一直以来的研究热点。理论上讲，确实更好的特征会产生更好的检测器，但其内在原因还有待研究。如果明白了怎样的特征更好那么便可以指导我们进行更进一步的设计更好的特征，从而获得更好的结果。

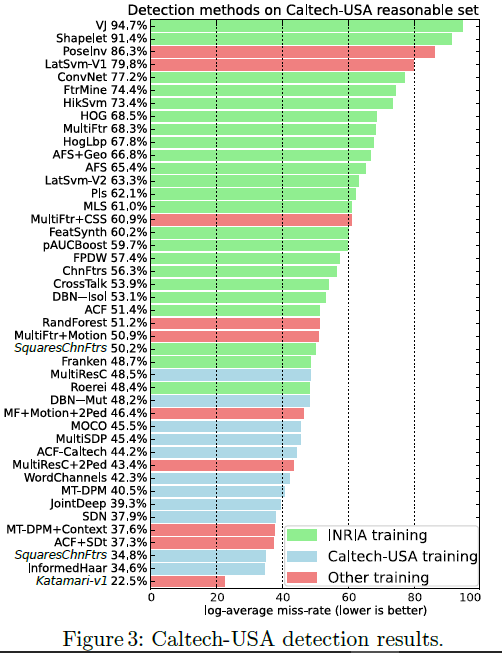
**3. 提高行人检测的主要方法**

图3和表1一起提供了一个对于40+种在Caltech行人检测标准上发布过结果的方法定性和定量的综述。被标记为斜体的方法是我们新训练的模型。我们参考了所有使用Caltech标准速记的方法。我们找出每个方法最独特的方面并据此将他们分类，而不是特别的讨论某一个方法。我们将会在下一小节讨论这些方面。

**简明的年表** 2003年，Viola and Jones在行人检测任务上应用了VJ检测器。2005年，Dalal and Triggs 介绍了具有里程碑意义的HOG检测器，后在2008年被现在Felzenswalb经典的可变形部件模型DPM所使用。2009年Caltech 行人检测基准被提出，并比较了七个行人检测器。在这时，每个窗口评估的缺陷被提出，评估标准也从FPPW变换到了FPPI。在这个评估标准之下，一些早期的检测器表现的并不好。这里考虑的方法大约三分之一都是在2013年发布的，反映出人们对该问题重新燃起了兴趣。相似的，在KITTI上一半的行人检测结果都是在2014年被提交的。

**3.1训练数据**

图 3展示了检测器的不同表现，不出意料的，都是由于对训练数据选择的不同。在Caltech-USA上训练的方法比从INRIA概括的方法表现要好。表1给出了所使用的训练数据的详情。其他训练表现较好的方法使用了Caltech-USA的扩展版本。例如MultiResC+2Ped使用了Caltech-USA加上一个在INRIA上拓展的标注集，MT-DPM+Context使用了一个额外的汽车训练集，ACF+SDt使用了原始Caltech-USA视频上附加的帧。



**3.2 Solution families**

总体来说，在40+种方法中我们可以识别出三个族：1）DPM变体。2）深度网络。3）决策树。在表1中我们使用DPM，DN和DF来分别代表这三个族。

基于原始数据单提升的决策树似乎特别适合行人检测，在“INRIA上训练, Caltech上测试”和“Caltech上训练，Caltech上测试”两方面任务都达到了最好的表现。不知道是什么给了他们优势。探索的深度网络同样在检测质量上展现出有趣的属性和快速的发展。

**结论** 总的来说，目前DPM变体，深度网络和决策树在行人检测上都达到了最好的表现。

**3.3 更好的分类器**

由于最初HOG+SVM的提议，线性和非线性的内核都被考虑了。HikSvm考虑了非线性内核的快速接近。这个方法在使用有缺陷的FPPW评估标准时会获得提高，但在适当的评估方法下（FPPI）却表现的不好。在MultiFtrs的工作中，有人认为提供足够的特征，Adaboost和线性SVM在行人检测上表现大致相同。

最近，越来越多的检测器组件和“决策组件”联合起来优化。这也导致特征和分类器之间的区别不再有那么明显。

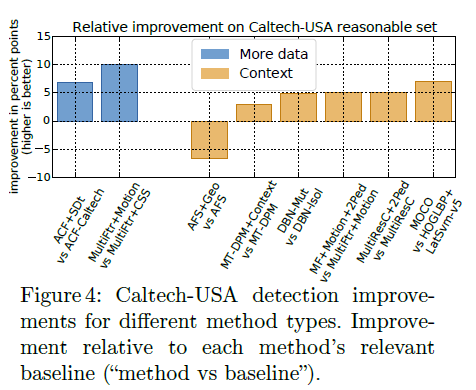
**结论** 没有确凿的经验证明是否非线性内核比线性内核提供更有意义的增益。相似的，是否有某个类型的分类器在行人检测上比其他的更好也是不确定的。

**3.4 附加数据**

行人检测的核心问题主要在单个单目彩色图像帧。一些方法探索在训练和测试时间上的附加信息来改善检测器。他们考虑立体图像，光流，追踪或从其他传感器获取的数据。

对于单目标方法，目前还不清楚追踪可以改善多少单帧检测。如图4利用光流提供了在基线上不小的改进。奇怪的是，目前最好的结果（ACF-SDt）是从粗糙的光流而不是高质量的光流中受益。在4.2节中，我们探讨与流互补的一些其他要素。在Daimler数据集上利用流和立体获得了很大的成功，但是类似的结果在新的数据集例如KITTI上还未见到。

**结论** 使用附加数据提供了有意义的提高，虽然在现代的数据集上立体和流的线索还没有被充分利用。目前来说，几乎基于单目图像帧的方法已经可以和使用了附加信息的方法表现的差不多。



**3.5 利用上下文**

滑窗检测器使用该窗口内的内容来评估可能的检测窗口。借助于检测窗口的上下文信息，即窗口周围的图像内容，可以提高检测性能。利用上下文信息的策略包括：地平面限制，自动上下文的变体，其他种类的检测器和人对人模式。

图4展示了合并了上下文信息的方法在性能上的提高。总体来说，可以看到3~7MR百分点的提高。有趣的是，+2Ped获得了一个比现有方法稳定的2~5个百分点的提高，即使是表现最好的方法。

**结论** 上下文为行人检测提供了一个稳定的提高，虽然提高的比例比附加的测试数据和深度结构的提高要低。检测质量的大部分来自其他来源。

**3.6 可变形部件**

DPM检测器最初是为行人检测而设计的。这个思想已经十分流行并且已经有数十种变体被探索过了。

对于行人检测来说结果具有竞争力，但不显著。在对深度结构部件和内部变形进行建模时获得了一些有趣的结果。

DPM和他的变体系统的通过使用单组件和无部件的方法而较优，造成了对是否需要部件产生了疑惑。最近的研究已经探究过完全不使用组件而捕获变形的方法。

**结论** 对于行人检测，除了解决遮挡，目前还没有一个明确的证据表明组件和部件的必要性。

**3.7 多尺度模型**

对于检测，一般来说高分辨率和低分辨率的候选窗口在提取特征之前都会被重采样成一个共同的大小。最近的研究表明为不同的分辨率训练不同的模型可以系统的提高检测性能1~2个百分点，因为检测器可以获得每个窗口大小的全部信息。这个技术不会影响在检测时的计算开销，但是训练时间会增加。

**结论** 多尺度模型提供了一个对现有检测器的简单及一般化拓展。尽管有持续的改进，但他们对最终检测质量的贡献相当少。

**3.8 深度结构**

大量的训练数据和计算性能的提高促进了最近深度结构在多种计算机视觉任务上的成功。这个结果也启发了深度结构在行人检测上的应用。

ConvNet使用了监督和非监督训练的混合来创建在INRIA上的卷积神经网络。这个方法在INRIA，ETH和TUD-Brussels上获得相同的结果，却无法将其推广到Caltech设置。这个方法学习直接从原始像素值中提取特征。

另一类研究专注于对部件和遮挡共同建模。这种整合的性能提升在1.5到14个百分点。值得注意的是，这些工作使用了边缘和颜色特征，或将边缘敏感过滤器的网络权值初始化，而不是使用通常的做法从原始像素值中获取特征。目前还没有使用在ImageNet上预训练特征的结果被报道。

**结论** 尽管有一个共同的描述，但目前还没有明确的证据表明深度网络擅长在行人检测中学习特征。大部分成功的方法使用这个结构来对部件的高层面，遮挡和上下文进行建模。获得的结果和DPM以及决策树方法的结果处于同等水平，使得使用该结构的优点尚不明确。

**3.9 更好的特征**

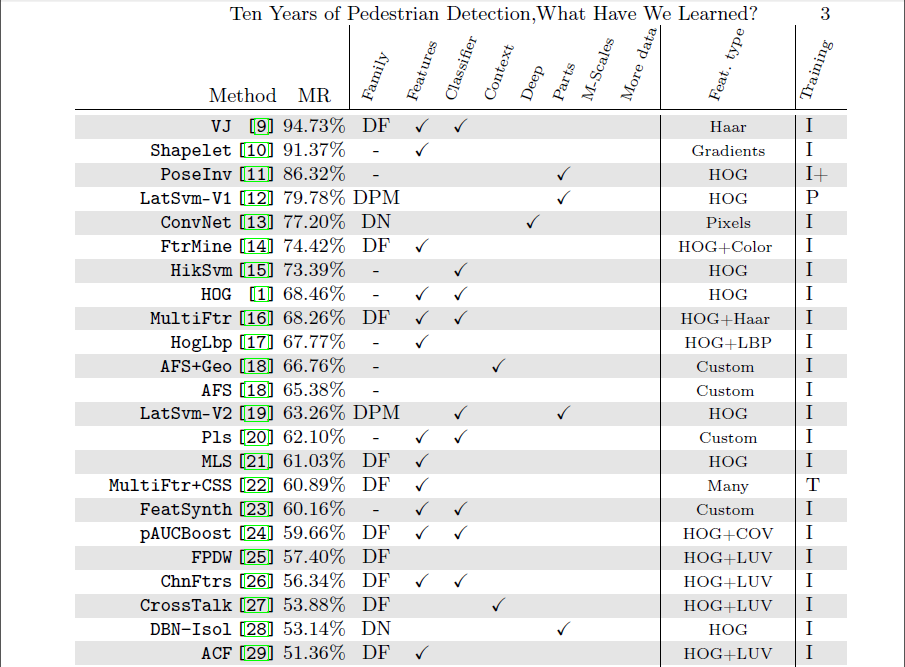
提高监测质量最普遍的方法是增加/多样化在输入图像上计算的特征。通过更丰富和更多维的表示，分类任务变的更容易，也使得结果有所提高。大量的特征类型已经被研究过：边缘信息，颜色信息，纹理信息，局部形状信息，协方差特征等。越来越多样化的特征被用来系统的提高检测性能。

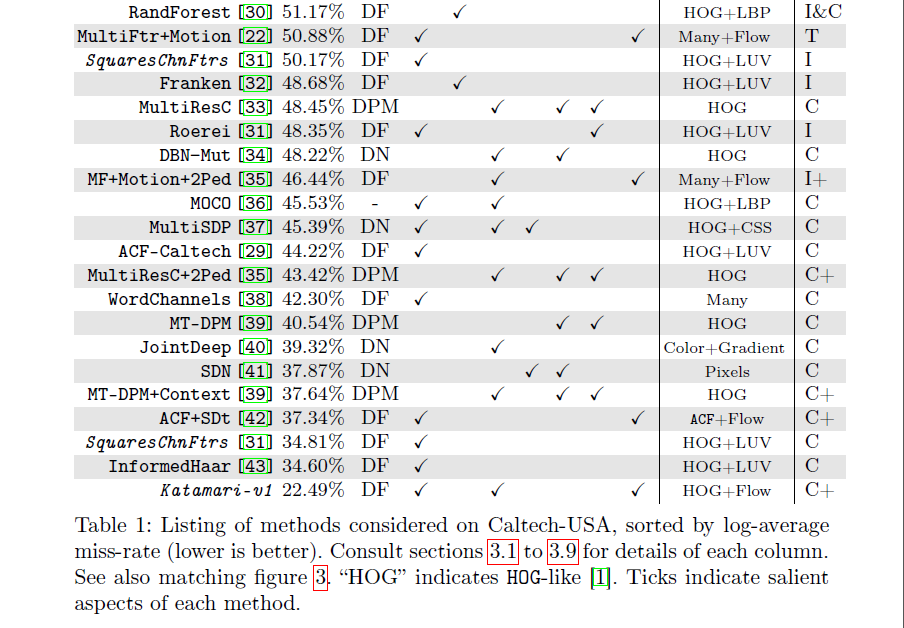
虽然很多决策森林方法使用了10个特征通道（ChnFtrs,ACF,Roerei,SquaresChnFtrs等），但有些论文考虑了更多的通道数量级。尽管通过增加通道数可以提高性能，最好表现的也只是使用了10个通道；见表1和图3.在4.1节中我们将详细研究更多不同特征的组合。

从我们的角度来看，从VJ到ChnFtrs，再到SquaresChnFtrs-Intia，改进的特征推动了发展。切换训练集使SquaresChnFtrs-Caltech在Caltech-USA数据集上达到了最好的表现；显著的提高更加复杂的方法。InformedHaar通过使用一系列为行人检测手工设计的Haar-like特征获得了最好的结果。相比之下，SquaresChnFtrs-Caltech没有使用该手工制作的特征而是数据驱动也获得了相似的结果。

最近的研究表明使用更多（更好的特征）可以获得更多的提升。值得注意的是对于行人检测更好的特征还未从深度学习方法中被获取。

**结论** 在过去的十年中，改进的特征一直推动着检测质量提高，并且似乎在接下来的几年依然如此。大多数的提高都是从广泛的试验和错误中获得的。下一个科学步骤将是建立一个对是什么让好的特征表现好的深刻认知以及如何设计更好的特征。





【本文的工作】:

[3]:对近年发展的再造/量化实验（使用了最新的20个模型来进行实验）。

本文所进行的实验主要是从三个方面进行的，即特征的影响、方法的补充以及跨数据集泛化。

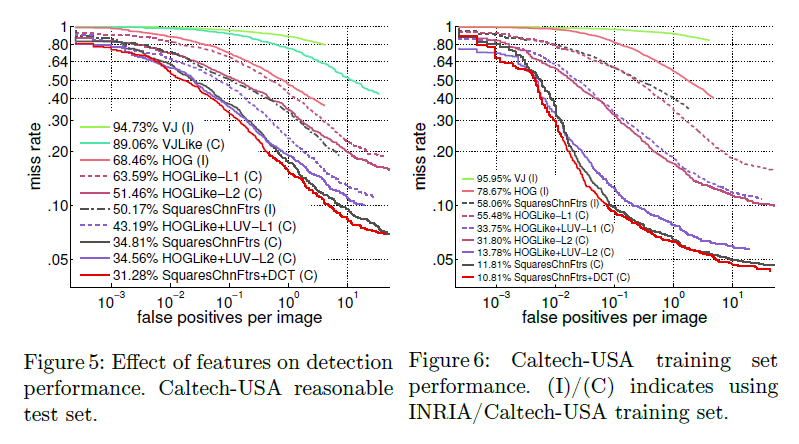
【思考】: 本文中的实验结果都与前文的分析结果一致，同时作者也在实验中获得了一些意外收获。如对已知特征的简单调整就可以获得不错的效果、多种行人（INRIA）比很多相似的行人（Caltech/KITTI）要好。这就表明了单纯的理论分析还不足以支持完整的工作，完备的实验是必不可少的。同时，仔细的分析实验结果可能会发现意想不到的结果。

**4. 实验**

基于我们前几节的分析，有三个方面在影响检测性能方面是最有希望的：更好的特征，附加的数据以及上下文信息。因此我们就这些方面进行补充性实验。

在讨论过的三个解决方案家族之中，我们选择Integral Channels Features框架来进行我们的实验。该家族的方法表现出良好的性能，训练时间在分钟到小时，并且适用于我们分析的目标。

特别的，我们使用了在[31]中描述的SquaresChnFtrs基线：2048level-2决策树在HOG+LUV通道上，组成一个64x128像素模板，通过vanilla AdaBoost和一些bootstrapping rounds of hard negative mining.



**4.1 回顾特征的影响**

在这一节中，我们评估了提升特征复杂度的影响。我们调整了所有在INRIA测试集上的方法，并在Caltech-USA测试集上进行了证明。在INRIA上的结果以及实现细节可以在补充材料中找到。

第一系列实验旨在模仿里程碑级的检测技术，如VJ，HOG+linearSVM和ChnFtrs。VJLike只使用了亮度色彩通道，仿照从原始特征中使用level 2 决策树的Haar wavelet like特征。HOGLike-L1/L2使用8x8像素池化区域，1梯度幅度和6定向梯度通道以及level 1/2决策树。我们还报告了添加LUV颜色通道HOGLike+LUV的结果。SquaresChnFtrs是第四节开头部分的基线，它和HOGLike+LUV很相似但却有着不同大小的方形池化区域。

受到[60]的启发，我们还通过用三个DCT（离散余弦变换）基函数（7x7像素）来对每个通道进行卷积并将过滤器响应的绝对值存储起来作为附加的特征通道把10HOG+LUV通道扩展到了40个通道。我们将该变体命名为SquaresChnFtrs+DCT.

**结论** 从VJ后大部分的进展可以通过使用更好的特征，基于定向梯度和颜色信息来解释。对这些已知的特征做简单的调整依然可以产生不错的提高。

**4.2 方法的补充**

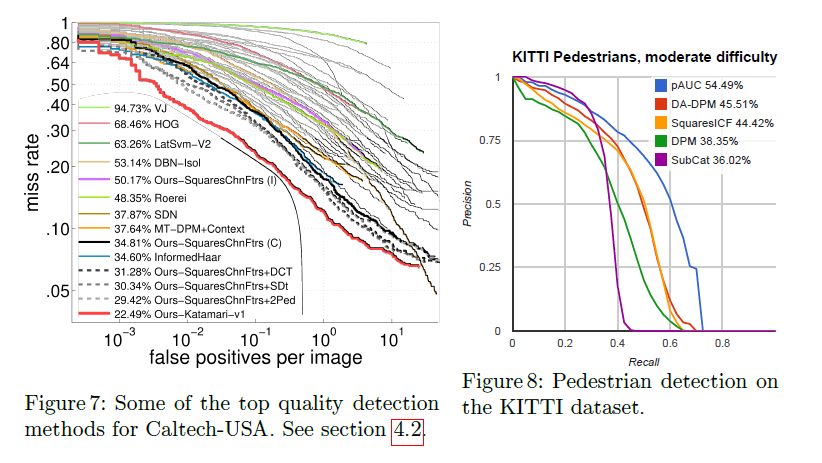
在4.1节对单帧特征影响进行回顾后，我们现在考虑更好的特征（HOG+LUV+DCT）、附加数据（通过光流）和上下文信息（通过人对人相互作用）的补充。

我们使用从ACF+SDt中获取的同样的SDt特征对光流进行编码（当前帧T和粗略对齐的T-4和T-8的图像差异）。使用+2Ped 重新加权策略来对上下文信息注入（检测器的评分和“2person”DPM检测器的评分相结合）。在所有实验中，DCT和SDt特征都被集中在8x8的区域中，而不是HOG+LUV特征的“所有方形大小”。

SquaresChnFtrs+DCT+SDt+2Ped的组合叫做Katamari-v1。不出意料的，Katamari-v1在Caltech-USA数据集上达到了最好的结果。图7中我们展示了所有表现最好的方法每个训练集和解决方案家族。补充材料中包含了所有元素之间组合的结果。

**结论** 我们的实验表明添加附加的特征，流和上下文信息是很大的补充，即使最开始使用的是强检测器。

通过对核心算法的进一步了解（从而进一步减少附加组件的相对影响）或者扩展一个系统中使用技术的多样性是否能使检测质量进一步发展是有待观察的。



**4.3 需要多少模型容量？**

检测的主要任务就是从训练集到测试集的泛化。在我们分析泛化能力之前，我们考虑一个对好质量检测器必须的条件：是否学习到的模型在训练集上表现的很好？

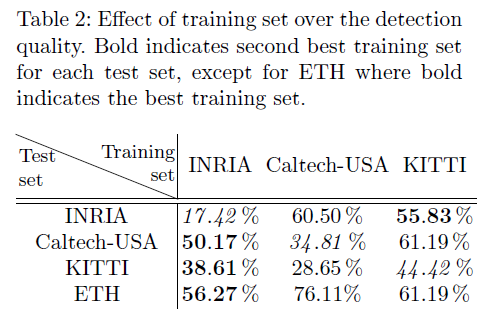
在图6中可以看到4.1节中所考虑模型在他们训练集上的检测质量。这些方法没有一个在训练集上表现的很好。 事实上，这和测试集上的表现趋于相似并且我们并没有观察到过拟合的现象。

**结论** 我们的结果表明增加检测器的识别能力有可能进一步提高检测质量。更好的识别能力可以从更多和更好的特征或复杂的分类器中获取。

**4.4 跨数据集泛化**

对于超越特定基准的现实应用，模型的泛化能力是关键。在这个意义上，在INRIA上训练，Caltech-USA上测试的模型比在Caltech-USA上训练和测试的模型结果更有意义。

表2展示了当使用不同训练集时SquaresChnFtrs在Caltech-USA上的表现。这些实验表明在Caltech或KITTI上训练提供了对INRIA的一些泛化能力，而相反则不行。出乎意料的是，尽管KITTI和Caltech看上去相似，INRIA是对KITTI和Caltech第二好的训练集选择。这表明了Caltech-USA行人有他们 “自己的种类”，INRIA数据集很有效是由于其多样性。换句话说，多种行人（INRIA）比很多相似的行人（Caltech/KITTI）要好。



好消息是最好的方法似乎在跨数据集以及在各自的训练数据上训练都表现的很好。图8展示了在KITTI上训练和测试的方法，可以看出SquaresChnFtrs比vanilla DPM 要好，并且与最好的DPM变体水平相当。目前在KITTI上最好的方法，Pauc,是使用了250个特征通道的ChnFtrs的变体。这两个观察结果与我们在3.9和4.1中讨论的一致。

**结论** 虽然在一个数据集上训练的检测器不能很好的转移到其他检测器，他们的排名在数据集之间是稳定的，这也表明了可以从表现好的方法中学到insights，而不管基准。

**5. 总结**

我们的实验表明过去十年行人检测的大部分进展可以归结于对特征的提高。证据表明该趋势会继续下去。尽管部分的特征是通过学习驱动，但他们主要是通过试验和错误手工制作的。

我们结合回顾性分析中表现良好的检测器成分的实验表明这些成分大部分是互补的。他们的组合会产生在Caltech-USA上最好的表现。

虽然三个行人检测器大家族是基于不同的学习技术，他们的结果却意外的相似。

目前主要的挑战似乎是对是什么让好的特征表现良好的更深层次理解，并依此可以设计出更好的特征。

【文末总结】: 作者从横跨十年的发展进程来看，大部分的进展归功于对特征的提高。这也符合了机器学习的相关理论。从近十年的发展我们可以看出未来行人检测领域的趋势。本文中提出的是“对是什么让好的特征表现良好的更深层次理解，并依此可以设计出更好的特征”。这篇文章发表于2014年，但近两年的深度学习浪潮似乎打破了该推断。深度网络作为黑箱模型却达到了最好的结果。越来越多的工作倾向于对该结构的探索及优化。但不可否认的是，对深度网络解释性的研究同样是为了设计更好的特征而所做的前期工作。