实验结果

1. Pascal VOC上的实验结果

我们在PASCAL VOC 2007测试集上对我们的方法进行了验证。表2显示了基于不同层特征的检测精度。对于我们的方法，我们分别使用单一尺度（s=688）以及5个尺度的特征。这里R-CNN的结果是[18]中报告的使用5个卷积层AlexNet[29]基础网络的版本。只使用pool5层特征，我们的结果（44.9%）与R-CNN的（44.2%）精度相差无几。但是，如果使用没有经过微调的fc6层特征，我们的方法在精度上会略差，一种可能的解释是，在预训练中全连接层是基于图像区域训练的，但是在检测中，它们的输入是特征图像的区域。特征图像区域在靠近边界的地方可能会有强响巧，而基于图像区域的特征一般不会。这种用法上的不同带来的精度损失可以通过微调网络来弥补。使用微调之后的全连接ftfc67，我们的结果与微调网络之后R-CNN的精度相似或更好一些。在使用框回归之后，我们5个尺度的结果（59.2%）比R-CNN的（58.5%）好（0.7%），而我们单尺度的版本（58.0%）略差（0.5%）。

在表1中，我们进一步比较了在使用同一个预训练网络（ZF）时，R-CNN及SPP-net的精度。在这种情况下，我们的方法R-CNN有着相同的性能。R-CNN的性能得益于这个预训练网络，有一定的提升。表3比较了每类的检测精度。

表3同时也增加了其他检测方法的结果。选择性搜索算法（SS）[32]在尺度不变特征变换特征图像上使用空间金字塔匹配。可形变部件模型（DPM）[23]和Regionlet[74]使用方向梯度直方图（HOG）特征[13]。Regionlet同结合包括conv5特征在内的多种特征将检测精度提升到46.1%[87]。DetectorNet[53]训练了一个深度网络用来输出像素级的物体掩码（mask）。这种方法与我们的方法一样，只需要对于整张图像使用一次深度卷积网络，但是他们的检测精度比较低，只有30.5%。

表 1 PASCAL VOC 2007上的检测结果（mAP）。使用与SPP(ZF)相同的基础网络结构

|  |  |  |  |
| --- | --- | --- | --- |
|  | SPP(1-sc) | SPP(5-sc) | R-CNN |
| (ZF) | (ZF) | (ZF) |
| ftfc7  ftfc7bb | 54.5 | 55.2 | 55.1 |
| 58.0 | **59.2** | **59.2** |
| 卷积时间（GPU)  全连接时间（GPU） | 0.053s | 0.0293s | 14.37s |
| 0.089s | 0.089s | 0.089s |
| 总时间（GPU)  加速比（vs.RCNN） | 0.142s | 0.382s | 14.46s |
| **102** | **38** | - |

表 2 PASCAL VOC 2007上的检测结果（mAP）。"ft"和"bb"表示网络微调和框回归。

|  |  |  |  |
| --- | --- | --- | --- |
|  | SPP(1-sc) | SPP(5-sc) | R-CNN |
| (ZF) | (ZF) | (ZF) |
| pool5  fc6 | 43.0 | 44.9 | 44.2 |
| 42.5 | 44.8 | 46.2 |
| ftfc6  ftfc7 | 52.3 | 53.7 | 53.1 |
| 54.5 | 55.2 | 54.2 |
| ftfc7bb | 58.0 | **59.2** | 58.5 |
| 卷积时间（GPU)  全连接时间（GPU） | 0.053s | 0.293s | 8.96s |
| 0.089s | 0.089s | 0.07s |
| 总时间（GPU)  加速比（vs.RCNN） | 0.142s | 0.382s | 9.03s |
| **64** | **24** | - |

图片包含 游戏机

描述已自动生成

1. 多模型融合的提升

多模型融合在基于卷积神经网络的图像分类任务中是提升精度的重要策略[29]。这里我们为物体检测任务提出了一种简单的模型融合方法。

我们使用相同的网络结果但是不同的随机初始化在ImageNet分类任务上预训练了另外一个网络。然后在这个新训练的网络上，我们重复以上检测任务的训练过程。表4（SPP-net（2））显示了这个网络的检测结果。它的mAP与第一个网络相当（59.1% vs. 59.2%），并且在11个子类上比第一个网络的效果好。

有了这两个网络之后，我们首先分别使用它们对测试图片上的生成框进行打分。然后我们对于两个模型检测结果的并集（包含分数）使用非极大值抑制。某个模型输出的高置信度的输出会抑制其他模型低置信度的结果。在模型融合之后，mAP提升到了60.9%（表4）。对于20个子类别，融合模型在其中17上比单个模型的精度要高。这说明了这两个模型是一定程度上互补的。

进一步，我们发现这种互补性主要来源于卷积层。我们尝试融合两个使用相同卷积层，以及在微调时分别随机初始化的网络，并没有观察到提升。

手机屏幕截图

描述已自动生成