Keep Calm and Carry On

RCNN (Regions with CNN) 目标物检测 Fast RCNN的基础

Abstract:
贡献主要有两点1：可以将卷积神经网络应用region proposal的策略，自底下上训练可以用来定位目标物和图像分割 2：当标注数据是比较稀疏的时候，在有监督的数据集上训练之后到特定任务的数据集上fine-tuning可以得到较好的新能，也就是说用Imagenet上训练好的模型，然后到你自己需要训练的数据上fine-tuning一下，检测效果很好。现在达到的效果比目前最好的DPM方法 mAP还要高上20点，目前voc上性能最好。

着篇文章主要是介绍RCNN，跟后面的，Fast RCNN和Faster RCNN比较关联，这篇文章是后两个的基础
1.介绍
在开始他说到LeCun对卷积神经网络中采用的SGD（通过反向传播的随机梯度下降算法）对网络训练很有效，也直接促进了利用CNN来做检测。
其实CNN的算法在90年代就已经出现了，可惜当时被SVM取代了，主要原因就是当时训练不动。2012年的时候Krizhevsky复燃了CNN，其在Imagenet的数据集上训练达到了非常好的效果，主要是用了LeCun中的一些技巧如（rectifying non-linearities and “dropout” regularization）

后来就有了讨论说把CNN方到目标检测上能达到什么样的效果。因此RossGirshick把问题主要聚集在了2个点上：
1一个是用深度网络来做一个检测，并且在整个high-capacity model中用较少的标注数据来training，比如几万张图像，（毕竟Imagenet上有上千万的图像数据）。不像图像分类任务，检测是需要定位的。因此RCNN里把这

个定位转换成一个regression problem（即回归问题）。当然他们在当时也想采用最经典的也就是sliding window，在卷积层增加了较大的感受野。但是他们最后没有采用，因为之前的DPM中也已经不采用这种方法了，无效的操作太多（PS.这里是我个人感觉，而且会增加复杂度）。他们最后采用的是Recognition using region的策略（这种paradigm已经在目标识别和semantic segmention中取得了较好的成功）。在测试阶段，他们提取约2000K预选框，从预选框中通过CNN提取出fixed-length的特征，最后通过特定类别的SVM来分类。对于不同大小的ROI采用了（affine image warping）来调整到固定的size，这种方法是不考虑region的形状的。整个系统的overview

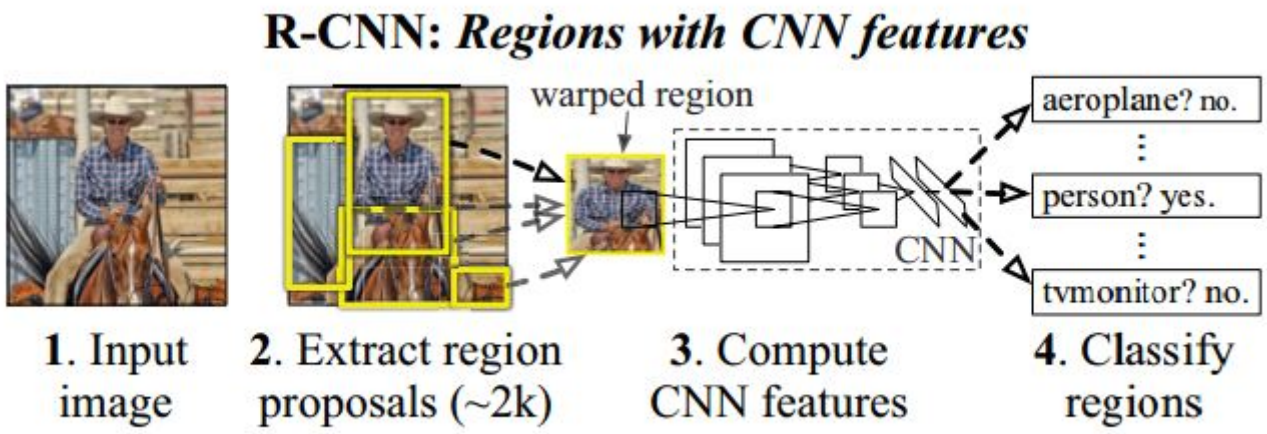

Figure 1: Object detection system overview. Our system (1) takes an input image, (2) extracts around 2000 bottom-up region proposals, (3) computes features for each proposal using a large convolutional neural network (CNN), and then (4) classifies each region using class-specific linear SVMs. R-CNN achieves a mean average precision (mAP) of **53.7% on PASCAL VOC 2010**. For comparison, [34] reports 35.1% mAP using the same region proposals, but with a spatial pyramid and bag-of-visual-words approach. The popular deformable part models perform at 33.4%.

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签

随笔档案

2016年5月 (2)
2016年3月 (2)
2016年2月 (2)
2016年1月 (7)
2015年12月 (2)
2015年11月 (5)
2015年10月 (5)
2015年9月 (1)
2015年8月 (2)
2015年7月 (1)
2015年6月 (1)
2015年5月 (1)
2015年4月 (3)
2014年12月 (1)

最新评论

1. Re:Caffe Python MemoryDataLayer Segmentation Fault

2.在实际检测中，训练的样本肯定是scarce，不足以训练一个大型的CNN网络。解决这个问题的方法是，首先通过无监督的预训练unsupervised pretraining,然后再进行supervised training，在实验中他们提到经过fine-tuninig，检测的mAP有8个点的提高。Ross提到Donahue的同时期的工作，其直接拿了krizhevsky的CNN网络用来做一个blackbox feature的extractor，这也在识别任务中表现出了较好的性能，如场景识别，细粒度的子分类，领域适应。分类计算中只有整个的分类工作只是一个矩阵相乘和非极大值抑制。

在错误分析中，可以发现bounding box的regression 可以明显的减少mislocalization。同理，作者说因为RCNN是工作在Region上的，因此其也可以较好的应用到semantic segmentation，最后也在voc2011上取得了较好的效果，比最好的高出1个点，（PS.我认为应该会有更好的性能，应该还没有做透，原来的那些分割仍然依赖浅层的特征）
特征提取：在网络之前，ROI不管大小形状都被缩放到一个固定的尺寸以适应网络。

2.2
测试时检测
在RCNN中，为每一类都训练了一个SVM，最后根据输出的特征类判断，每一个区域都有一个得分，最后通过greedy non-maximun supperssinon(for each class independently)来接受或者拒绝一个region，主要是看他这个有IoU的region是否比学习到的阈值有更高的得分。
对于运行分析
1.所有的CNN参数都在各个类别分享参数
2.CNN计算出来的参数是low dimensional 低维的，与其他方法比起来如空间金字塔，以及视觉词带模型

1. The only class-specific computations are dot products between features and SVM weights and non-maximum suppression
   4.RCNN可以应付类别很多情况并且不需要借助一些额外的近似手段，比如哈希什么的，别的方法在类别增长时，整个复杂度会上升很多，比之前的DPM的方法也要好很多
   2.3 训练过程
   supervised-pretraining ----> domain-specifi fine-tuning ---->object categroy classier
   1.supervised-pretraining是在imagenet上训练好的模型
   2.domian-specific fine-tuning 首先需要修改类别数目，并且在文中，Ross将IoU和GT大于0.5的看成是正样本，在SGD中 lr为pre-training rate的十分之一为0.001，这样不会影响预训练。在SGD中，每一次迭代，mini-batch大小是128，总共有32个postive window，96个negtive window.`

你好，添加的代码中变量 num_tasks_ 在哪儿定义和初始化啊？

--Joey Tang

2. Re:Fast RCNN 训练自己的数据集（3训练和检测）
您好！请问准确率是怎么得到的呢？

--DecMarryli

3. Re:Caffe fine-tuning 微调网络
@卉卉是爱学习的小青年兄弟 你找到文中所说的431类车型数据集了么 能分享下么...

--暴力的轮胎

4. Re:Linux的交叉编译 及configure配置
大神

--一年一度

5. Re:faster_rcnn c++版本的 caffe 封装（1）
各位博友们，有遇到下面的错误吗？求解答！[libprotobuf FATAL google/protobuf/stubs/common.cc:61] This program requires ver.....

--我该怎么办

阅读排行榜

3可视化学习的特征

在可视化学习特征中中，采用了一个很大的局部感受野的数据集。这了主要对卷基层进行可视化，输入region图像，根据unit激活之的大小排序，来看他对什么样的输入敏感。下图可以看到有一些Unit对人脸敏感如1，有一些对点阵，狗敏感如2行，第三行，对红色敏感，对第四行对文字敏感，也能将一些特征融合到一起入颜色、纹理、形状，如5行的屋子。。等等

这里很关键！5层之后为全连接层，全连接层可以将这些丰富的特征进行组合建模！

Figure 3: Top regions for six pool$_5$ units. Receptive fields and activation values are drawn in white. Some units are aligned to concepts such as people (row 1) or text (4). Other units capture texture and material properties, such as dot arrays (2) and specular reflections (6).

3.2 关于参数消除的研究

1.performace没有fine-tuning

从表中可以看到fc7产生的特征比fc6澶色恒的特征要差，这也就是说29%差不多1.68million个数据是可以从CNN网络中去掉的，并且几乎对mAP没什么影响。更加惊讶的是，如果把f6和f7都去掉的话，只用pool5层的参数也就是大于整个网络6%的参数也可以取得不错的结果如下图所示：可以看到大部分representational的能力主要是来自于CNN的卷积层，而不是主要的全连接层。这个发现可以用在稠密的特征map中，比如说Hog。这种表现能力也就是说我们有可以将其应用到一些滑动窗检测子中如DPM，在pool5的特征基础之上。作者原文（Much of the CNN’s representationalpower comes from its convolutional layers, rather than fromthe much larger densely connected layers. This finding suggests potential utility in computing a dense feature map, in the sense of HOG, of an arbitrary-sized image by using only the convolutional layers of the CNN. This representation would

enable experimentation with sliding-window detectors, including DPM, on top of pool5 features)

2.经过fine-tuning的性能

可以看到fine-tuning的效果还是很明显的，几乎提高了8个点，并且对于fc67的效果更明显，这也就是说从imagenet中学习到的pool5的特征比较general，并且对于性能的提升主要是来自对于在他们基础上的domain-specific具体应用场景的non-linear分类器的训练。

VOC 2007 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mA
R-CNN $pool_5$	51.8	60.2	36.4	27.8	23.2	52.8	60.6	49.2	18.3	47.8	44.3	40.8	56.6	58.7	42.4	23.4	46.1	36.7	51.3	55.7	44.
R-CNN fc_6	59.3	61.8	43.1	34.0	25.1	53.1	60.6	52.8	21.7	47.8	42.7	47.8	52.5	58.5	44.6	25.6	48.3	34.0	53.1	58.0	46.
R-CNN fc_7	57.6	57.9	38.5	31.8	23.7	51.2	58.9	51.4	20.0	50.5	40.9	46.0	51.6	55.9	43.3	23.3	48.1	35.3	51.0	57.4	44.
R-CNN FT $pool_5$	58.2	63.3	37.9	27.6	26.1	54.1	66.9	51.4	26.7	55.5	43.4	43.1	57.7	59.0	45.8	28.1	50.8	40.6	53.1	56.4	47.
R-CNN FT fc_6	63.5	66.0	47.9	37.7	29.9	62.5	70.2	60.2	32.0	57.9	47.0	53.5	60.1	64.2	52.2	31.3	55.0	50.0	57.7	63.0	53.
R-CNN FT fc_7	64.2	69.7	50.0	41.9	32.0	62.6	71.0	60.7	32.7	58.5	46.5	56.1	60.6	66.8	54.2	31.5	52.8	48.9	57.9	64.7	54.
R-CNN FT fc_7 BB	**68.1**	**72.8**	**56.8**	**43.0**	**36.8**	**66.3**	**74.2**	**67.6**	**34.4**	**63.5**	**54.5**	**61.2**	**69.1**	**68.6**	**58.7**	**33.4**	**62.9**	**51.1**	**62.5**	**64.8**	**58.**
DPM v5 [18]	33.2	60.3	10.2	16.1	27.3	54.3	58.2	23.0	20.0	24.1	26.7	12.7	58.1	48.2	43.2	12.0	21.1	36.1	46.0	43.5	33.
DPM ST [26]	23.8	58.2	10.5	8.5	27.1	50.4	52.0	7.3	19.2	22.8	18.1	8.0	55.9	44.8	32.4	13.3	15.9	22.8	46.2	44.9	29.
DPM HSC [28]	32.2	58.3	11.5	16.3	30.6	49.9	54.8	23.5	21.5	27.7	34.0	13.7	58.1	51.6	39.9	12.4	23.5	34.4	47.4	45.2	34.

Table 2: Detection average precision (%) on VOC 2007 test. Rows 1-3 show R-CNN performance without fine-tuning. Rows 4-6 sho results for the CNN pre-trained on ILSVRC 2012 and then fine-tuned (FT) on VOC 2007 trainval. Row 7 includes a simple bounding bo regression (BB) stage that reduces localization errors (Section 3.4). Rows 8-10 present DPM methods as a strong baseline. The first us only HOG, while the next two use different feature learning approaches to augment or replace HOG.

3.4 关于BBOX

首先需要明确的是，RCNN并不是从预选框里选择一个判断一下那么简单，在论文中的错误分析，大部分的检测错误的主要成分都是localize error 也就是定位错误，IoU在0.1和0.5之间。与别的类别以及背景confusion比例非常小，在这里作者根据最后输出的feature 进一步做了regression， 采用的是之前在DPM检测中的用的Linear regression model，这个让mAP大概提高了4个点。

4 Semantic segmentation

文中也提到了将RCNN网络用语分割，但是效果与目前较好的O2P的方法没有本质的提高约0.9。我认为主要还是网络学习过程并不足，其对于细粒度的特征没有一个整体的学习过程，目前在semantic segmentation上性能最好的是《Learning Deconvolution Network for Semantic Segmentation》目前在pascal-voc数据集上是第一的性能，他的网络中有一个对称的deconvolutoin network。

结束语：最近这几年确实，在目标物检测的性能上是停滞不前了，现在最好的DPM算法都是结合好多low-level的feature，并且这些feature都是手工设计的加上一些high-level context from detector和scene classifier。这篇

文章给出了基于Region proposal和CNN网络极大的提高了mAP。有监督的预训练在特定场合的fine-tuning这一模式会针对很多数据系稀疏的是视觉问题有效。作者这里的意思是说拿Imagenet上训练好的那个模型，然后根据自己的特定应用场景，把模型用自己的数据fine-tuning一下，这样的做法是挺有效的。

Hello~again
关注 - 1
粉丝 - 165
+加关注

4 1

« 上一篇：车脸检测 Adaboost 检测过程
» 下一篇：Linux与Windows 解压乱码 UTF8BOM读取问题

posted @ 2015-09-26 01:11 Hello~again 阅读(15757) 评论(2) 编辑 收藏

评论列表

#1楼 2016-01-27 16:02 **kltsyn**

您好楼主
看了您的分析明白了不少
一直不明白最后的SVM是根据什么做的分类，每个proposal提完特征以后，svm是怎么知道把这些特征分为哪一类的呢

支持(0) 反对(0)

#2楼[楼主] 2016-02-19 13:07 **Hello~again**

@ kltsyn
首先训练的时候是有GroundTruth的，跟这个GroundTruth重合的比例在文中默认是50%就归为这个GT对应的类别，否则小于50%就算作Background了

支持(0) 反对(0)

刷新评论 刷新页面 返回顶部

注册用户登录后才能发表评论，请 登录 或 注册，访问网站首页。