

灰巧克力爱松露

这个人很懒，什么也没写~

目录视图

摘要视图

RSS 订阅

个人资料



灰巧克力爱松露

访问：116771次

积分：1656

等级：BLOG C4

排名：千里之外

原创：53篇

转载：0篇

译文：1篇

评论：142条

文章分类

Deep Learning (24)

Python OpenCV (12)

FPGA Vision (7)

Detection (1)

Matlab (1)

Unity 3D (1)

Leap Motion (1)

Mobile Robotics (4)

Arduino (1)

Android (3)

Tracking (1)

文章存档

2017年03月 (1)

2017年02月 (2)

2016年11月 (1)

2016年10月 (4)

2016年07月 (1)

展开

阅读排行

R-FCN：基于区域的全卷

循环和递归神经网络 (RN

Caffe学习——使用自己

Theano入门——CIFAR-10

OpenCV Using Python——

OpenCV Using Python——

对抗的深度卷积生成网络

与判别网络对抗的生成网

Torch学习——使用CUDA

(20397)

(6865)

(6649)

(5887)

(5282)

(3293)

(3010)

(2501)

(2417)

福利 | 百度 AI 开发者大会免费门票领取

CSDN日报20170627——《回眸后那流泪的脸——从高考到程序猿》

6月书讯 | 最受欢迎的 SQL 入门书重磅升级

R-FCN：基于区域的全卷积网络来检测物体

标签：R-FCN 基于区域的全卷积网络 Jifeng-Dai 物体检测 CVPR

2016-06-29 14:07 20438人阅读 评论(17) 收藏 举报

分类： Deep Learning (23)

版权声明：本文为博主原创文章，未经博主允许不得转载。

目录 (?) [+]

原文标题为“R-FCN: Object Detection via Region-based Fully Convolutional Networks”，作者代季，清华博士到微软亚洲研究院的视觉计算组，CVPR 16 两篇一作的会议主持人~ 同时公布了源码~ 2

后面主要内容为原文随便的翻译或概括。必有不紧贴原文原意之处，曲解请指出，否则求放过~

1. 简介

物体检测的深度网络按感兴趣区域 (RoI) 池化层分为两大主流：共享计算的全卷积子网络 (每个子网络与 RoI 无关) 和 不共享计算的作用于各自 RoI 的子网络。工程分类结构 (如 Alexnet 和 VGG Nets) 造成这样的分流。而工程上的图像分类结构被设计为两个子网络——1个后缀1个空间池化层的卷积子网络和多个全连接层。因此，图像分类网络中最后的空间池化层自然变成了物体检测网络中的 **RoI** 池化层。

近年来，诸如残差网络和 GoogLeNets 等先进的图像分类网络为全卷积网络。类似地，自然会想到用在物体检测中用全卷积网络 (隐藏层不包含作用于 RoI 的子网络)。然而，物体检测工作中的经验表明，这样天真的解决方案的检测效果远差于该网络的分类效果。为弥补尴尬，更快 R-CNN 检测器不自然地在两卷积层间插入 RoI 池化层，这样更深的作用于各 RoI 的子网络虽精度更高，但各个 RoI 计算不共享所以速度慢。

尴尬在于：物体分类要求平移不变性越大越好 (图像中物体的移动不用区分)，而物体检测要求有平移变化。所以，ImageNet 分类领先的结果证明尽可能有平移不变性的全卷积结构更受青睐。另一方面，物体检测任务需要一些平移变化的定位表示。比如，物体的平移应该使网络产生响应，这些响应对描述候选框覆盖真实物体的好坏是有意义的。我们假设图像分类网络的卷积层越深，则该网络对平移越不敏感。

我曾看到的尴尬包括：

a) Kaggle 中的白鲸身份识别。刚开始很多人尝试从图像到坐标的直接回归，到后面有几位心善的大哥分享了自己手动标定后白鲸的图像坐标，后来显著的进展大多是因为把白鲸的位置检测和身份识别问题简化为白鲸的身份识别问题。

b) Caffe 用于物体检测时的均值收敛问题。

为消除尴尬，在网络的卷积层间插入 RoI 池化层。这种具体到区域的操作在不同区域间跑时不再有平移不变性。然而，该设计因引入相当数目的按区域操作层 (region-wise layers) 而牺牲了训练和测试效率。

本文，我们为物体检测推出了基于区域的全卷积网络 (R-FCN)，采用全卷积网络结构作为 FCN，为给 FCN 引入平移变化，用专门的卷积层构建位置敏感分数地图 (position-sensitive score maps)。每个空间敏感地图编码感兴趣区域的相对空间位置信息。在FCN上面增加1个位置敏感 RoI 池化层来监管这些分数地图。

2. 方法

(1) 简介

评论排行

| | |
|-----------------------|------|
| Caffe学习——使用自己的数据 | (53) |
| OpenCV Using Python—— | (20) |
| R-FCN：基于区域的全卷积网络 | (17) |
| 与判别网络对抗的生成网络 | (10) |
| 用核化的相关滤波器来高 | (7) |
| UFLDL教程——Softmax | (7) |
| OpenCV Using Python—— | (5) |
| 检测9000类物体的YOLO | (4) |
| OpenCV Using Python—— | (4) |
| Caffe与Lasagne使用—— | (4) |

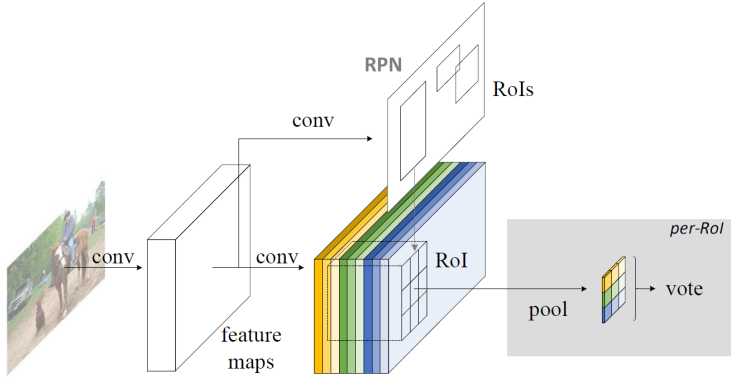
最新评论

| |
|--|
| OpenCV Using Python——单目测距：博主做过线结构光的测量吗 |
| OpenCV Using Python——单目测距：qiongjutong3175: 还有还有，那个检查R矩阵是不是正交矩阵的函数能不能也给我发一下啊，... |
| OpenCV Using Python——单目测距：qiongjutong3175: @shadow_guo: 太好啦谢谢博主。我的邮箱是 ZQ_INS_SEU@163.com. 祝博主天天开... |
| OpenCV Using Python——单目测距：灰巧克力爱松露: @qiongjutong3175: 很久没弄这个啦。之前用的是OpenCV 2.4.9，还是自行解决吧... |
| OpenCV Using Python——基于SIFT：qiongjutong3175: 博主，求问 from match import * 中的 match 提示没有怎么办？在官网上也没有搜到ma... |
| Caffe学习——使用自己的数据（灰巧克力爱松露: @tenghaiwander: 命令行安装：sudo pip install pydot sudo ap... |
| Caffe学习——使用自己的数据（tenghaiwander: @shadow_guo: 楼主，我又回来了。。。我按照文中的代码跑的时候在 print_network 函... |
| Caffe学习——使用自己的数据（xdqingsfm: 好的，万分感谢楼主！ |
| Caffe学习——使用自己的数据（灰巧克力爱松露: @xdqingsfm: |

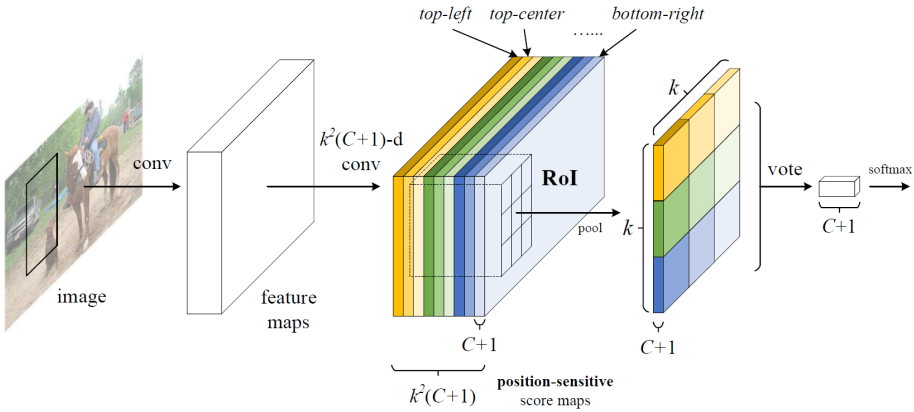


R-FCN：基于区域的全卷积网络来检测物体 - 灰巧克力爱松露 - 博客频道 - CSDN.NET

效仿 R-CNN，采用流行的物体检测策略，包括区域建议和区域分类两步。不依赖区域建议的方法确实存在 (SSD 和 Yolo 弟兄)，基于区域的系统在不同 benchmarks 上依然精度领先。用更快 R-CNN 中的区域建议网络 (RPN) 提取候选区域，该 RPN 为全卷积网络。效仿更快 R-CNN，共享 RPN 和 R-FCN 的特征。



RPN 给出感兴趣区域，R-FCN 对该感兴趣区域分类。R-FCN 在与 RPN 共享的卷积层后多加1个卷积层。所以，R-FCN 与 RPN 一样，输入为整幅图像。但 R-FCN 最后1个卷积层的输出从整幅图像的卷积响应图像中分割出感兴趣区域的卷积响应图像。



R-FCN 最后1个卷积层在整幅图像上为每类生成 k^2 个位置敏感分数图，有C类物体外加1个背景，因此有 $k^2(C + 1)$ 个通道的输出层。 k^2 个分数图对应描述位置的空间网格。比如， $k \times k = 3 \times 3$ ，则9个分数图编码单个物体类的 {top - left, top - center, top - right, ..., bottom - right}。

R-FCN 最后用位置敏感 RoI 池化层，给每个 RoI 1个分数。选择性池化图解：看上图的橙色响应图像 (top - left)，抠出橙色方块 RoI，池化橙色方块 RoI 得到橙色小方块 (分数)；其它颜色的响应图像同理。对所有颜色的小方块投票 (或池化) 得到1类的响应结果。

选择性池化是跨通道的，投票部分的池化为所有通道的池化。而一般池化都在通道内。

R-FCN 最后1个卷积层的输出为什么会具有相对空间位置这样的物理意义 (top-left, top-center, ..., bottom-right)？

原文为“With end-to-end training, this RoI layer shepherds the last convolutional layer to learn specialized position-sensitive score maps.”。所以，假设端到端训练后每层真有相对位置的意义，那么投票前的输入一定位置敏感。投票后面的内容用作分类。

端到端训练先自行脑补：
假设已知原图像与真实物体的边界框中心坐标和宽高，把1个物体的边界框中心坐标分成 k^2 个网格的中心坐标，宽高缩放为物体宽高的 $\frac{1}{k}$ 倍，得到每个网格的掩码。用原图像和每类物体的网格在整幅图像中的掩码端到端训练全卷积网络。挺像图像分割~

(2) 基础结构

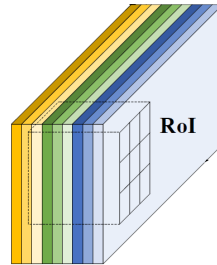
ResNet-101 网络有100个卷积层，1个全局平均池化层和1个1000类的全连接层。仅用ImageNet预训练的该网络的卷积层计算特征图。

(3) 位置敏感分数图

对 R-FCN 的卷积响应图像按 RPN 的结果分割出来感兴趣区域，对单通道的感兴趣区域分成 $k \times k$ 个网格，每个网格平均池化，然后所有通道再平均池化。

其实不是这样的~ 因为 RoI 覆盖的所有面积的橙色方片都是左上位置的响应。

“To explicitly encode position information into each RoI, we divide each RoI rectangle into $k \times k$ bins by a regular grid.” 这句话应对应下图 (对应后面效果图的黄色虚线部分)：



对1个大小为 $w \times h$ 的 RoI，1个桶 (bin) 的大小为 $\frac{w}{k} \times \frac{h}{k}$ ，最后1个卷积层为每类产生 k^2 个分数图 ($0 \leq i, j \leq k-1$)，定义1个位置敏感 RoI 池化操作：

$$r_c(i, j|\Theta) = \frac{1}{n} \sum_{(x,y) \in \text{bin}(i,j)} z_{i,j,c}(x + x_0, y + y_0|\Theta)$$

其中， $r_c(i, j|\Theta)$ 为第 c 类第 (i, j) 个箱子的池化响应， $z_{i,j,c}$ 为 $k^2(C+1)$ 个分数图中的输出， (x_0, y_0) 为 RoI 的左上角坐标， n 为桶里的像素总数，且 Θ 为网络的参数。

桶对应后面效果图的黄色实线部分，1个桶只抠了每类的每个相对空间位置通道中 RoI 的对应相对空间位置的分数图，其它的部分丢弃。

(4) 分类

对该 RoI 每类的所有相对空间位置的分数平均池化 (或投票)。

$$r_c(\Theta) = \sum_{i,j} r_c(i, j|\Theta)$$

Softmax 回归分类。

(5) 定位

$k^2(C+1)$ 维的卷积层后，增加1个 $4k^2$ 维的卷积层来回归边界框。每个 RoI 产生的 $4k^2$ 维向量经平均投票后，用快速 R-CNN 的参数化得到1个4维向量 (t_x, t_y, t_w, t_h) 。

(6) 训练

每个 RoI 的损失函数为交叉熵损失与边界框回归损失的和。

$$L(s, t_{x,y,w,h}) = L_{cls}(s_{c^*}) + \lambda[c^* > 0]L_{reg}(t, t^*) = -\log\left(\frac{e^{r_{c^*}(\Theta)}}{\sum_{c'=0}^C e^{r_{c'}(\Theta)}}\right) + \lambda[c^* > 0]L_{reg}(t, t^*)$$

其中， $c^* = 0$ 说明 RoI 的真实标签为背景。 L_{reg} 与快速 R-CNN 中的边界框损失回归相同。RPN 产生的区域建议当 RoI 与真实边框的 IoU 超过0.5时，标定为正样本。

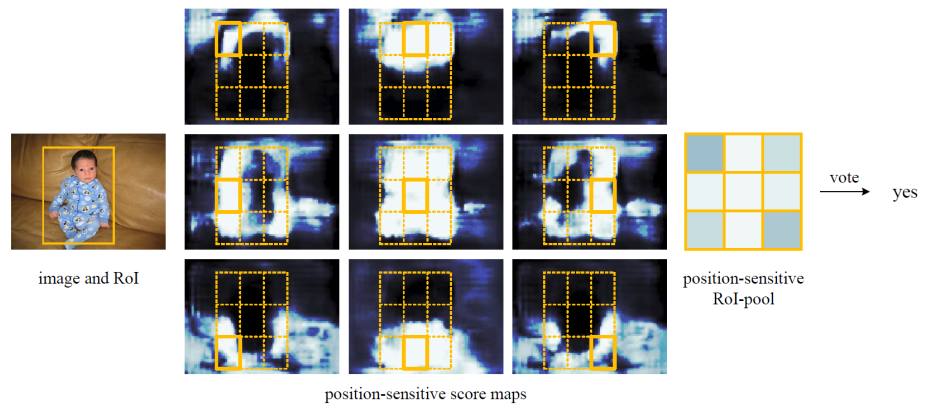
在线难例挖掘 (OHEM)。假设每个图像前向产生 N 个区域建议，计算所有建议的损失。按损失排序所有 RoIs，选择损失最高的 B 个 RoIs。

(7) 可视化

RoI 分类的可视化。RPN 刚好产生包含 person 类的 RoI。经过 R-FCN 的最后1个卷积层后产生9个相对空间位置的分数图，对 person 类的每个相对空间位置通道内的 RoI 桶平均池化得到 3×3 的池化分数，投票后送入分类



器判断属于 person 类。当分类正确时，该类通道的位置敏感分数图 (中间) 的大多数橙色实线网格内的响应在整个 RoI 位置范围内最强。



3. 相关工作

R-CNN 证实用深度网络产生区域建议是有效的。R-CNN 在剪切变形的区域上评价卷积网络， $\int_{\text{region}} \text{score} \, dx$ 计算。SPP 网络，快速 R-CNN 和更快速 R-CNN 为“半卷积” (卷积子网络在整幅图像上共享计算， $\int_{\text{region}} \text{score} \, dx$ 评价各个区域)。

一些物体检测器被认为是“全卷积”模型。OverFeat 在共享卷积特征图上滑窗操作来检测物体。类似地，快速 R-CNN 等也用滑动窗口，它们的1个单尺度的滑动窗口可看作1个卷积层。更快 R-CNN 的 RPN 部分为1个预测关于多尺寸参考盒 (锚) 的边界框的全卷积检测器。更快 R-CNN 的 RPN 未知区域建议的类，但 SSD 该部分已知特定的类。

另一类物体检测器采用全连接层，在整个图像上产生整体物体的检测结果。

4. 实验

(1) PASCAL VOC

训练 VOC 07 trainval 和 VOC 12 trainval，测试 VOC 07 test。

a. 与其它全卷积策略比较

朴素更快 R-CNN

ResNet-101 的共享特征图，最后1个卷积层后用 RoI 池化。每个 RoI 上用21类全连接层。

ResNet-101 (conv4 与 conv5 间插入 RoI 池化层)，朴素更快 R-CNN (conv5 后插入 RoI 池化层)。mAP 升 7.5%。经验证实更快 R-CNN 系统的卷积层间插入 RoI 池化层能提高相关空间信息的重要性。

特定类 RPN

训练 RPN 与 更快 R-CNN 部分相同，2类卷积分类层 (物体或背景) 改为 21类卷积分类层 (20类物体+1背景)。

特定类 RPN 类似于快速 R-CNN 的特殊形式 (用稠密的滑窗替换区域建议)。mAP 跌 8.8%。效果不如2类 RPN。

无位置敏感的 R-FCN

$k = 1$ 时，位置不敏感。相当于每个 RoI 全局池化。

位置敏感的 R-FCN 相对于 ResNet-101，mAP 升 0.02 %~

b. 与用 ResNet-101 的更快 R-CNN 比较

所要比较的检测器为各大 Benchmark 上的最强竞争者。名字太长，后面简称暂时最强检测器。

原作者的意思可能是这样：结合 MS COCO 训练后，R-FCN 仅需多尺度训练 mAP 就能到 82%，而暂时最强检测器，除了多尺度训练，还要迭代盒回归和上下文才比 R-FCN 多 2.0% 和 1.8%；而且，即使不带 MS COCO 训练，没有上述附加的“+++”工作，R-FCN 也比暂时最强检测器的 mAP 还要至少高 3%。而且，R-FCN 快得多。

带 MS COCO 训练数据后，mAP 大涨~ P.S. 该数据集由微软发起~



c. 深度的影响

深度为50~101时 mAP 会增加，到152层时饱和。

d. 区域建议的影响

RPN 即使用选择搜索和边缘盒 (Edge Boxes) 也有 77% 以上的 mAP。

(2) MS COCO

IoU 为0.5时，R-FCN和暂时最强检测器的 mAP 仅有刚过50%。说明 MS COCO 与 PASCAL VOC 相比有更大的挑战性~

5. 小结

R-FCN 在数据集 VOC 07 和 12 上的 mAP 分别为 **83.6%** 和 **82%**，测试时每张图片耗时 170ms。微软的视觉计算组其实每年在领先的成果上改进了一点点，但原理简单，分析角度又新，实验规模也不小。该团队不仅明星云集，且力往一块使，容易出大片~(￣▽￣)

顶 踩
3 0

上一篇 CRAFT：级联的区域建议网络与快速RCNN结合

下一篇 循环和递归神经网络 (RNN) 与 长短时记忆 (LSTM)

相关文章推荐

- R-FCN：基于区域的全卷积网络来检测物体
- OpenCV学习——物体跟踪的粒子滤波算法实现之...
- OpenCV 使用光流法检测物体运动
- (论文+代码) R-FCN：基于区域的全卷积神经网...
- 无线连接提示“检测到无线网络”但打开无线列表显...

- [转]一个Collision类，其中的block方法可以实现两...
- C++图形边缘检测与区域提取
- 卷积神经网络的基础知识及应用——识别、检测、...
- OpenCV编程案例：使用轮廓函数检测连通区域
- 论文翻译 基于R-FCN的物体检测

猜你在找

- 深度学习基础与TensorFlow实践
- 【在线峰会】前端开发重点难点技术剖析与创新实践
- 【在线峰会】一天掌握物联网全栈开发之道
- 【在线峰会】如何高质高效的进行Android技术开发
- 机器学习40天精英计划
- Python数据挖掘与分析速成班
- 微信小程序开发实战
- JFinal极速开发企业实战
- 备战2017软考 系统集成项目管理工程师 学习套餐
- Python大型网络爬虫项目开发实战（全套）



查看评论

7楼 jinmingxinz 2017-01-05 17:31发表



博主您好，这篇文章的一些图示有点不懂。想问下：

“最后一个feature map通过conv生成这么多个score map” 这一步是怎么进行的？？如何通过conv能生成所谓的score map？？图示中不同颜色的score map就代表着不同位置吗？？意思是每个位置一张score map？一共9张？

Re: 灰巧克力爱松露 2017-01-05 21:47发表



回复jinmingxin: 生成分数图可参考2楼评论（但不一定和原作者本意一样），或翻翻源码，进一步可联系原作者。而后3个问题是这样的，其实你只是不确定而已~ \((^ω^)\)

6楼 不破楼兰终不还 2016-12-29 11:34发表



博主您好，我看了一下论文实现部分，发现是不是他用的方法非常想DPM的做法，就是分成K*K个网格之后，计算这些网格的总得分才能决定最后的得分，但是这样感觉网络学习到的就不是整体的目标信息，而是分开的信息了，这样对于分类的准确性更好么？

2.还有一个问题就是，作者在最后一个卷积层之后又接了一个position-sensitive score maps，然后又接了一个position-sensitive ROI pooling层?是这样的结构么？那最后这个得分图和池化图有什么关系呢？感觉最后的这个池化图起了分类的作用？

Re: 灰巧克力爱松露 2016-12-29 15:30发表



回复fx409494616: 看过代码了呀，那么效果上你有发言权。因为目前每个人的经验性结论都是靠自己的实验结果来支撑的，所以我的观点不代表普适性的正确~

1. 个人感觉所有同时考虑物体整体和局部的模型都像DPM，但我并不会去关心它像什么，因为部件模型反映的只是人看待物体检测问题的一种方式：考虑整体与局部的联系。

正如你说，分数图对应的卷积层得到的是分开的信息，但这只能说明分数图那一层并未考虑整体。

但是，不看代码你也会清楚：从分数图后的池化层（假设为输入）到最终的分类或检测结果（假设为输出）这一中间部分，输出节点的值一定考虑了所有输入节点的值（以某些权重连接），这个才是考虑整体的地方。

此时，只需要理解到局部后面考虑了整体就够了。因为“考虑整体和局部关系”这一观点是可以理解的，但对自己任务影响的大小是由自己的实验结果决定的。所以，哪个效果就支持哪个。

2. 得分图与池化的关系就是指卷积与池化的关系，只是卷积输出的特征图含义不一样。而池化的作用如原文所说，消除平移不变性，毕竟分类问题也不需要考虑平移特征，自然对分类任务有贡献。但池化和考虑整体也无关系，因为池化操作是固定操作（平均或最大池化等），并没有针对不同的任务在池化操作上对网格特征采用不同的加权~

5楼 zj443108444 2016-11-19 17:00发表



楼主，你好，我想请问一个问题，就是位置得分的map为什么要设置成k*k*(c+1)个，相比直接设置成c+1个，他的优势到底体现在哪里呢？？我每个roi里面的bin去不同的map里面取比我直接在一个map里面取，到底有啥好处啊？？

Re: B-Kalasiky 2016-12-23 12:55发表



回复zj443108444: 我理解的是k*k是分成k*k个块，每个块都对应C+1个，论文里就设k=3了，所以是49* (C+1) 。

Re: 灰巧克力爱松露 2016-11-19 19:16发表



回复zj443108444: 小伙纸你好~ 问得挺认真嘛~

(1) 以识别人脸表情来理解你的问题（但愿没理解歪）：识别人脸表情到底是识别人脸整体运动好，还是把人脸图像块网格化成运动单元再综合起来好？

(2) 如果说一定要给出个为什么，嘿嘿~这只是个解决问题的思路呢。并且，原作者给出了k=1时的实验比较，见“无位置敏感的 R-FCN”部分，这个效果提升一定很难说服你。但是，不代表你把这个思路用在具体场合一定不会有大的提升。

(3) 如果还不满意，阔以直接邮件原作者~

\(^_~^)/———>

Re: zj443108444 2016-11-20 15:15发表



回复shadow_guo: hah~也不是说不能说服我呐。我看了原文中的说法，总觉得说的不是太清楚，也不直观。不过博主的说法还是让我略微透彻了点。。

感谢

Re: 灰巧克力爱松露 2016-11-20 17:04发表



回复zj443108444: 不客气~ φ(° ▽ °)~

4楼 KCDKFT 2016-10-14 00:16发表



好吧，原来你是翻译的.....难怪看起来怪怪的

3楼 KCDKFT 2016-10-14 00:11发表



感觉楼主对于全卷积网络的定义有问题吧，全卷积网络默认指的是做分割的那套框架，相当于多level feature的融合，而low level feature对于位置肯定是敏感的

Re: 灰巧克力爱松露 2016-10-14 11:20发表



回复kuaitoukid: 倒是提醒我了~ 就按论文“Fully Convolutional Networks for Semantic Segmentation”的内容来说吧：

(1) 全卷积网络的定义

FCN的定义：This functional form is maintained under composition, with kernel size and stride obeying the transformation rule

[公式] (该公式将卷积，池化和激活函数合并入一个统一形式的函数)。



While a general deep net computes a general nonlinear function, a net with only layers of this form computes a nonlinear filter, which we call a deep filter or fully convolutional network.

由上述统一形式的函数计算的网路称为全卷积网路。

(2) 全卷积网路与多 level 特征的关系

分割的原文这样说：We address this by adding skips [1] that combine the final prediction layer with lower layers with finer strides. This turns a line topology into a DAG, with edges that skip ahead from lower layers to higher ones (Figure 3).

所以呢，全卷积网路的定义看图2比较清楚。这里，多 level 特征融合只是结合的全卷积网路的中间 level 的特征，也就是说最后分割用的网路是 DAG 网路。

(3) 翻译问题

其实并不推荐大家看翻译，仅仅是看论文时的副产品。我自己也很少看别人的翻译。水平有限，怪怪的见谅~

2楼 神谷姬爱ML 2016-08-30 18:12发表



R-FCN 最后1个卷积层的输出为什么会具有相对空间位置这样的物理意义 (top-left,top-center,...,bottom-right)？这个也没看懂

Re: 灰巧克力爱松露 2016-08-31 09:28发表



回复kakakakaka111111：卷积层的输出内容原作者是这样定义的。至于怎么用标定的数据去训练使得最后的卷积层输出有这样的物理意义，原文没有提及。所以后面脑补的内容是为了说明如何使输出层具有相对空间位置的意义。假设图像已经标好边界框，那么理想的卷积层输出是边界框内响应大，边界框外响应小；如果标定了边界框内左上角网格的响应大，图像其它地方响应小。直接的标定方法为响应大的灰度为255，响应的灰度为0。那么边界框内左上角网格的响应大，图像其它地方响应小。直接的标定方法为响应大的灰度为255，响应的灰度为0。可以得到9个相对空间位置的标定图像。相当于标定图像可以从标定对象整体细分到它的9个位置。把它看成分割这样会比较好理解：输入是原图像，输出是相对空间位置的标定图像，训练这样的网路使得最后1个卷积层输出有空间意义。实际操作时我感觉不是分割，可以选择对9个相对位置的标定图像分别训练，这样每个残差网路（或权值共享成1个网路）最后1个卷积层输出为1个相对位置的热度图。拼起来后再训练池化后面的部分。（这一段算我猜的~）池化部分的计算注意颜色：橙色对应左上角，那么系统框图中所有相同颜色的部分都是左上角，可以看到池化后的橙色剩下了1/9，那是从卷积输出层按ROI网格的位置抠出的橙色部分。证据是后面的婴儿图像，看中间那张图，每个婴儿图像为9个位置的热度图，但却只抠了1/9出来，左上角的婴儿热度图对应左上角的卷积层输出，所以抠了左上角的1/9。

Re: Sworddddd 2016-11-15 11:56发表



回复shadow_guo：这个确实是关键，但是如果在每次训练的时候，只取9个层的固定区域，那么每个层也就是对一个固定位置响应比较大

Re: 灰巧克力爱松露 2016-11-15 14:33发表



回复Sworddddd：(o^▽^o)☆[BINGO!]

1楼 神谷姬爱ML 2016-08-30 18:10发表



我还是搞不懂这个选择性pooling具体是怎么算的。。。

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

