

空洞卷积(dilated convolution)理解



realloing

关注

赞赏支持

空洞卷积(dilated convolution)理解



realloing

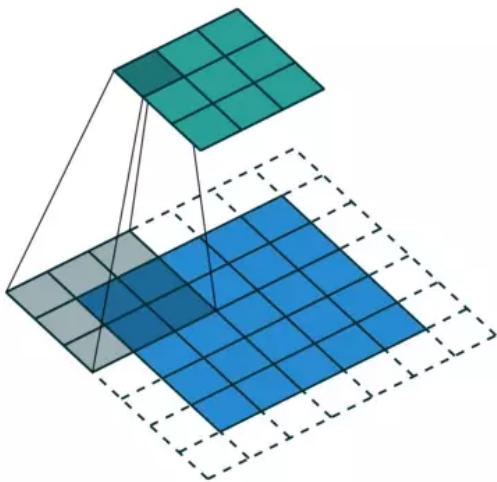
关注



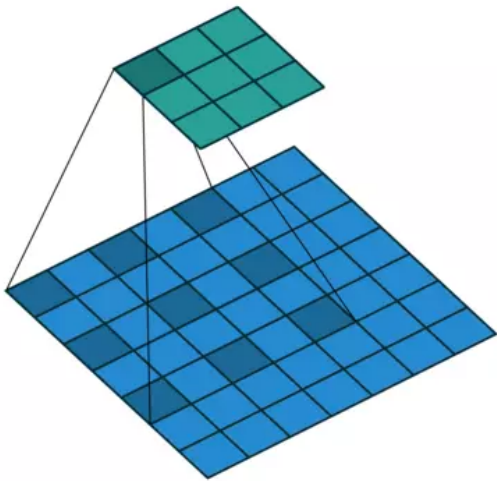
2

2018.10.24 10:13:51 字数 2,075 阅读 76,824

Dilated/Atrous Convolution(中文叫做空洞卷积或者膨胀卷积) 或者是 Convolution with holes 从字面上就很好理解，是在标准的 convolution map 里注入空洞，以此来增加 reception field。相比原来的正常convolution，dilated convolution 多了一个 hyper-parameter 称之为 dilation rate 指的是kernel的间隔数量(e.g. 正常的 convolution 是 dilatation rate 1)。



Standard Convolution with a 3 x 3 kernel (and padding)



Dilated Convolution with a 3 x 3 kernel and dilation rate 2

Deep CNN 对于其他任务还有一些致命性的缺陷。较为著名的是 up-sampling 和 pooling layer 的设计。
主要问题有：



推荐阅读

对卷积神经网络CNN的理解，一文读懂卷积神经网络。

阅读 212

推荐系统召回算法之——图模型 (Personal Rank)

阅读 442

机器学习高频面试题(41道)

阅读 3,716

YOLOv2网络

阅读 375

python用线性回归预测时间序列股票价格

阅读 1,223

广告

小鹅通大班课上线
千万补贴 政策助力商家

写下你的评论...

评论2

赞37

...

空洞卷积(dilated convolution)理解



realloing

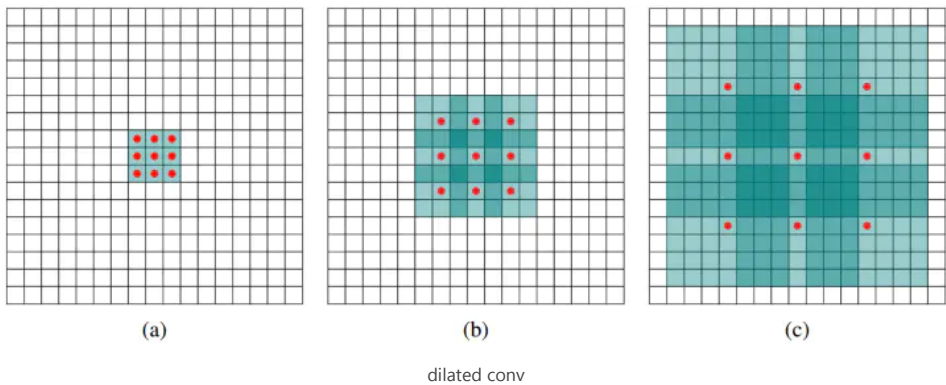
关注

赞赏支持

在这样问题的存在下，语义分割问题一直处在瓶颈期无法再明显提高精度，而dilated convolution的设计就良好的避免了这些问题。

在图像分割领域，图像输入到CNN（典型的网络比如FCN[3]）中，FCN先像传统的CNN那样对图像做卷积再pooling，降低图像尺寸的同时增大感受野，但是由于图像分割预测是pixel-wise的输出，所以要将pooling后较小的图像尺寸upsampling到原始的图像尺寸进行预测（upsampling一般采用deconv反卷积操作，deconv可参见知乎答案[如何理解深度学习中的deconvolution networks?](#)），之前的pooling操作使得每个pixel预测都能看到较大感受野信息。因此图像分割FCN中有两个关键，一个是pooling减小图像尺寸增大感受野，另一个是upsampling扩大图像尺寸。在先减小再增大尺寸的过程中，肯定有一些信息损失掉了，那么能不能设计一种新的操作，不通过pooling也能有较大的感受野看到更多的信息呢？答案就是dilated conv。

下面看一下dilated conv原始论文[4]中的示意图：



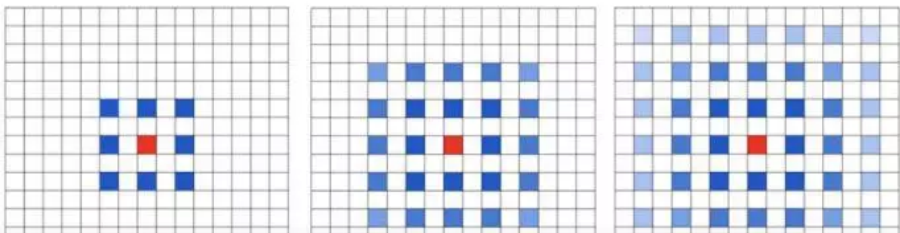
(a)图对应3x3的1-dilated conv，和普通的卷积操作一样，(b)图对应3x3的2-dilated conv，实际的卷积kernel size还是3x3，但是空洞为1，也就是对于一个7x7的图像patch，只有9个红色的点和3x3的kernel发生卷积操作，其余的点略过。也可以理解为kernel的size为7x7，但是只有图中的9个点的权重不为0，其余都为0。可以看到虽然kernel size只有3x3，但是这个卷积的感受野已经增大到了7x7（如果考虑到这个2-dilated conv的前一层是一个1-dilated conv的话，那么每个红点就是1-dilated的卷积输出，所以感受野为3x3，所以1-dilated和2-dilated合起来就能达到7x7的conv），(c)图是4-dilated conv操作，同理跟在两个1-dilated和2-dilated conv的后面，能达到15x15的感受野。对比传统的conv操作，3层3x3的卷积加起来，stride为1的话，只能达到(kernel-1)*layer+1=7的感受野，也就是和层数layer成线性关系，而dilated conv的感受野是指数级的增长。

dilated的好处是不做pooling损失信息的情况下，加大了感受野，让每个卷积输出都包含较大范围的信息。

空洞卷积存在的问题

潜在问题 1: The Gridding Effect

假设我们仅仅多次叠加 dilation rate 2 的 3 x 3 kernel 的话，则会出现这个问题：



写下你的评论...

评论2

赞37

...

小鹅通大班课上线
千万补贴

政策助力商家

广告

空洞卷积(dilated convolution)理解



realloing

关注

赞赏支持

我们发现我们的 kernel 并不连续，也就是并不是所有的 pixel 都用来计算了，因此这里将信息看做 checker-board 的方式会损失信息的连续性。这对 pixel-level dense prediction 的任务来说是致命的。

潜在问题 2: Long-ranged information might be not relevant.

我们从 dilated convolution 的设计背景来看就能推测出这样的设计是用来获取 long-ranged information。然而**光采用大 dilation rate 的信息或许只对一些大物体分割有效果，而对小物体来说可能则有弊无利了**。如何同时处理不同大小的物体的关系，则是设计好 dilated convolution 网络的关键。

通向标准化设计：Hybrid Dilated Convolution (HDC)

对于上个 section 里提到的几个问题，图森组的文章对其提出了较好的解决的方法。他们设计了一个称之为 HDC 的设计结构。

- 第一个特性是，叠加卷积的 dilation rate 不能有大于1的公约数。比如 [2, 4, 6] 则不是一个好的三层卷积，依然会出现 gridding effect。
- 第二个特性是，我们将 dilation rate 设计成锯齿状结构，例如 [1, 2, 5, 1, 2, 5] 循环结构。
- 第三个特性是，我们需要满足一下这个式子：

其中 r_i 是 i 层的 dilation rate 而 M_i 是指在 i 层的最大 dilation rate，那么假设总共有 n 层的话，默认 $M_n = r_n$ 。假设我们应用于 kernel 为 $k \times k$ 的话，我们的目标则是 $M_2 \leq k$ ，这样我们至少可以用 dilation rate 1 即 standard convolution 的方式来覆盖掉所有洞。

一个简单的例子: dilation rate [1, 2, 5] with 3×3 kernel (可行的方案)

image

而这样的锯齿状本身的性质就比较好的来同时满足小物体大物体的分割要求(小 dilation rate 来关心近距离信息，大 dilation rate 来关心远距离信息)。

这样我们的卷积依然是连续的也就依然能满足VGG组观察的结论，大卷积是由小卷积的 regularisation 的叠加。

以下的对比实验可以明显看出，一个良好设计的 dilated convolution 网络能够有效避免 gridding effect。

推荐阅读

对卷积神经网络CNN的理解，一文读懂卷积神经网络。

阅读 212

推荐系统召回算法之——图模型 (Personal Rank)

阅读 442

机器学习高频面试题(41道)

阅读 3,716

YOLOv2网络

阅读 375

python用线性回归预测时间序列股票价格

阅读 1,223



写下你的评论...

评论2

赞37

...

空洞卷积(dilated convolution)理解



realloing

关注

赞赏支持

image

多尺度分割的另类解：Atrous Spatial Pyramid Pooling (ASPP)

在处理多尺度物体分割时，我们通常会有以下几种方式来操作：

image

然仅仅(在一个卷积分支网络下)使用 dilated convolution 去抓取多尺度物体是一个不正统的方法。比方说，我们用一个 HDC 的方法来获取一个大（近）车辆的信息，然而对于一个小（远）车辆的信息都不再受用。假设我们再去用小 dilated convolution 的方法重新获取小车辆的信息，则这么做非常的冗余。

基于港中文和商汤组的 PSPNet 里的 Pooling module（其网络同样获得当年的SOTA结果），ASPP 则在网络 decoder 上对于不同尺度上用不同大小的 dilation rate 来抓去多尺度信息，每个尺度则为一个独立的分支，在网络最后把他合并起来再接一个卷积层输出预测 label。这样的设计则有效避免了在 encoder 上冗余的信息的获取，直接关注与物体之间之内的相关性。

总结

Dilated Convolution 个人认为想法简单，直接且优雅，并取得了相当不错的效果提升。他起源于语义分割，大部分文章也用于语义分割，具体能否对其他应用有价值姑且还不知道，但确实是一个不错的探究方向。有另外的答主提到WaveNet, ByteNet 也用到了 dilated convolution 确实是一个很有趣的发现，因为本身 sequence-to-sequence learning 也是一个需要关注多尺度关系的问题。则在 sequence-to-sequence learning 如何实现，如何设计，跟分割或其他应用的关联是我们可以重新需要考虑的问题。

推荐阅读

图森组和 Google Brain 都对于 dilated convolution 有着更细节的讨论，推荐阅读：

- [Understanding Convolution for Semantic Segmentation](#)
- [Rethinking Atrous Convolution for Semantic Image Segmentation](#)

参考

空洞卷积

推荐阅读

对卷积神经网络CNN的理解，一文读懂卷积神经网络。

阅读 212

推荐系统召回算法之——图模型 (Personal Rank)

阅读 442

机器学习高频面试题(41道)

阅读 3,716

YOLOv2网络

阅读 375

python用线性回归预测时间序列股票价格

阅读 1,223



广告

写下你的评论...

评论2

赞37

...

空洞卷积(dilated convolution)理解



realloing

关注

赞赏支持

赏员又行

还没有人赞赏，支持一下



realloing Gmail:liaopanblog@gmail.com
总资产5 (约0.43元) 共写了1.8W字 获得103个赞 共48个粉丝

关注

推荐阅读

Deformable Convolutional Networks论文翻译——中英文对照

文章作者: Tyan|博客: noahsnail.com | CSDN | 简书 声明: 作者翻译论文仅为学习, 如有侵权请...

SnailTyan 阅读 2,894 评论 1 赞 3

卷积神经网络中不同类型的卷积方式介绍

之前在文章《<模型汇总_1>牛逼的深度卷积神经网络CNN》详细介绍了卷积神经网络的基本原理, 以及常见的基本模型, 如...

Arthurchsh 阅读 2,412 评论 0 赞 9

OCR如何读取皱巴巴的文件? 深度学习在文档图像形变矫正的...

一、背景 随着集团业务的高速发展以及集团对用户群体信用要求的提高, 证件审核成为业务中必不可少的一个环节。譬如: 支付...

阿里云云栖号 阅读 1,677 评论 0 赞 7

变形卷积核、可分离卷积? 卷积神经网络中10大拍案叫绝的操作

姓名: 周雪宁 学号: 1702110196 转载: https://mp.weixin.qq.com/s/4-9SHF...

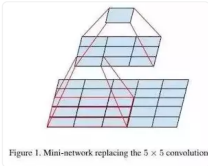
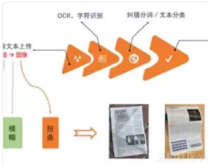
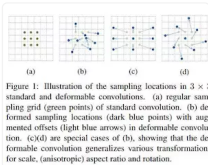
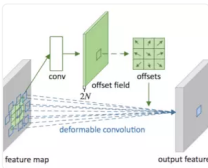
周雪宁 阅读 1,751 评论 1 赞 1

蜜糖罐日记---CC用心记录生活点滴 (433)

2018年6月26日 星期二 阵雨 (农历二零一八年五月十三) 女儿日记摘录:
(6/23) 今天这篇日记是我用钢笔...

去油可可豆 阅读 356 评论 5 赞 18

更多精彩内容>



推荐阅读

对卷积神经网络CNN的理解, 一文读懂卷积神经网络。

阅读 212

推荐系统召回算法之——图模型 (Personal Rank)

阅读 442

机器学习高频面试题(41道)

阅读 3,716

YOLOv2网络

阅读 375

python用线性回归预测时间序列股票价格

阅读 1,223

小鹅通大班课上线
千万补贴 政策助力商家
广告

写下你的评论...

评论2

赞37

...