Dilated Convolution

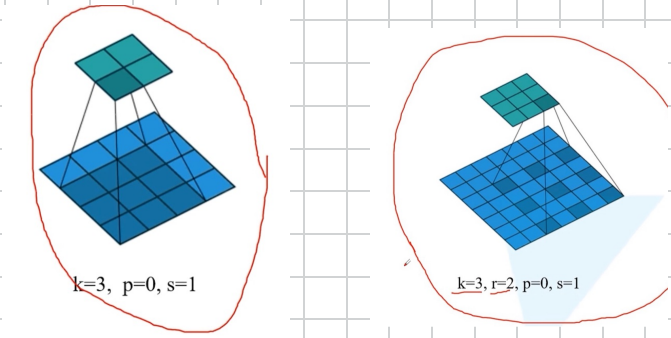
240906

(1) 也称 04法 Atrous convolution

(2) function { (1) 增大感受野
(2) 保持原输入特征图 W、H

✓ 什么是 dilated convolution?

○ How dilated Convolution? (怎么使用)



(3) Why Dilated Convolution? 为什么使用膨胀卷积?

① 在语义分割中，如FCN 通常使用分类网络 作为网络的backbone 使用backbone 对图片进行一系列的下采样，再使用一系列的上采样还原原图的大小

② 通常使用的分类网络将高度&宽度下采样 32倍，后面再使用上采样还原原图尺寸，而下采样倍率太大 我们再还原为原图有很大影响 如 VGG 由 maxpooling layer 池化、下采样 首先降低图片高&宽度 其次会失细节信息 以及比较小的目标 这些丢失的细节及目标 无法通过上采样还原 这会导致语义分割效果不理想

③ 那去掉 maxpooling 呢? 确实不会降低高&宽度了。但这样的话 特征图对应的感受野就变小了

④ SO dilated convolution

(4) 既然如此，那全用 dilated Convolution?

Ans = no!

<< Understanding convolution for Semantic Segmentation >>

gridding effect

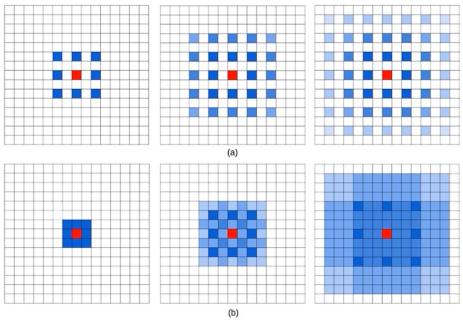
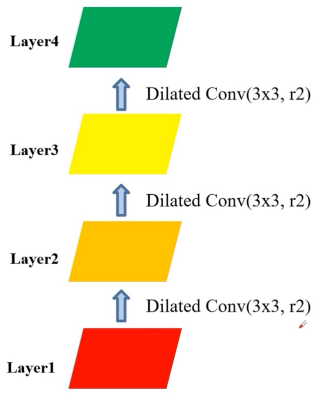


Figure 2. Illustration of the gridding problem. Left to right: the pixels (marked in blue) contributes to the calculation of the center pixel (marked in red) through three convolution layers with kernel size 3×3. (a) all convolutional layers have a dilation rate $r = 2$. (b) subsequent convolutional layers have dilation rates of $r = 1$, 2, 3, respectively.

Experiment 1

(1) 什么是 gridding effect?

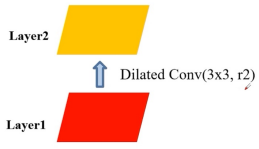


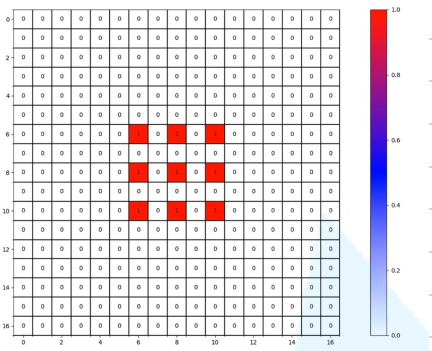
3个 3×3 的 dilated convolution

r=2

r=1 ordinary convolution

(1) 首先，看 Layer1 ⟶ Layer2 用到了 Layer1 的哪些元素

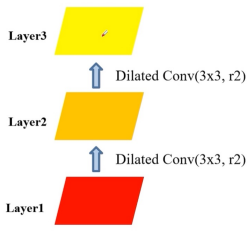


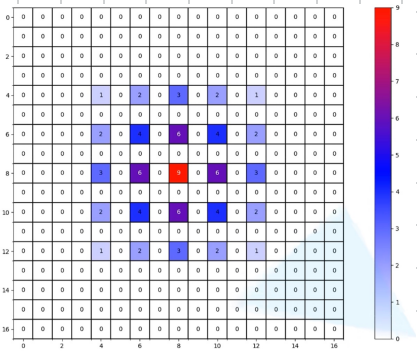


每2行 & 2列之间都间隔了1行/1列0

用到了 layer1 中 9个 pixel 信息

(2) 当我们连续使用3个 dilated convolution layer3 使用了 layer1 哪些元素?





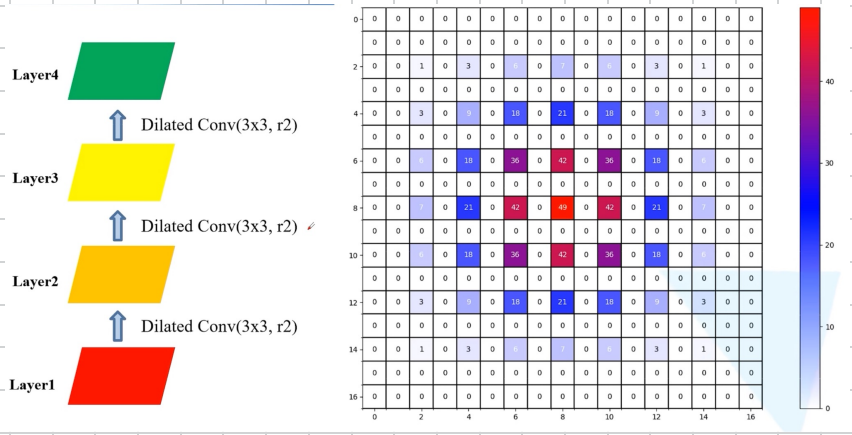
用到了 layer1 中 25个 pixel 信息

颜色越红 次数越多

colorbar

(3) when 我们连续堆叠3个 rate=2的 dilated convolution

layer4 用到的 layer1 上的 pixel



RF = 13x13

RF = 13x13

但有很多 pixel 未被使用

相比实验2

高亮区域使用低层的连续区域

(1) 首先 layer4 用到的 layer1 上的 pixel 并不是连续的

每个非零元素之间都有一定的间隔

↕

gridding effect

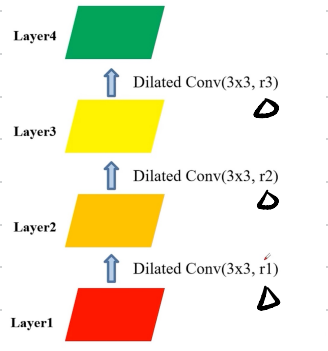
i.e. layer4 并没有用到 layer1 范围内的所有像素值

只用到了其中一部分 leading to 丢失像素信息

so 我们应尽量避免这样的情况

now 看另外一组实验

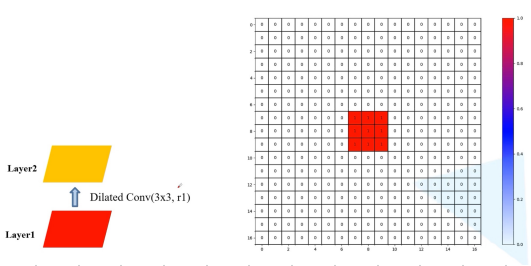
Experiment 2



- 观察 膨胀系数 r=1、2、3

r=1 ordinary convolution

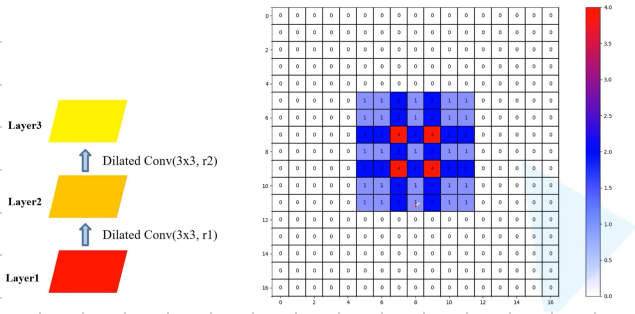
(1) Layer2 使用了 Layer1 的哪些 pixel?



(1) 用到了9个 pixel

(2) 且连续

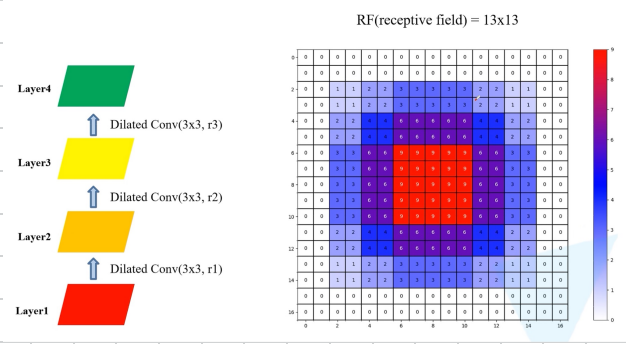
(2) Layer3 上的 pixel 对应 Layer1 上的哪些 pixel?



(1) 1个7x7的 region

(2) 且连续

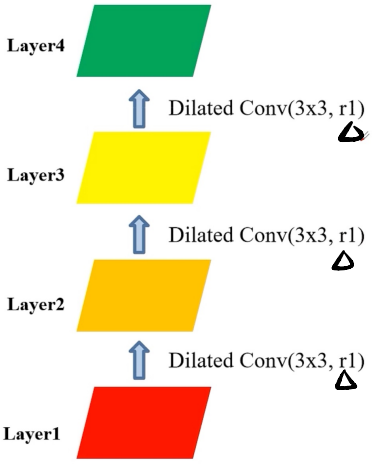
(3) Layer4 上的 pixel 对应 Layer1 上的哪些 pixel?



(1) RF = 13x13

(2) 连续 没有间隙

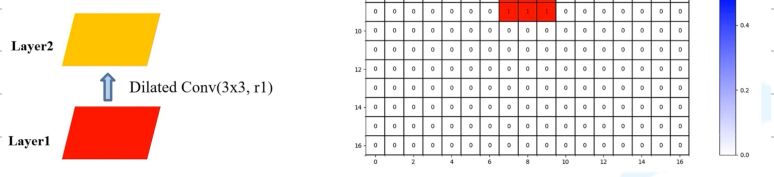
Experiment 3



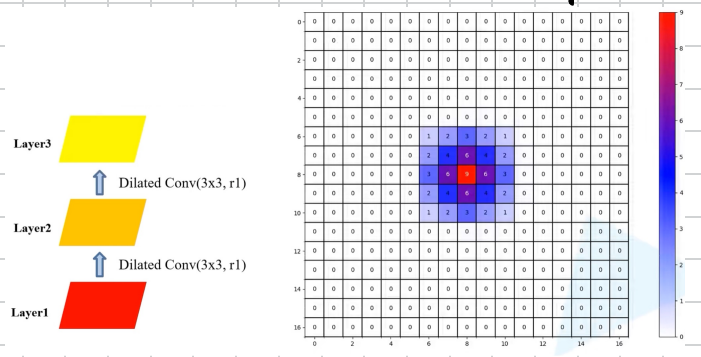
当我们全部 r=1

连续使用3个 3x3 的 ordinary conv

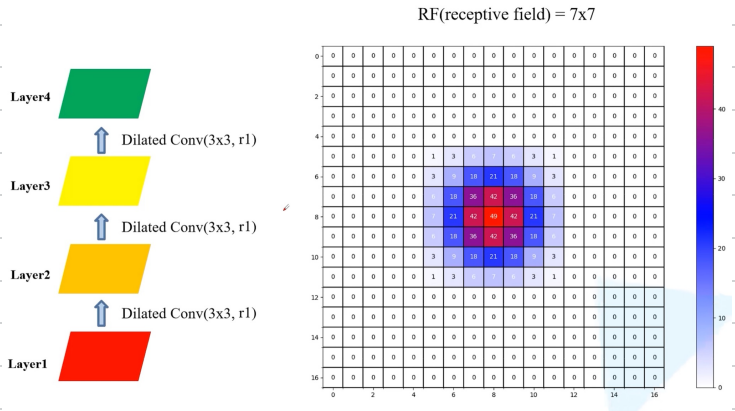
(1) Layer2 使用 Layer1 哪些 pixel, RF=?



(2) Layer3 使用 Layer1 { pixel RF



(3) Layer4 使用 Layer1



Layer4 RF = 7x7

首先 注意到 3组实验 全部是 3x3 的 conv ⟹ 参数数量是一样的

case1: r=1 —3个3x3 conv→ RF = 7x7

case2: r=1、2、3 ⟶ RF = 13x13

且连续 consecutive

summary: Hybrid Dilated Convolution

感受野可以变大且不损失信息

ref <<Understanding convolution for semantic segmentation>>

讨论如何设置 dilated rate

(6) About HDC & dilated rate 如何设计 dilated rate?

Understanding Convolution for Semantic Segmentation

Hybrid Dilated Convolution (HDC)

Here we propose a simple solution- *hybrid dilated convolution (HDC)*, to address this theoretical issue. Suppose we have N convolutional layers with kernel size $K \times K$ that have dilation rates of $[r_1, ..., r_i, ..., r_n]$, the goal of HDC is to let the final size of the RF of a series of convolutional operations fully covers a square region without any holes of missing edges. We define the "maximum distance between two nonzero values" as

$M_2 <= K$

←HDC的目的

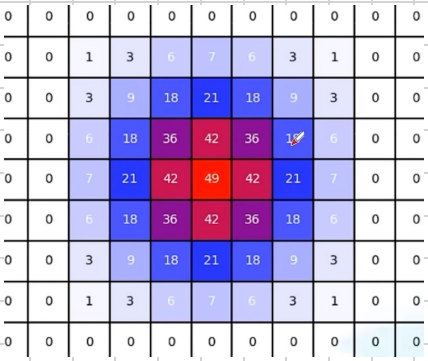
$$M_i = max[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i], \quad (2)$$

with $M_n = r_n$. The design goal is to let $M_2 \leq K$. For example, for kernel size $K = 3$, an $r = [1, 2, 5]$ pattern works as $M_2 = 2$; however, an $r = [1, 2, 9]$ pattern does not work as $M_2 = 5$. Practically, instead of using the same dilation rate for all layers after the downsampling occurs, we

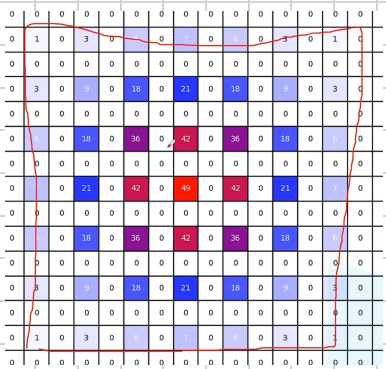
← HDC Target

(第一个 r = 1，r = 1为普通卷积)

① 厘清 什么是 maximum distance



maximum distance = 1



maximum distance = 2

间隔 1 行 / 1 列的零

② 看公式

相反数

$$M_i = \max[M_{i+1} - 2r_i,\ M_{i+1} - 2(M_{i+1} - r_i),\ r_i]$$

meaning:

M_i: 第 ith 层 两个非0元素 (pixel) 之间最大距离 (maximum distance)

r_i: ith 的 dilated rate

M_n: 最后1层 maximum distance $= r_n$ n th dilated distance

区分 M_i & r_i

Target: $M_2 \le K$ K: kernel_size

问题: 那 Tm 是 M_i or r_i?

For example (举例说明)

eg. kernel_size = 3 $r = [1, 2, 5]$ $M_2 = 2$ $M_n = M_3 = 5$

$M_2 = 2 \le K =$ kernel_size $= 3$

Now discuss M_2 怎么来的?

公式 $M_i = \max[M_{i+1} - 2r_i,\ M_{i+1} - 2(M_{i+1} - r_i),\ r_i]$

$M_2 = \max[M_3 - 2r_2,\ M_3 - 2(M_3 - r_2),\ r_2]$

$M_3 = r_3 = 5$ $(M_n = r_n)$ $\Rightarrow M_3 - 2r_2 = 5 - 4 = 1$

$r_2 = 2$

$M_3 - 2(M_3 - r_2) = 5 - 2(5-2) = -1$

$r_2 = 2$

$\therefore M_2 = 2$

me: r 是我们设计的 设计出来的 r (决定 M)

s.t. $M_2 \le K$ (先有 r 再有 M)

eg2. In: $r=[1,2,9]$ Out: $M_2=?$

deal:

$M_T=\max[M_{T+1}-2r_T, M_{T+1}-2(M_{T+1}-r_T), r_T]$

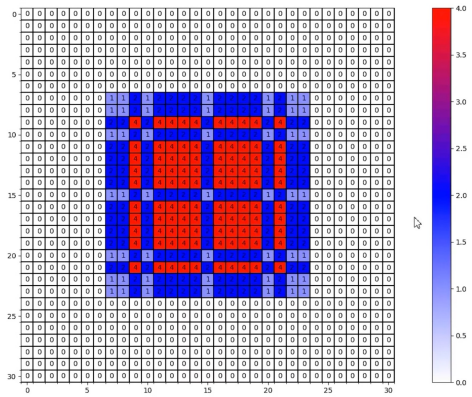
$M_2=\max[M_3-2r_2, M_3-2(M_3-r_2), r_2]$

$M_3=r_3=9$ $M_3-2r_2=9-4=5$
$r_2=2$

$M_3=9$
$M_3-r_2=9-2=7$ $M_3-2(M_3-r_2)=9-14=-5$
$r_2=2$

∴ $M_2=5>K=3$ 不满足设计要求 这组参数不合适

eg1 & eg2 visualization observe 高层conv用到低层conv哪些pixel

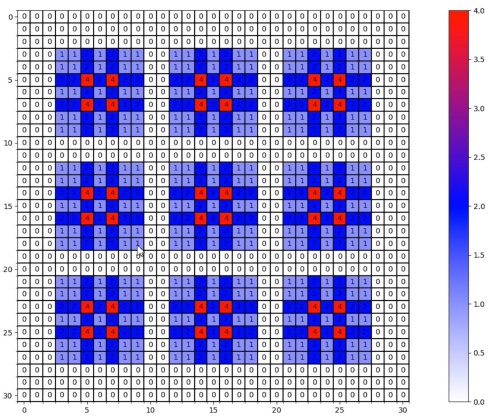


$r=[1,2,5]$

整个区域所有pixel

consecutive

没有 gridding effect 问题



$r=[1,2,9]$

从图中可以看到 maximum distance = 3

i.e. row/column 之间有 2行 12列 0

由 eg2 $M_2=5$

现求 M_1

$M_1=\max[M_2-2r_1, -(M_2-2r_1), r_1]$

$M_2-2r_1=5-2\times1=3, -3, 1$

∴ $M_1=3=$ maximum distance

(7) discuss [1,2,5] [1,2,9] 为什么都从1开始?

r=1 ⇔ ordinary convolution

(8) principle1 : $M_2 \le k$

principle2 dilation rate 锯齿结构 重复

Understanding Convolution for Semantic Segmentation

Hybrid Dilated Convolution (HDC)

use a different dilation rate for each layer. In our network, the assignment of dilation rate follows a sawtooth wave-like heuristic: a number of layers are grouped together to form the "rising edge"of the wave that has an increasing dilation rate, and the next group repeats the same pattern. For example, for all layers that have dilation rate $r = 2$, we form 3 succeeding layers as a group, and change their dilation rates to be 1, 2, and 3, respectively. By doing this, the top layer can access information from a broader range of pixels, in the same region as the original configuration (Figure 2 (b)). This process is repeated through all layers, thus making the receptive field unchanged at the top layer.

锯齿形状

将dilation rates设置成锯齿结构，例如：[1, 2, 3, 1, 2, 3]

{ 锯齿结构
公约数≤1

(9) principle3 公约数≤1

Understanding Convolution for Semantic Segmentation

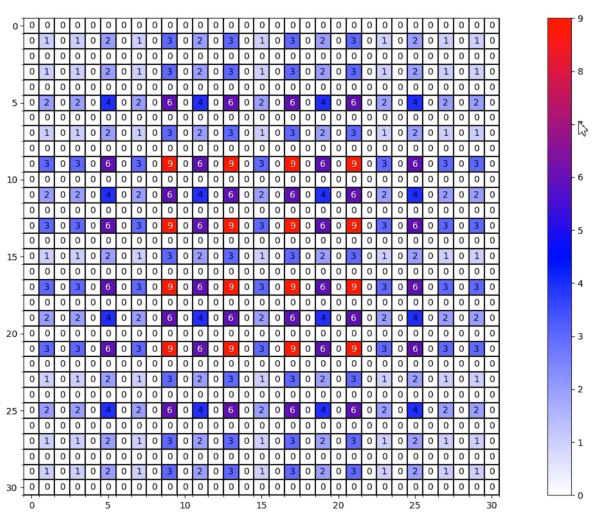
Hybrid Dilated Convolution (HDC)

Another benefit of HDC is that it can use arbitrary dilation rates through the process, thus naturally enlarging the receptive fields of the network without adding extra modules [29], which is important for recognizing objects that are relatively big. One important thing to note, however, is that the dilation rate within a group should not have a common factor relationship (like 2,4,8, etc.), otherwise the gridding problem will still hold for the top layer. This is a key difference between our HDC approach and the atrous spatial pyramid pooling (ASPP) module in [3], or the context aggregation module in [29], where dilation factors that have common factor relationships are used. In addition, HDC is naturally integrated with the original layers of the network, without any need to add extra modules as in [29, 3].

公约数不能大于1

otherwise gridding problem

visulization: r=[2,4,8]



Summary HDC 3个 principle
(1) $N_D \leq K$
(2) 锯齿状
(3) 公约数 ≤ 1