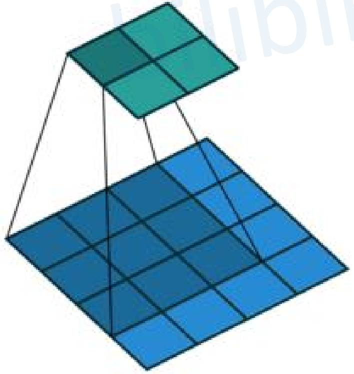
Dilated Convolution

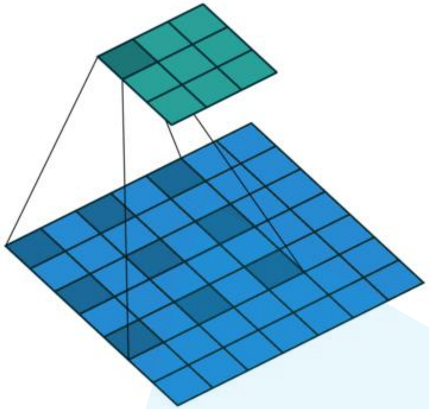
Dilated convolution

(Atrous convolution)

- 增大感受野
- 保持原输入特征图W、H



k=3, p=0, s=1



k=3, r=2, p=0, s=1

Dilated Convolution

Understanding Convolution for Semantic Segmentation

gridding effect

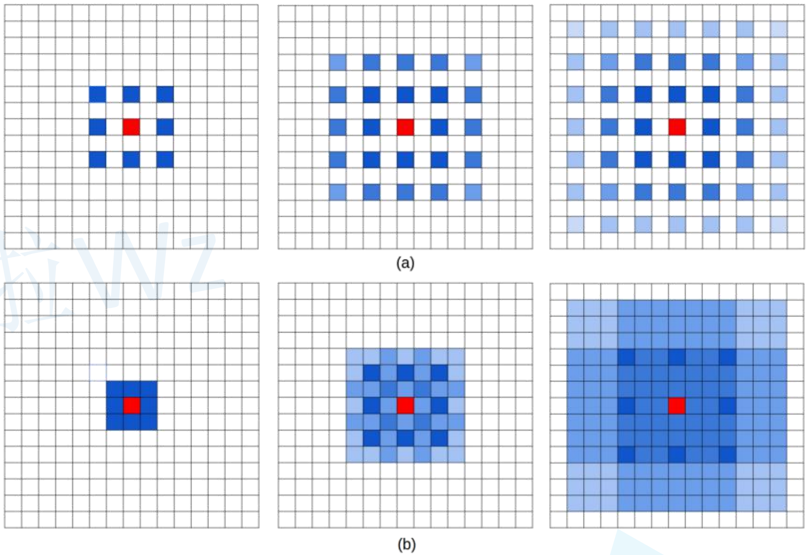
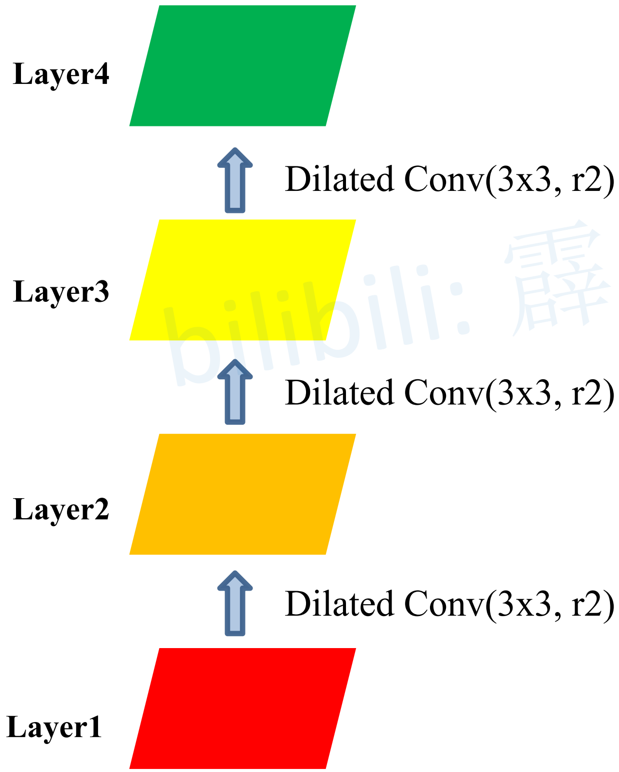


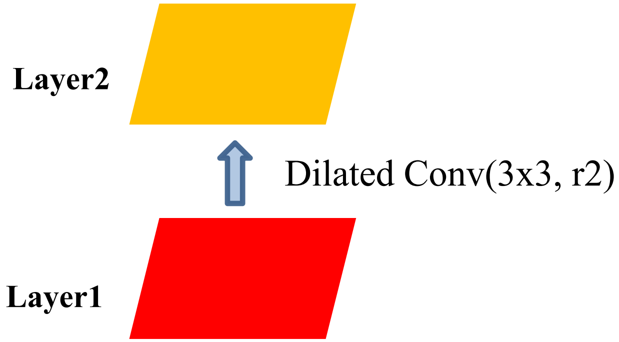
Figure 2. Illustration of the gridding problem. Left to right: the pixels (marked in blue) contributes to the calculation of the center pixel (marked in red) through three convolution layers with kernel size 3×3. (a) all convolutional layers have a dilation rate $r = 2$. (b) subsequent convolutional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

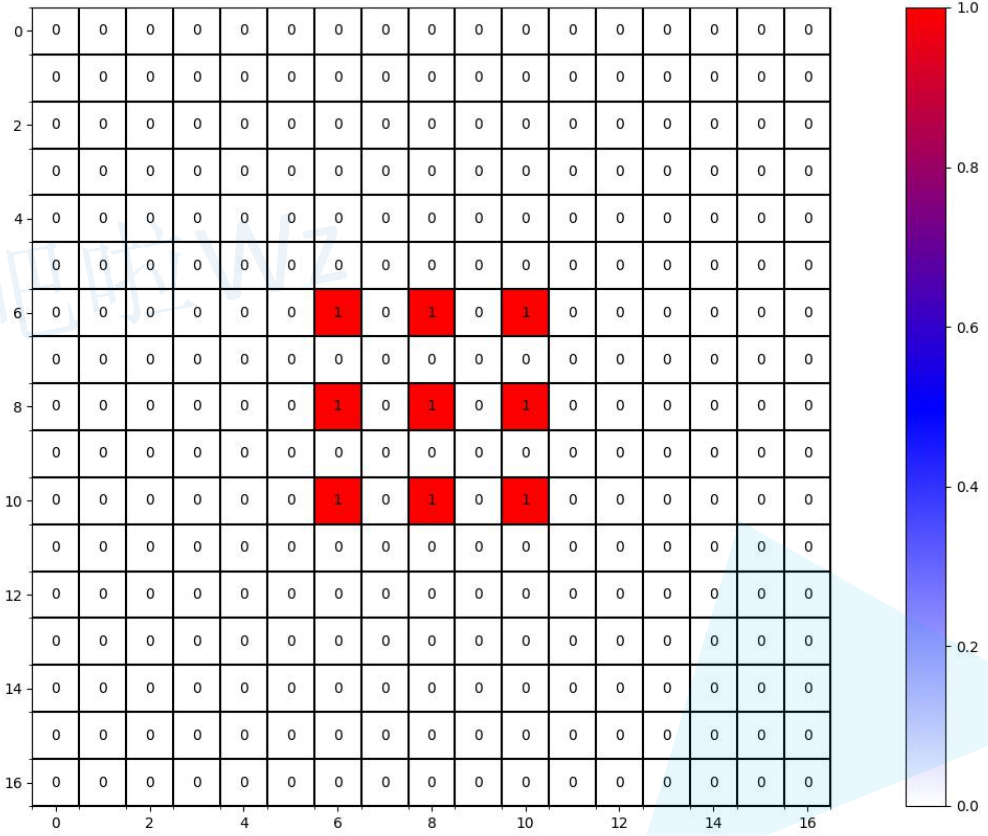
https://arxiv.org/abs/1702.08502

Dilated Convolution

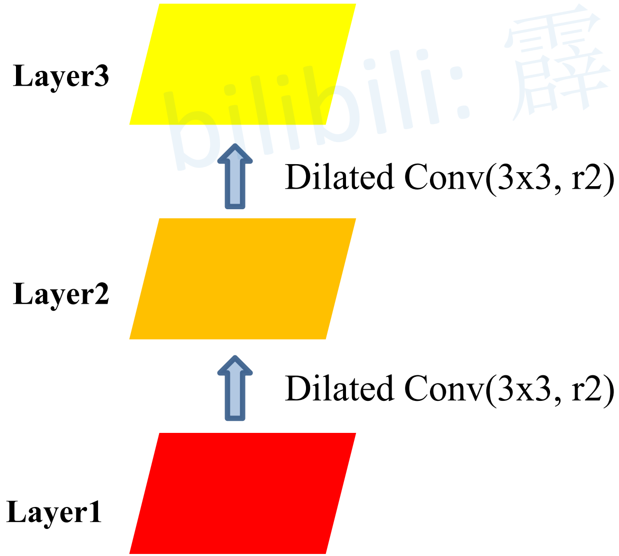


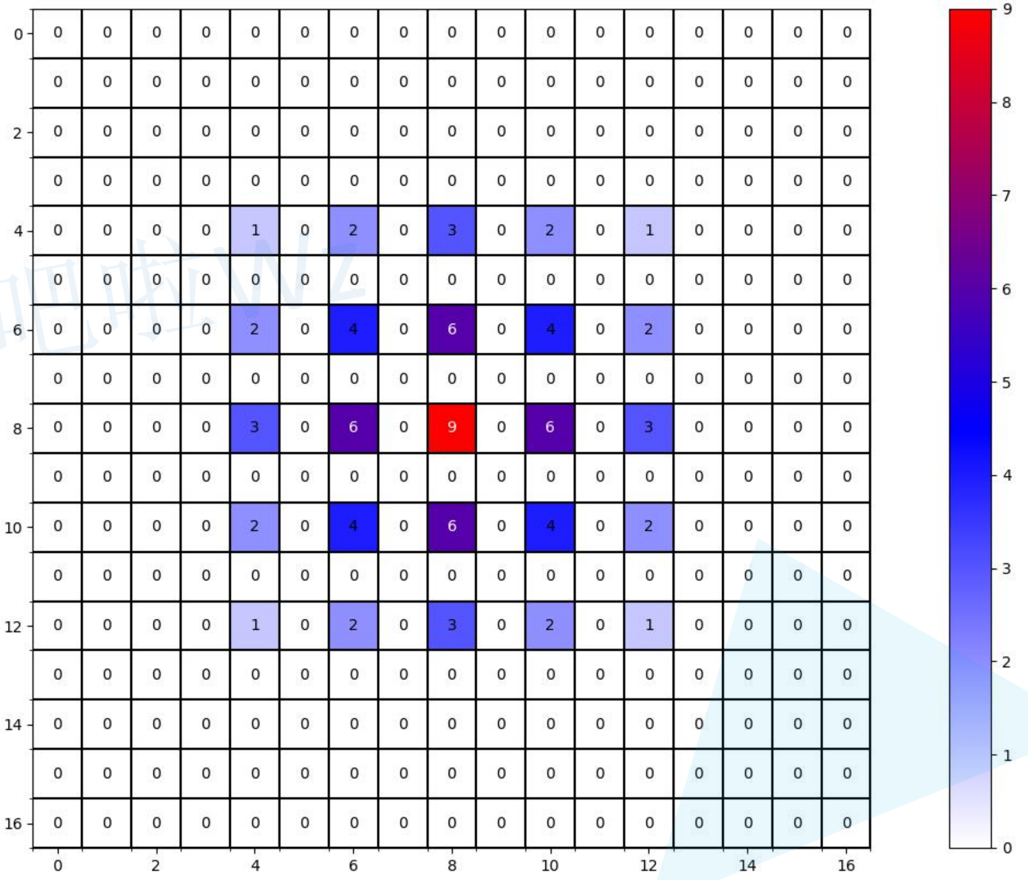
Dilated Convolution



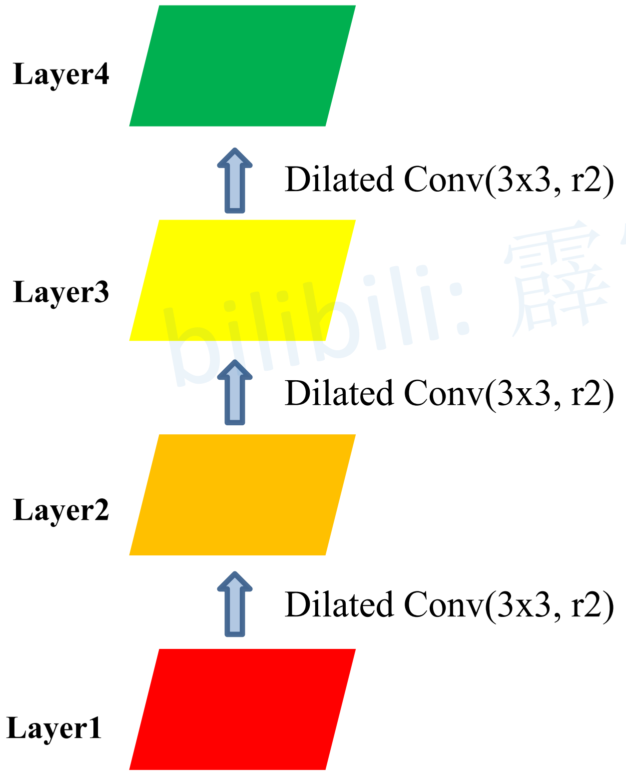


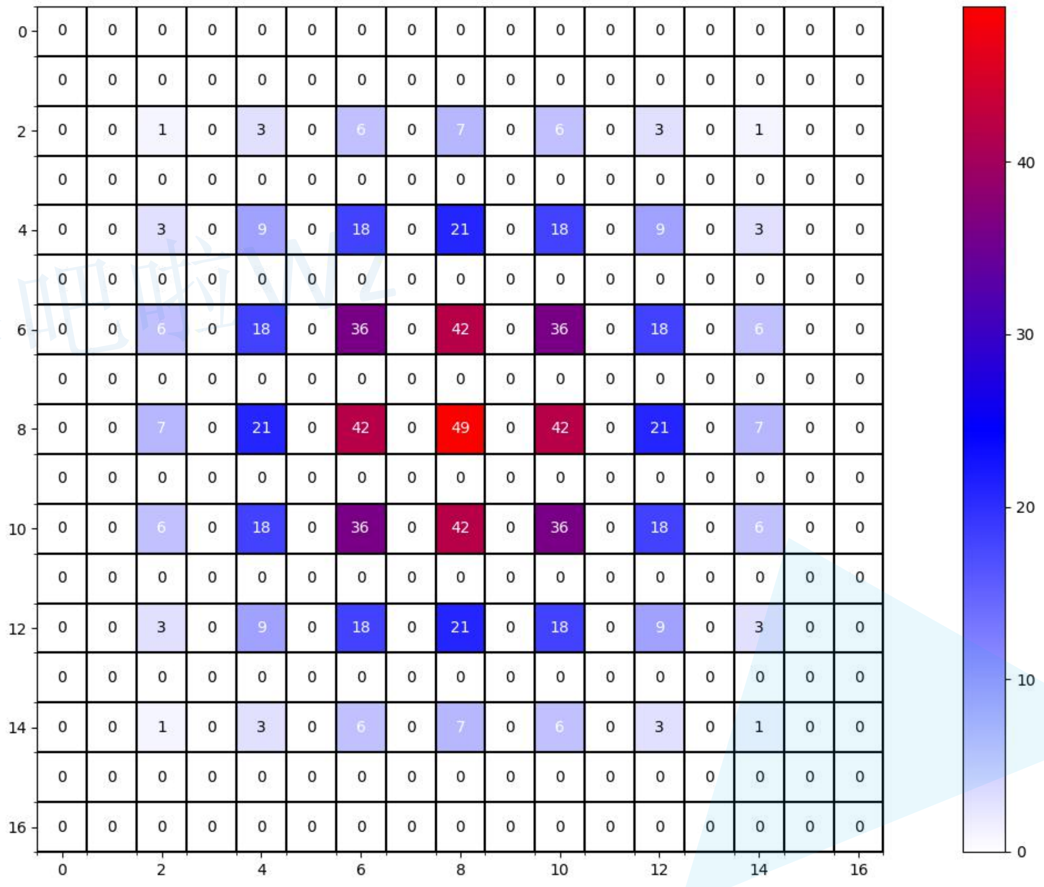
Dilated Convolution



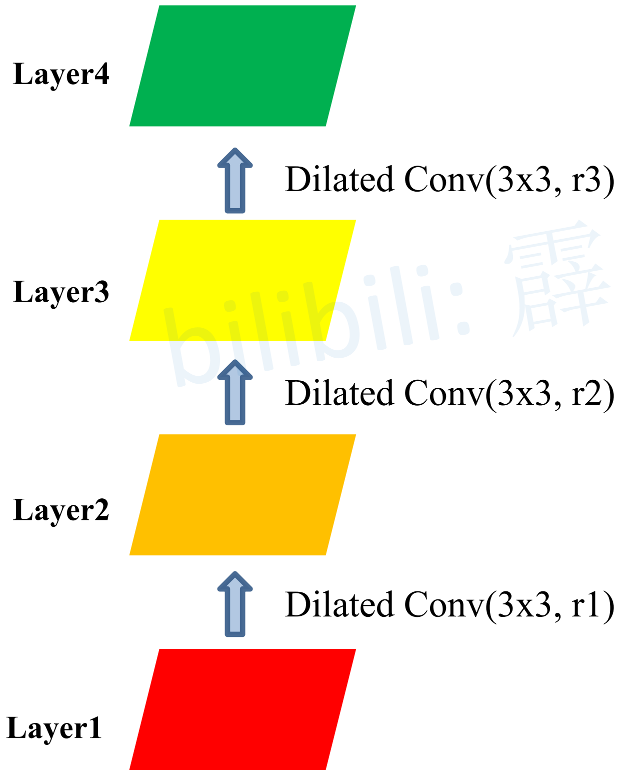


Dilated Convolution

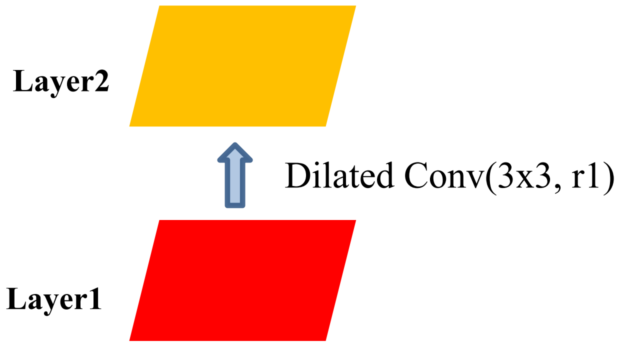


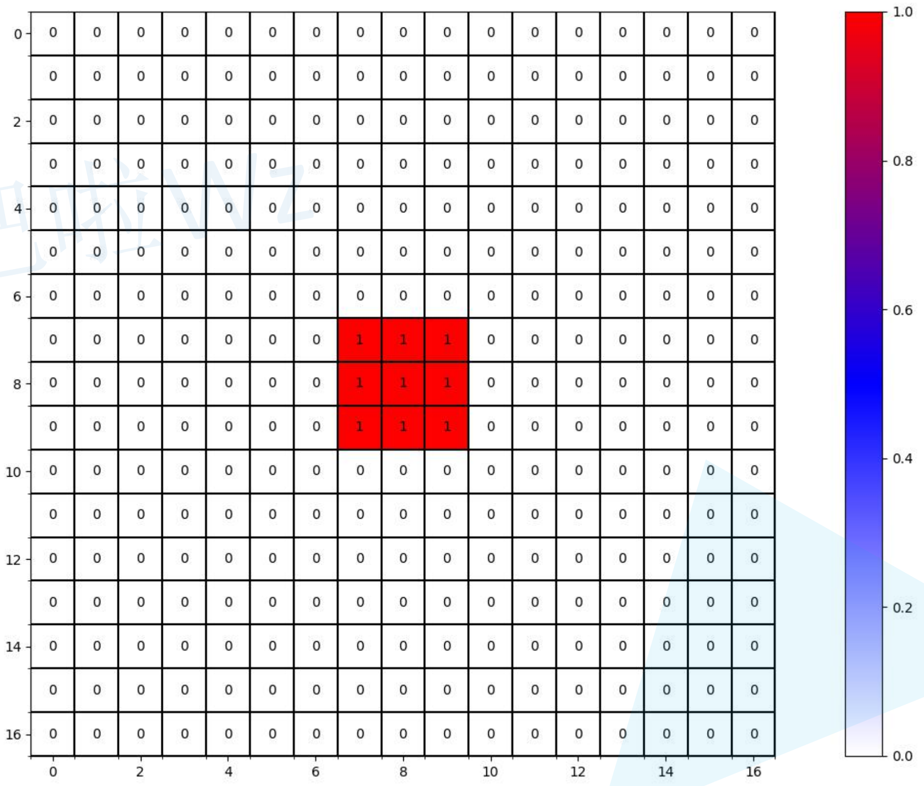


Dilated Convolution

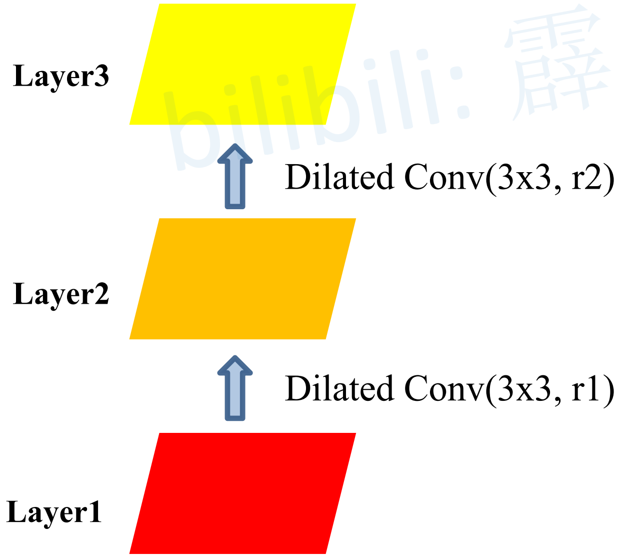


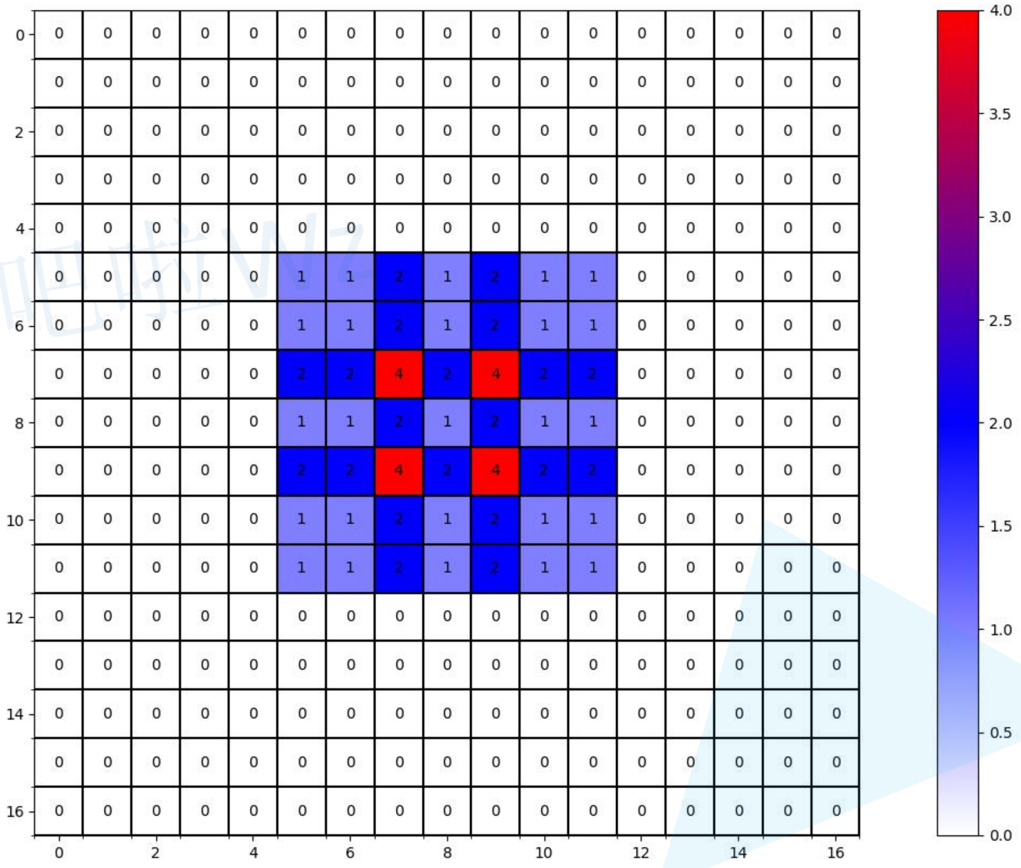
Dilated Convolution



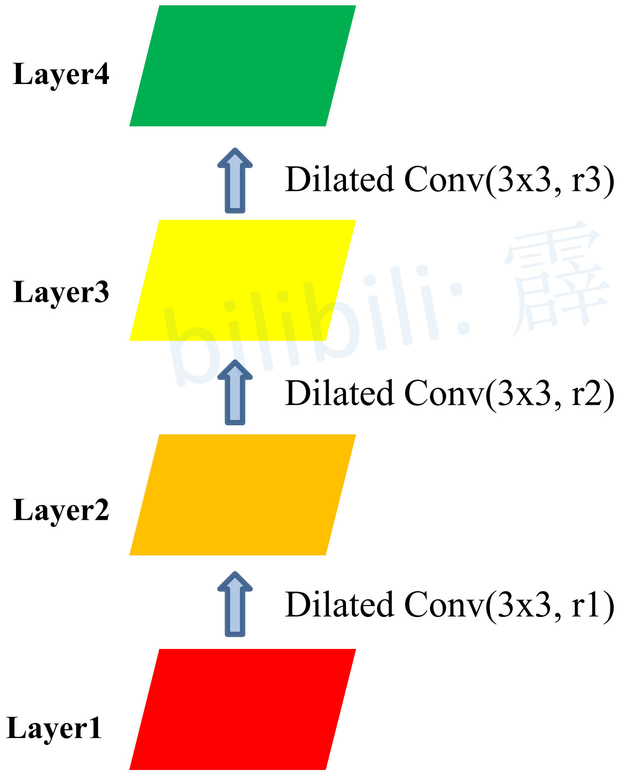


Dilated Convolution

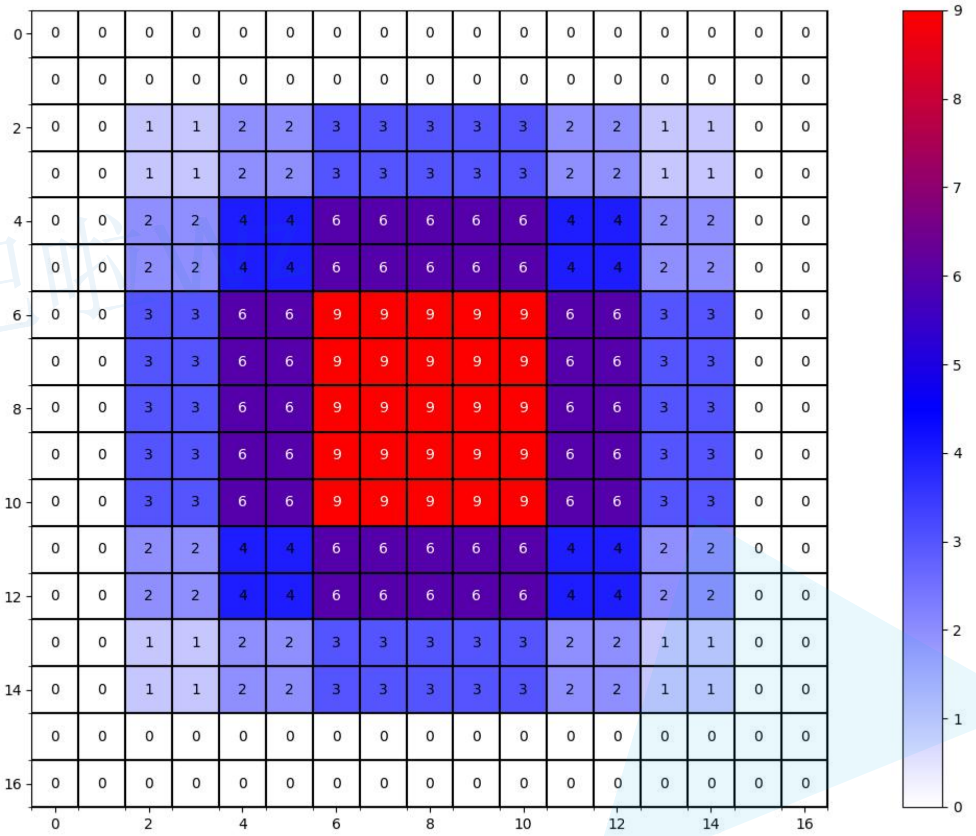




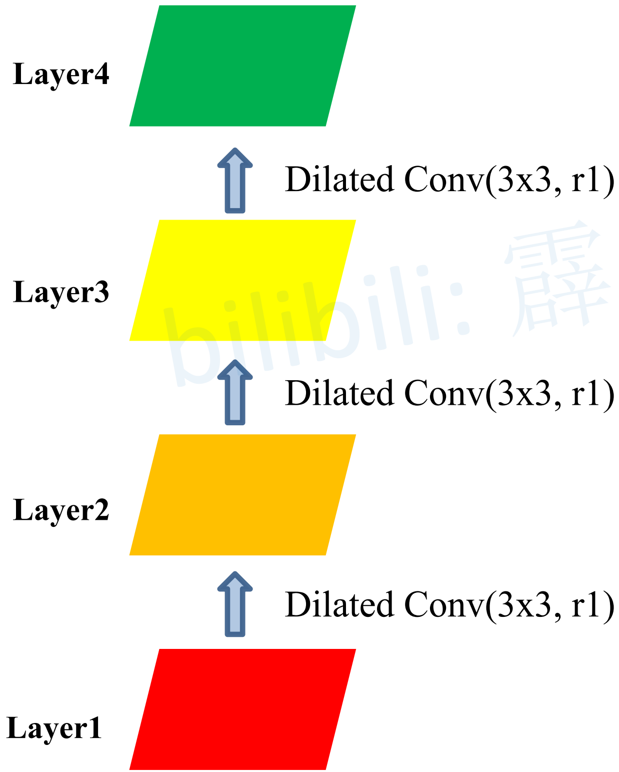
Dilated Convolution



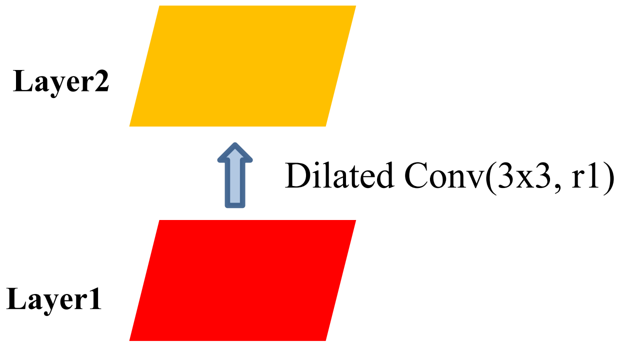
RF(receptive field) = 13x13

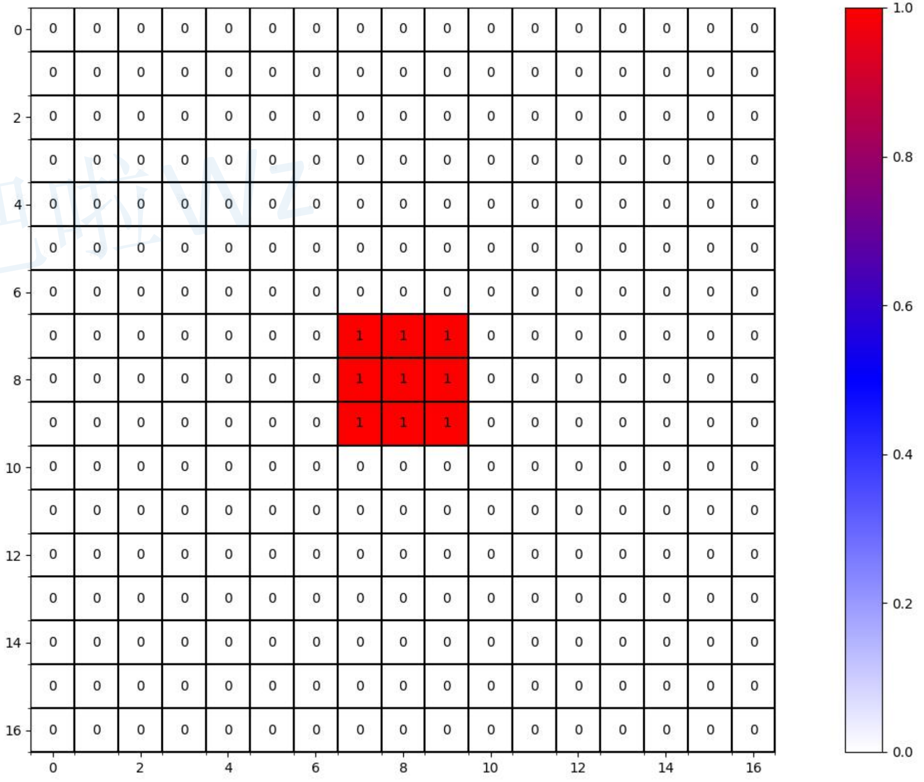


Dilated Convolution

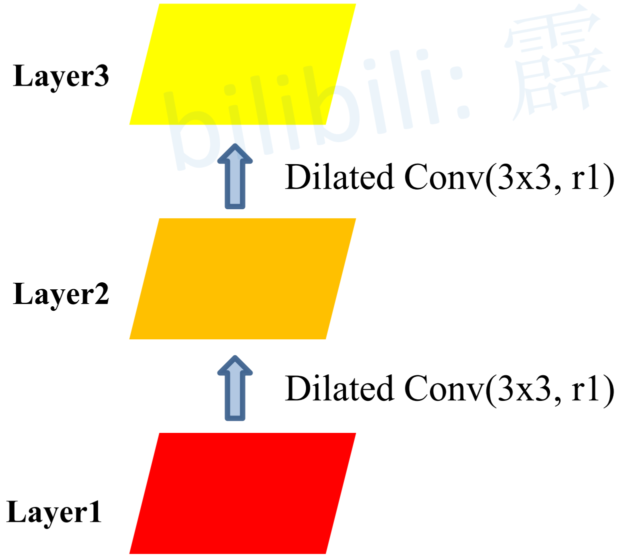


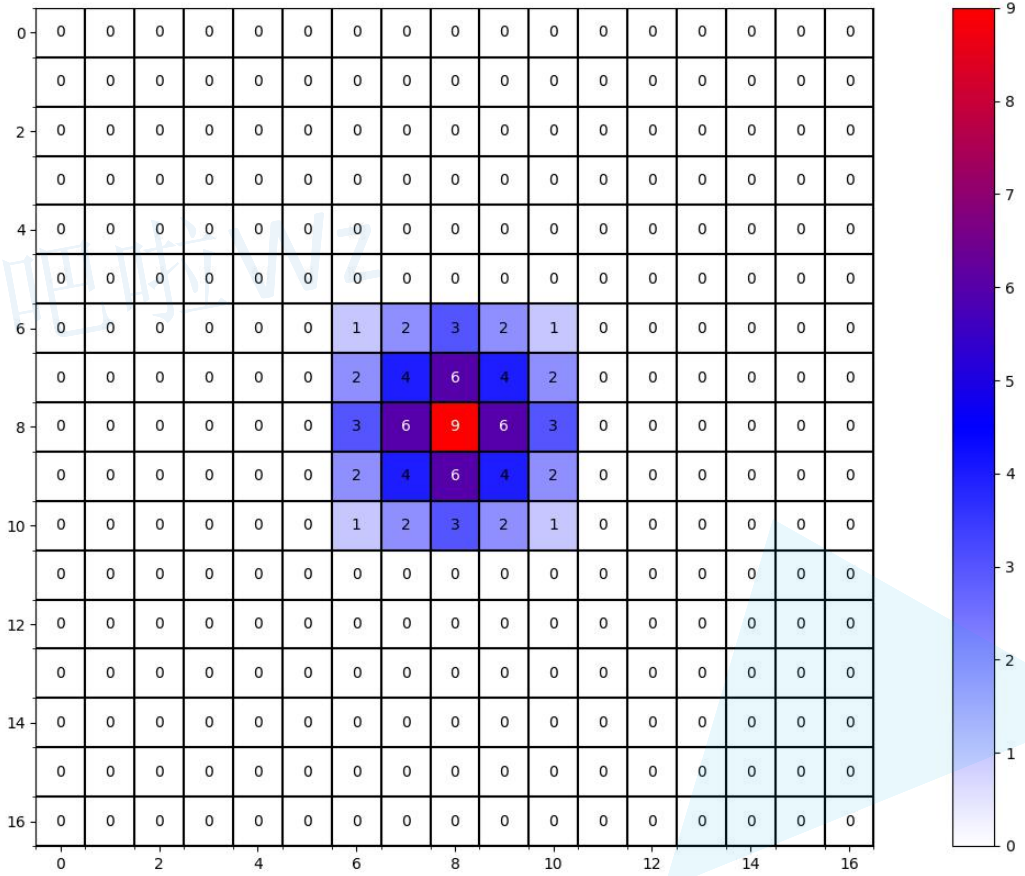
Dilated Convolution



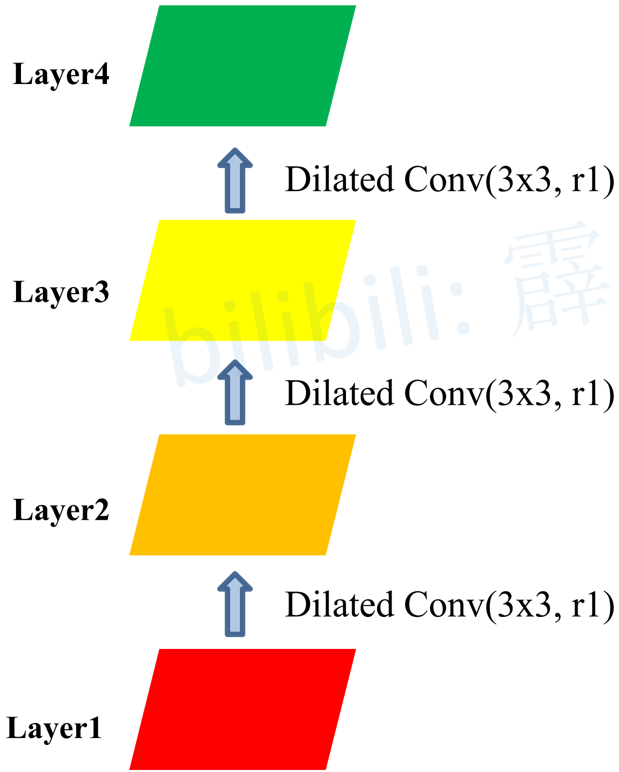


Dilated Convolution

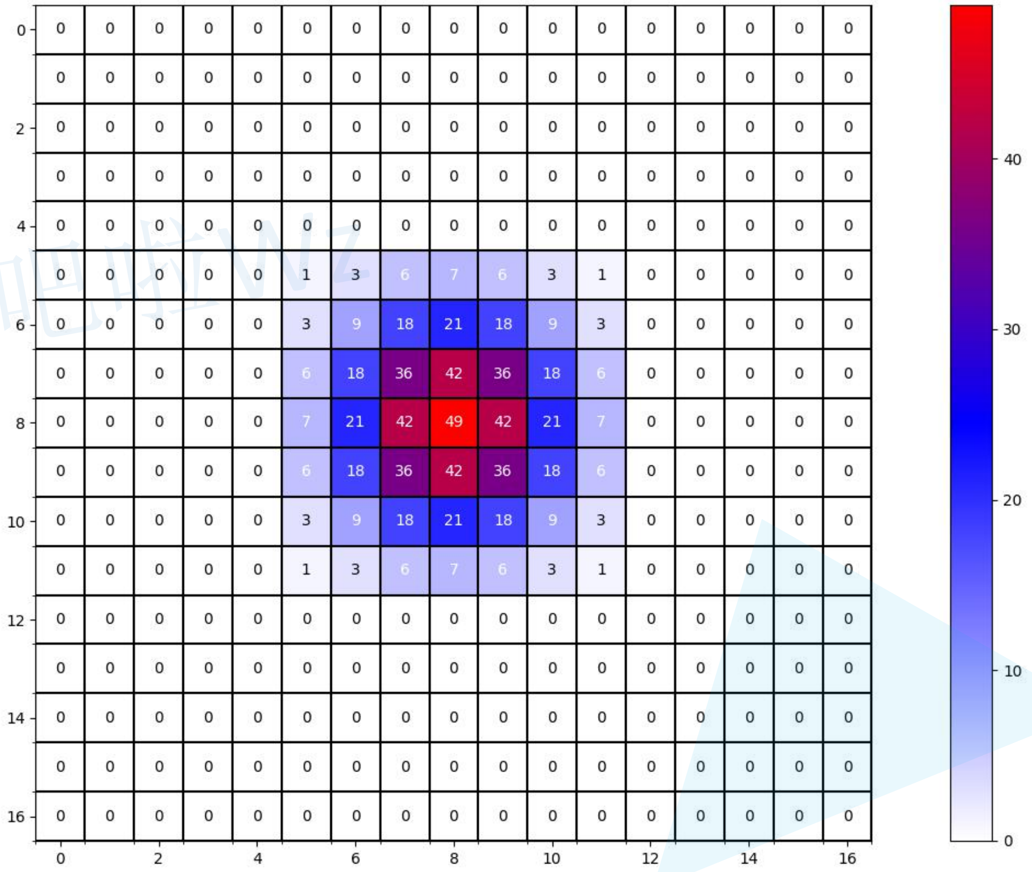




Dilated Convolution



RF(receptive field) = 7x7



Dilated Convolution

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 or missing edges. We define the "maximum distance between two nonzero values" as

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

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

$M_2 <= K$

(第一个 r = 1， r = 1为普通卷积)

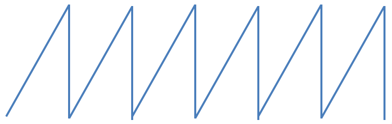
Dilated Convolution

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.

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



Dilated Convolution

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

Dilated Convolution

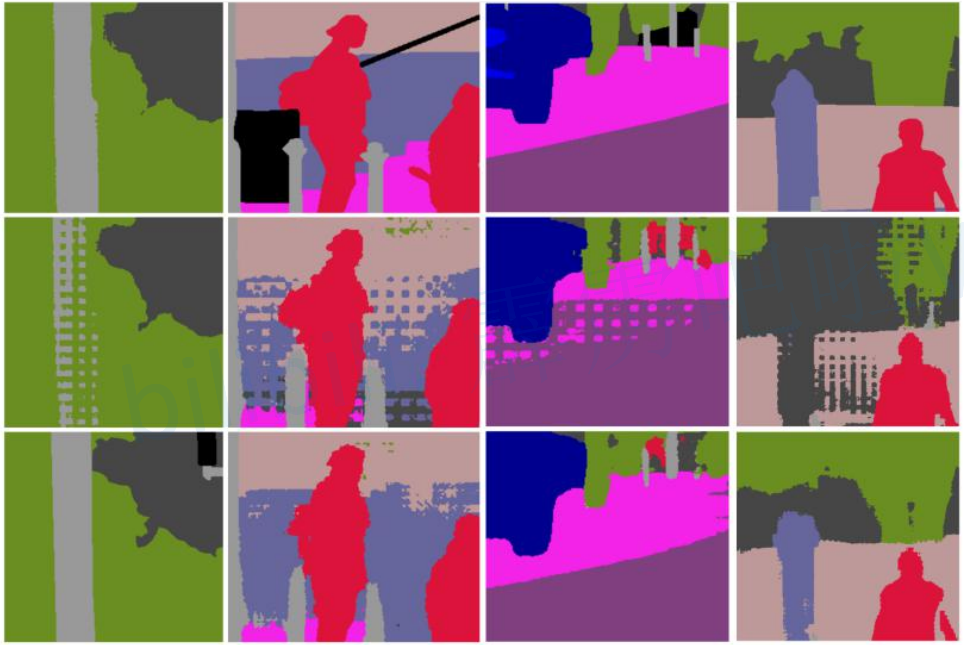


Figure 5. Effectiveness of HDC in eliminating the gridding effect. First row: ground truth patch. Second row: prediction of the ResNet-DUC model. A strong gridding effect is observed. Third row: prediction of the ResNet-DUC-HDC (Dilation-RF) model.

Hybrid Dilated Convolution (HDC)