

《新一代人工智能：从深度学习到大模型》

常见卷积操作和卷积神经网络

主讲人 张重生





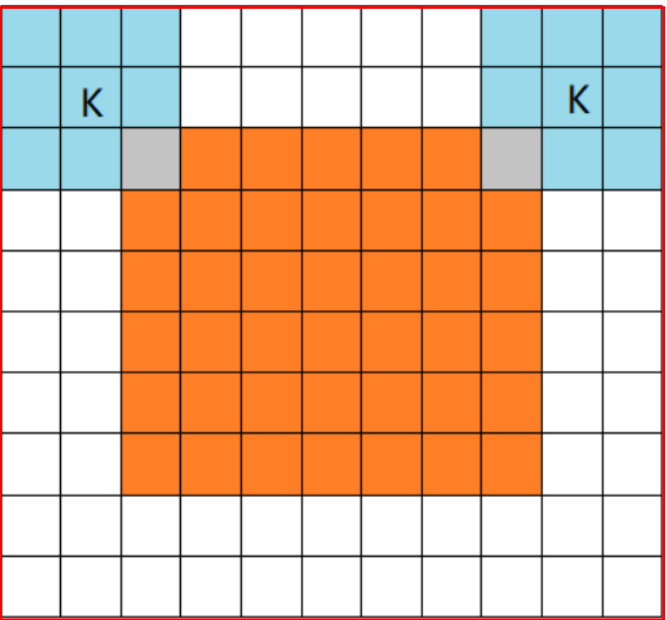
Part 05-1

各种类型的卷积操作

讲解各种常见的卷积操作的原理及其特点

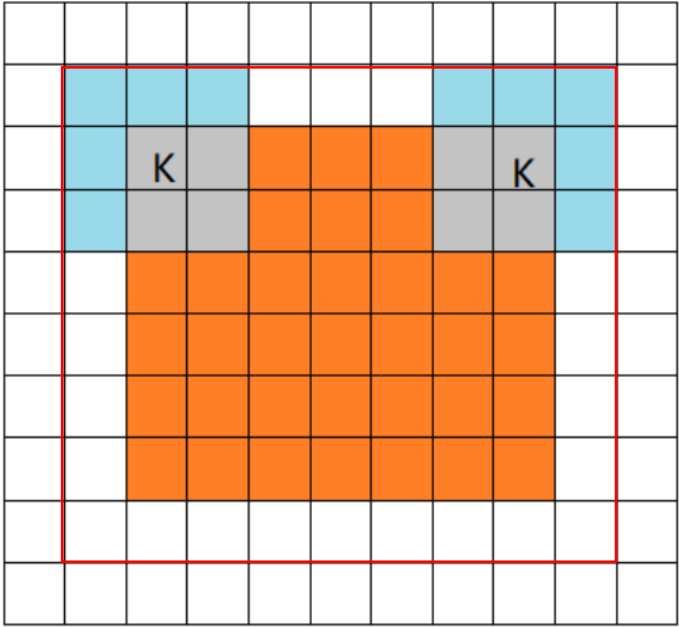
0、卷积操作的三种模式及输出矩阵尺寸计算（1）

卷积的三种padding模式： Full, Same, Valid



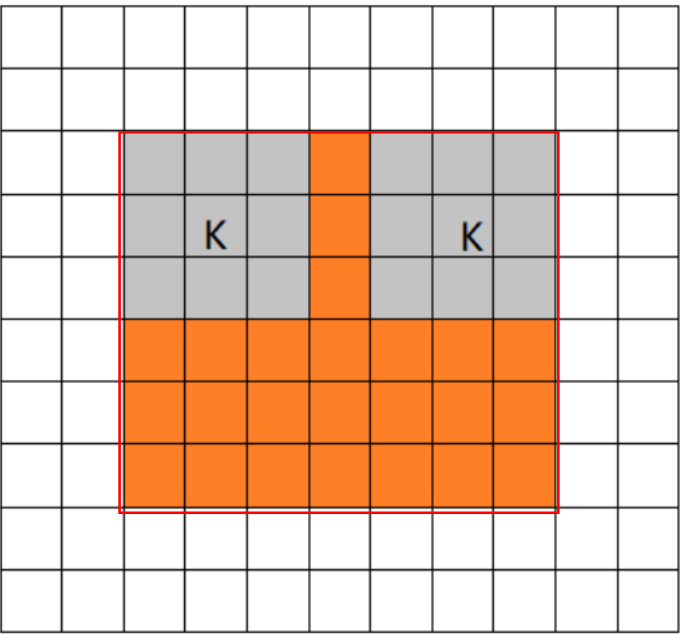
Full
补k-1圈

$$w_o = \frac{|W| - w + 2p}{stride} + 1$$



Same
补若干圈

$$w_o = \text{ceil}(\frac{|W|}{stride})$$



Valid
不补

$$w_o = \text{ceil}(\frac{|W| - w + 1}{stride})$$

输出矩阵
尺寸计算

0、卷积操作的三种模式及输出矩阵尺寸计算（2）

卷积后的输出矩阵的尺寸计算

$|W|$ 是输入矩阵宽， w 是卷积核的宽， p 是padding的数值，默认=0； $stride$ 是滑动步幅，默认是1；

分为两种参数情形

- 1) Valid, 无padding, 舍弃多出的部分（不足一个窗口）， w_o 下取整
- 2) Same, 有padding, 补零，不足一个窗口的部分补够一个窗口， w_o 上取整

Stride=1时，valid的意思是滑动窗口一直在实际的矩阵内，而没有超出输入矩阵
same的意思是，经过padding，输出尺寸和输入

Valid: $w_o = \text{ceil}(\frac{|W| - w + 1}{stride})$ 无padding

Same: $w_o = \text{ceil}(\frac{|W|}{stride})$ 有padding

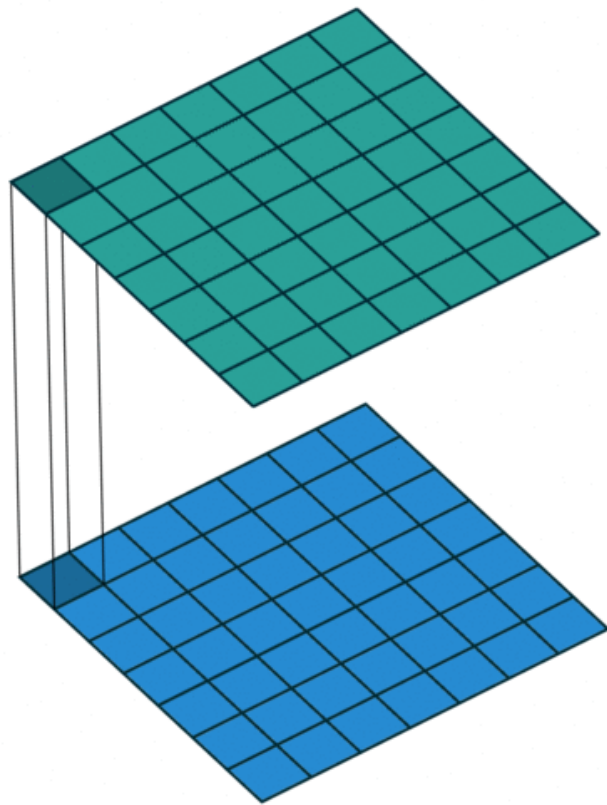
如，输入图像为2x3，2x2 kernel，输出尺寸

{	[1, 1]	无padding
	[1, 2]	有padding

一、1x1 卷积操作 (1)

1x1卷积，又称为NiN (Network in Network)

1x1卷积中，如果输入只有一个通道，则每个神经元乘以相同的权重weight，因此相当于scaling操作

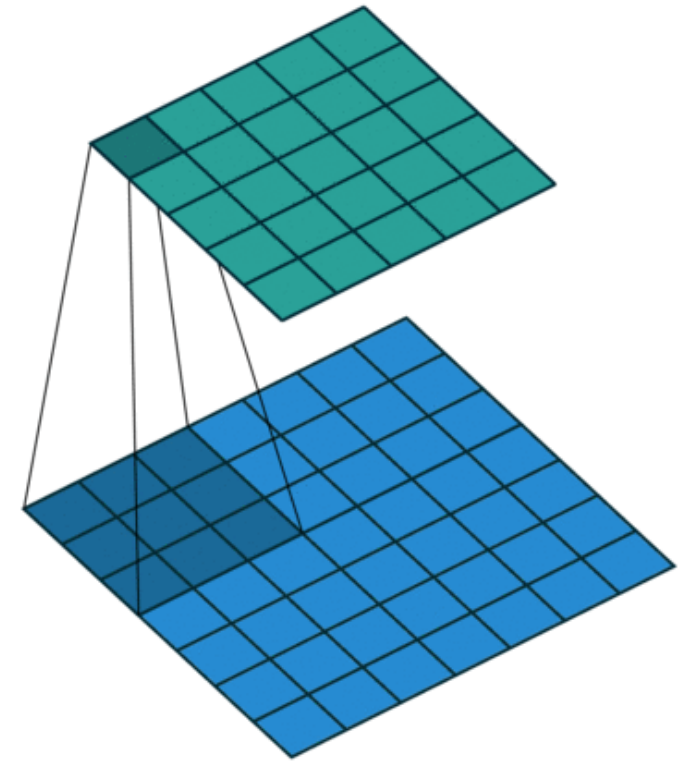


1x1 卷积

1	2	1	0	0
0	3	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

$$* \begin{array}{|c|} \hline 2 \\ \hline \end{array} =$$

2	4	2	0	0
0	6	2	2	0
0	0	2	2	2
0	0	2	2	0
0	2	2	0	0



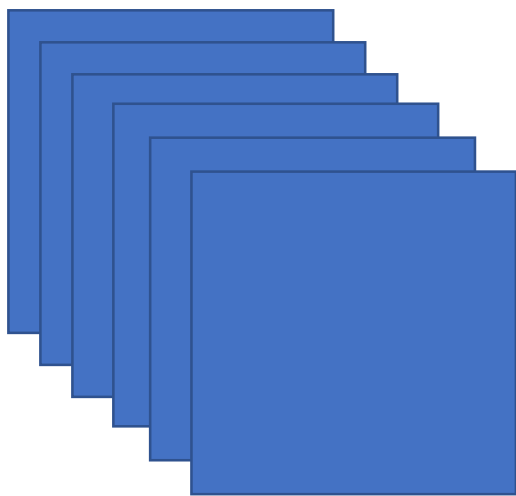
一般卷积

一、1x1 卷积操作 (2)

1x1卷积中，如果输入有多个通道，则相当于降维操作，如下面例子

1x1卷积，主要作用在于：

- 1) 降维。例如，从 $26 \times 26 \times 6$ ，到 26×26 ；
- 2) 增加非线性激励。在输入数据分辨率/尺度不损失的前提下，大幅增加神经网络的非线性特性



六个 26×26 矩阵

*



=



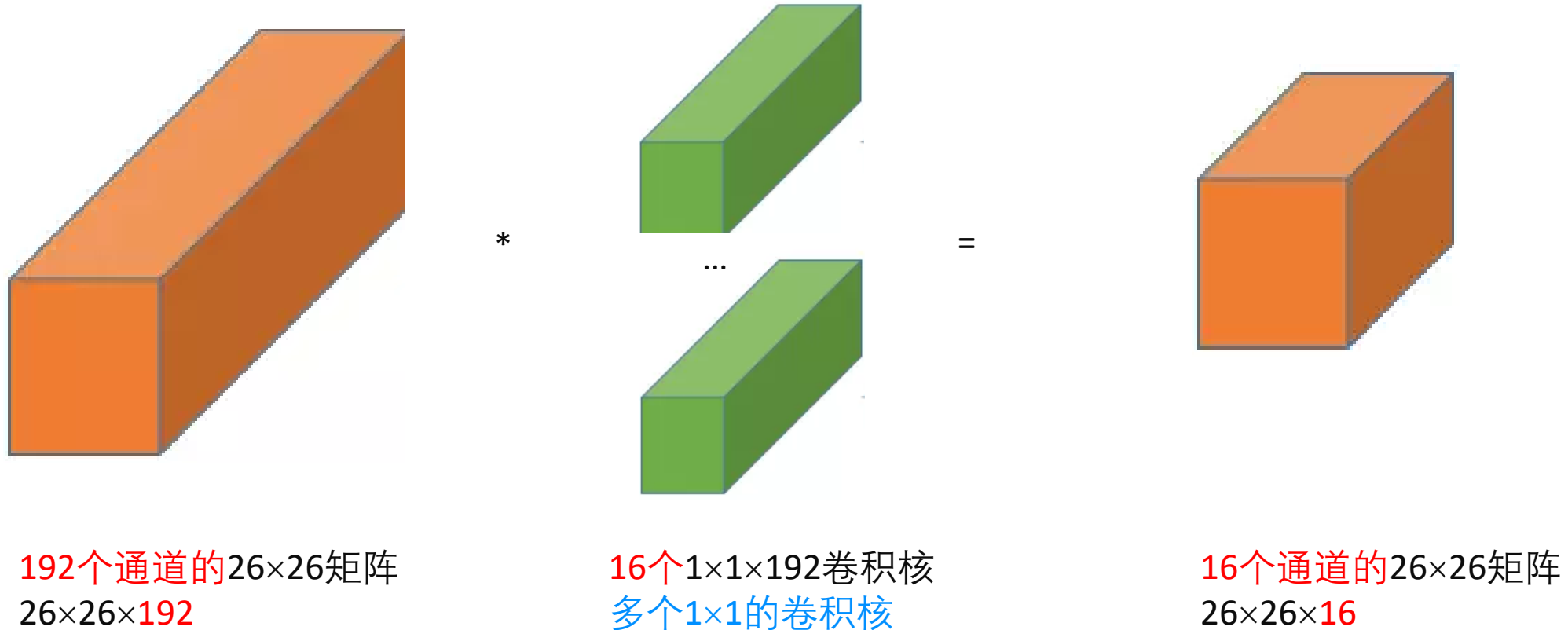
一个 26×26 矩阵

一、1x1 卷积操作 (3)

1x1卷积中，如果输入有多个通道，则相当于降维操作，同时提高了神经网络的表达能力。

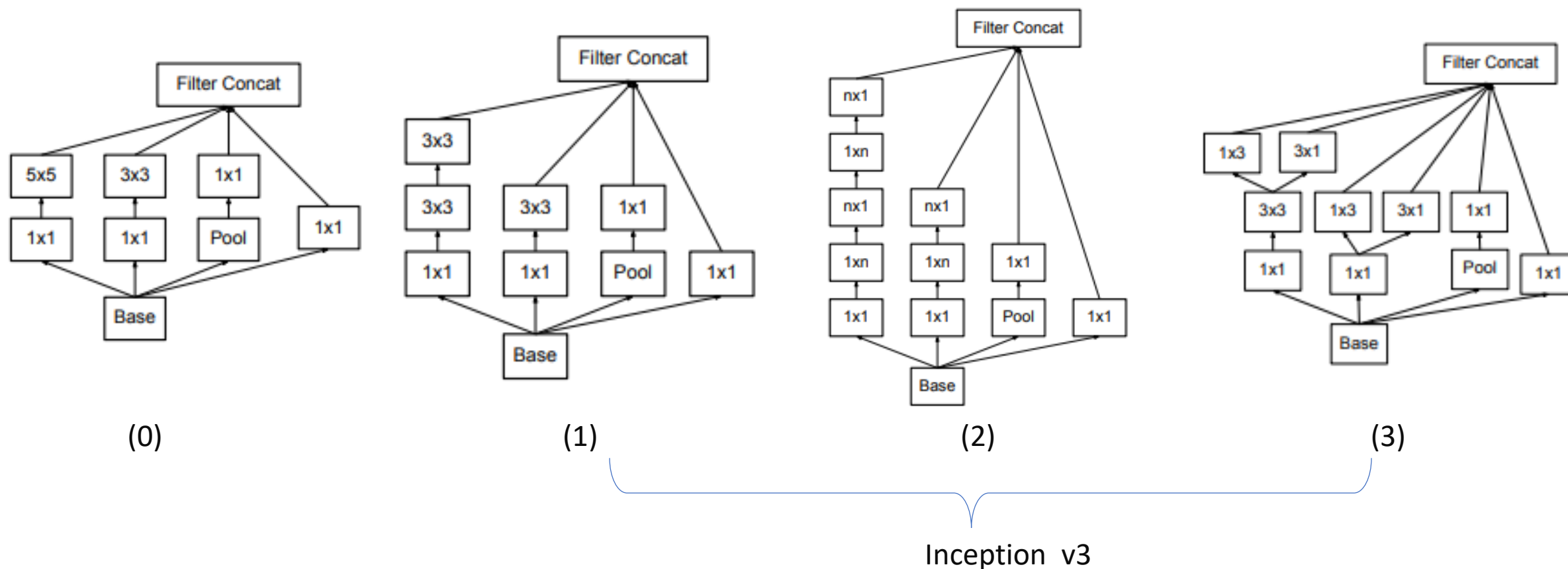
1x1卷积，主要作用在于：

- 1) 降维。例如，从 $26 \times 26 \times 192$ ，到 $26 \times 26 \times 16$ ；1x1卷积是跨通道线性组合，是通道间的信息交互。
- 2) 增加非线性激励。只改变通道数，不改变图像的分辨率，大幅增加神经网络的非线性特性；
- 3) 减少权重个数。



一、1x1 卷积操作 (4) Inception v3

1x1卷积，在经典神经网络GoogLeNet (即InceptionNet v1)和InceptionNet (v2, v3, v4)中得到了应用

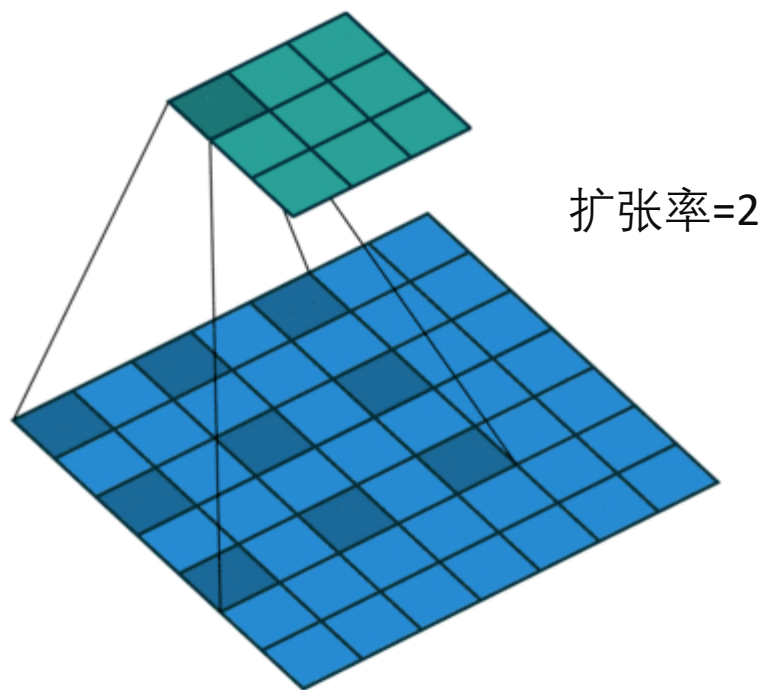


Christian Szegedy et al. Rethinking the Inception Architecture for Computer Vision, CVPR 2016.

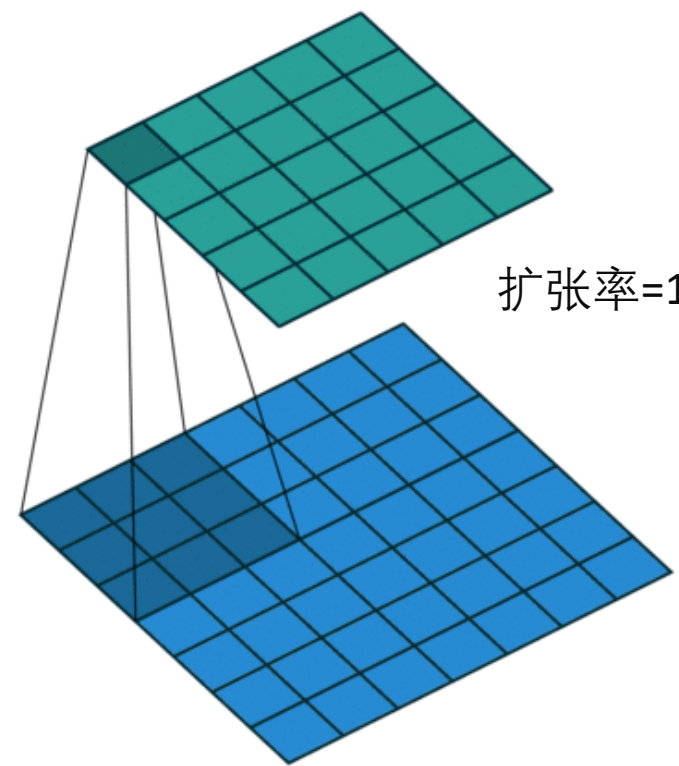
二、扩张卷积/膨胀卷积/空洞卷积 (1)

扩张卷积，即空洞卷积、膨胀卷积 (Dilated Convolution, Atrous Convolution)，**在标准卷积中注入空洞**
扩张卷积常被用以低成本地增加输出单元上的感受野，同时还不需要增加卷积核大小

空洞卷积的实际卷积核大小/宽： $K = w + (w-1)*(r-1)$ ， w 为原始卷积核的宽， r 为扩张率



空洞卷积

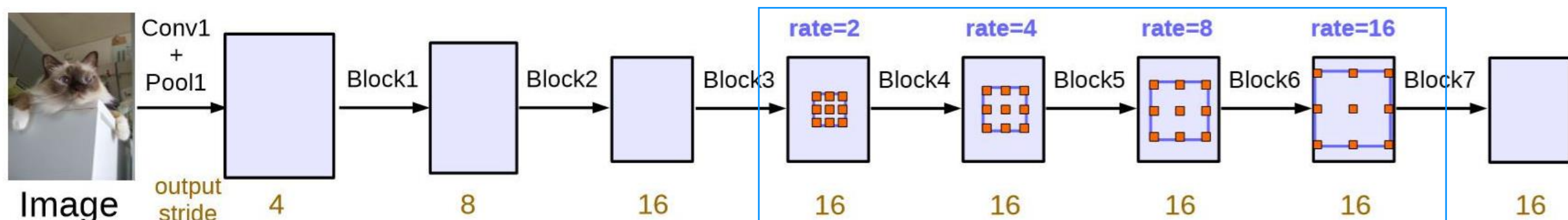


一般卷积

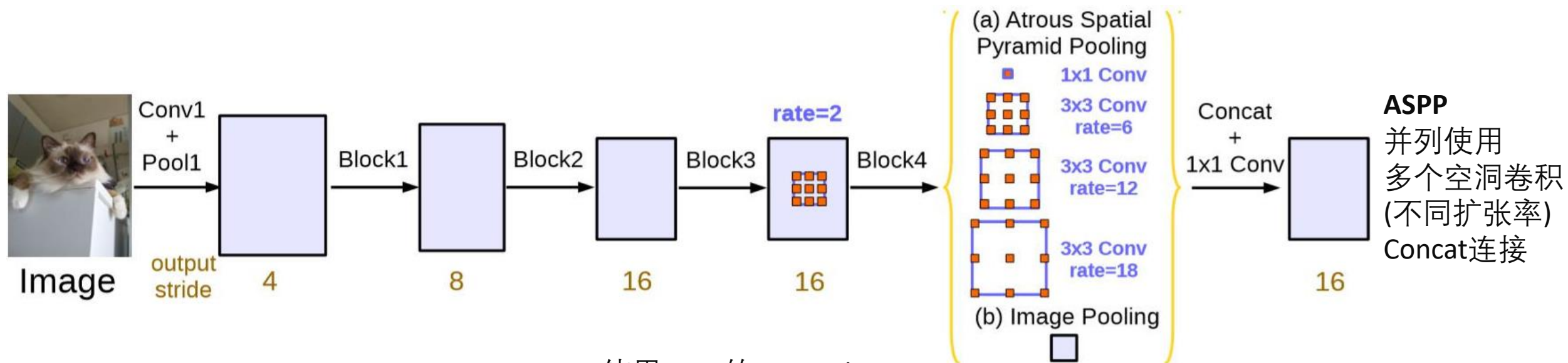
二、扩张卷积/膨胀卷积/空洞卷积 (2)

扩张卷积/空洞卷积，增加了神经网络模型的感受野

扩张卷积广泛应用于语义分割和目标检测中，如DeepLab v2等采用了ASPP（金字塔型的空洞池化）



DeepLab v2



使用ASPP的DeepLab v2

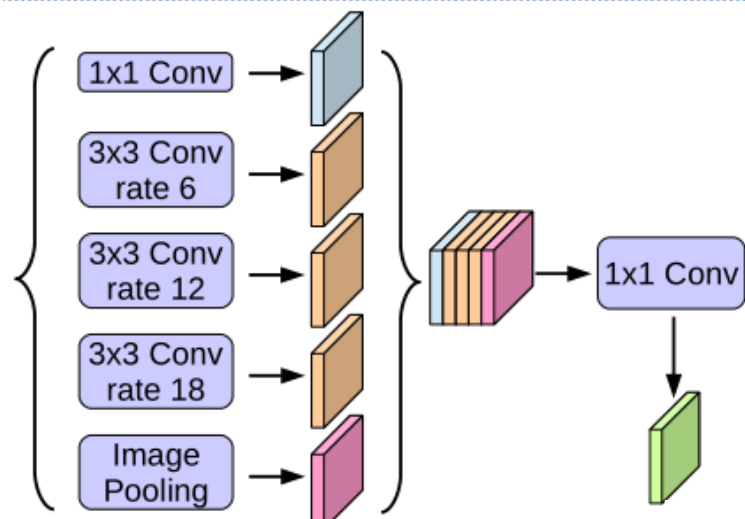
ASPP
并列使用
多个空洞卷积
(不同扩张率)
Concat连接

二、扩张卷积/膨胀卷积/空洞卷积 (3)

扩张卷积/空洞卷积,

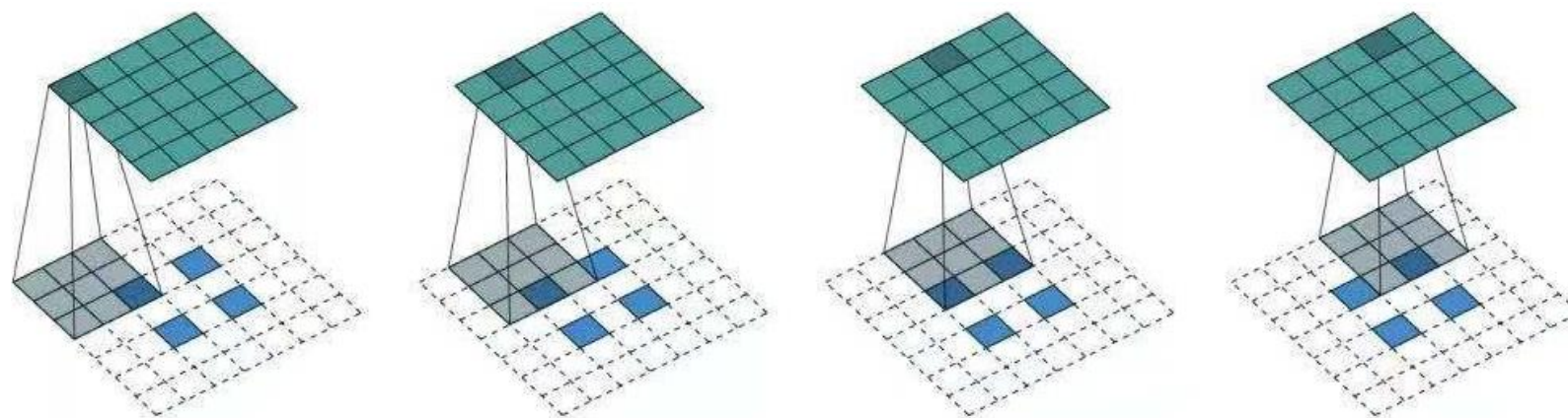
优点：扩大了神经网络模型的感受野，捕获更多上下文信息，尤其对大尺寸的物体分割有用

缺点：1) 局部信息丢失，因为kernel不连续，损失了信息的连续性；
2) 使得远距离卷积得到的信息之间没有相关性，影响分类结果；
3) 存在棋盘问题/网格效应，空洞卷积的结果中，邻近的像素相互之间的依赖减少。
反卷积后生成的图像，放大后往往会呈现棋盘外形，深色部分尤为明显。
解决方法之一：确保卷积核能被步长(stride)整除



三、反卷积/逆卷积 (1)

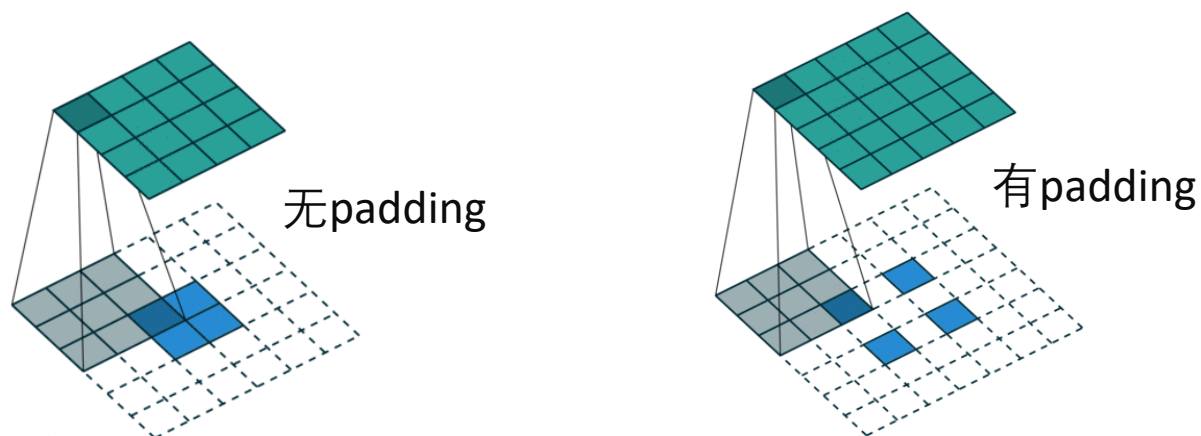
反卷积又称为转置卷积或逆卷积，英文为 Deconvolution/Transposed Convolution，是一种上采样的方法。反卷积是中间填0再卷积；普通的上采样如果用双线性插值，中间填相邻元素的差值。



输入：2x2矩阵 (padding为7x7矩阵)

经过3x3卷积核

输出：5x5矩阵



反卷积/转置卷积

反卷积是指根据当前位置的信息，决定周围区域的数据，是上采样方法。

三、反卷积/逆卷积 (2)

反卷积/转置卷积的主要应用是语义分割FCN、UNet等网络中，及GAN中的DCGAN

将反卷积/转置卷积的计算过程转换为普通卷积 stride=1 padding=0

0	0	0	0	0	0
0	0	0	0	0	0
0	0	12	12	0	0
0	0	10	17	0	0
0	0	0	0	0	0
0	0	0	0	0	0

X

i	h	g
f	e	d
c	b	a

a	b	c
d	e	f
g	h	i

逆时针旋转180度

=

从2x2矩阵，经反卷积到4x4矩阵

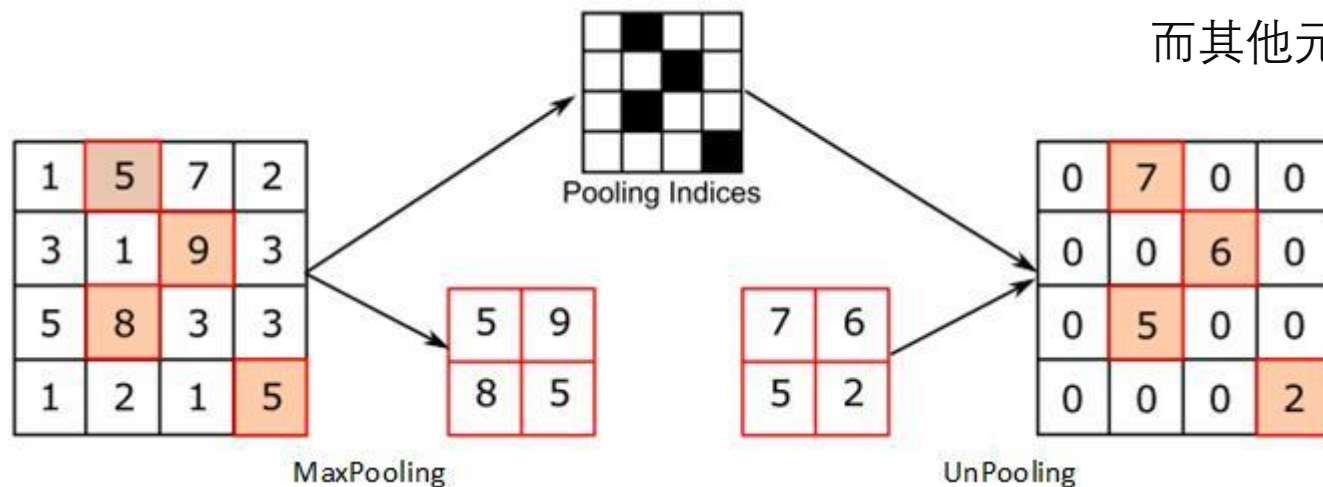
$w_o = \text{ceil}(\frac{|W|-w+1}{stride})$

无padding

四、反池化/上池化 Unpooling 与上采样 Upsampling

反池化/上池化，Unpooling，专指Maxpooling、Avgpooling等的逆过程。

池化时，保存了最大值在输入数据中的位置信息矩阵；反池化时，将对应位置上的值置为输出矩阵的对应值而其他元素置0。如 2x2 池化



上采样，即Upsampling，本质上是插值法，可以采用重复采样和插值法(如双线性插值)
它没有Unpooling的每个池化窗口中的最大值在输入矩阵中的位置信息

广义上说，Upsampling也包含了Unpooling和Deconvolution / Transposed Convolution两种方式。

上采样 反池化 反卷积

五、PixelShuffle 像素重排列上采样 ESPCN 超分辨率方法

ESPCN只在模型末端进行上采样，可以使得在低分辨率空间保留更多的纹理区域。使用亚像素卷积的方式进行上采样。 r 为上采样倍数， c 为最终的通道数，若RGB输出 $c=3$ 。如 $r=3, c=1$ ，单通道图的3倍上采样图

特征图通道数中连续的 c 个通道作为一个整体，再然后进行像素重排列，得到多通道的上采样图。前面都是卷积，只有倒数第二层才是像素重排列，将一个 $H \times W \times C \cdot r^2$ 的特征图，重排列(顺序交替排列)为一个 $rH \times rW \times C$ 的特征图， r 个通道作为一组，该组中的像素在一起交替顺序排列，一共有 r 个这样的组。

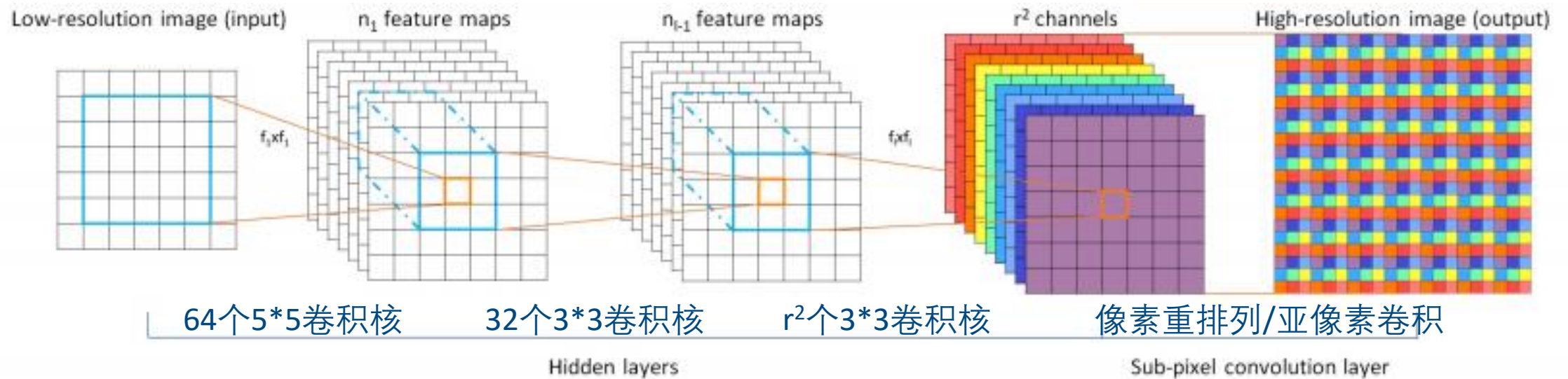


Figure 1. The proposed efficient sub-pixel convolutional neural network (ESPCN), with two convolution layers for feature maps extraction and a sub-pixel convolution layer that aggregates the feature maps from LR space and builds the SR image in a single step.

Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network, CVPR 2016

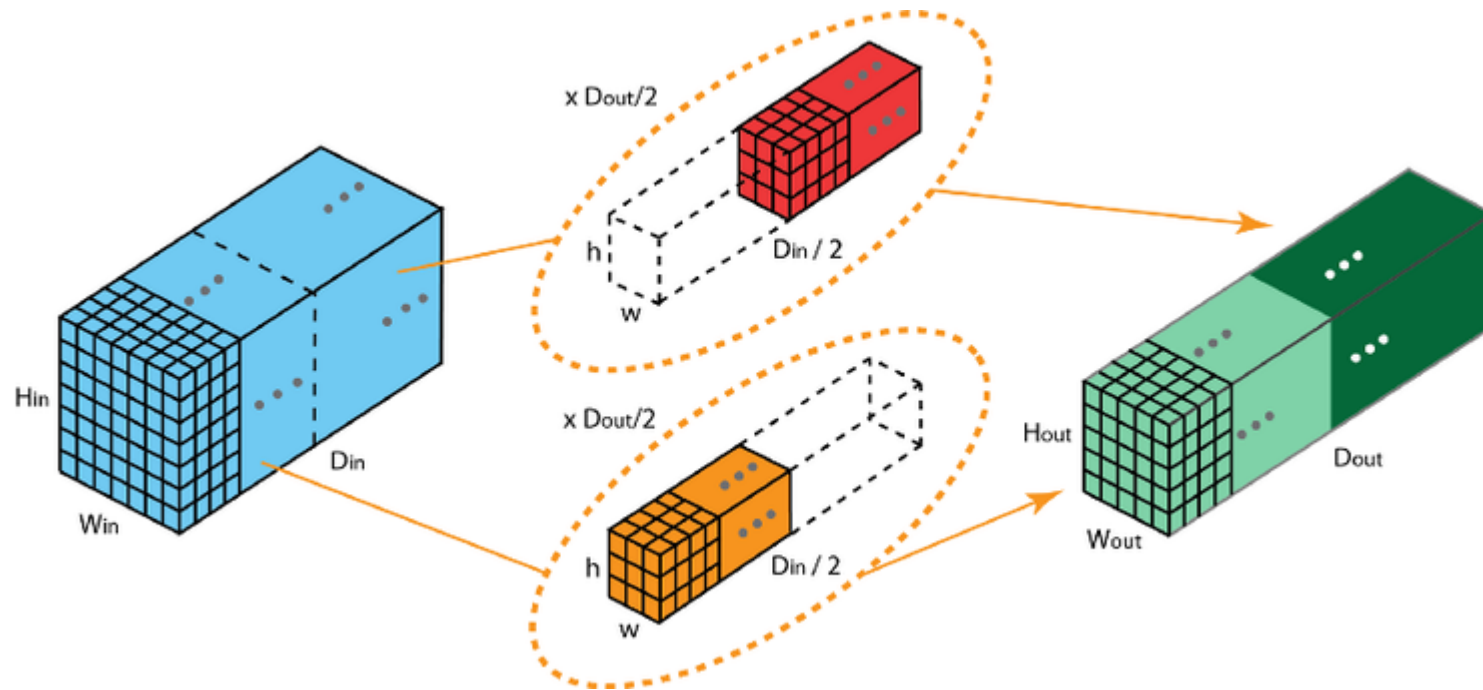
六、分组卷积 (AlexNet) Grouped Convolution

分组卷积，如将所有通道分成2组，则每一组通道分别使用各自对应的 $D/2$ 个卷积核，最后两组卷积后特征堆叠

作用：1) 减少参数量，参数为原来的 $1/G$

2) 加快训练速度

3) 有时可以起到正则化效果



七、深度可分离卷积 Depthwise Separable Convolutions

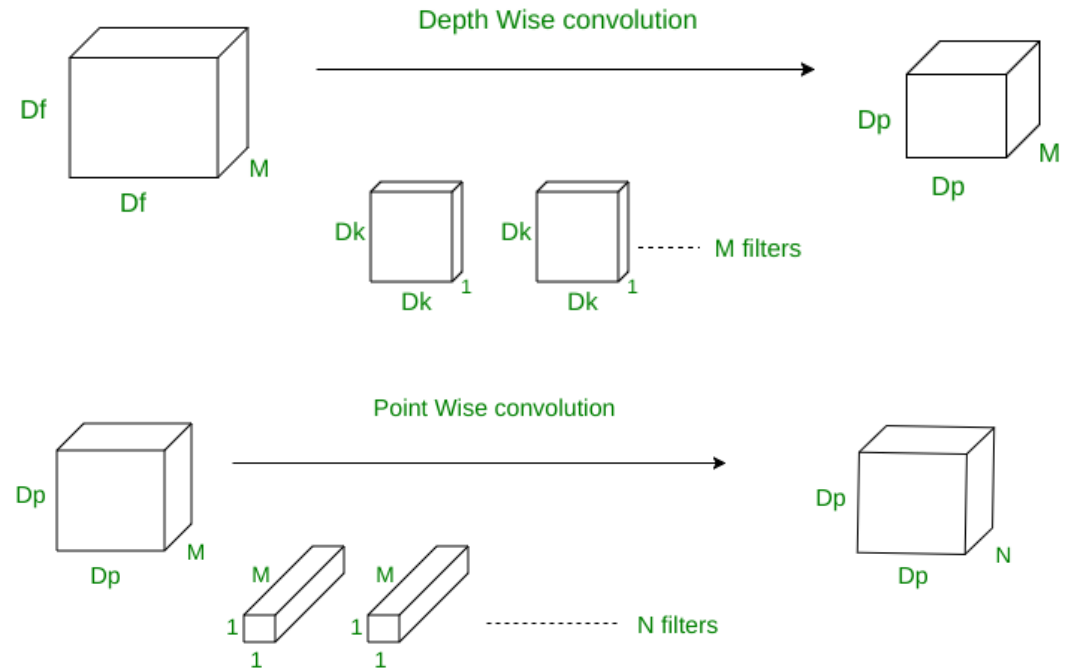
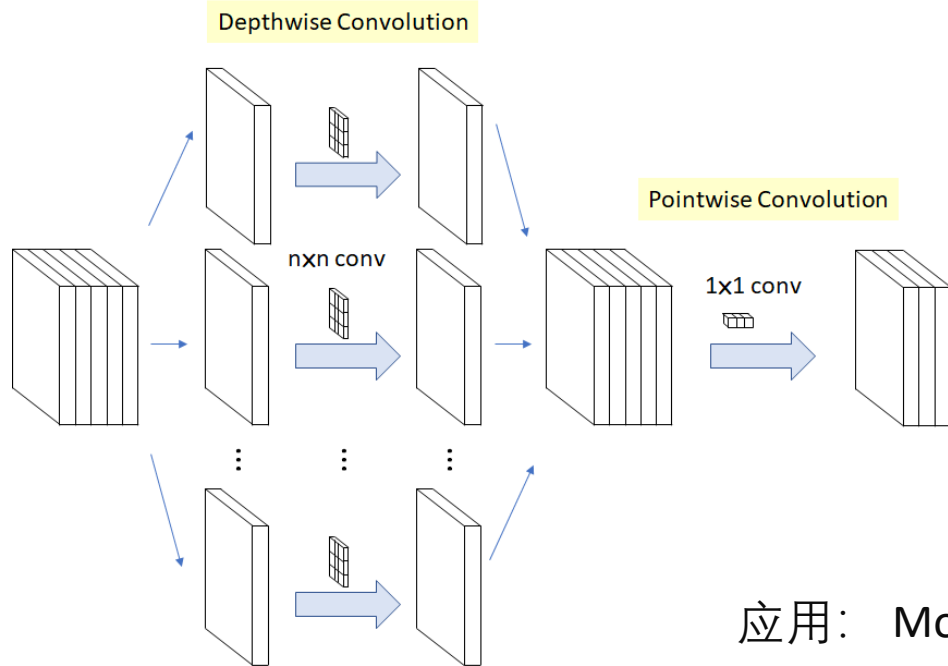
逐通道卷积

Depthwise Convolution的一个卷积核负责一个通道，一个通道只被一个卷积核卷积（如 3×3 ）输出的通道数与输入相同，但是没有利用不同通道在相同空间位置上的特征关系信息

逐点卷积

执行 1×1 卷积， M 为输入通道数，进行单点上的特征提取

深度可分离卷积 = 逐通道卷积 + 逐点卷积，前后两个步骤



深度可分离卷积将分组卷积推向了极端，此时：
分组数是输入通道数目，即每个输入通道单独卷积

应用： MobileNet 移动端轻量化网络，参数少，占用资源少，运算快

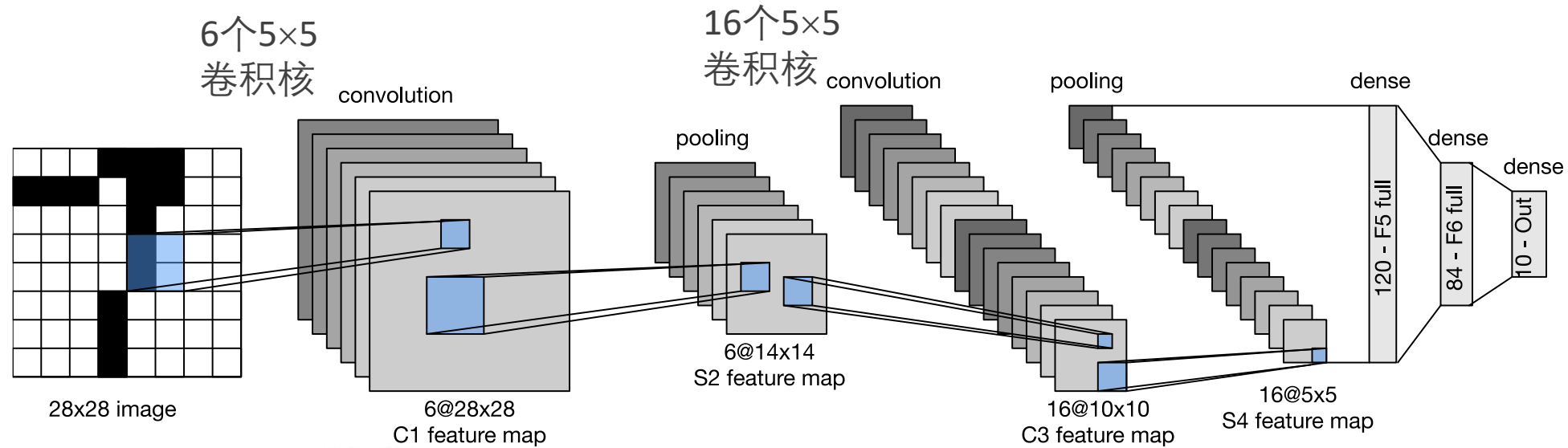


Part 05-2

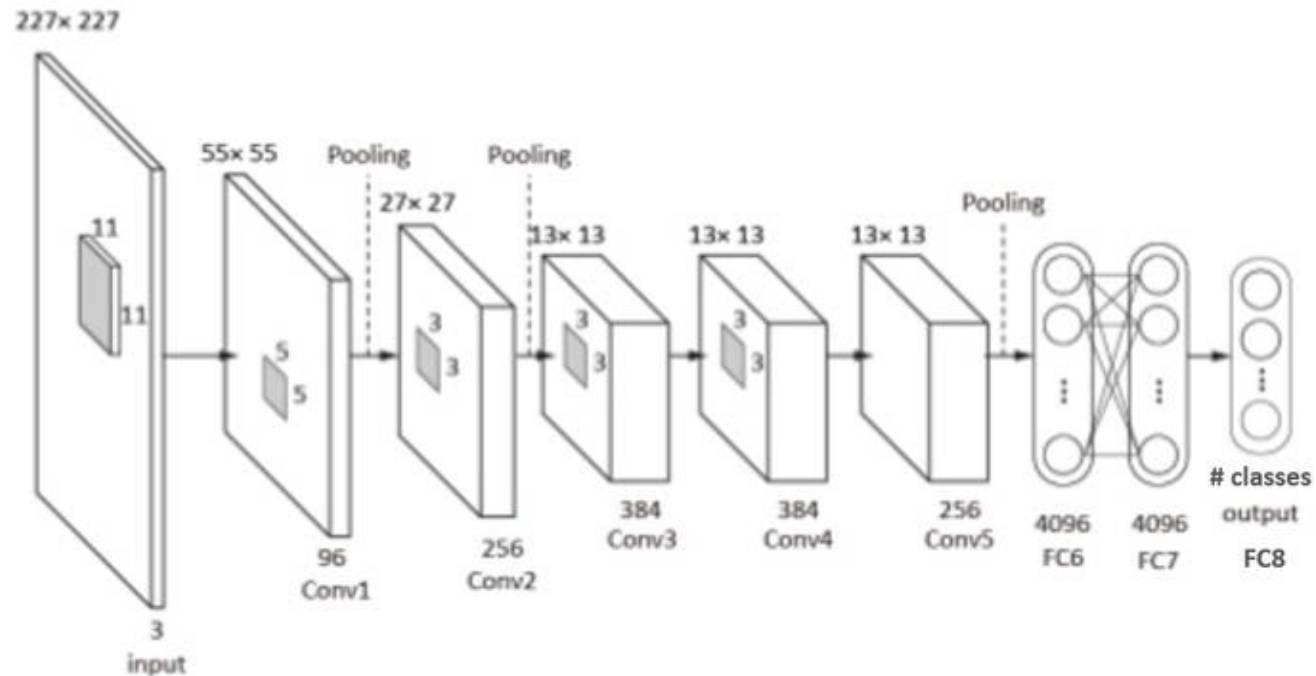
常见卷积神经网络讲解 从LeNet到ResNet

各种通用的卷积神经网络讲解

一、LeNet和AlexNet卷积神经网络 (1)



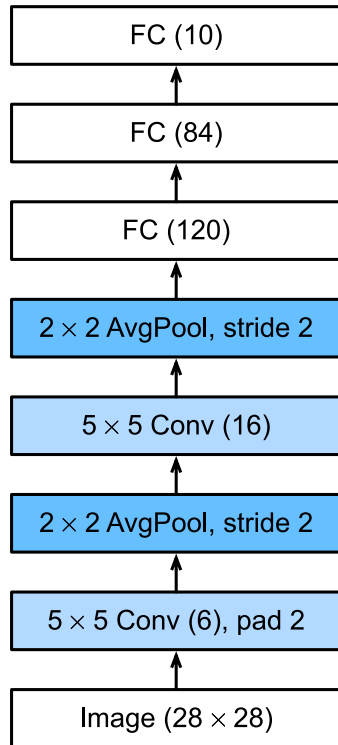
LeNet神经网络



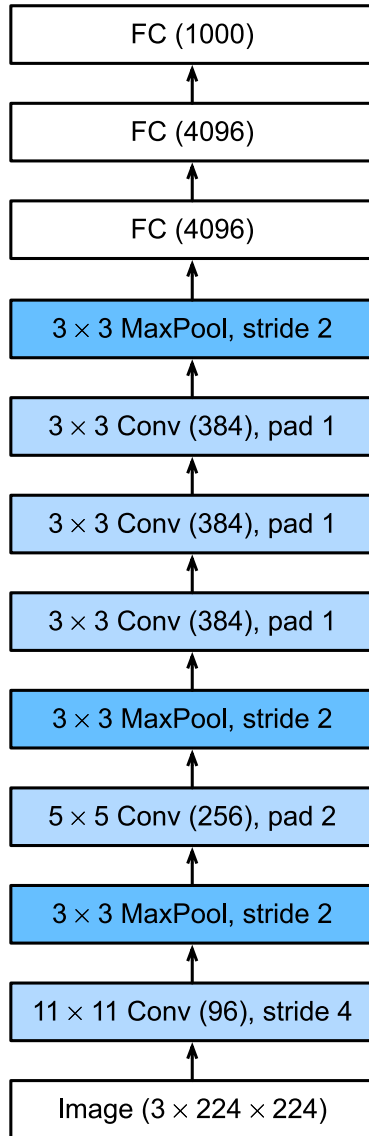
AlexNet神经网络

一、LeNet和AlexNet卷积神经网络（2）

LeNet



AlexNet



LeNet (1989):

两个卷积层，两个池化层，两个全连接层，一个输出层
AvgPool, Sigmoid激活函数

两个卷积层都是5×5窗口
两个全连接层的神经元数量分别为120和84

创新性：卷积神经网络的开山之作。完成了CNN从0到1的过程。

AlexNet (2012年ImageNet竞赛冠军):

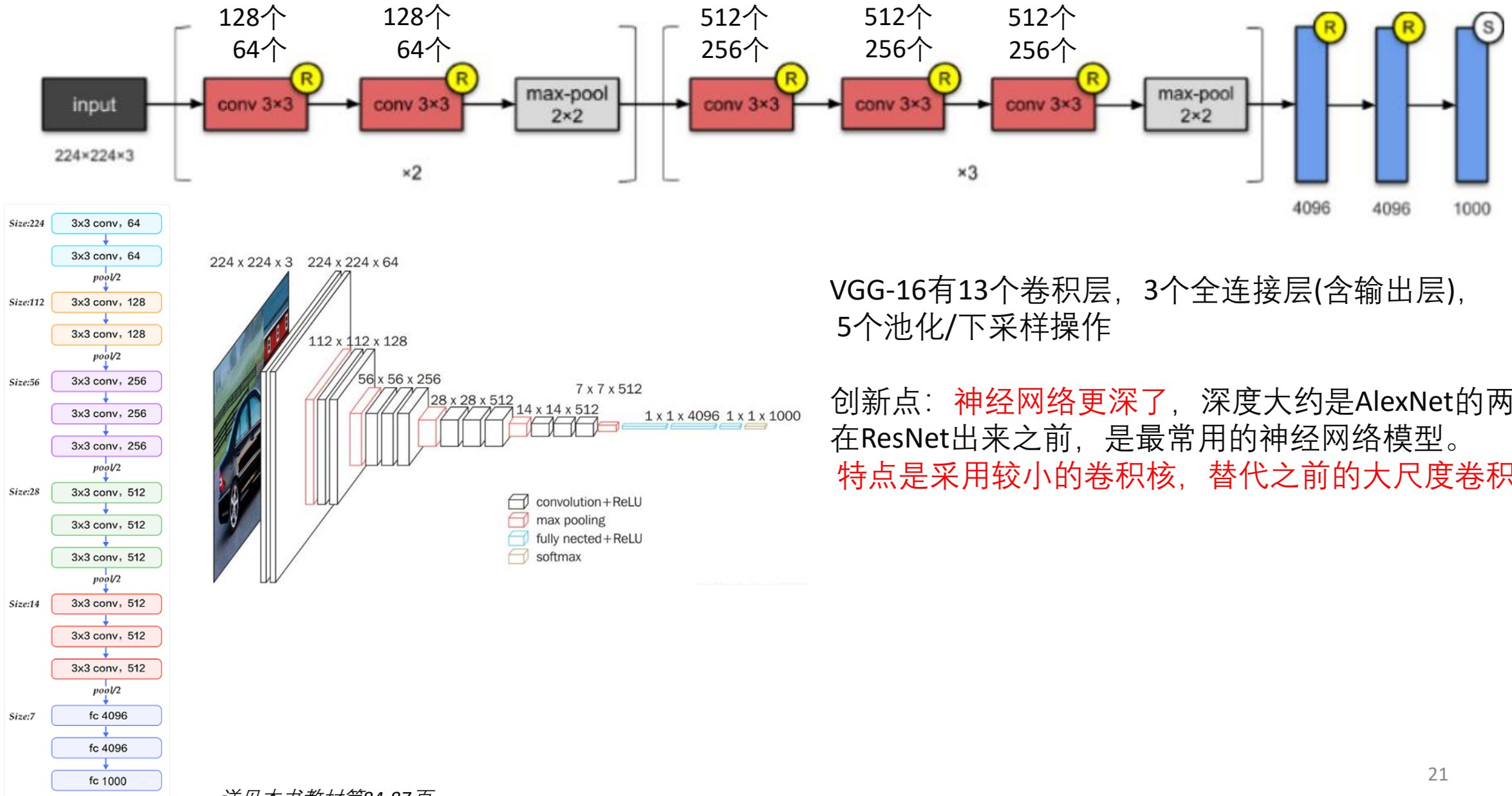
五个卷积层，三个池化层，两个全连接层，一个输出层
MaxPool, ReLU激活函数

第一层卷积11×11窗口，以后5×5, 3×3
两个全连接层的神经元数量均为4096

两个全连接层之间使用了Dropout技术，随机一半隐层节点值为0
(因为当模型的参数太多，而训练样本太少时，易产生过拟合)

创新性：AlexNet比LeNet更深一些，但是架构和流程没变²⁰

二、VGG-16 卷积神经网络



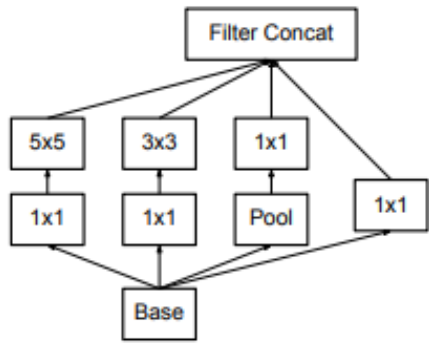
VGG-16有13个卷积层，3个全连接层(含输出层)，5个池化/下采样操作

创新点：神经网络更深了，深度大约是AlexNet的两倍
在ResNet出来之前，是最常用的神经网络模型。
特点是采用较小的卷积核，替代之前的大尺度卷积核

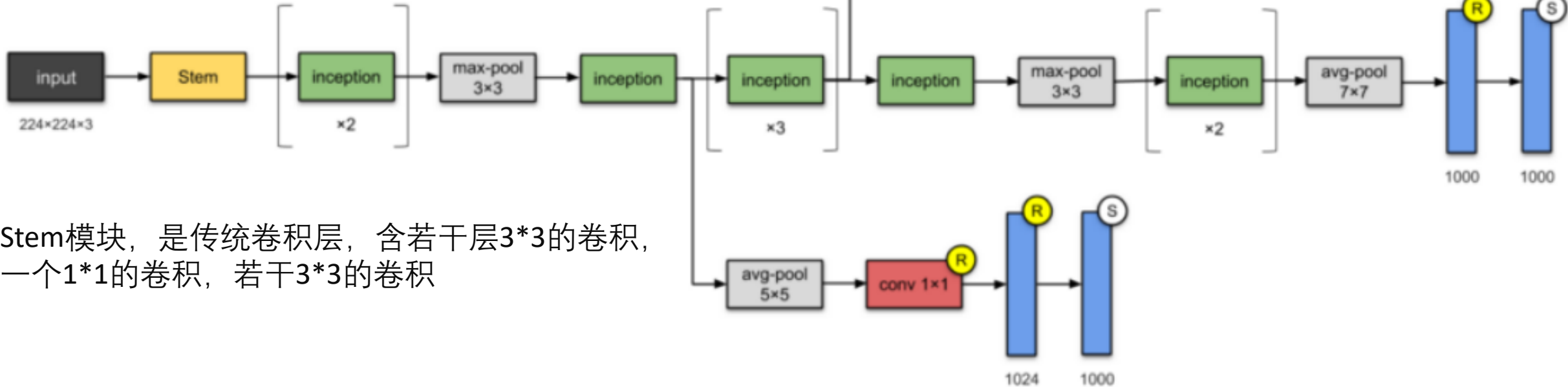
详见本书教材第84-87页

三、GoogLeNet 卷积神经网络 Inception v1

2014年ImageNet竞赛冠军

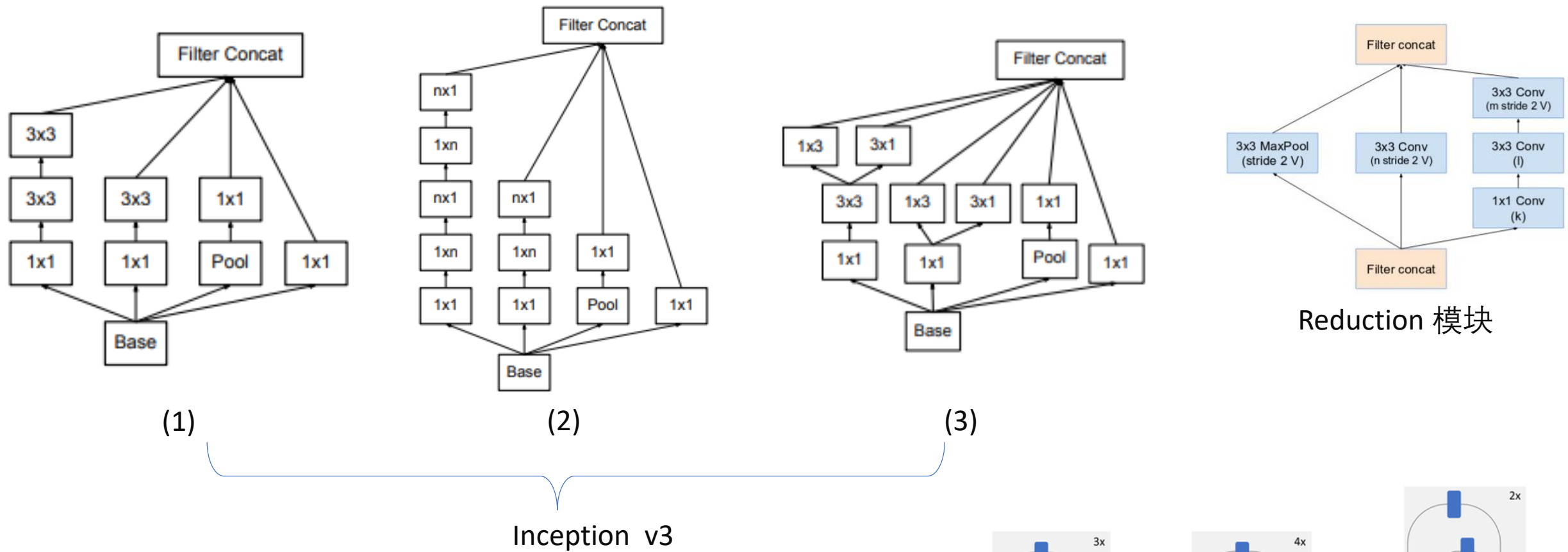


创新点：Inception v1 引入了1*1卷积和多个Inception块
神经网络结构设计上进行大胆突破和尝试

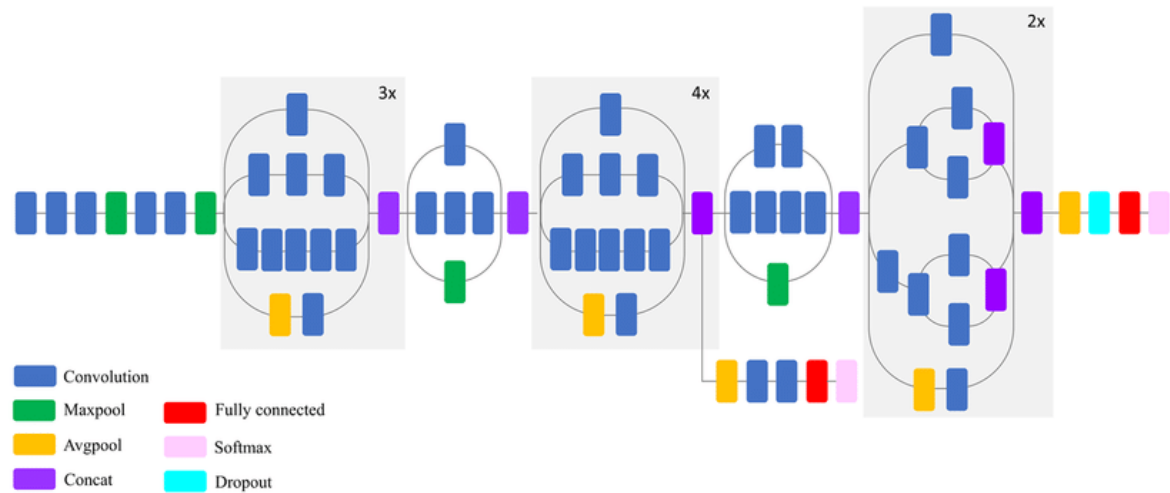


Stem模块，是传统卷积层，含若干层3*3的卷积，
一个1*1的卷积，若干3*3的卷积

四、Inception v3及后续衍生版本

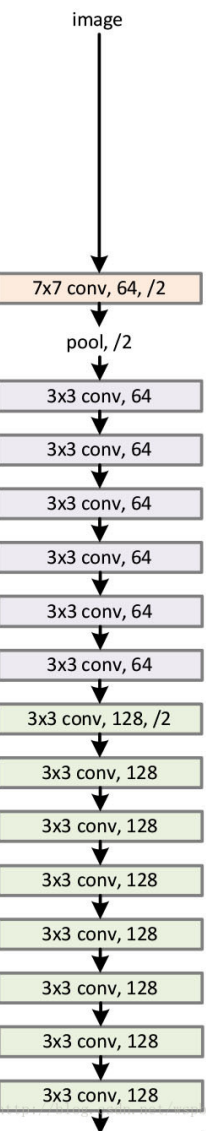


创新点：
使用了多种不同的Inception块
首次使用了批归一化

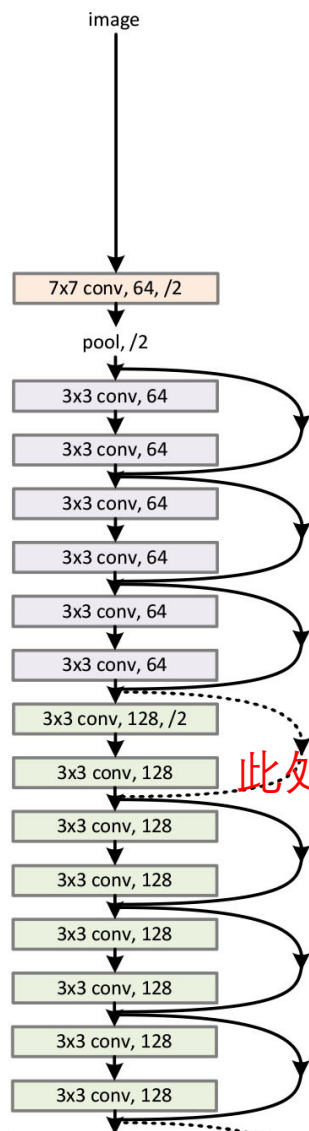


五、ResNet 及后续衍生版本

34-layer plain



34-layer residual

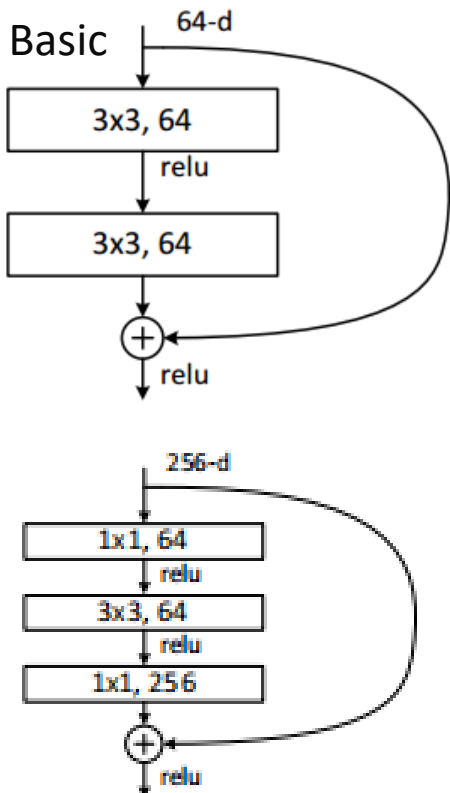


此处无捷径

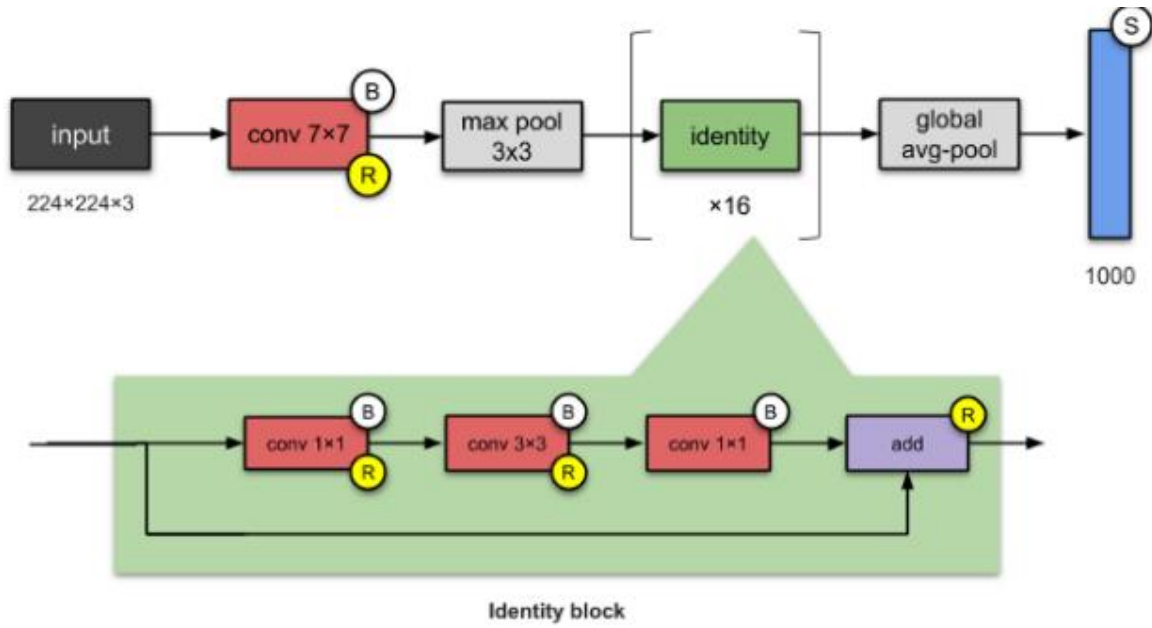
ResNet创新点：

使用了残差块--捷径连接，在输入激活函数前，将前层网络的输入与当前层网络层的输出进行结合，数据可以跨层连接。

因为：网络越深，梯度就越容易出问题，捷径连接的方式一定程度上缩短了损失的反向传播路径，减小了梯度风险。



Resnet-50/101/152



Resnet第一阶段是一个7x7的卷积，stride=2，然后再池化；后面，进入若干个残差块(basic 或 bottleneck block)，如上

六、其它神经网络 (1)

其它更先进的神经网络包括：

DenseNet
MobileNetV3
EfficientNet
EfficientDet
Xception
ResNeXt等

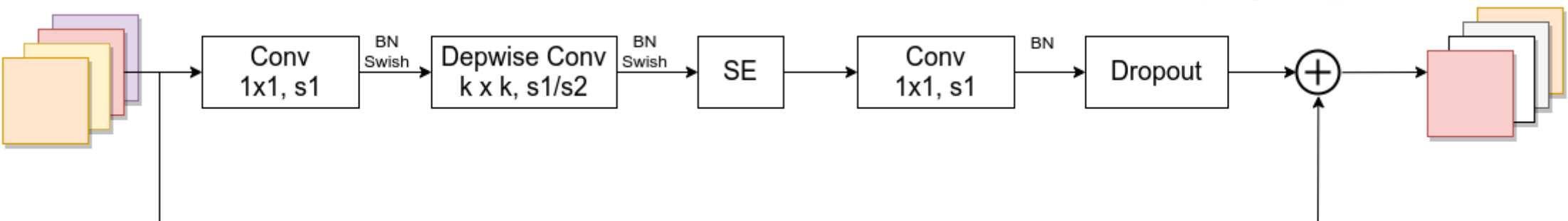
Table 1. **EfficientNet-B0 baseline network** – Each row describes a stage i with \hat{L}_i layers, with input resolution $\langle \hat{H}_i, \hat{W}_i \rangle$ and output channels \hat{C}_i . Notations are adopted from equation 2.

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels \hat{C}_i	#Layers \hat{L}_i
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

V1 (2019)复合缩放
按照一定比例对
宽度、深度和分辨率
进行缩放

V2(ICML 2021)
渐进学习

https://blog.csdn.net/qq_37541097

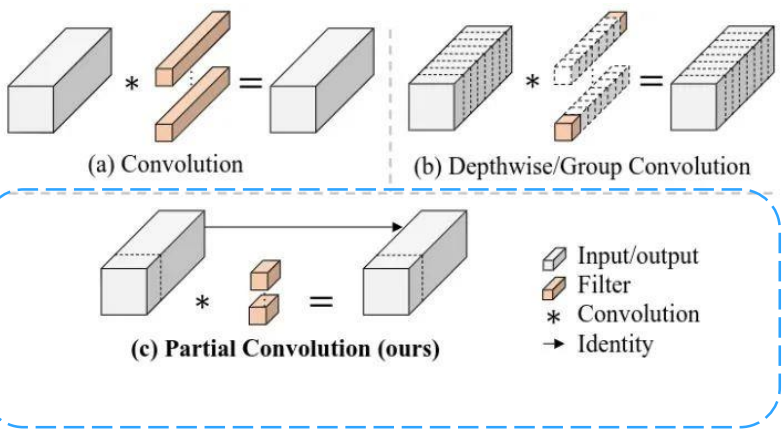
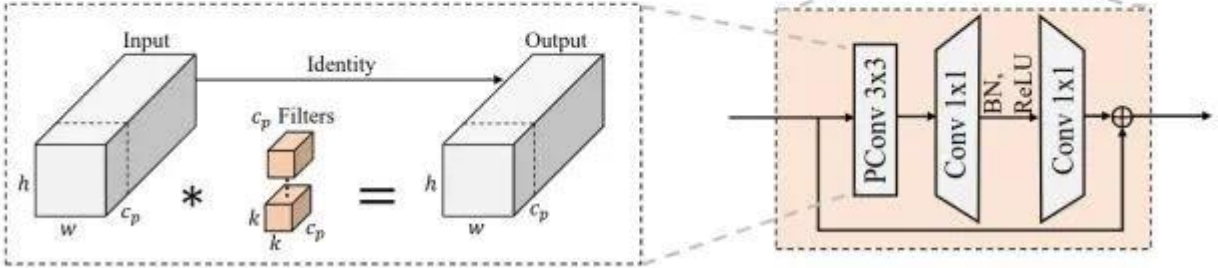
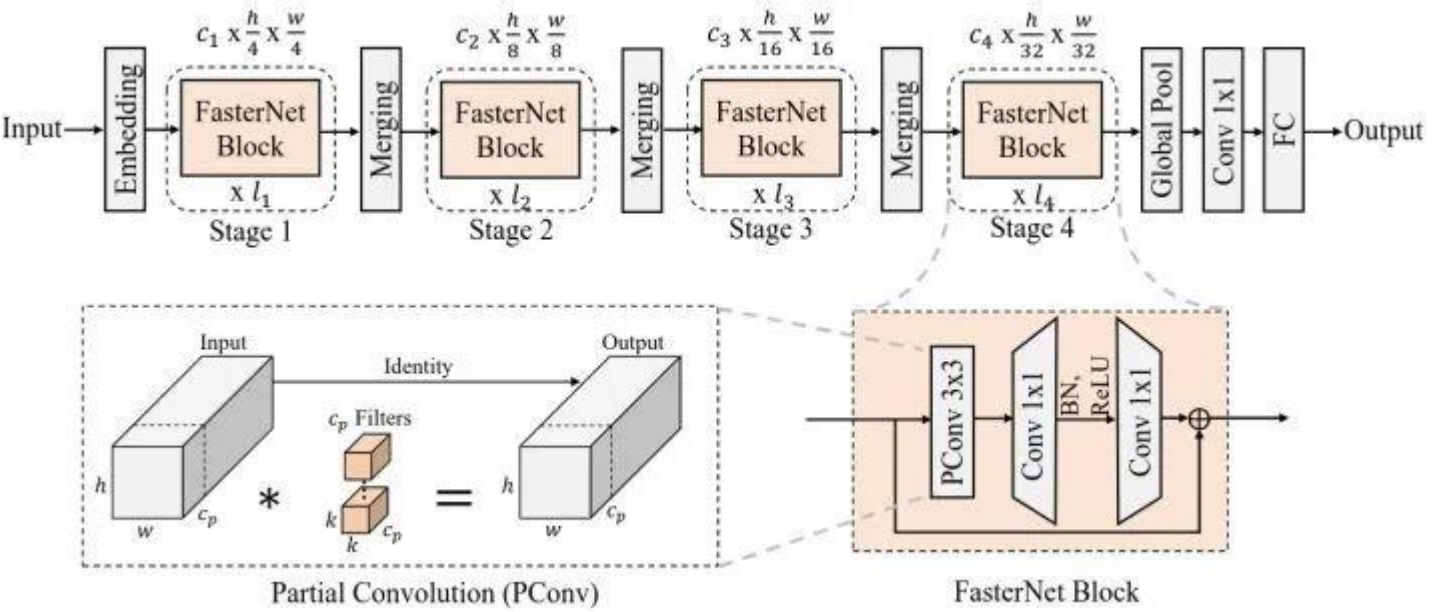
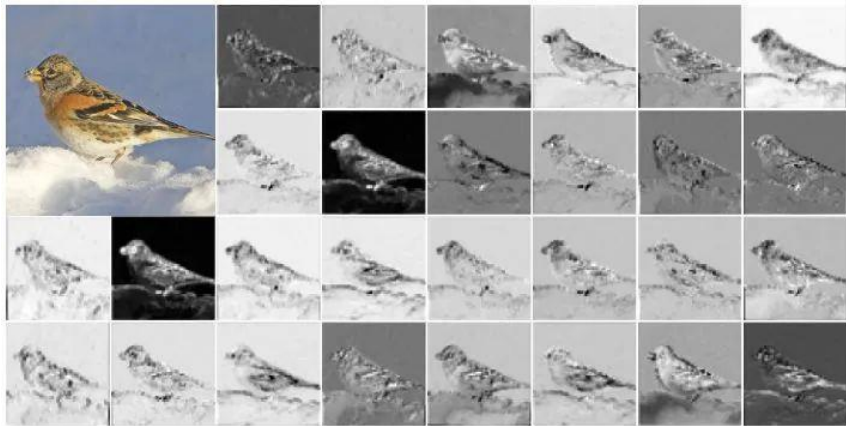


MobileNetV3中的逆残差块(但激活函数为Swish激活)

六、其它神经网络 (2)

其它更先进的神经网络：

FasterNet



FasterNet速度更快， 且与部分ViT(Transformer)的性能表现相当

因为特征图在不同通道之间具有很高的相似性，所以PConv 对部分输入通道应用常规的 Conv 来进行空间特征提取，而对其余通道保持不变。内存访问数量仅仅为常规卷积的1/4
PConv之后又增加逐点卷积， 详见FasterNet Block图。