



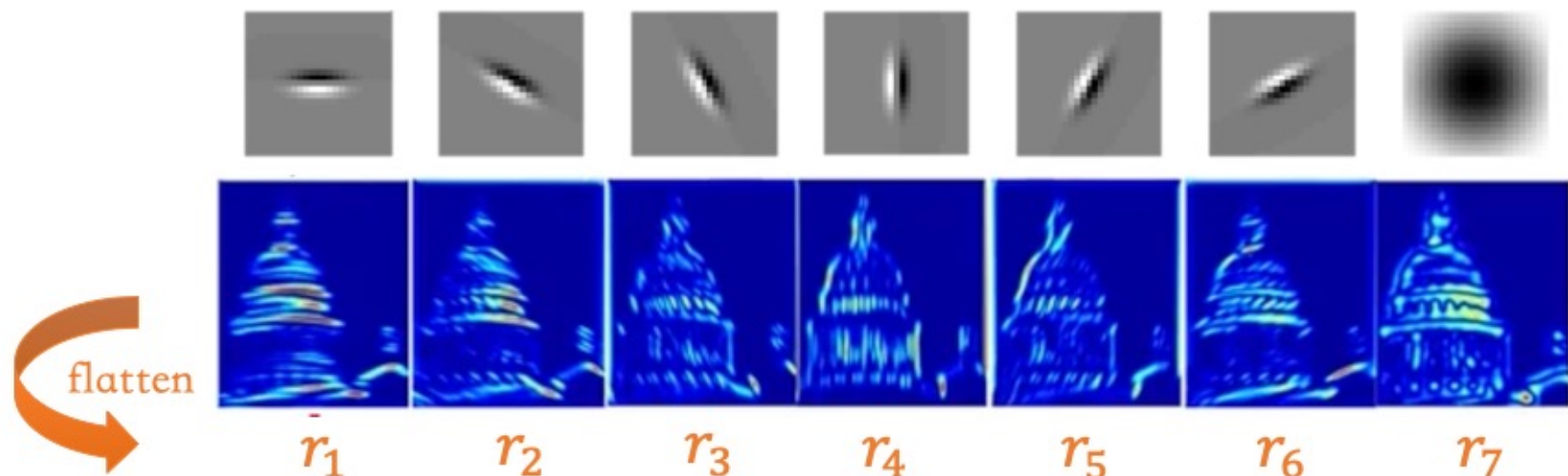
# 卷积神经网络 (2)

叶山 中国地质大学 (北京)

[yes@cugb.edu.cn](mailto:yes@cugb.edu.cn)

# 上期回顾

# 基于卷积核组的图像表示



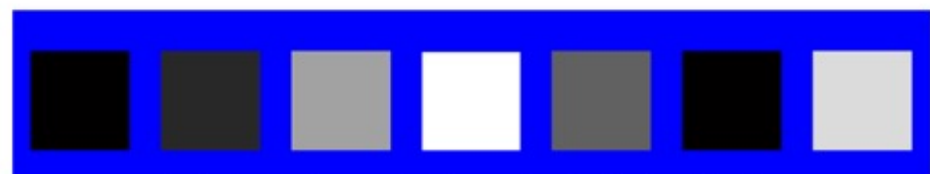
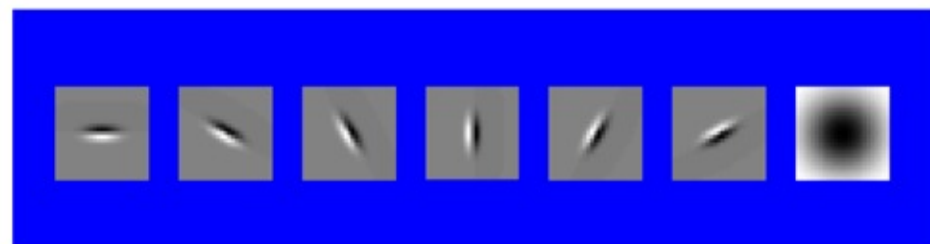
把每一个特征响应图做flatten操作，展开成向量。再把这些向量拼接起来，作为图像的表达。

$$R = [r_1, r_2, r_3, r_4, r_5, r_6, r_7]$$

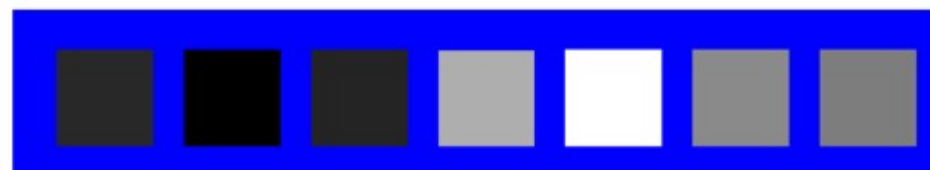
特征图角度



# 基于卷积核组的图像表示



第4个特征图最亮：竖直特征较多。



第5个特征图最亮：右上-左下的特征较多。



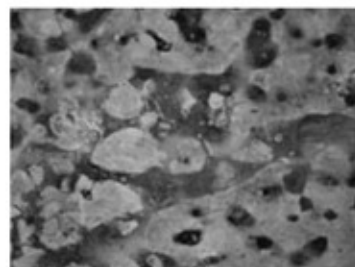
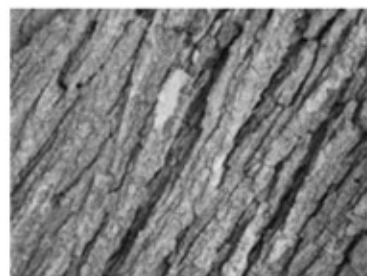
第7个特征图最亮：斑点状特征较多。

$\bar{r}_1$     $\bar{r}_2$     $\bar{r}_3$     $\bar{r}_4$     $\bar{r}_5$     $\bar{r}_6$     $\bar{r}_7$

A

B

C



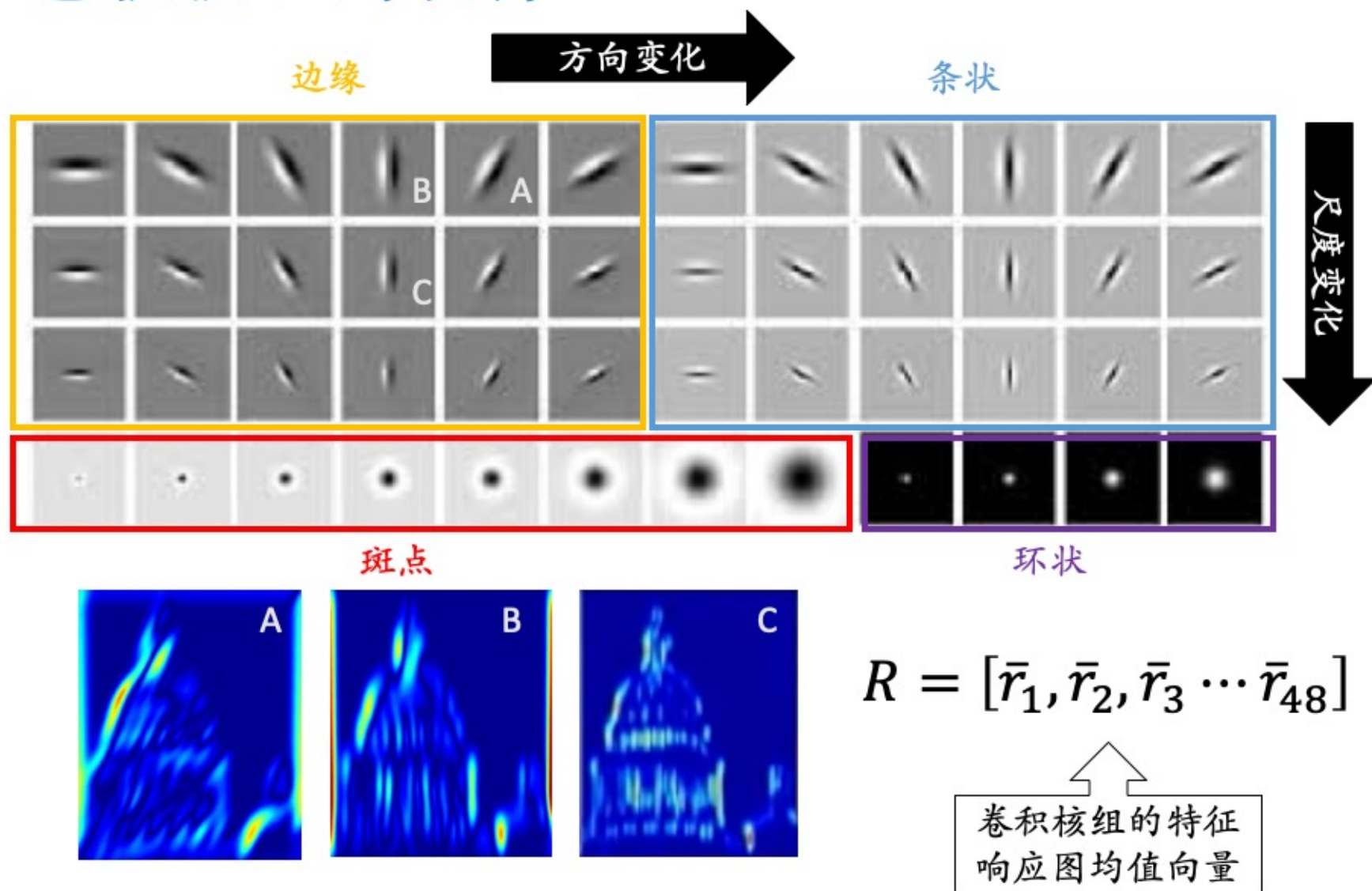
1

2

3

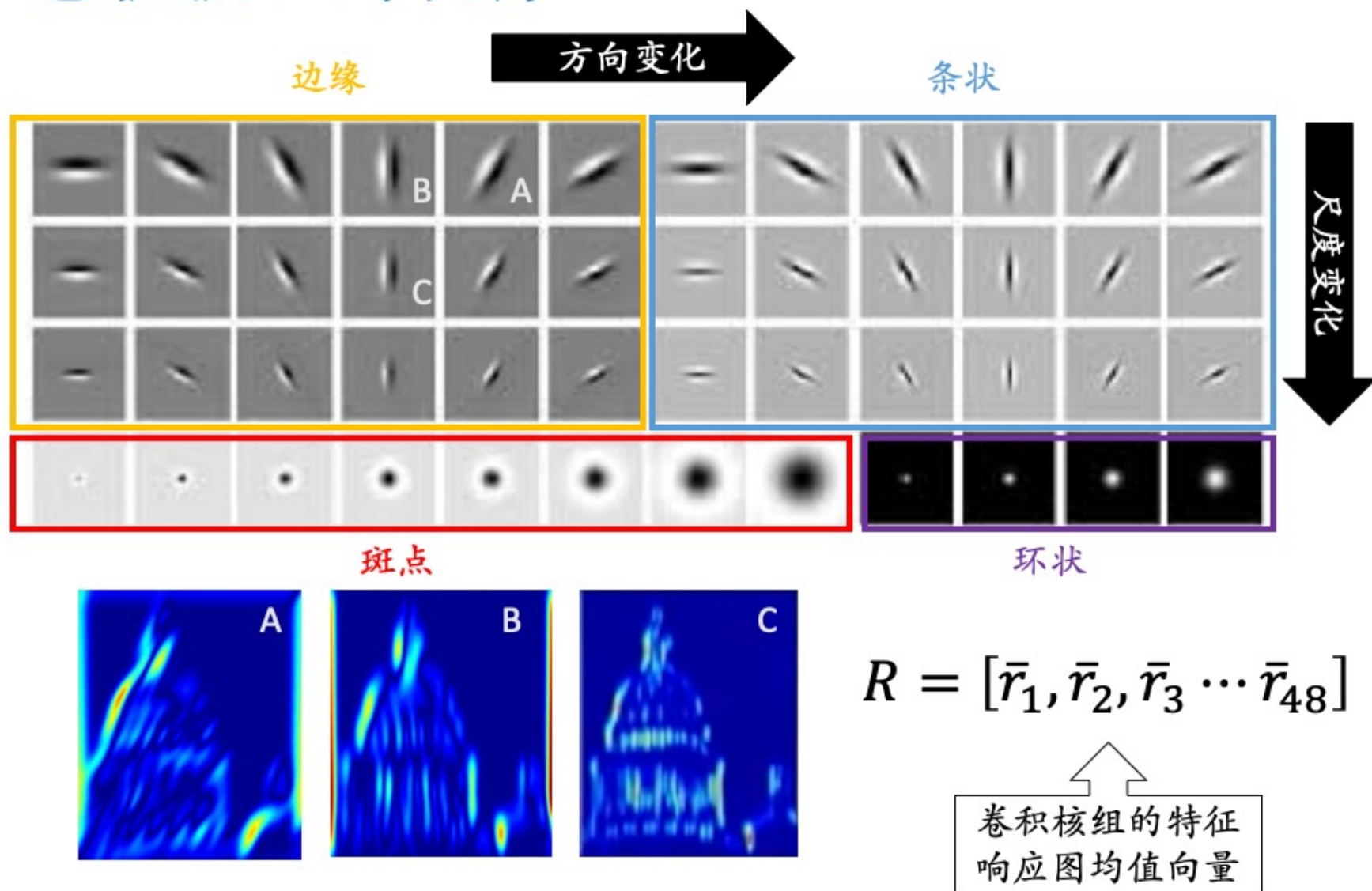
规律多于噪音  
忽略纹理基元的具体位置

# 卷积核组的设计



综合考虑三个需求：特征类型、尺度、方向

# 卷积核组的设计



综合考虑三个需求：特征类型、尺度、方向

# 纹理表示的步骤

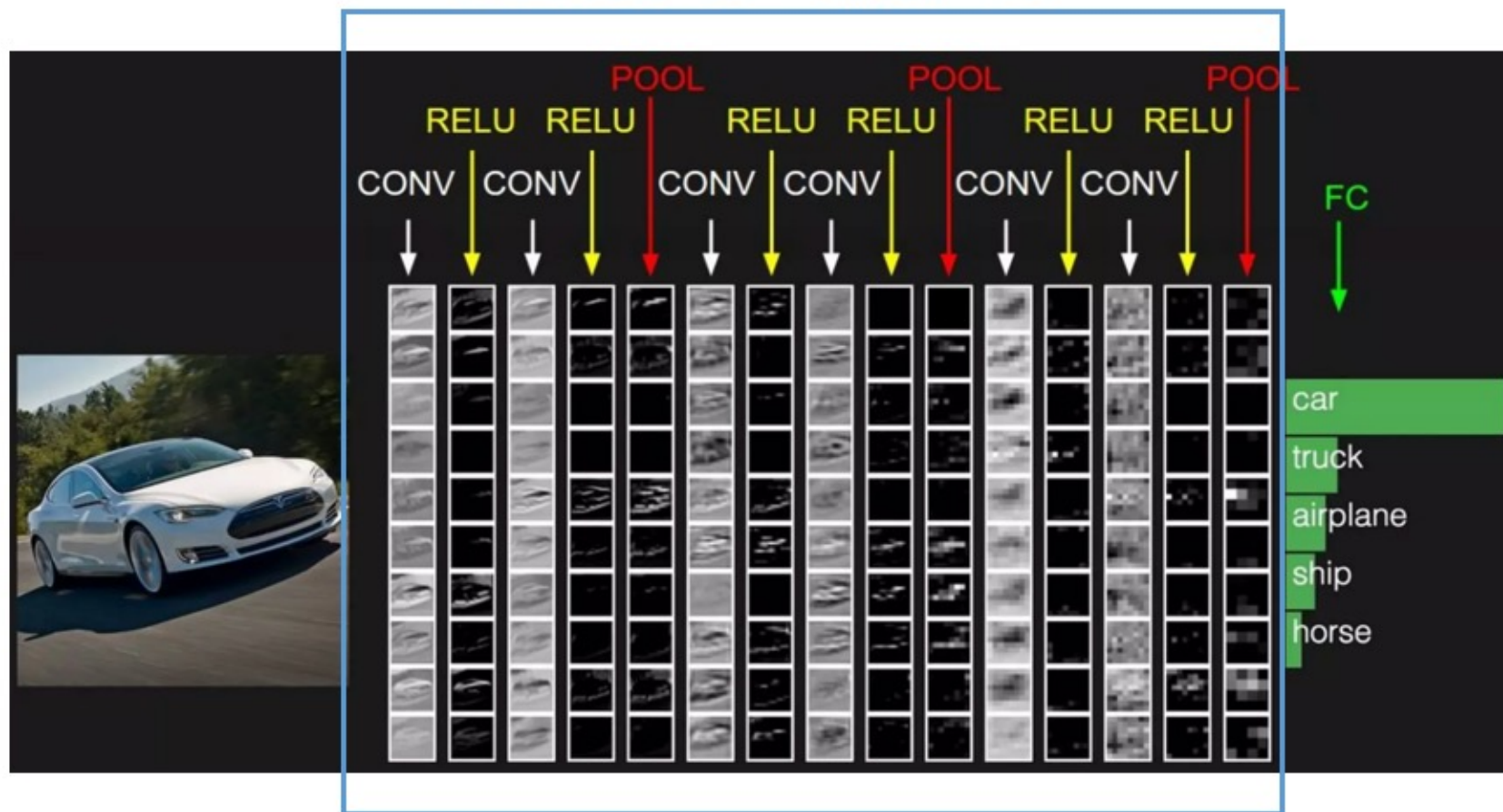
1. 设计卷积核组。
2. 对图像实施卷积操作，描述图像中每个位置存在的纹理基元。
3. 用特征响应图（或其统计值，比如均值）表示图像中的纹理特性。

# 卷积层



# 卷积层

卷积模块的目的：获得特征向量



$$R = [\bar{r}_1, \bar{r}_2, \bar{r}_3 \cdots \bar{r}_n]$$

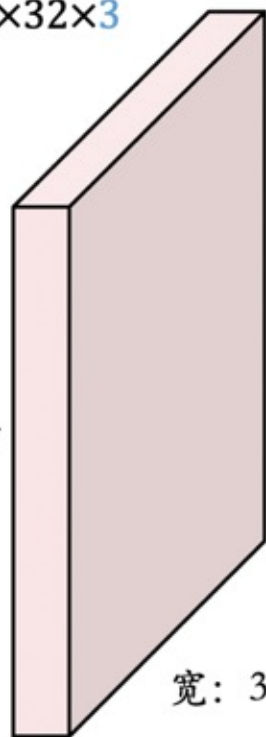
# 卷积层



CNN中，卷积核组的最终形态由训练结果决定（模型参数）。  
每一个卷积层可视为一个卷积核组，其中的卷积核数由人为设定（超参数）。

# 三维卷积核

RGB彩图  
 $32 \times 32 \times 3$



高: 32像素

宽: 32像素

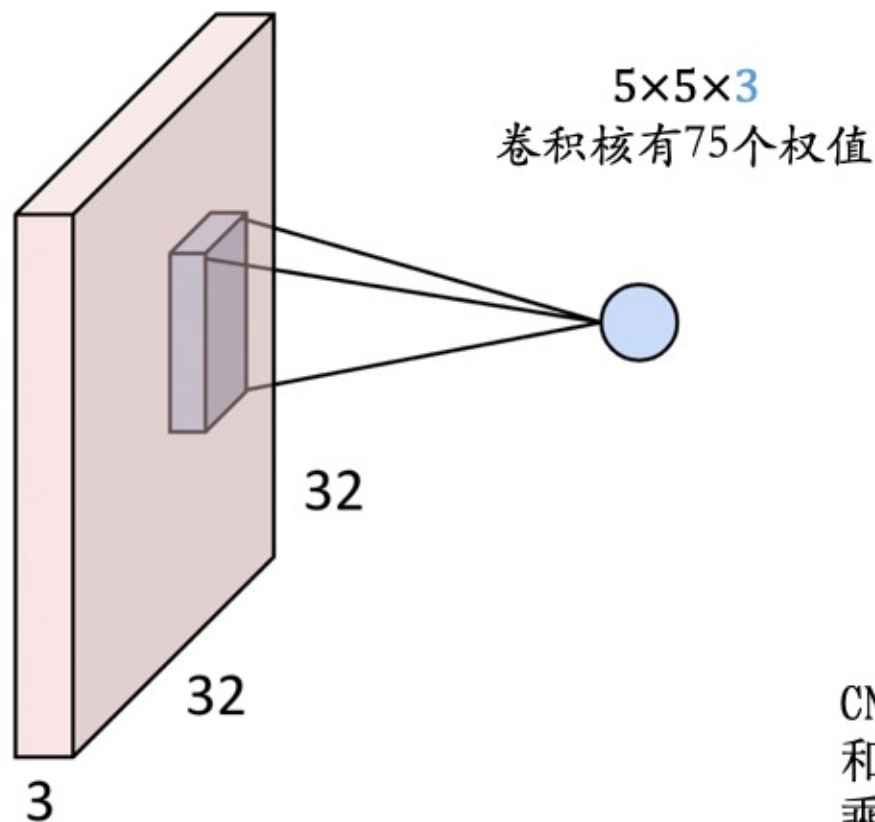
深: 3通道

卷积核  
 $5 \times 5 \times 3$



CNN中，卷积核是3维的，其深度和输入图片（或特征图）深度相同。

# 三维卷积核

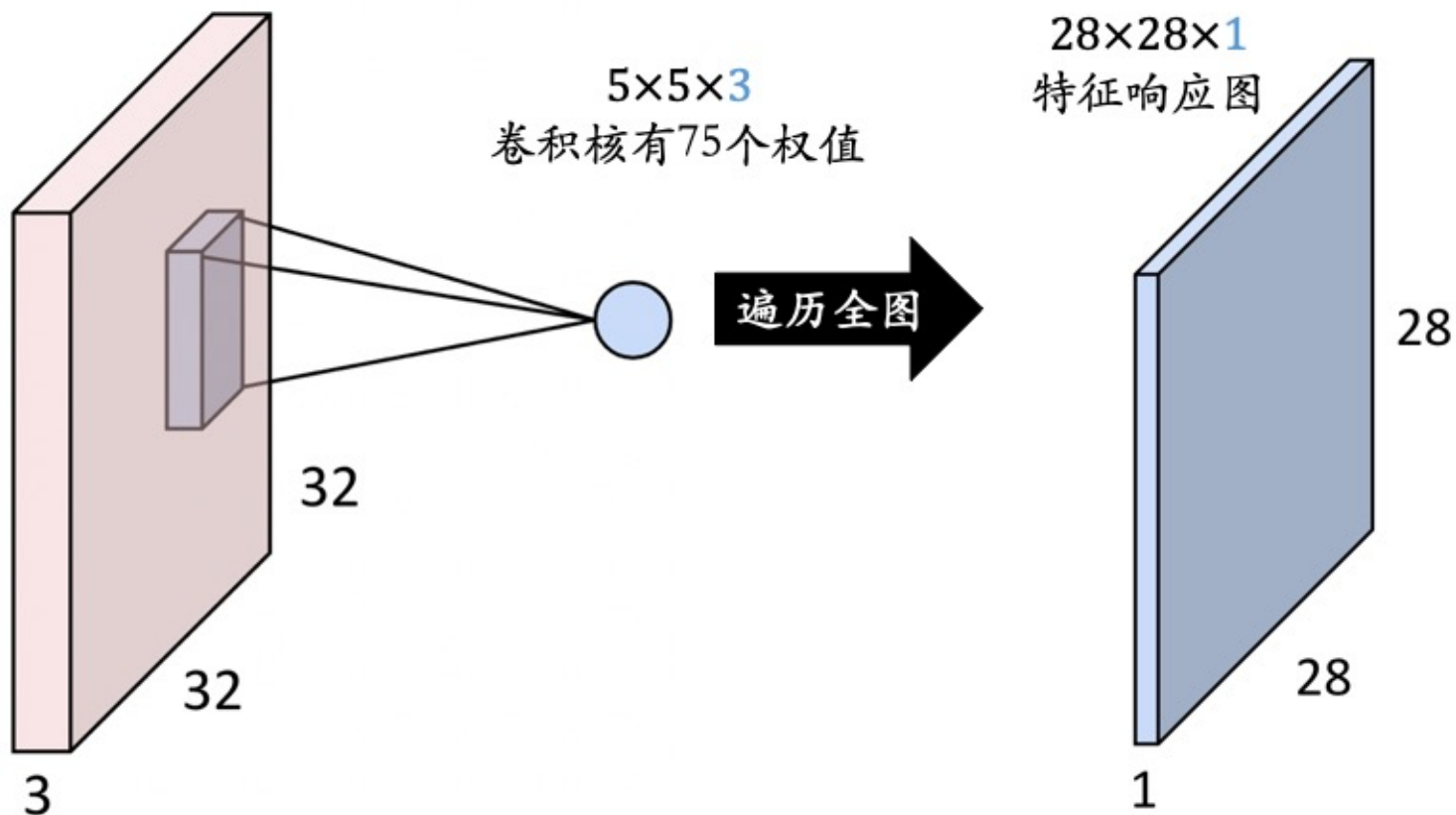


CNN中的三维卷积：卷积核的权值( $W$ )和图像的被卷积区域( $x$ )进行75维点乘，再加上偏置 $b$ ，获得卷积中心点对应位置的特征响应值 $f$ 。

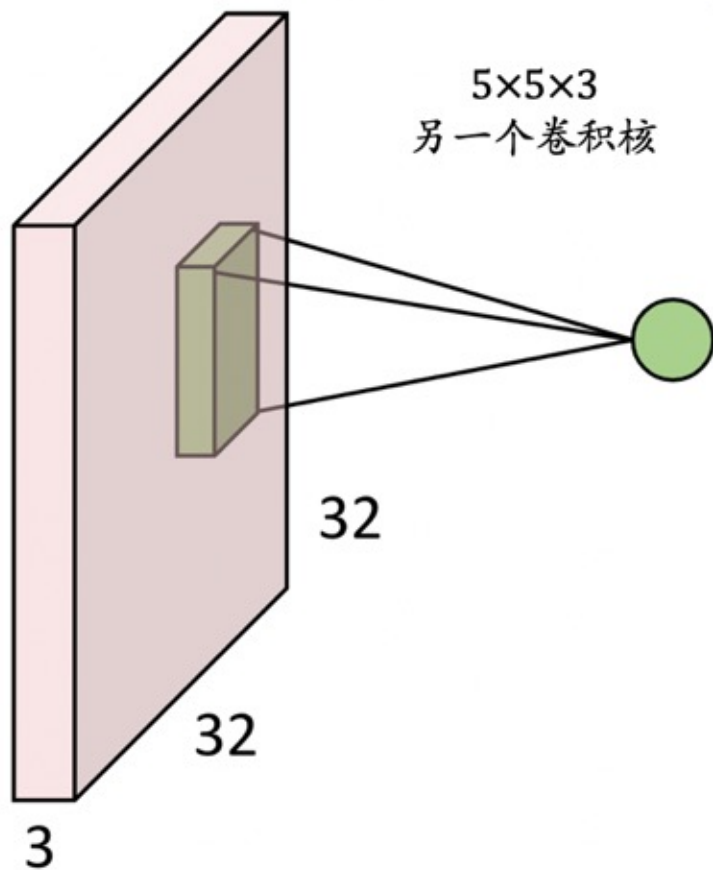
$$f = W^T x + b$$



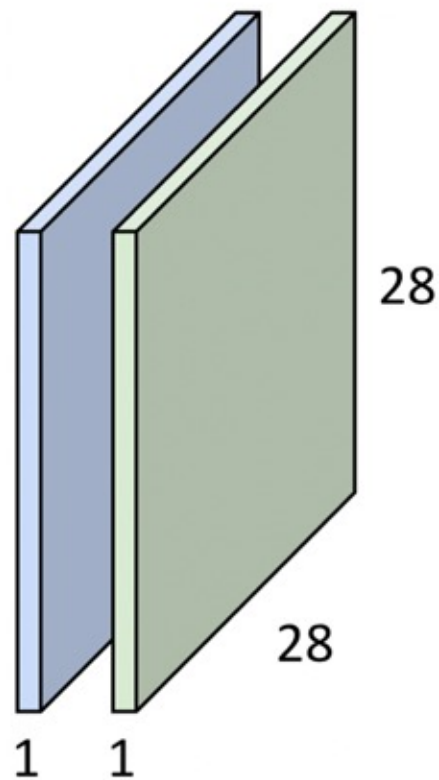
# 三维卷积核



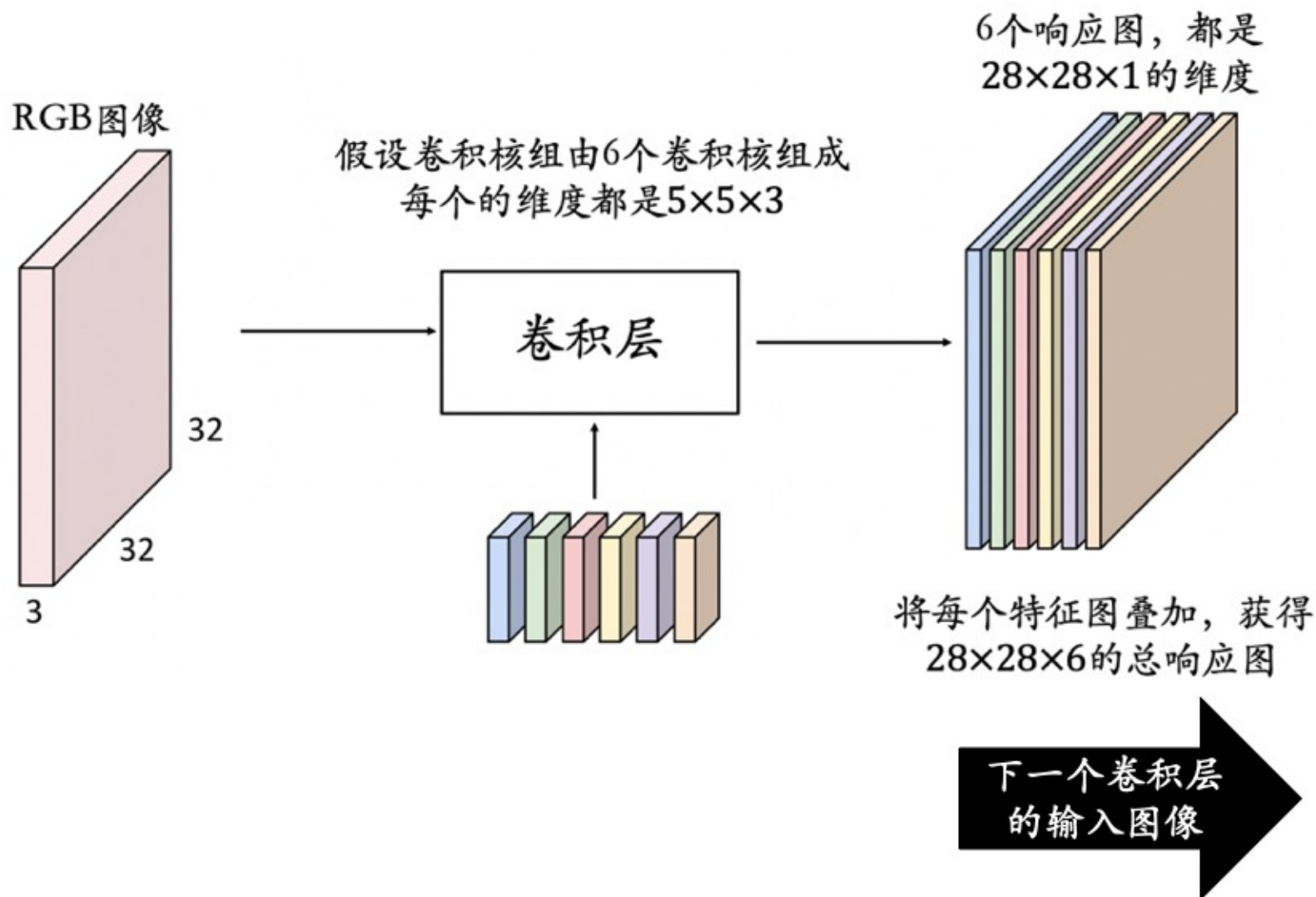
# 三维卷积核



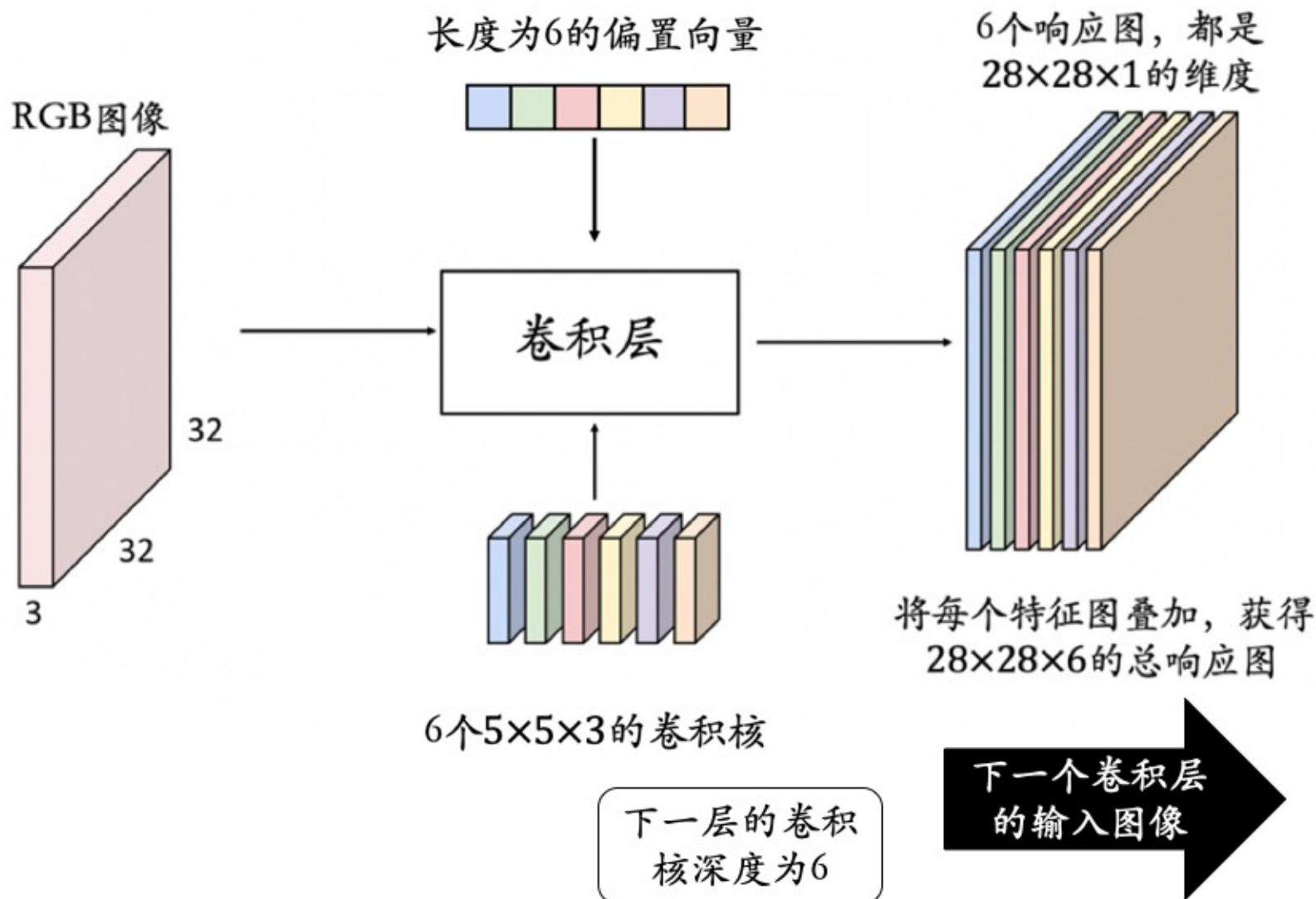
两个特征响应图  
每个卷积核对应一个响应图



# 卷积层的维度

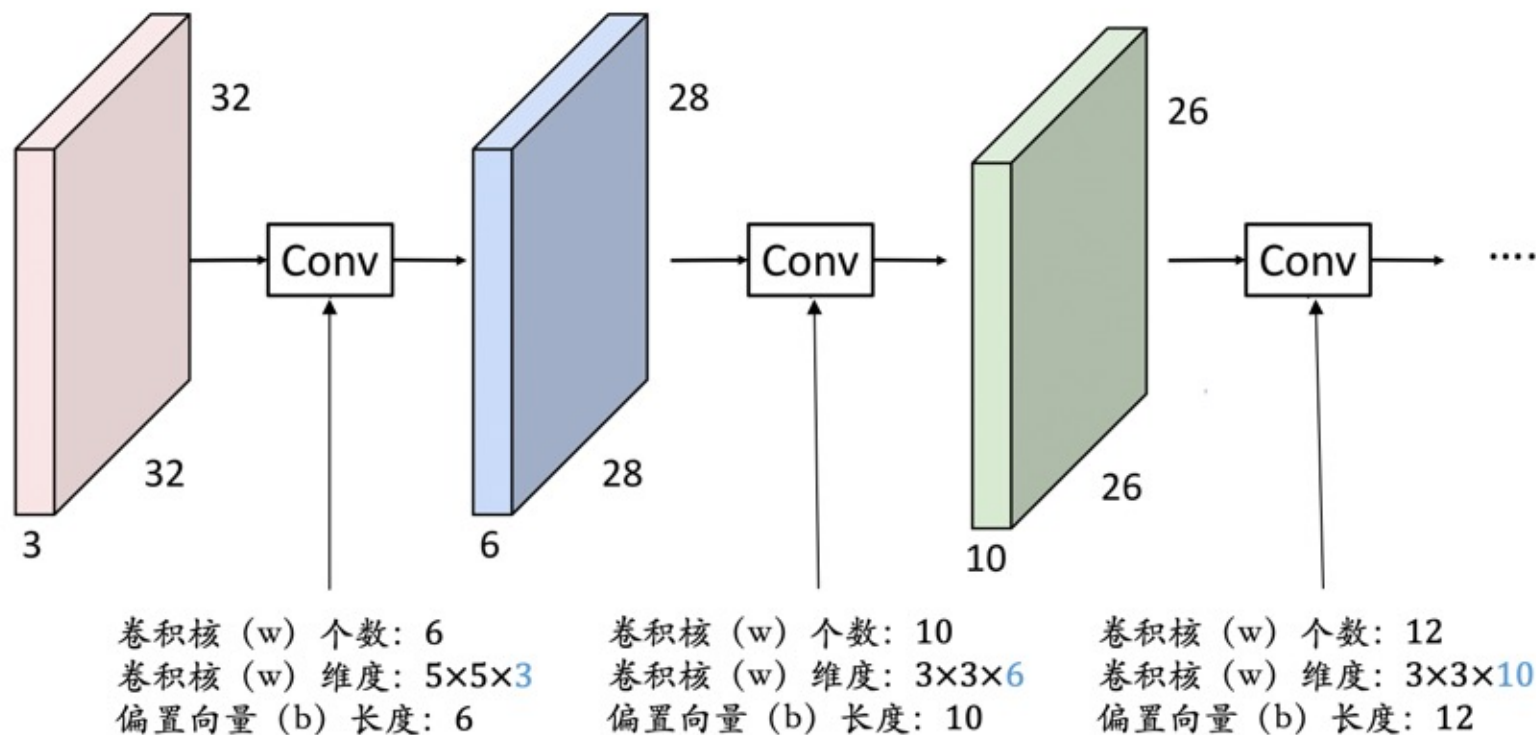


# 卷积层的维度





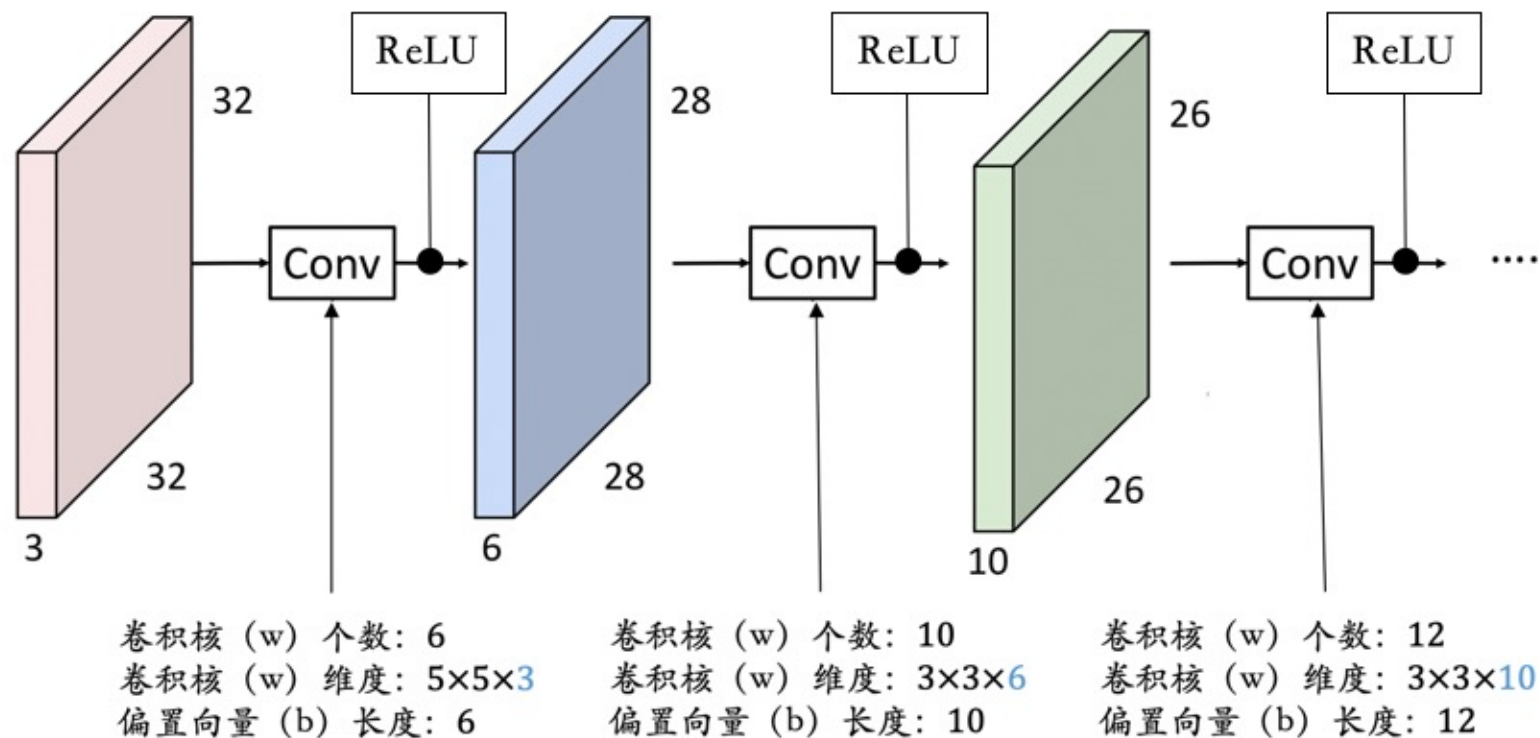
# 卷积层的维度



## 卷积核的深度由什么决定?

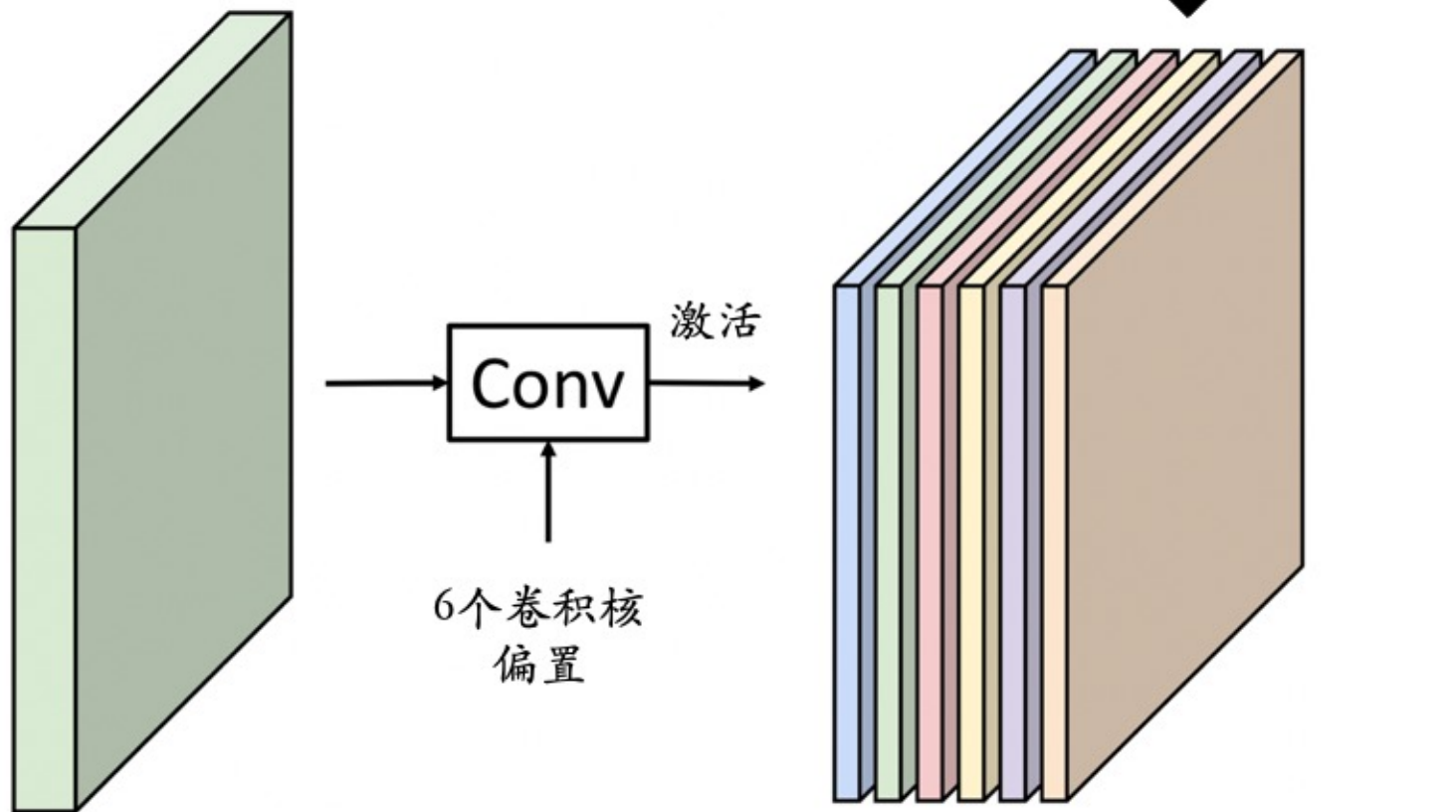
- 第一层: 传入图片的色彩通道 (灰度还是RGB)
- 其它层: 上一层使用了多少个卷积核 (即上一层的特征响应图的深度)

# 卷积层的连接

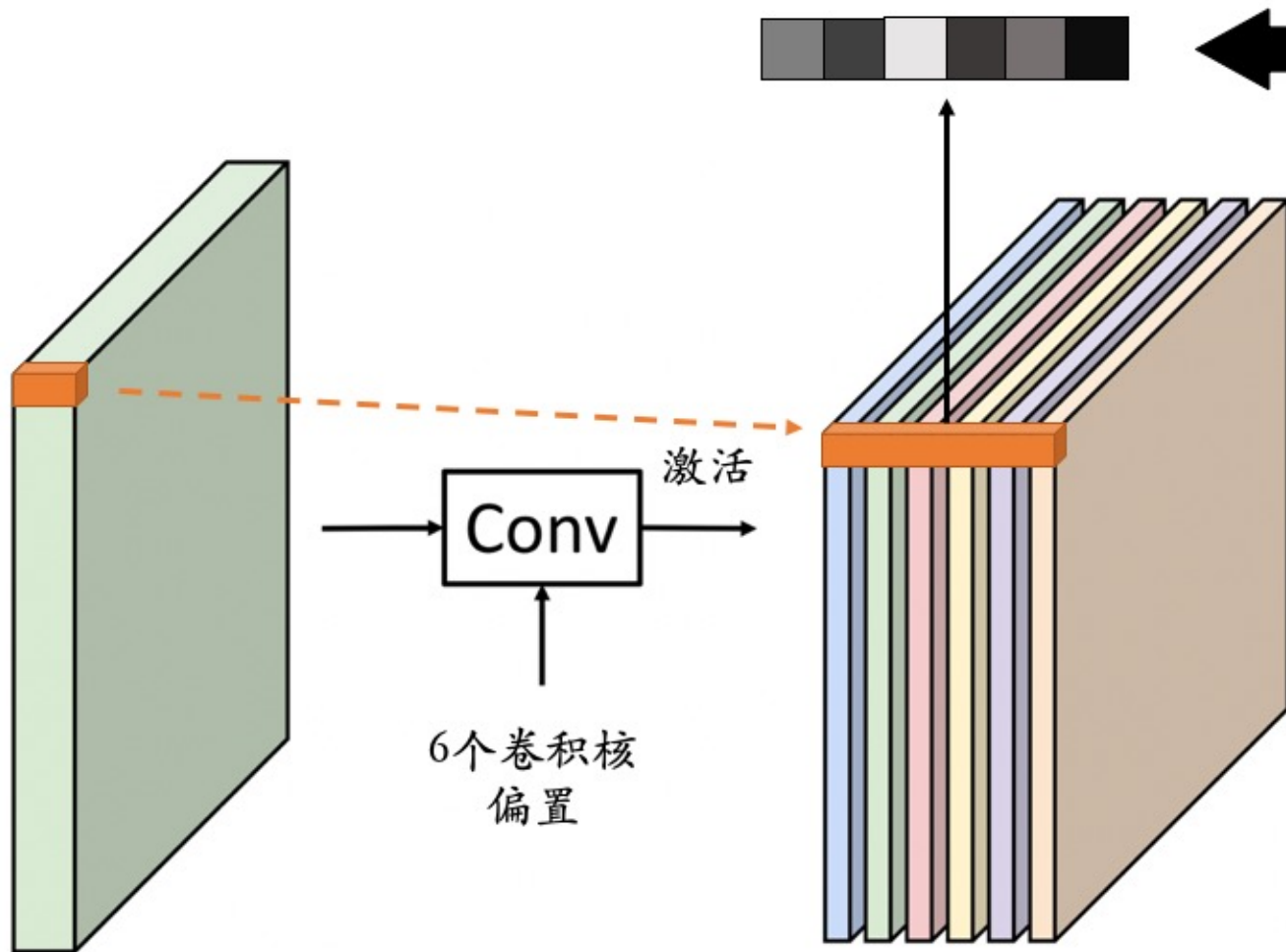


在卷积层之后添加非线性操作 (激活函数/激活层)

# 特征响应图组



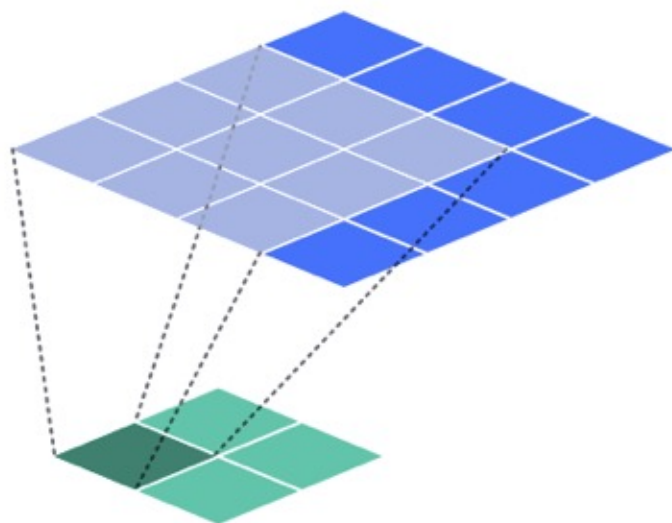
# 特征响应图组



同一特征响应图上不同位置的值表示输入图像上不同位置对同一卷积核的响应结果。



# 卷积步长 Stride



## Parameters

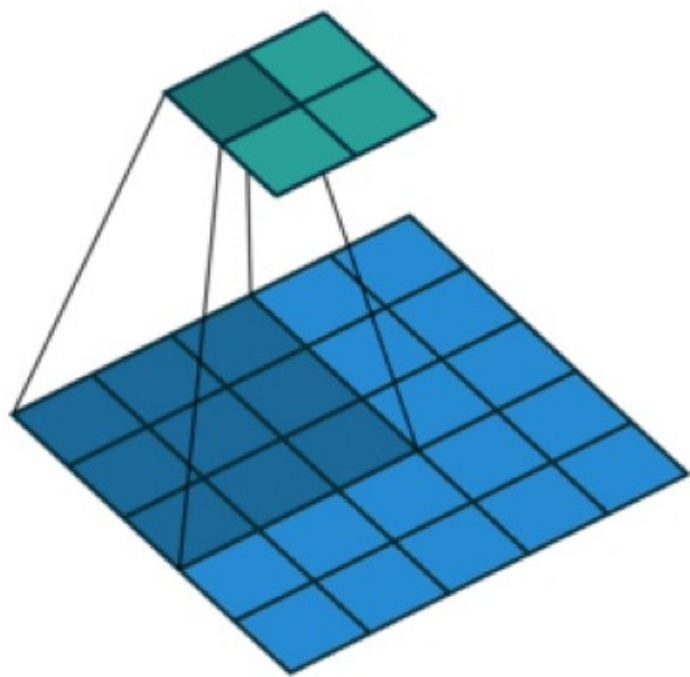
```
kernal: 3  
stride: 1  
padding: 0  
output_padding: 0  
dilation: 1
```

■ Kernel  
■ Output  
■ Input

决定卷积核在输入数据（如图像）上**每次滑动的距离**，是一个**超参数**，其主要作用包括：

- **控制输出特征图的大小**：步长越大，输出特征图的尺寸会越小（空间分辨率越小）。这是因为步长增大导致卷积核在输入数据上的采样点减少。
- **减小计算量**：通过增大步长，可以减少卷积操作的次数，从而降低模型的计算复杂度。
- **降采样**：在不添加池化层的情况下，直接通过调整步长也可以实现降采样的效果，进一步减小参数数量和计算量。

# 卷积步长 Stride



输入数据矩阵平面尺度:  $w_1 \times h_1$   
输出响应图的平面尺度:  $w_2 \times h_2$   
卷积核边长:  $k$   
卷积步长:  $s$

在CNN模型中，每一个卷积层需要指定卷积步长。

卷积步长、输入矩阵和输出响应图之间满足如下维度关系：

$$w_2 = \left\lfloor \frac{w_1 - k}{s} \right\rfloor + 1$$

$$h_2 = \left\lfloor \frac{h_1 - k}{s} \right\rfloor + 1$$

# 边缘填充 Padding

0	0	0	0	0	0	0	0	0
0								0
0								0
0								0
0								0
0								0
0								0
0								0
0	0	0	0	0	0	0	0	0

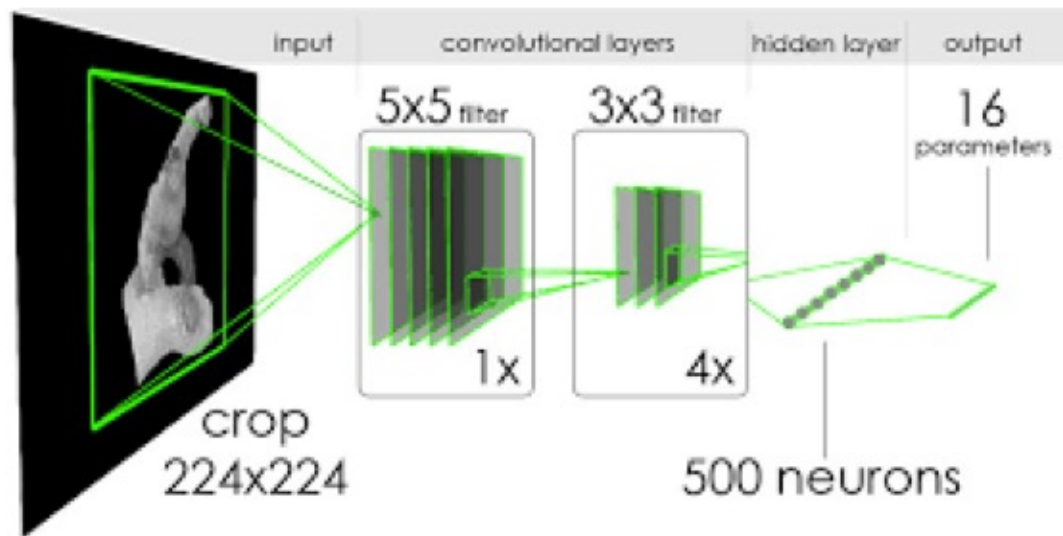
如果不希望卷积后图像变小，  
则要添加的Padding层数为：

$$p = \frac{k - 1}{2}$$

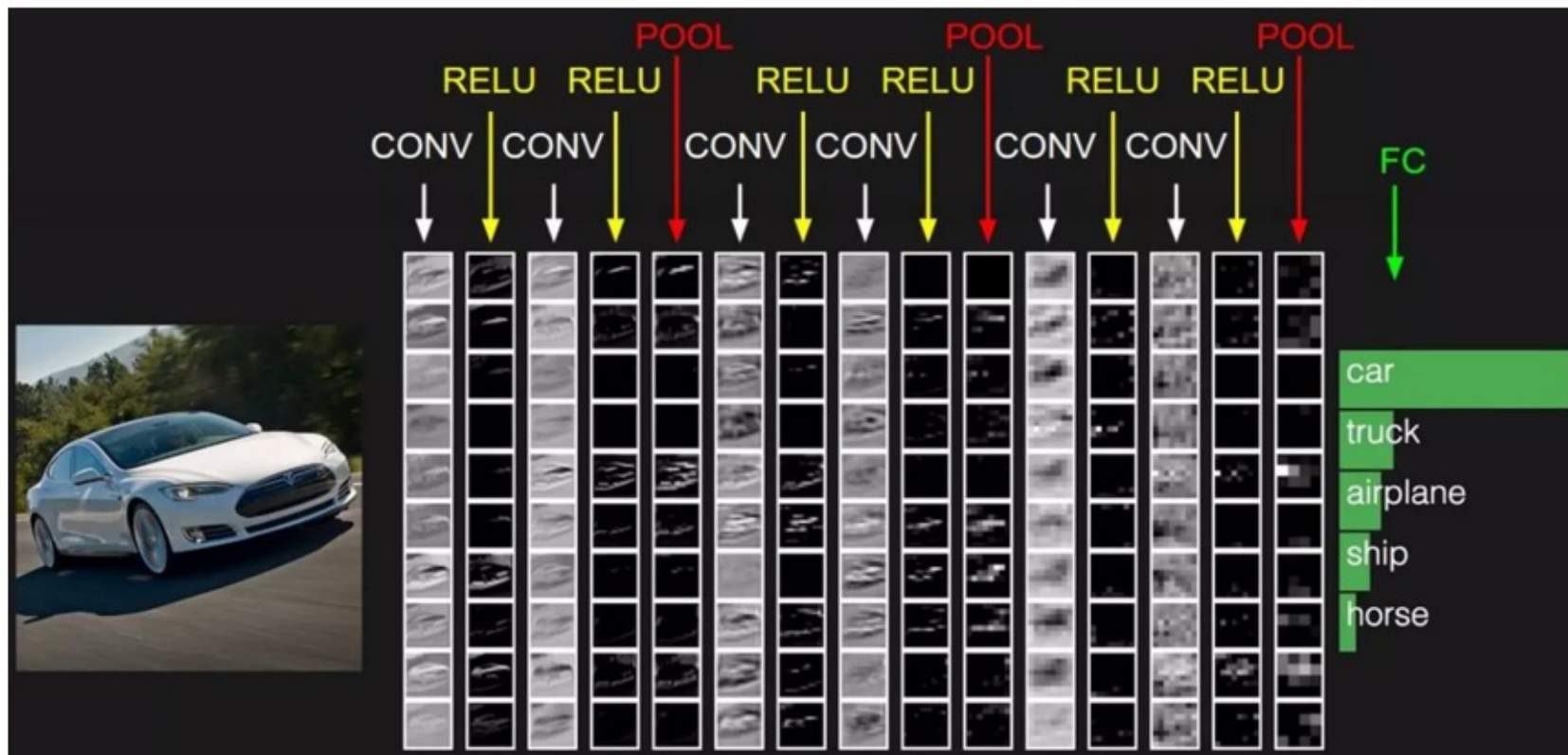
其中 $k$ 为卷积核的边长。

# 常见的图片尺寸

- CNN常见输入图像的平面边长：
  - 64、128、224、227等
- 如果数据集中图像尺寸不满足输入要求，通常需要对图像进行预处理（裁剪、降采样等）。







CONV: 卷积层

RELU: ReLU层 (激活层)

POOL: 池化层

FC: 全连接层

池化层

## 池化操作 Pooling

- 池化操作实际上是一种降采样的方法，主要用于减少数据的空间大小，从而减少计算量、减轻过拟合的风险，同时还可以提高特征的稳健性。
- 在卷积神经网络中，池化层通常位于卷积层后面。卷积层负责从输入数据中提取特征，而池化层则负责对这些特征进行降采样。

# 池化的作用

## 特征降维

- 通过减少数据的空间大小，可以降低模型的复杂性，加快计算速度，并减少所需的存储空间。

## 扩大卷积核的“视野”

- 在卷积核大小相同的情况下，通过池化缩小特征图组的空间尺寸，能让卷积核覆盖更宏观的特征，识别粗粒度的要素。

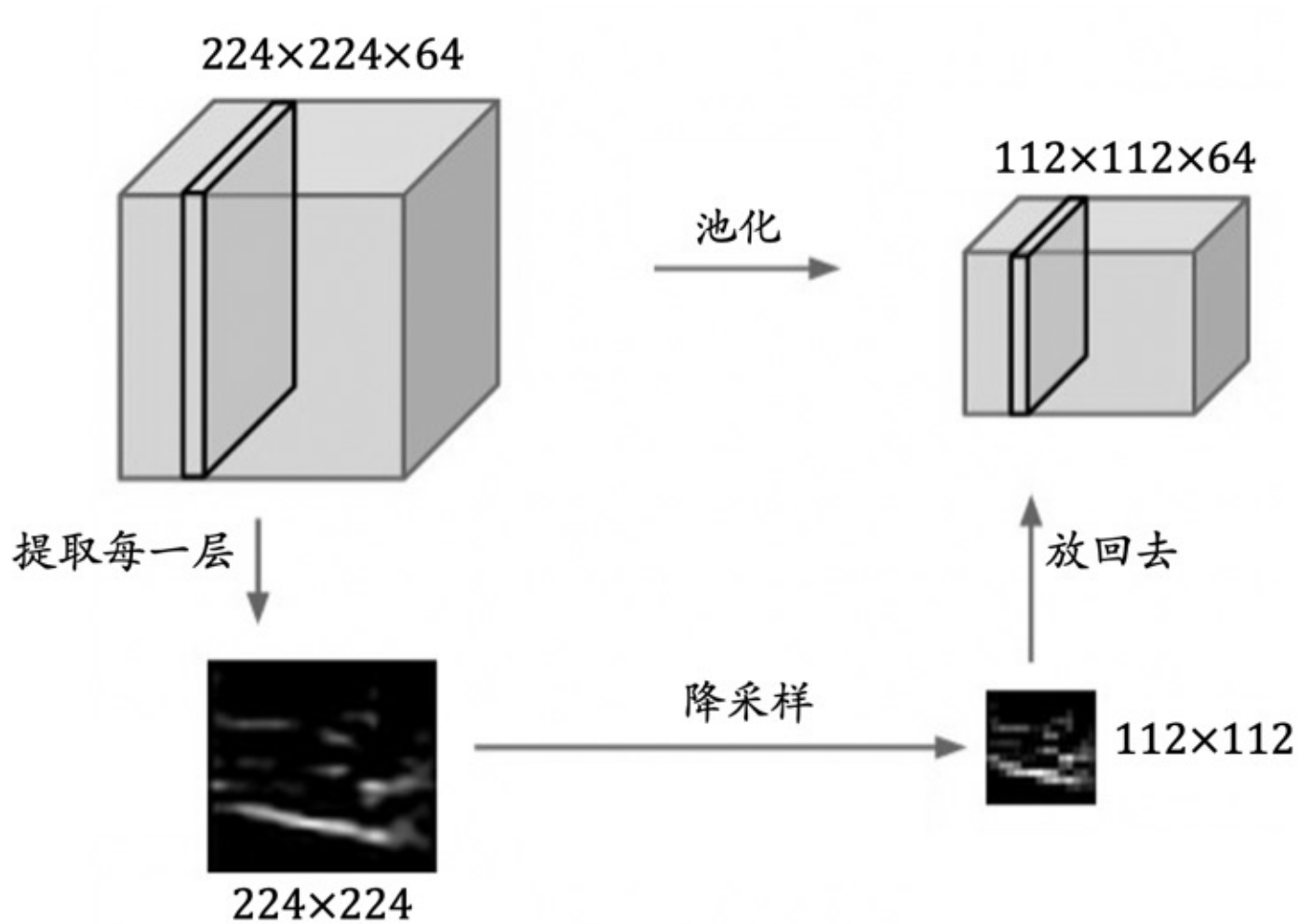
## 防止过拟合

- 池化操作可以视为一种正则化的形式，有助于减少模型对训练数据的过度拟合。

## 满足平移不变性

- 池化操作可以帮助模型学习到与特征位置无关的信息，从而提高模型的泛化能力。

# 池化的操作





# 池化的操作

- 最大池化 (**max pooling**)：在选定的池化窗口内，选择最大的值作为输出。这种方法可以突出最重要的特征，并去除一些噪声。
- 平均池化 (**average pooling**)：计算池化窗口内所有值的均值作为输出。这种方法可以提供更平滑的特征表示。

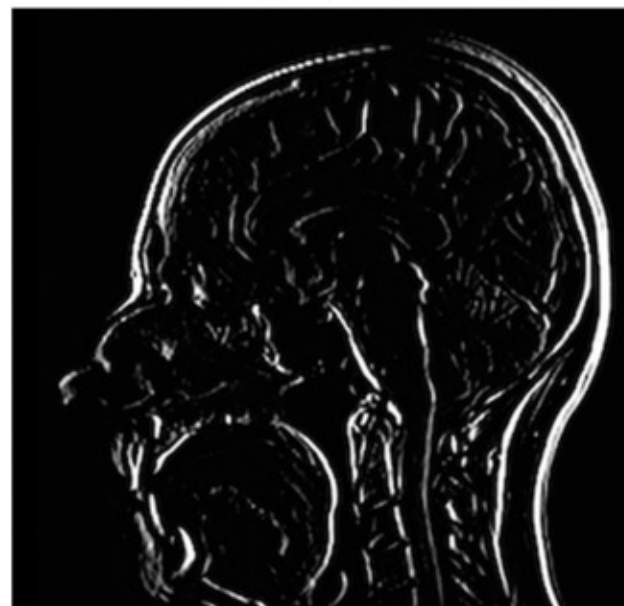


# 池化效果

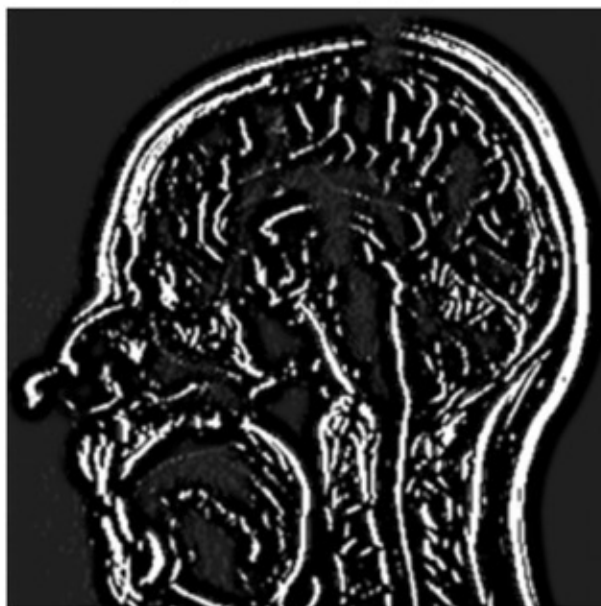
最大池化更常见，因为它能对特征图做**非最大抑制**（即：相邻位置的响应值“赢者通吃”）。

- 只保留最强烈的响应信号，突出局部重点。
- 最明显的特征有微小的空间位移，但不影响图像的总体结构。

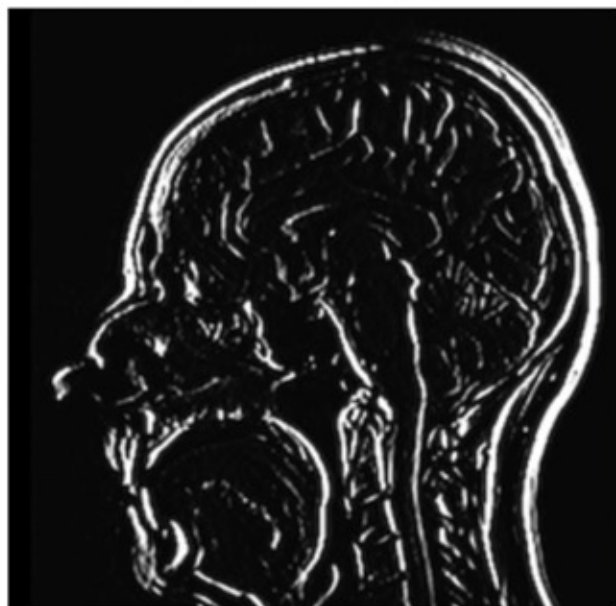
医用影像卷积特征图



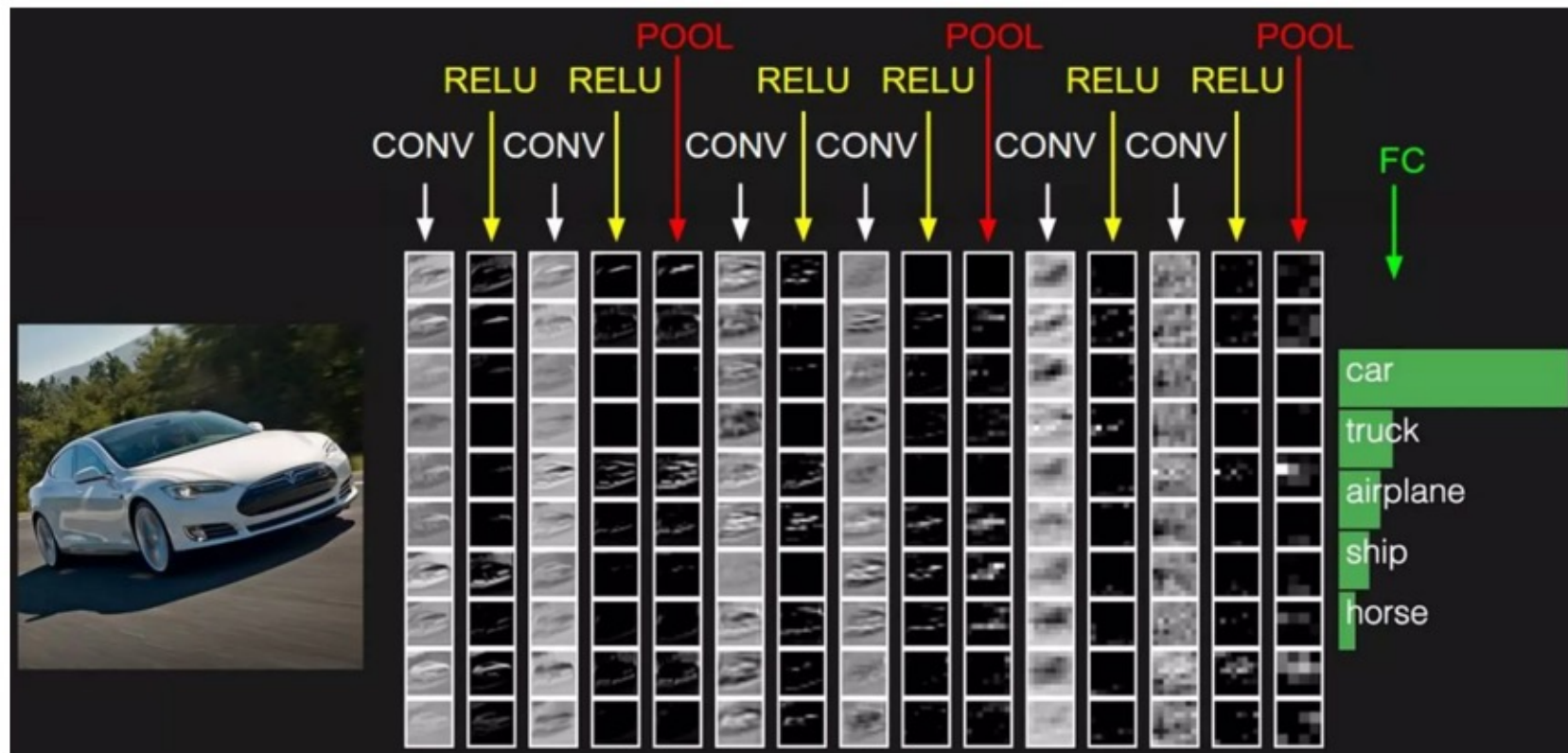
最大池化效果



平均池化效果



# 对接全连接层



CONV: 卷积层

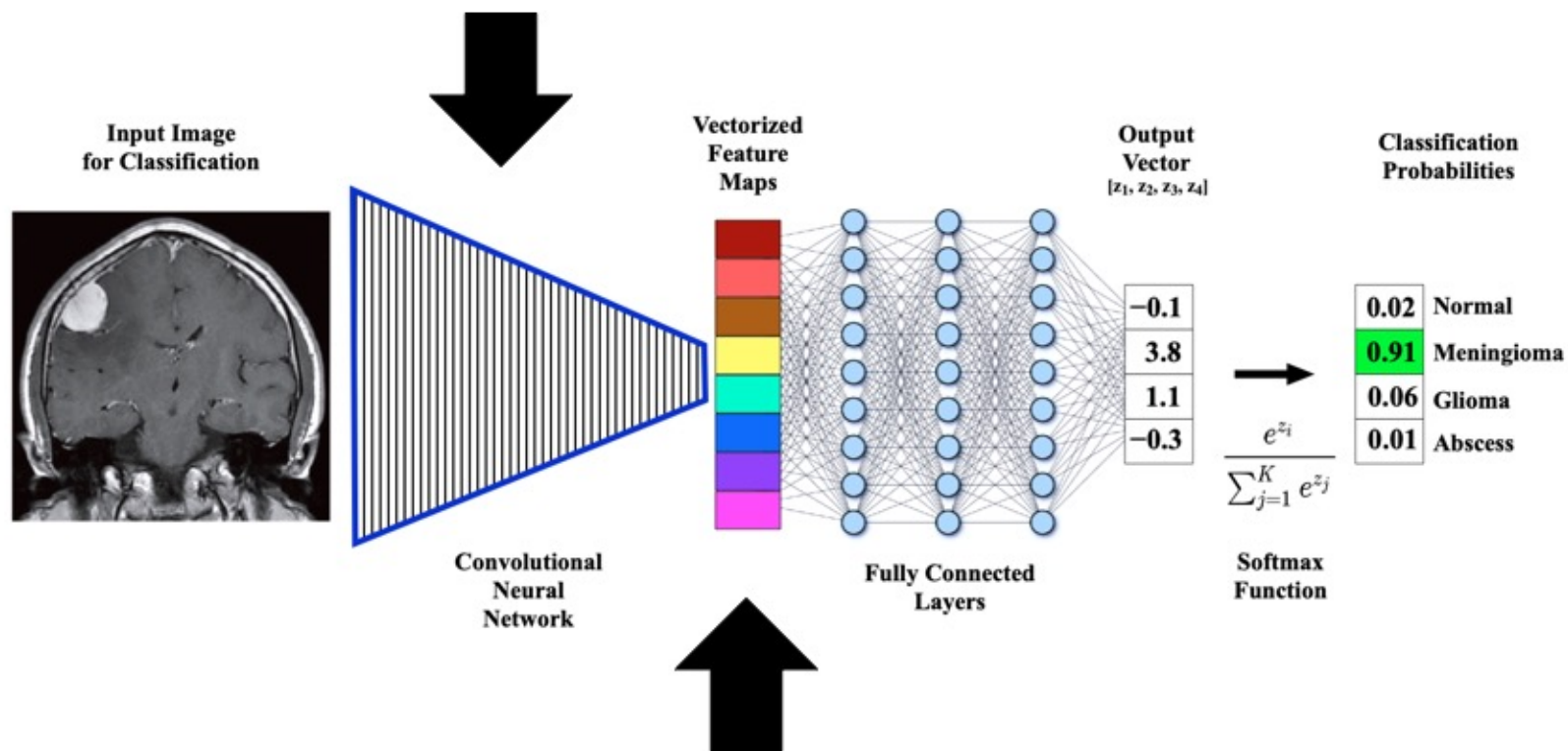
RELU: ReLU层 (激活层)

POOL: 池化层

FC: 全连接层

# 方案1

所有卷积层、激活层、池化层  
看作是一个大的卷积层

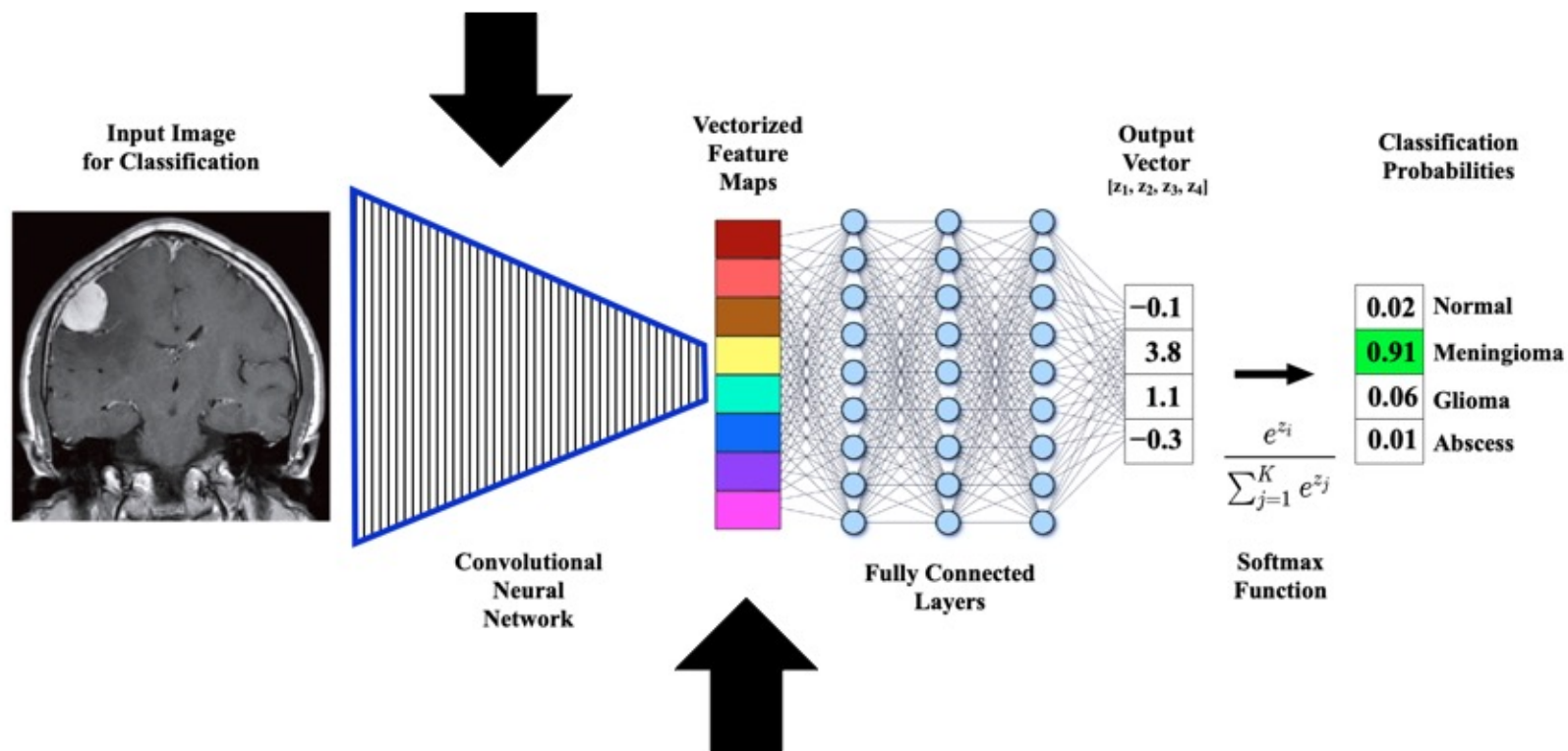


把最后一个池化层的输出结果 ( $w \times h \times d$ )  
直接展开为一维向量



# 方案2

所有卷积层、激活层、池化层  
看作是一个大的卷积层



得到最后一个池化层的输出结果 ( $w \times h \times d$ ) 后  
对特征图的每一层取均值  
把均值组织成一维向量

# 样本增强



# 样本增强 Data augmentation

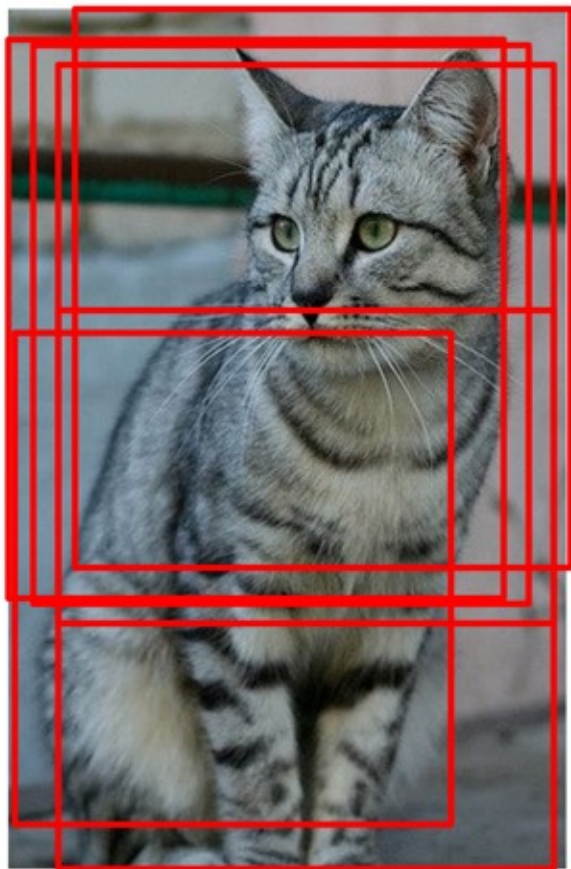
增加数据量、缓解过拟合！

## 镜像翻转 Flip



相似的：旋转

# 切割/缩放 Crop/Scale



# 色彩抖动 Color jitter



模拟强光、弱光、高光、低光  
改变白平衡、在RGB层面上添加色彩偏移等



# 综合不同的方法

- 镜像翻转、旋转
- 切割和缩放
- 色彩抖动
- 镜头畸变、模糊
- 拉伸和剪切
- .....

