



The Introduction To Artificial Intelligence

**Yuni Zeng yunizeng@zstu.edu.cn
2025-2026-1**

The Introduction To Artificial Intelligence

- Part I Brief Introduction to AI & Different AI tribes
- Part II Knowledge Representation & Reasoning
- Part III AI GAMES and Searching
- Part IV Model Evaluation and Selection
- Part V Machine Learning
- ✚ Part VI Neural Networks/ Deep Learning

深度学习

- 卷积神经网络



卷积神经网络

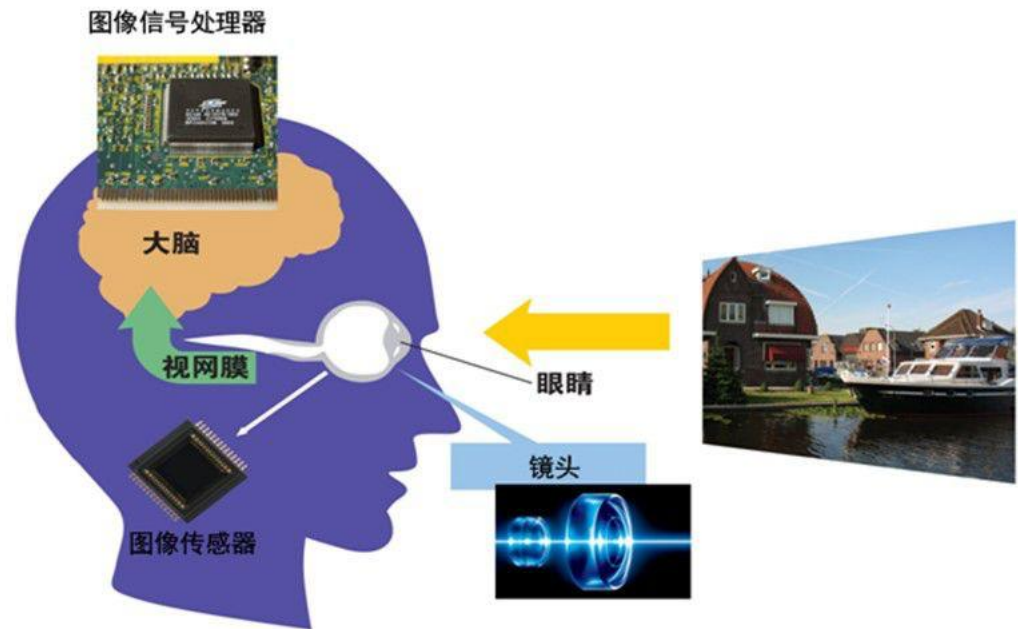
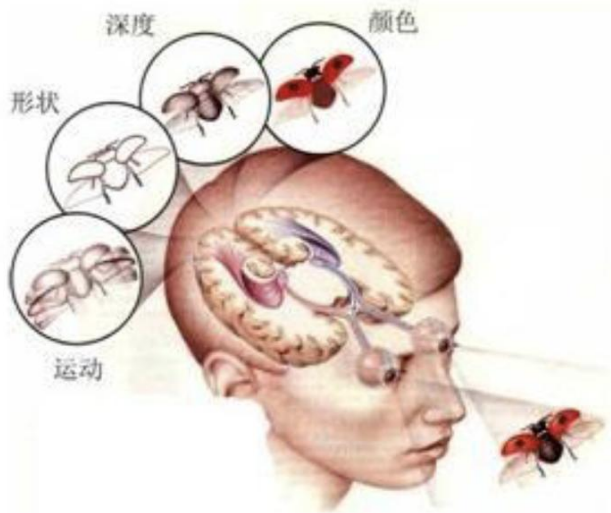
- 1989年，计算机科学家Yann LeCun提出了CNN；他是第一个通过卷积神经网络在MNIST数据集上解决手写数字问题的人

卷积神经网络

- 1962年Hubel和Wiesel通过对猫视觉皮层细胞的研究，提出了感受野(receptive field)的概念。视觉皮层的神经元就是局部接受信息的，只受某些特定区域刺激的响应，而不是对全局图像进行感知。
- 1984年日本学者Fukushima基于感受野概念提出神经认知机(neocognitron)。
- CNN可看作是神经认知机的推广形式。

卷积神经网络

- 人在认知图像时是分层抽象的。首先理解的是颜色和亮度，然后是边缘、角点和直线等局部细节特征；接下来是纹理、几何形状等更复杂的信息和结构，最后形成整个物体的概念

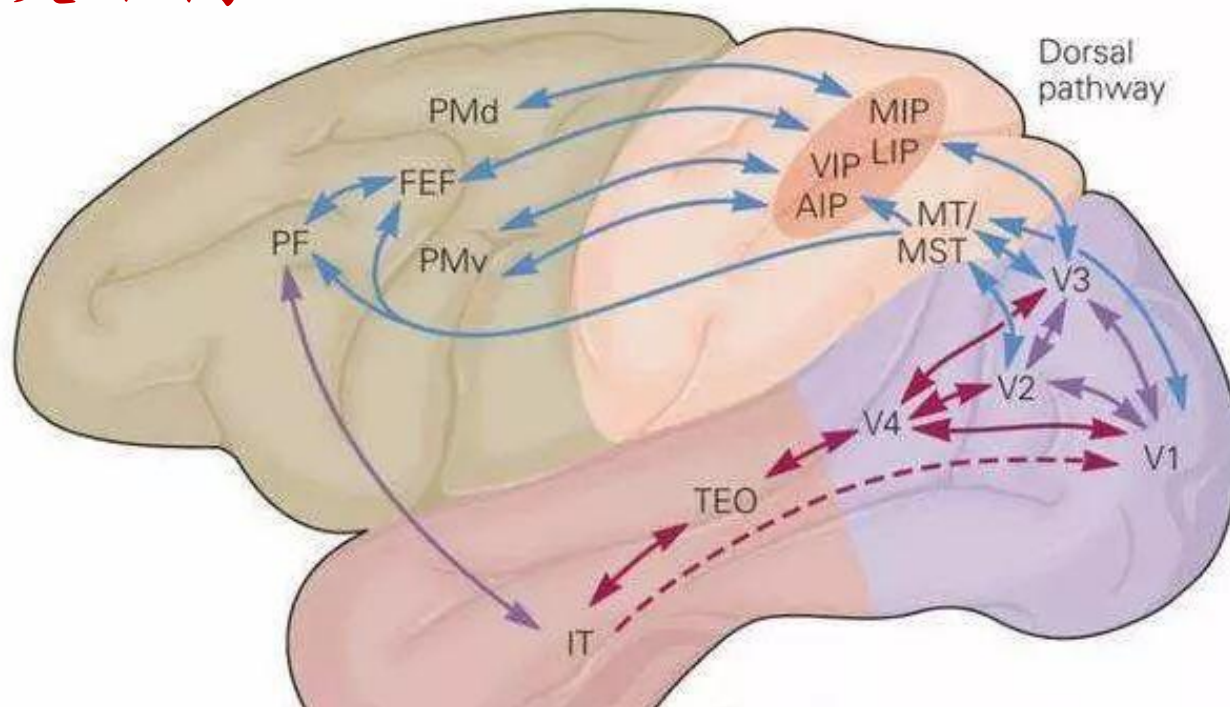


卷积神经网络

- 视觉神经科学（Visual Neuroscience）对于视觉机理的研究验证了这一结论，**动物大脑的视觉皮层具有分层结构**。眼睛将看到的景象成像在视网膜上，视网膜把光学信号转换成电信号，传递到大脑视觉皮层（Visual cortex），视觉皮层是大脑中负责处理视觉信号的部分

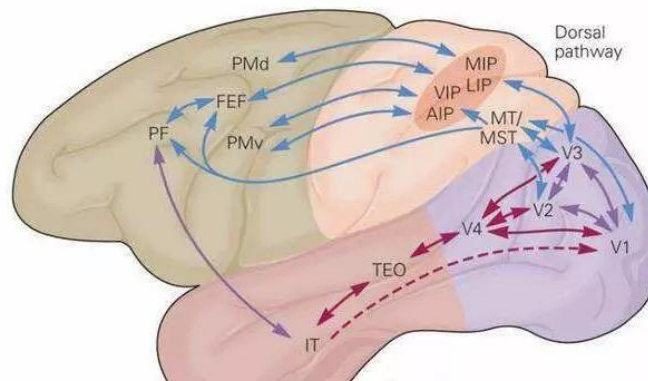
卷积神经网络

- 1981年，诺贝尔医学奖获得者、美国神经生物学家David Hubel 和Torsten Wiesel发现了视觉系统的信息处理是分级的，即视觉皮层具有层次结构。



卷积神经网络

- 从视网膜传来的信号首先到达初级视觉皮层，即V1皮层。V1皮层简单神经元对一些细节、特定方向的图像信号敏感。V1皮层处理之后，将信号传导到V2皮层。V2皮层将边缘和轮廓信息表示成简单形状，然后由V4皮层中的神经元进行处理，它颜色信息敏感。复杂物体最终在IT皮层（inferior temporal cortex）被表示出来。



卷积神经网络

- 卷积神经网络可以看成是上面这种机制的简单模仿。
- 典型的卷积神经网络由卷积层、池化层、全连接层构成。

卷积神经网络

- 卷积层是卷积神经网络的核心。卷积层的重要运算是**卷积运算**。
- **卷积**在图像处理领域是一种常用的运算。它被用于**图像去噪、增强、边缘检测**等问题，还可以**提取图像的特征**。**卷积运算**用一个称为**卷积核**的矩阵自上而下、自左向右在图像上滑动，将卷积核矩阵的各个元素与它在图像上覆盖的对应位置的元素相乘，然后求和，得到输出像素值。

卷积神经网络

■ 卷积操作

1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved
Feature

- 左图展示了一个3*3的卷积核在5*5的图像上做卷积的过程。
- 对图像用一个卷积核进行卷积运算，实际上是一个滤波的过程。每个卷积核都是一种特征提取方式，就像是一个筛子，将图像中符合条件的部分筛选出来。

卷积神经网络

□ Classical Architecture of CNN

These are the network parameters to be learned.

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

1	-1	-1
-1	1	-1
-1	-1	1

Kernel/Filter

Each filter detects a small pattern (3 x 3).

卷积神经网络

□ Classical Architecture of CNN

stride=1

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

Dot
product



3

-1

1	-1	-1
-1	1	-1
-1	-1	1

kernel

卷积神经网络

□ Classical Architecture

stride=1

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

1	-1	-1
-1	1	-1
-1	-1	1

kernel

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

卷积神经网络

□ Classical Architecture of CNN

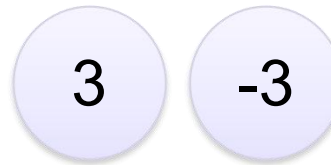
If stride=2

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

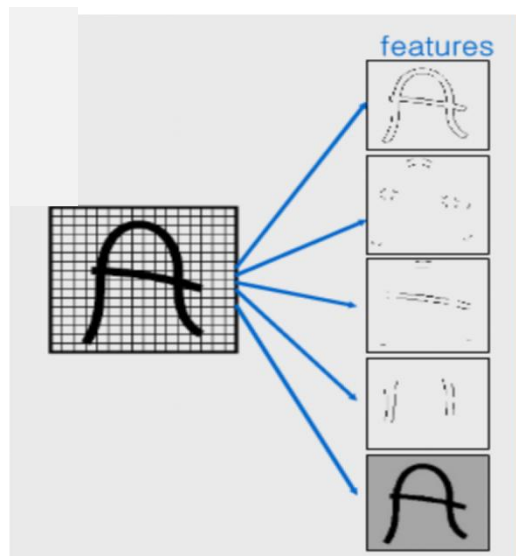
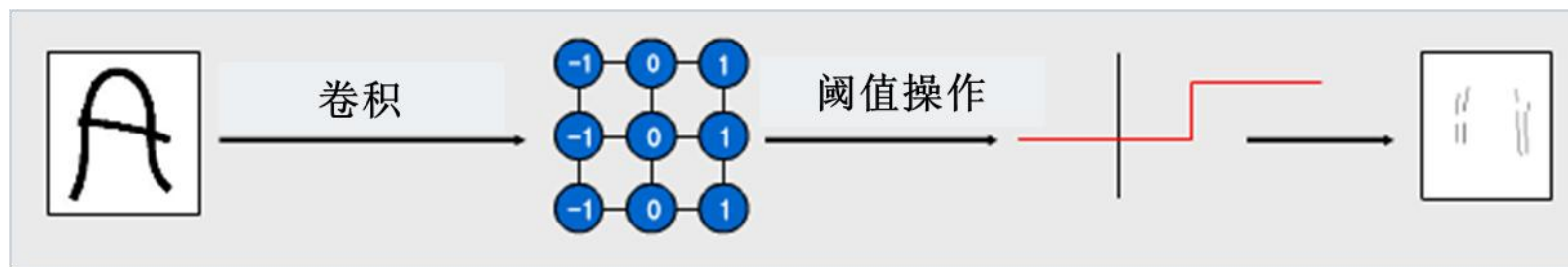
1	-1	-1
-1	1	-1
-1	-1	1

kernel



卷积神经网络

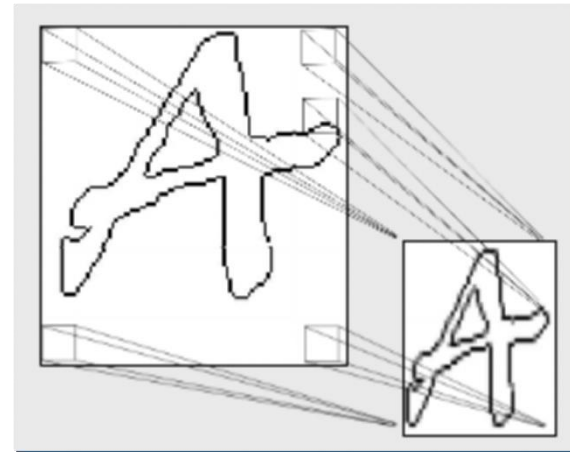
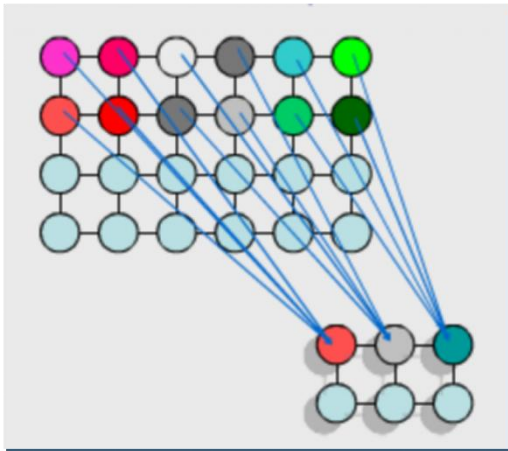
■ 卷积操作



- 大部分的特征提取都依赖于卷积运算
- 利用卷积算子对图像进行滤波，可以得到显著的边缘特征。

卷积神经网络

- 卷积运算主要是对输入图像进行降维和特征抽取，但特征图像的维数还是很高。维数高不仅计算耗时，而且容易导致过拟合。为此引入了下采样技术，也称为pooling即池化操作



卷积神经网络

■ 池化操作

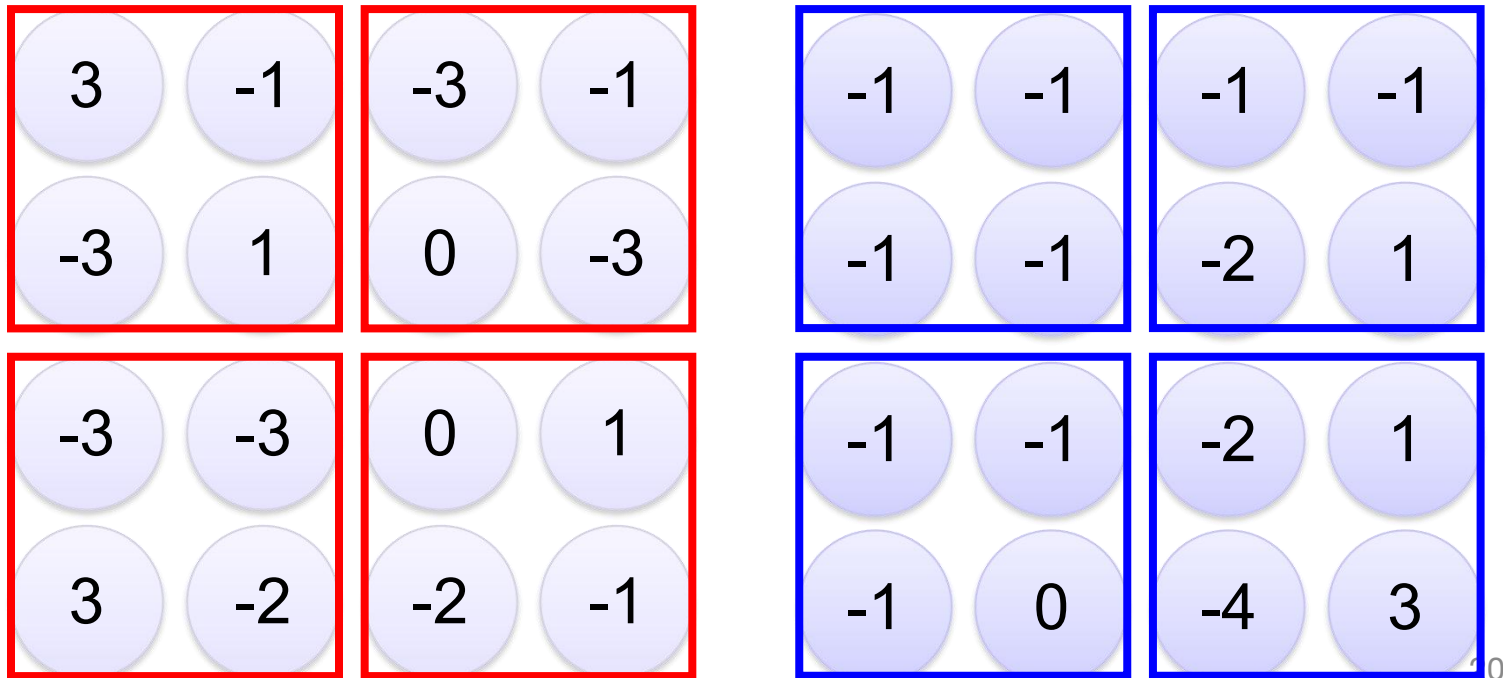
池化的做法是对图像的某一个区域用一个值代替，如最大值或平均值。如果采用最大值，叫做max池化；如果采用均值，叫做均值池化。



卷积神经网络

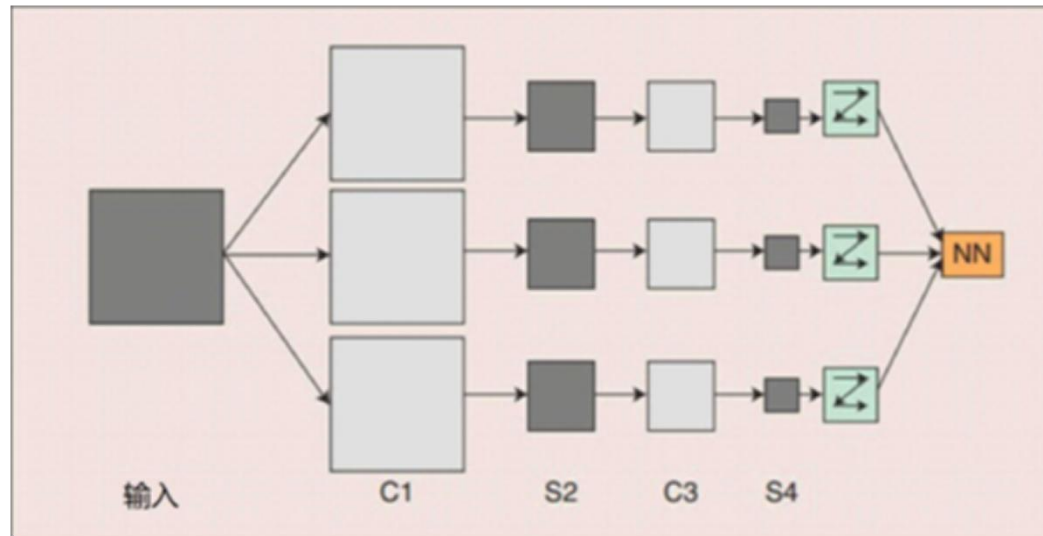
■ 池化操作

池化层的具体实现是在进行卷积操作之后对得到的特征图像进行分块，图像被划分成的不相交块，计算这些块内的最大值或平均值，得到池化后的图像



卷积神经网络

- CNN是一个**多层的神经网络**，每层由多个二维平面组成，而每个平面由多个独立神经元组成。
- C层为特征提取层（卷积层）
- S层是特征映射层（池化层）。
- CNN中的**每一个C层都紧跟着一个S层**。



卷积神经网络

给定以下输入X和filter,

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad w = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

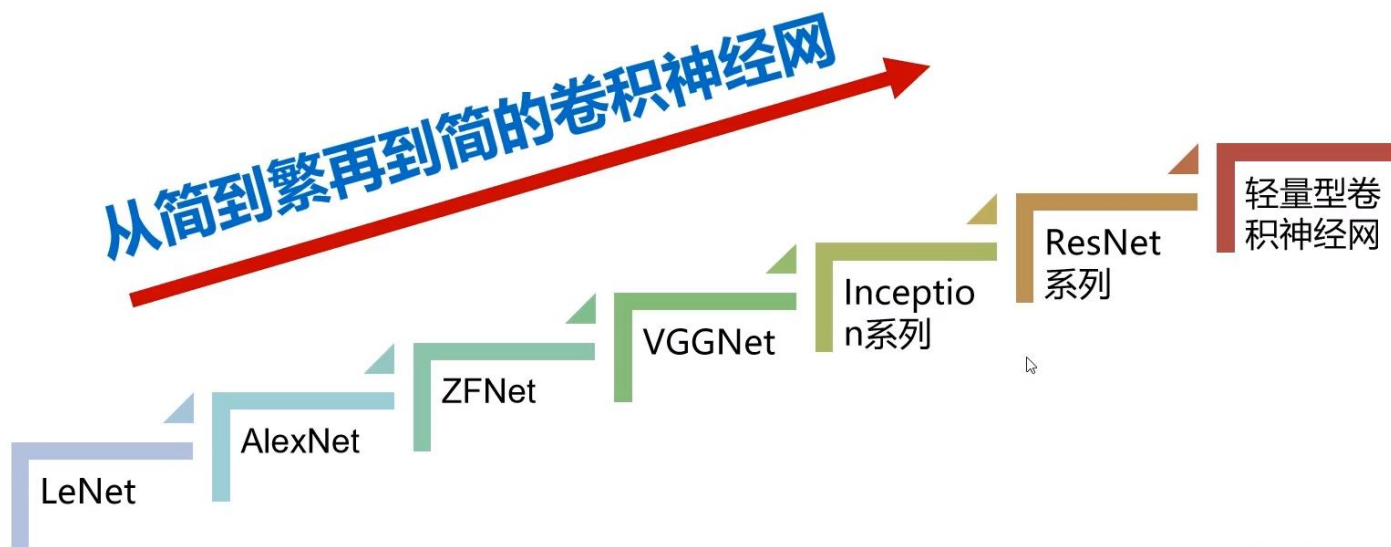
求通过w对X进行卷积操作得到的feature map, 其中, 步长=1, 即卷积过程中filter的移动步长为1格。

求使用一个3*3的窗口, 且stride=3时, 对X进行max pooling 和mean pooling 的结果。

卷积神经网络

- 卷积神经网络（Convolutional Neural Network, CNN）的提出与发展是深度学习领域一个重要里程碑之一。
 - 卷积神经网络一般是由卷积层、池化层和全连接层交叉堆叠而成。

常见的卷积神经网络结构

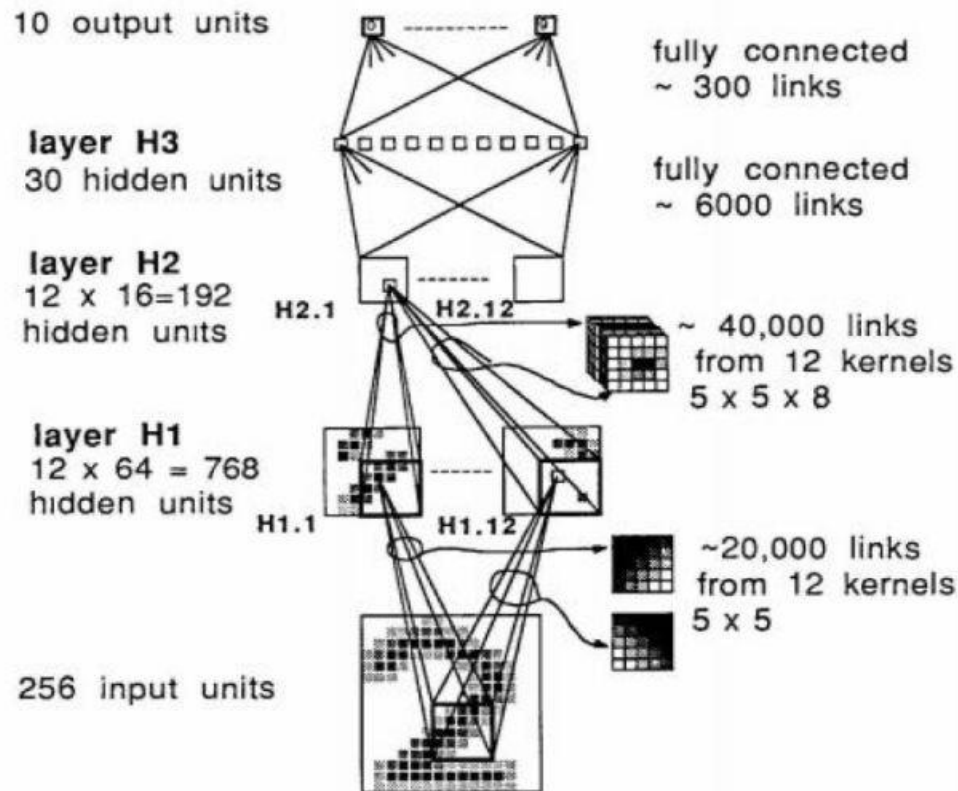


https://blog.csdn.net/sun_shine56

卷积神经网络

□ CNN网络结构

- LeCun在早期提出的3种卷积网络结构。

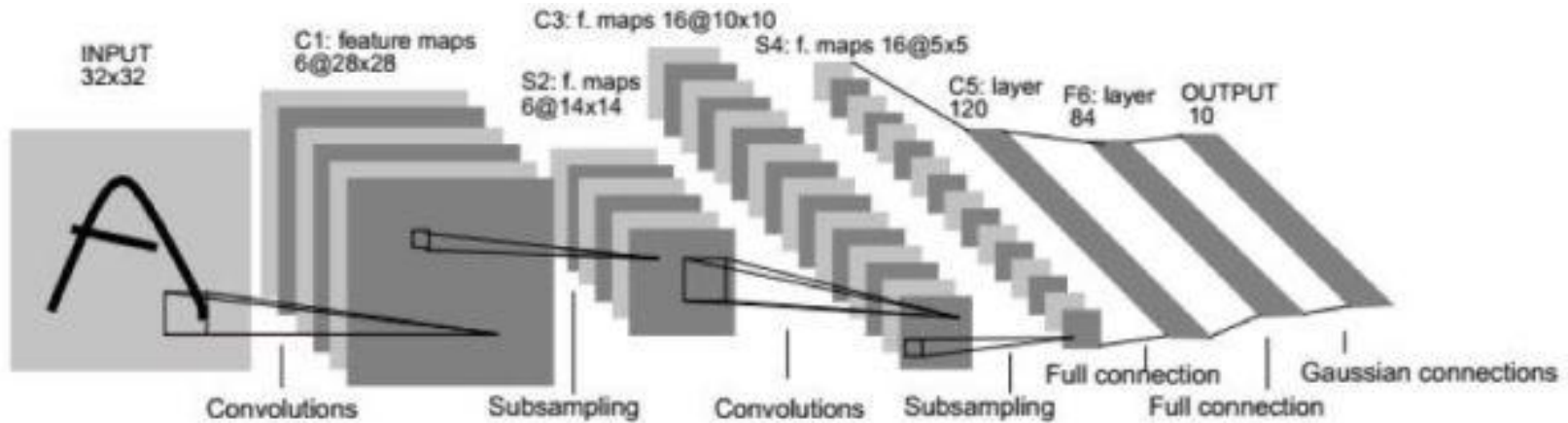


网络有1个输入层，1个输出层，3个隐含层构成，其中隐含层H1和H2是卷积层，H3是全连接层，网络的输入是16x16的归一化图像，输出为0-9这10个类，中间是3个隐含层。

卷积神经网络

□ CNN网络结构

- LeCun在早期提出的3种卷积网络结构。

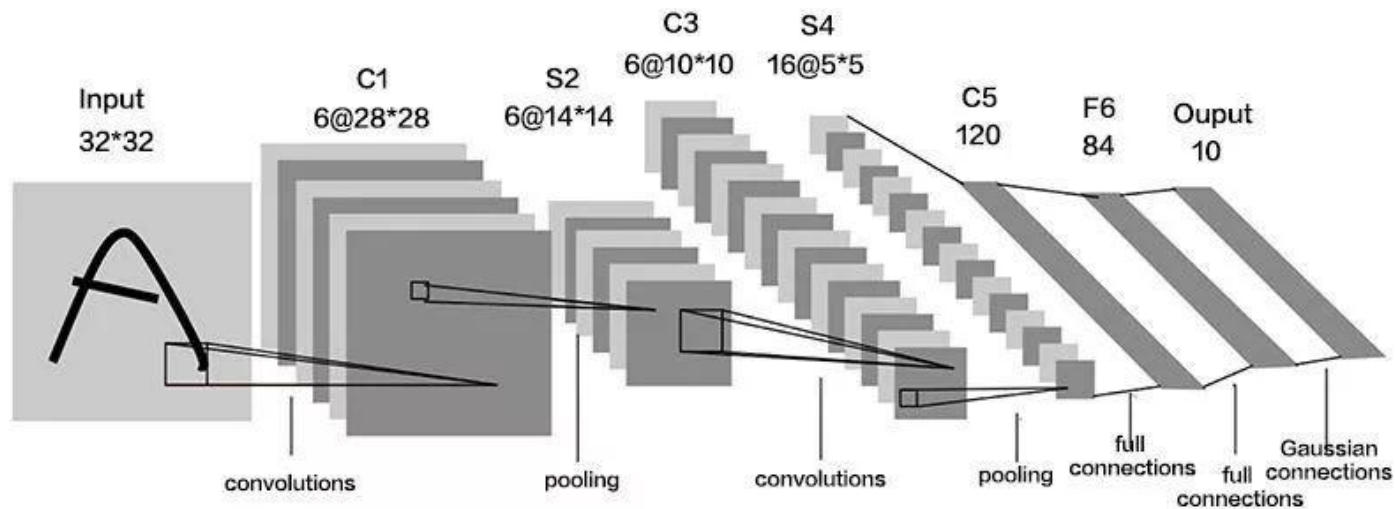


网络的输入为 28×28 的图像，整个网络有4个隐含层，其中H1为4个 5×5 的卷积核，输出为4张 24×24 的特征图像。H2为下采样层，对H1的输出结果进行 2×2 的下采样，得到4张 12×12 的图像。H3有12个 5×5 的卷积核，输出为12张 8×8 的图像。H4为下采样层，对H3的输出图像进行 2×2 的下采样，得到12张 4×4 的特征图像。最后为输出层，接收H4特征图像，输出10个类别的概率。

卷积神经网络

□ CNN网络结构

- LeCun在早期提出的3种卷积网络结构。下面是第一个广为流传的卷积网路-**LeNet5网络**:

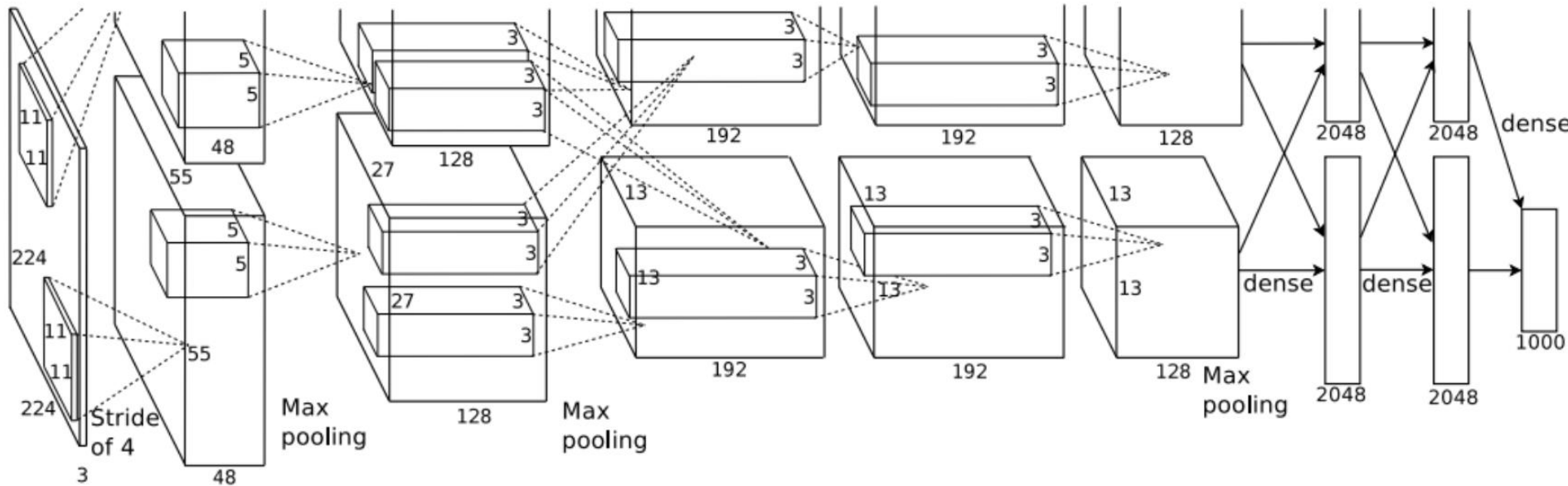


这个网络的输入为32x32的图像，整个网络有2个卷层，2个池化层，2个全连接层，一个输出层，输出层有10个神经元，代表10个数字类

卷积神经网络

□ CNN网络结构

- 现代意义上的深度卷积神经网络起源于AlexNet网络，它是深度卷积神经网络的鼻祖



这个网络有5个卷积层，它们中的一部分后面接着max-pooling层进行下采样；最后跟3个全连接层。最后一层是softmax输出层，共有1000个节点，对应ImageNet图集中1000个图像分类。网络中部分卷积层分成2个group进行独立计算，有利于GPU并行化以及降低计算量。

卷积神经网络

□ CNN网络结构

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

VGGNet:

2014年ImageNet竞赛亚军，增加网络深度，使用 3×3 卷积核 2×2 最大池化层。（VGG-16参数：138M）

卷积神经网络

□ CNN网络结构

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Features:

- 11-19 layers
- Same input, $3 \times 224 \times 224$
- 3 fully-connected layers
- Softmax output
- 5 stages, maxpool in between
- (most) 3×3 kernels
- Increasing kernel number

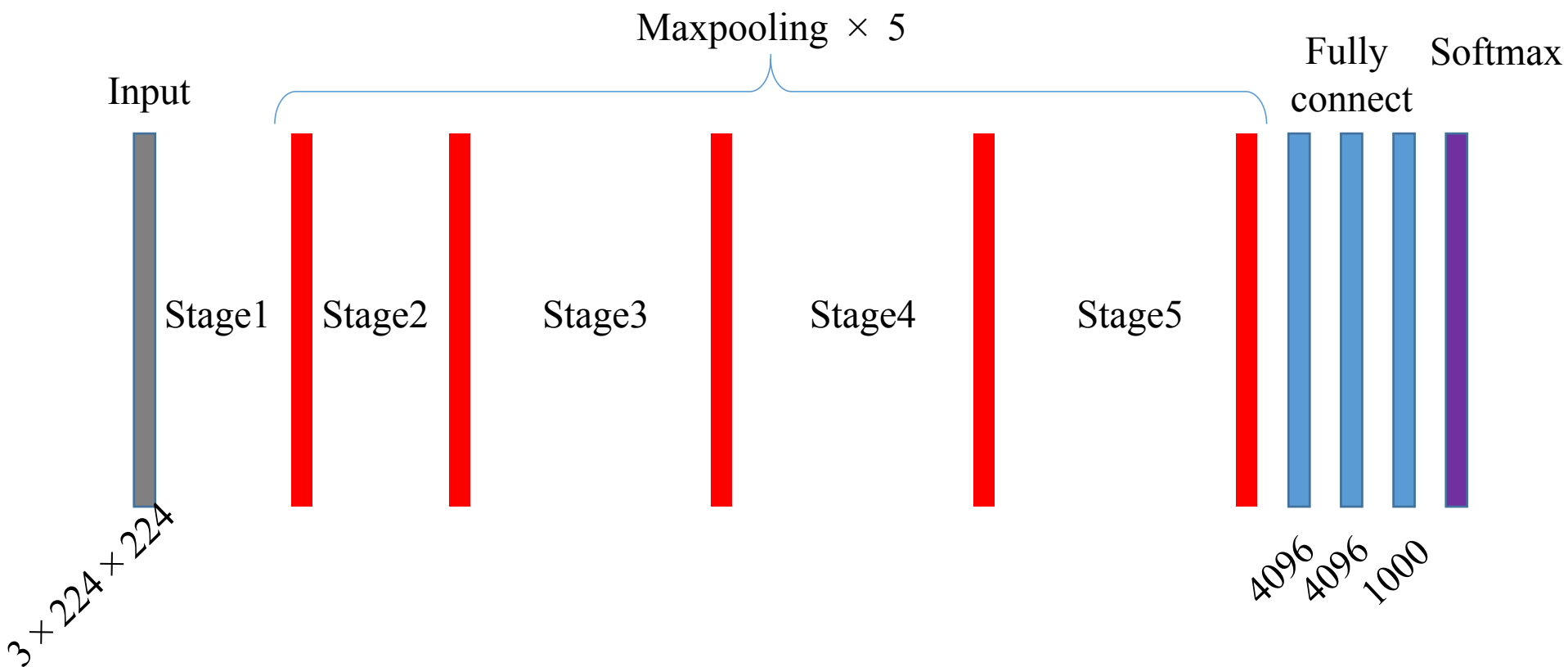
卷积神经网络

□ CNN网络结构-VGG



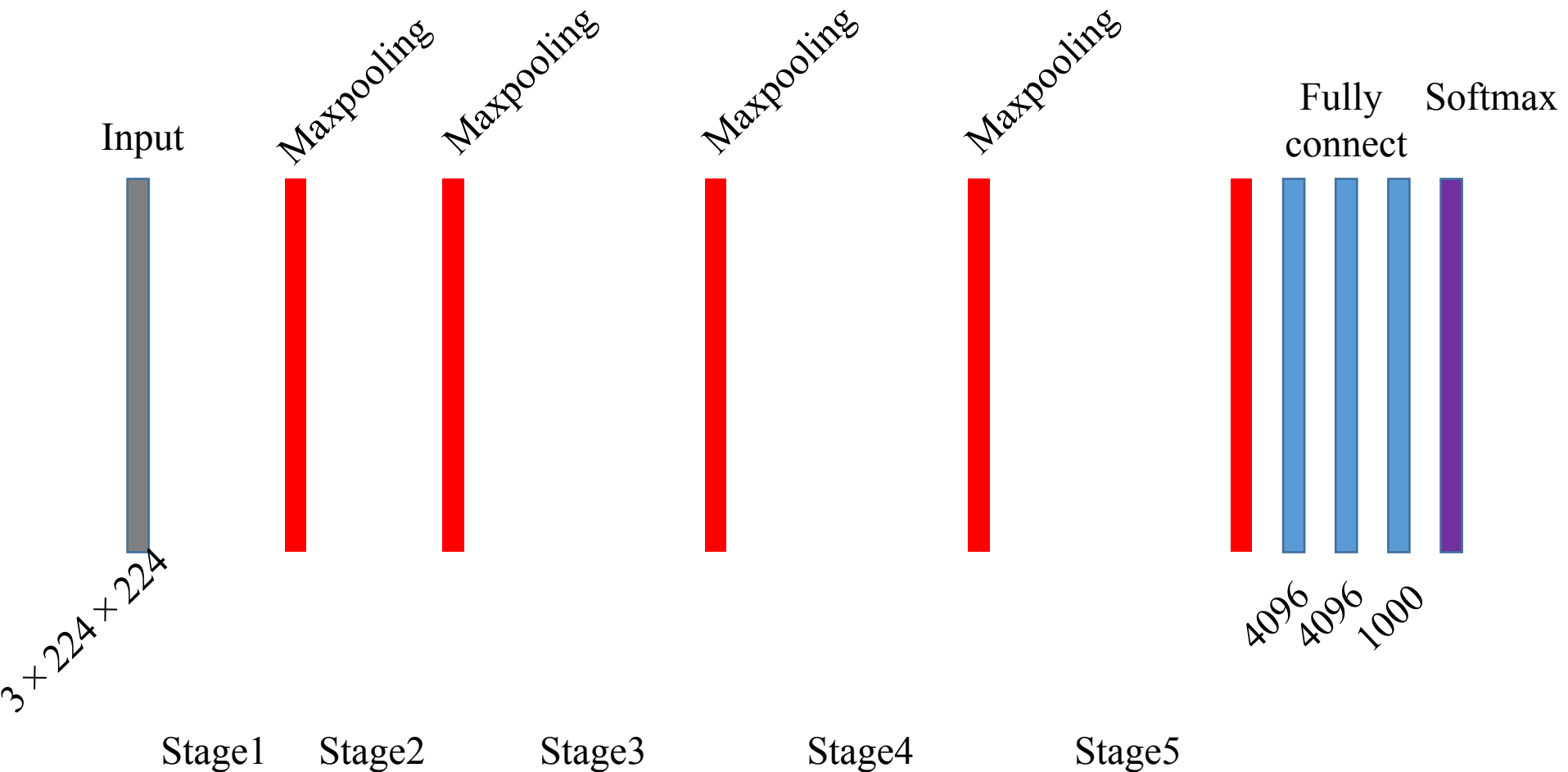
卷积神经网络

□ CNN网络结构-VGG



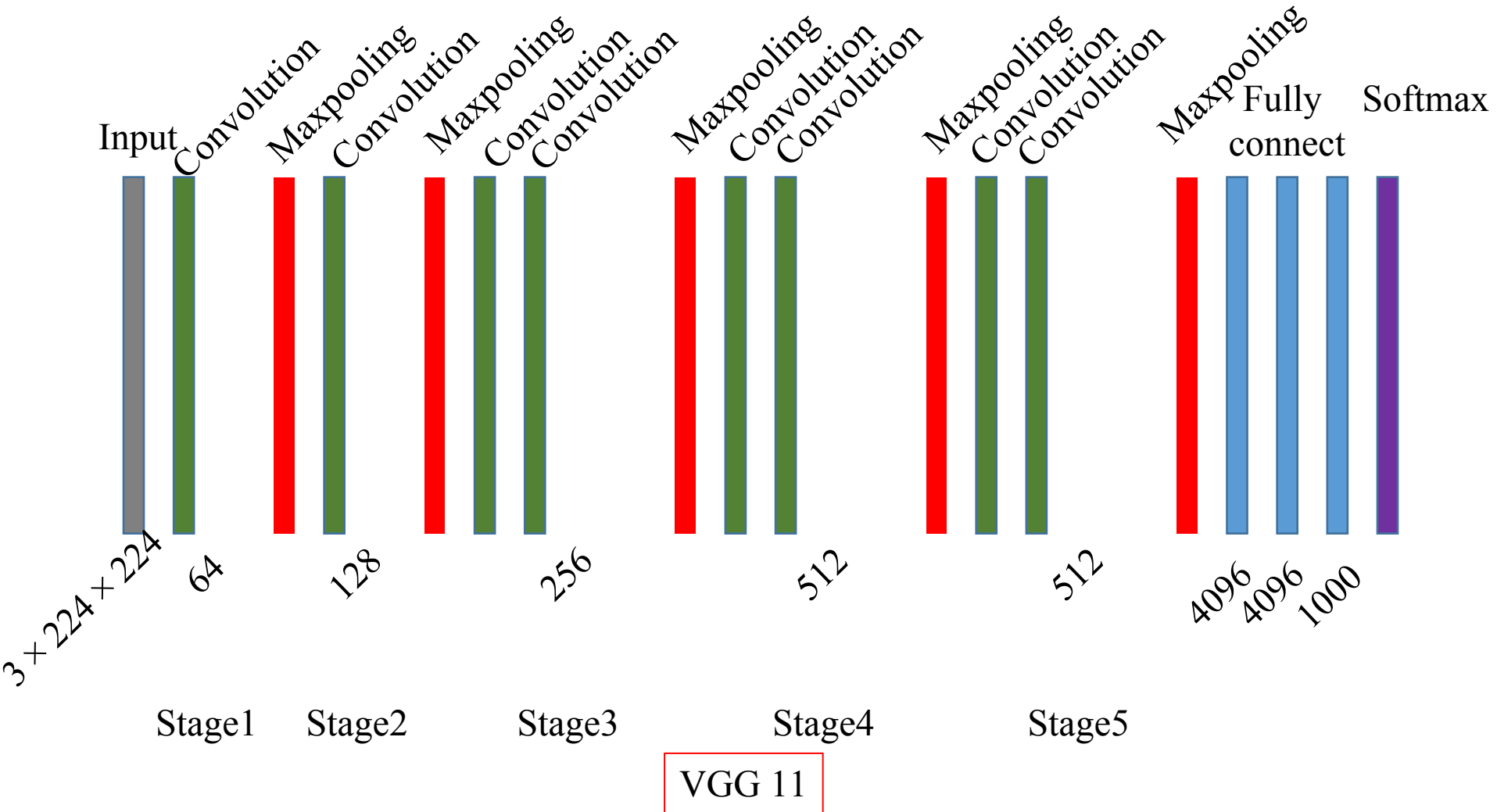
卷积神经网络

□ CNN网络结构-VGG



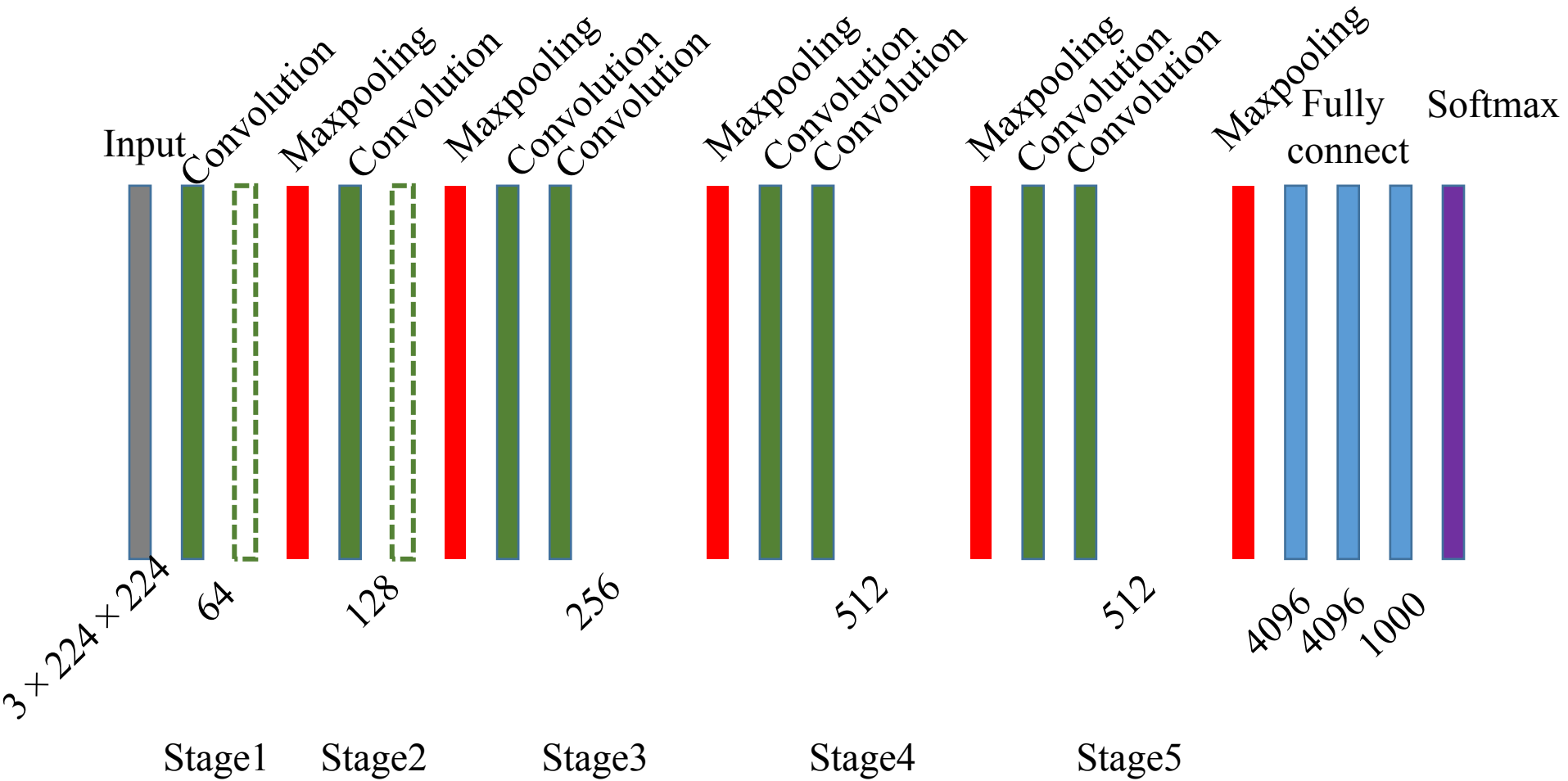
卷积神经网络

□ CNN网络结构-VGG



卷积神经网络

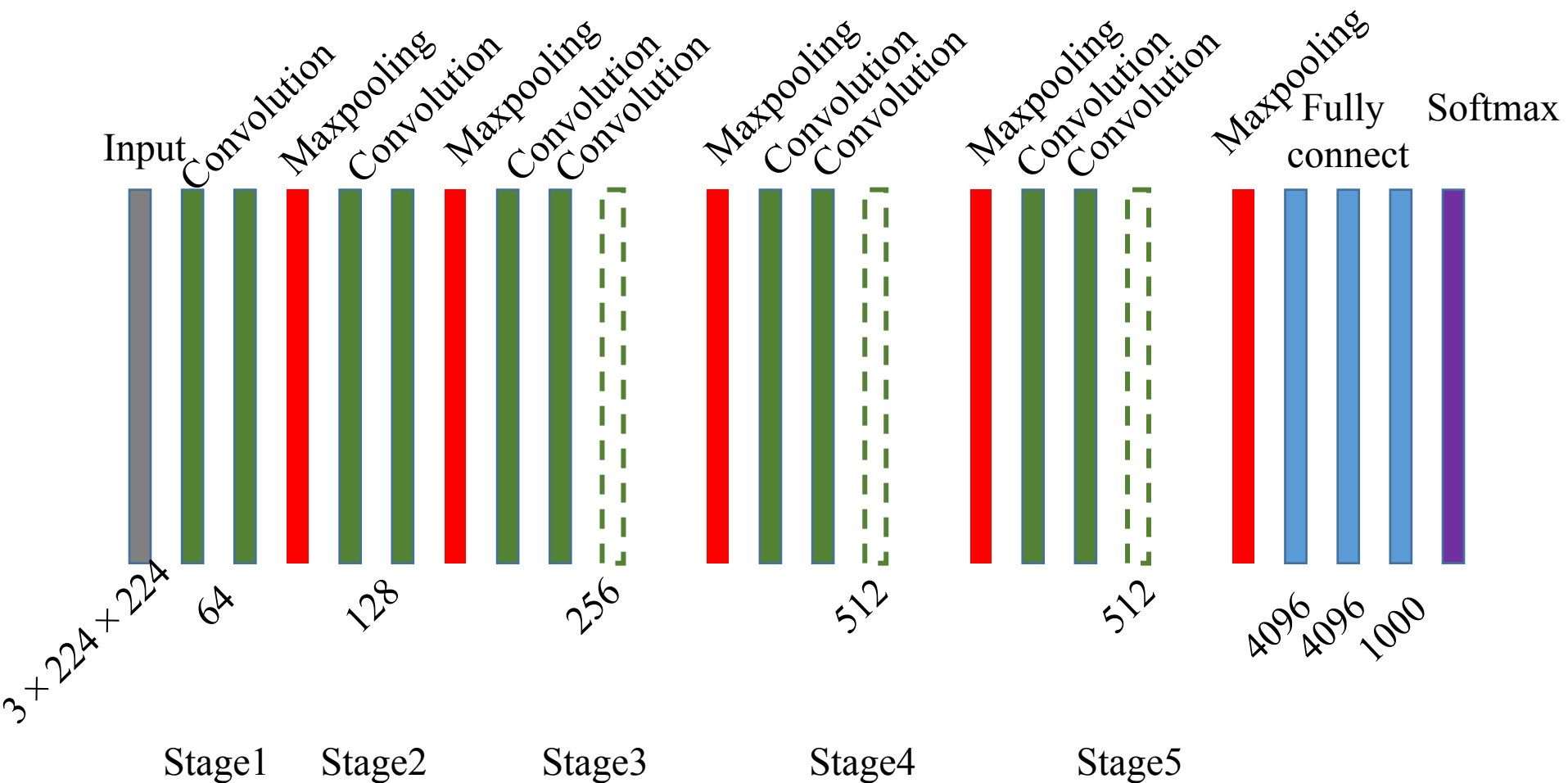
□ CNN网络结构-VGG



VGG 13

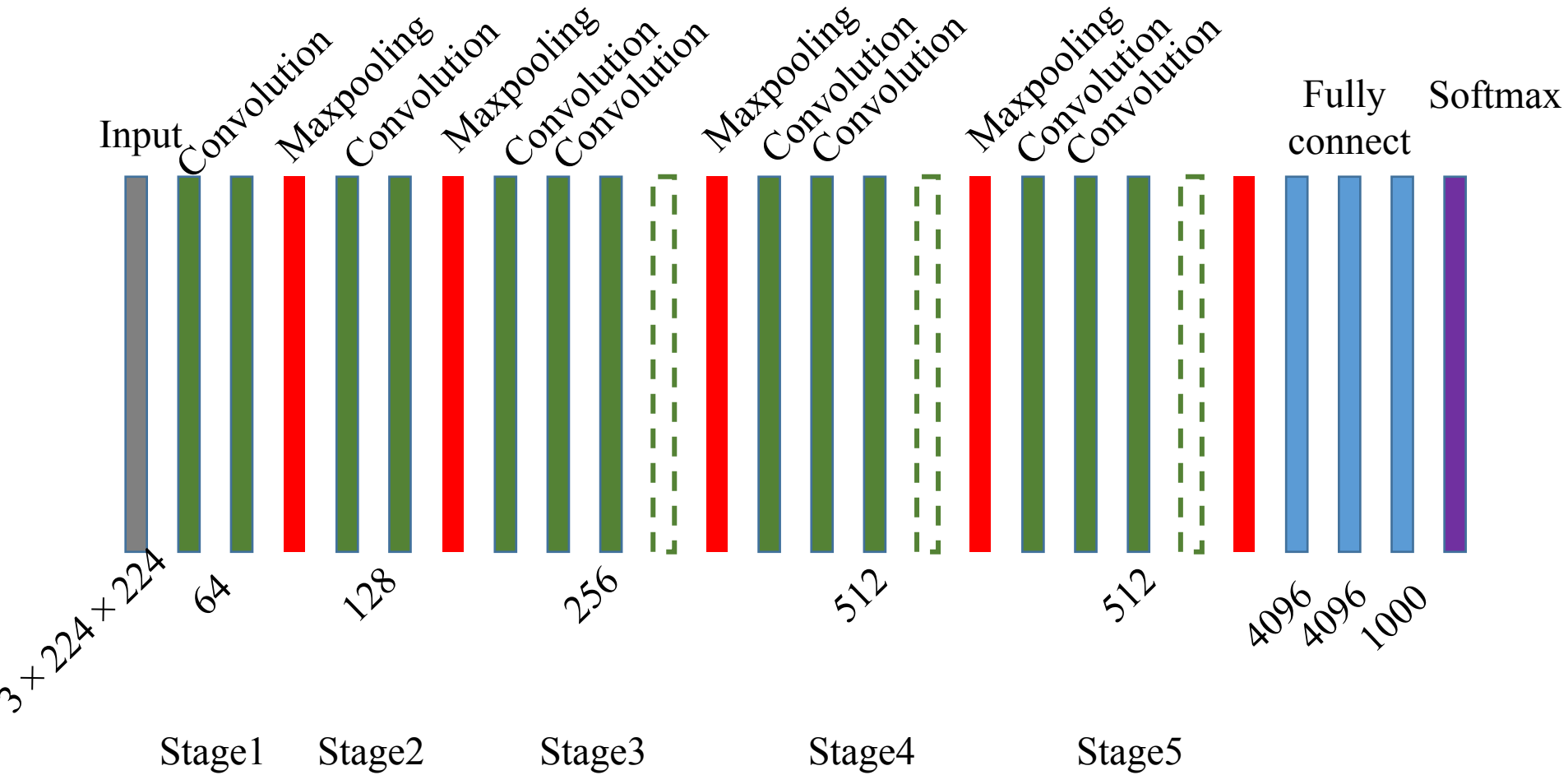
卷积神经网络

□ CNN网络结构-VGG



卷积神经网络

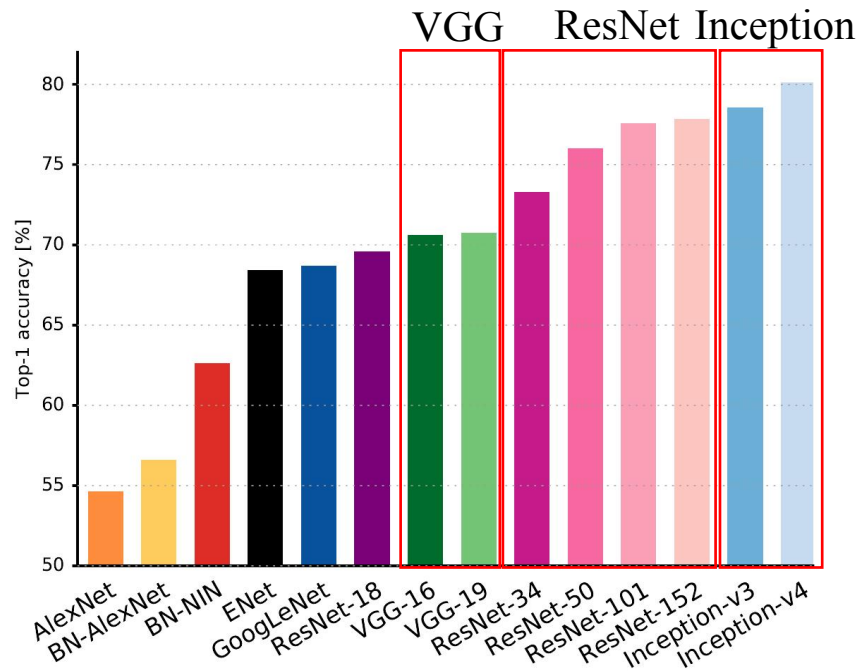
□ CNN网络结构-VGG



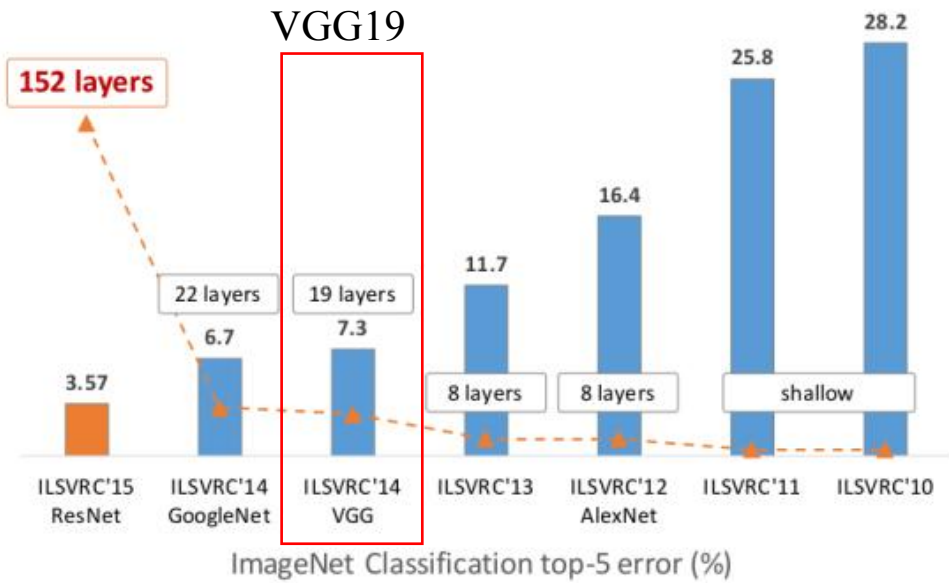
VGG 19

卷积神经网络

□ CNN网络结构-VGG



Single-crop Top-1 validation accuracies on ImageNet

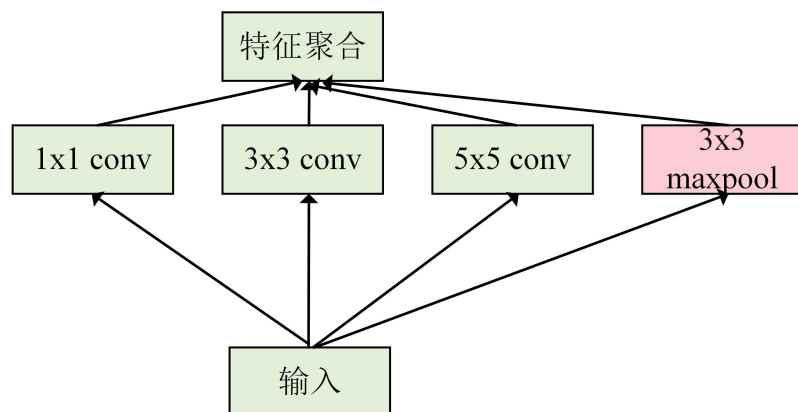


Reported Top-5 error on ImageNet

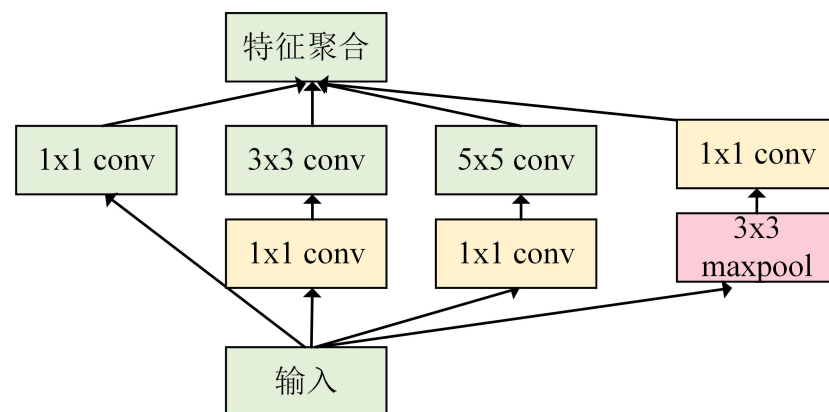
卷积神经网络

□ CNN网络结构

Inception V1: 又称GoogLeNet, 并行使用不同尺寸卷积核捕获多尺度特征。(参数: 5.58M)



(a) Inception结构朴素版本

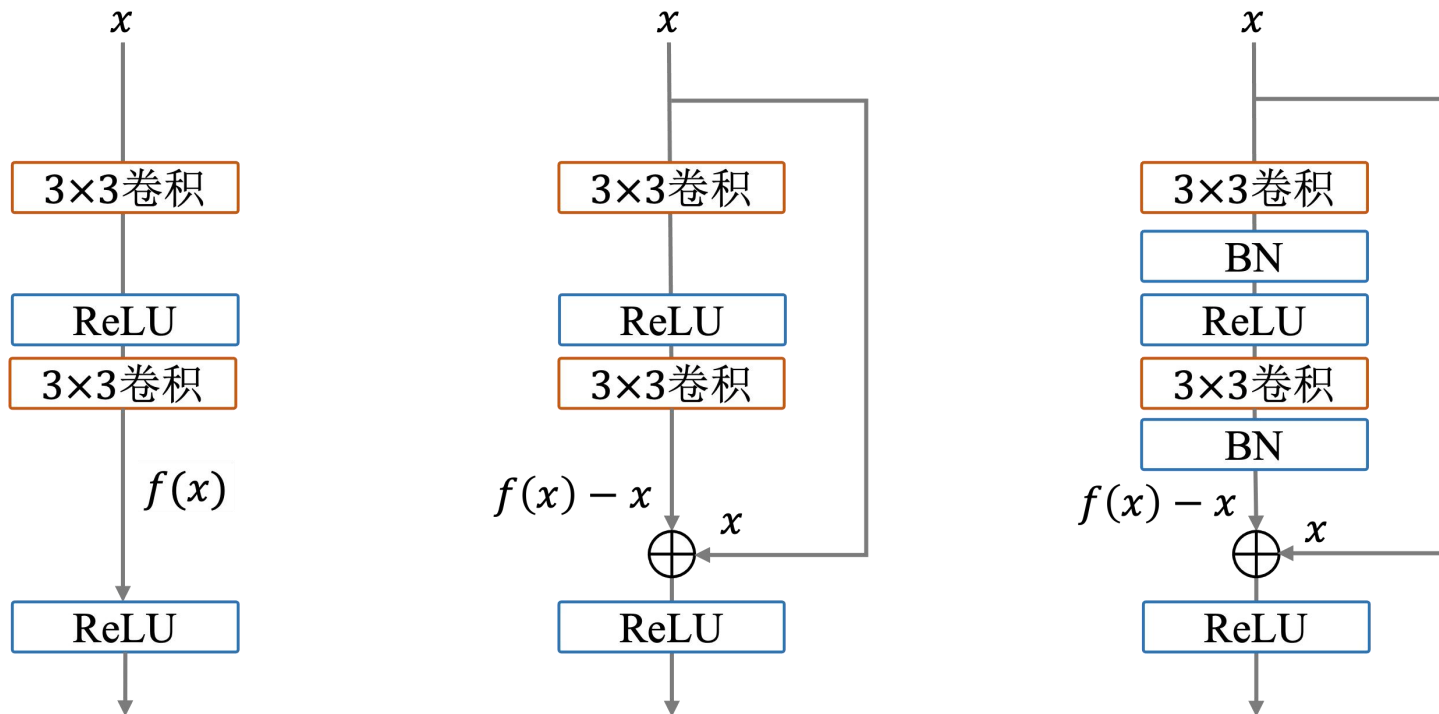


(b) 带降维的Inception结构

卷积神经网络

□ CNN网络结构

ResNet: 深度残差网络 (Deep Residual Network, ResNet) 利用残差学习方式, 允许信号跳过一层或多层。 (ResNet50参数: 25M)

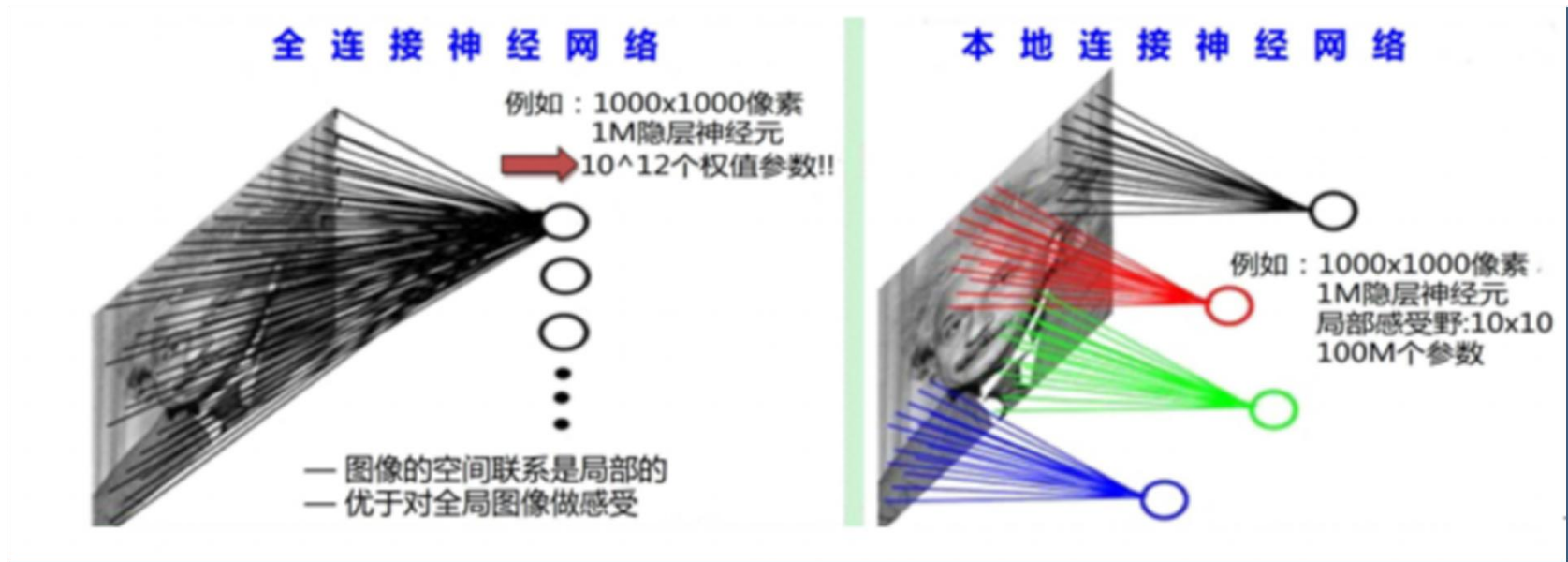


卷积神经网络

- 卷积神经网络（Convolutional Neural Network, CNN）的提出与发展是深度学习领域一个重要里程碑之一。
 - 卷积神经网络一般是由卷积层、池化层和全连接层交叉堆叠而成。
 - 卷积神经网络三个结构特性：局部连接、权重共享以及池化。
 - 和前馈网络相比，卷积神经网络的参数更少。

卷积神经网络

□ CNN的局部连接



- 视觉皮层中每个神经元不是对全局图像进行感知，而只对局部进行感知，然后在更高层将局部的信息综合起来得到全局信息。

卷积神经网络

□ CNN的权值共享



- 隐含层的每一个神经元如果只和 10×10 个像素连接，也就是说每一个神经元存在 $10 \times 10 = 100$ 个连接权值参数。如果将每个神经元的参数设置成相同，那么，不管隐层的神经元个数有多少，两层间的连接都只有100个参数，这就是卷积神经网络的**权值共享**。

卷积神经网络

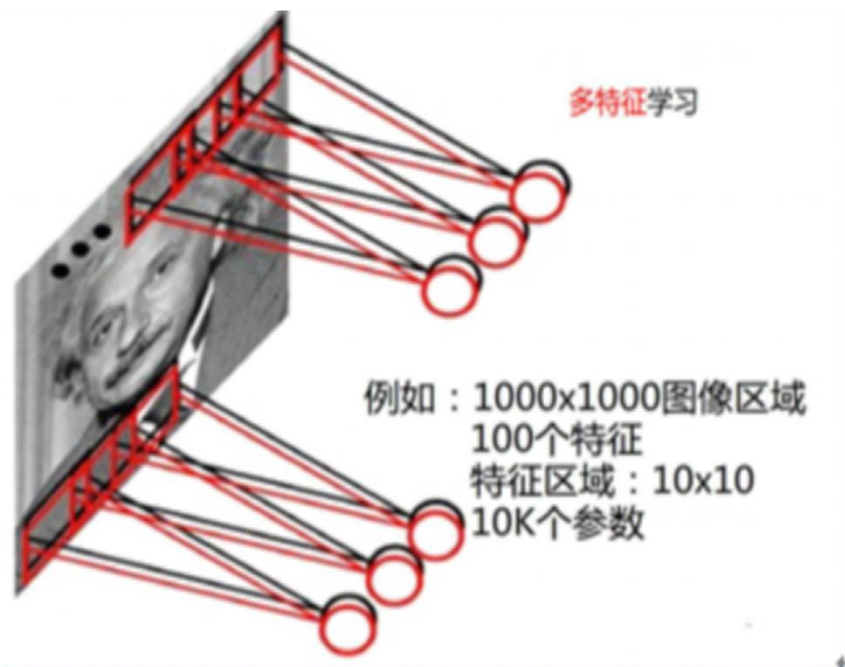
□ 减少参数的方法:

- **局部连接**: 每个神经元无需对全局图像进行感知, 而只需对局部进行感知, 然后在更高层将局部的信息综合起来得到全局信息。
- **权值共享**: 每个神经元参数设为相同, 即权值共享, 也即每个神经元用同一个卷积核去卷积图像。

卷积神经网络

□ CNN的多卷积核

- 下图中不同的颜色表示不同的卷积核，每个卷积核都会将图像生成为另一幅特征映射图（即：一个卷积核提取一种特征）。
- 为了使特征提取更充分，我们可以添加多个卷积核（滤波器）以提取不同的特征。



- ✓ 每层隐层神经元的个数按卷积核的数量翻倍。
- ✓ 每层隐层参数个数仅与特征区域大小、卷积核的多少有关。
例如：隐含层的每个神经元都连接**10x10**像素图像区域，同时有**100**种卷积核（滤波器）。则参数总个数为： $(10 \times 10 + 1) \times 100 = 10100$ 个

卷积神经网络

实例

- 一种典型的用来识别数字的卷积网络是[LeNet-5](#)。美国大多数银行当年用它识别支票上面的手写数字，达到了商用地步，说明该算法具有很高的准确性。

