北京交通大学
BEIJING JIAOTONG UNIVERSITY

高级数据建模技术

目录

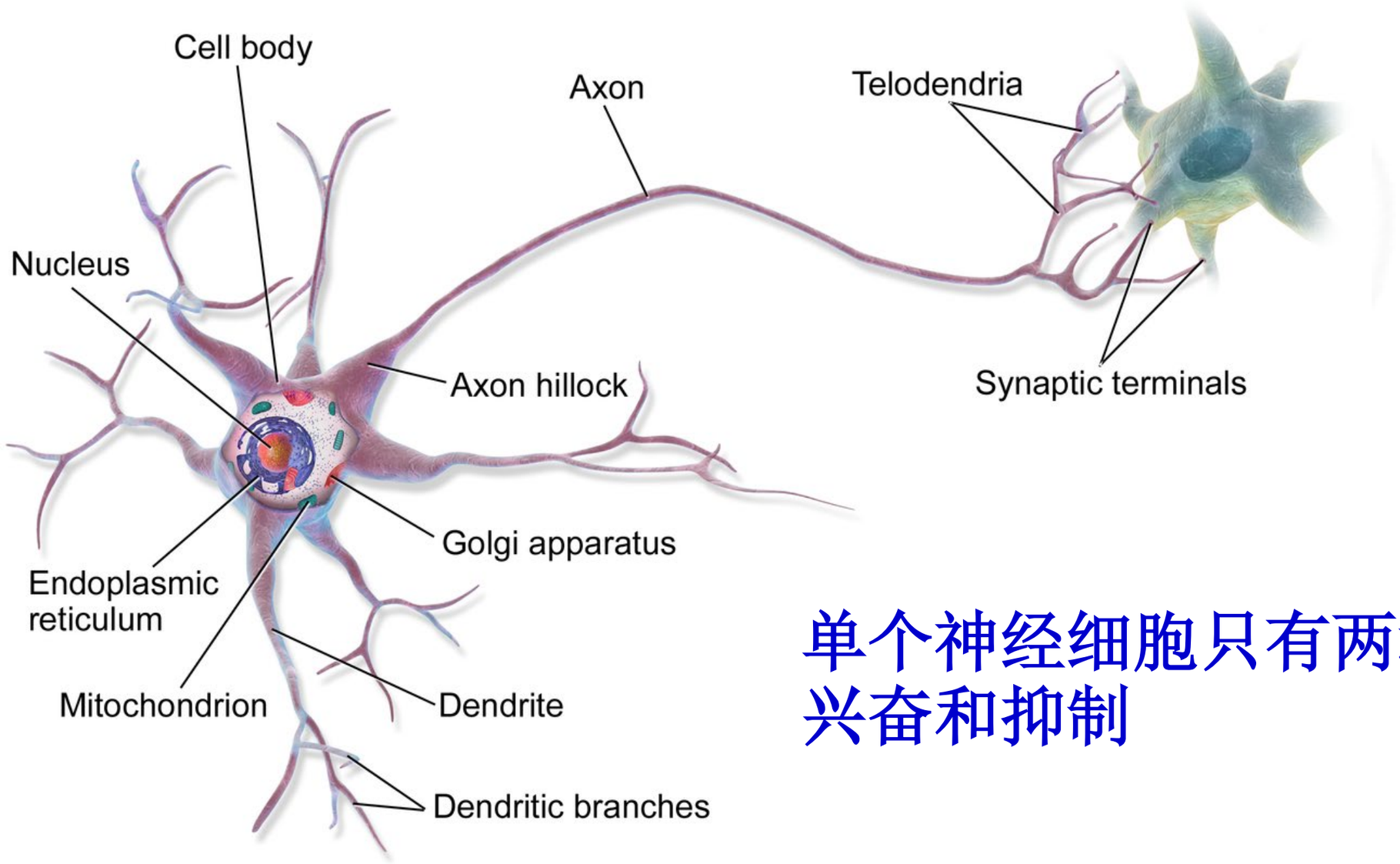
1. 从人工神经网络到深度神经网络
2. 卷积神经网络的若干改进
3. 循环神经网络若干改进
4. 生成式对抗神经网络

从人工神经网络到深度神经网络

人类大脑中的神经网络是由一个个神经元组成的，神经元由细胞体、轴突、、动作电位、突出末端、突触前神经元的轴突和树突等部分组成。身体的不同部位产生的信息通过不同的路径到达神经元，神经元处理它并产生一个输出。（神经元也可能被连接到另一个神经元。）

从人工神经网络到深度神经网络

生物神经元



单个神经细胞只有两种状态：兴奋和抑制

从人工神经网络到深度神经网络

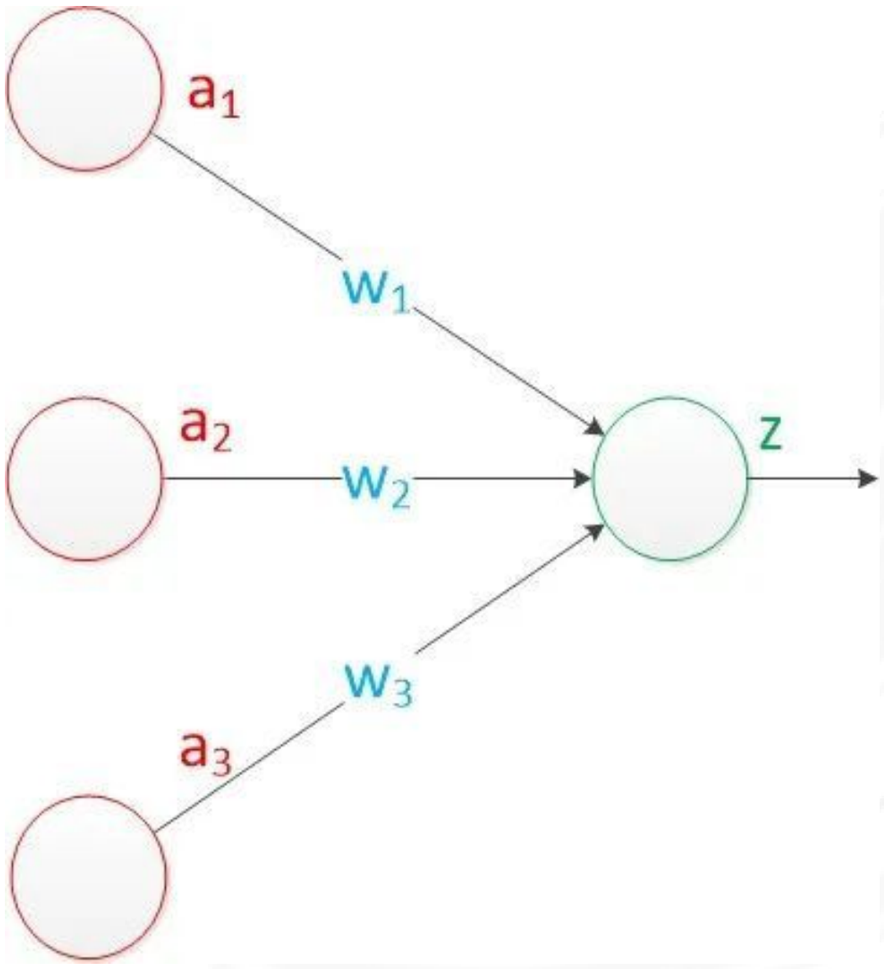
根据神经网络的基本生物学模型，建立如下数学模型：

$$y = f\left(\sum_{i=1}^{n} x_i w_i - \theta\right)$$

其中，x_i代表着从不同路径到达的信号值，w_i代表不同路径上的权重，θ是设置的阈值，f是激励函数，决定最后的响应内容。

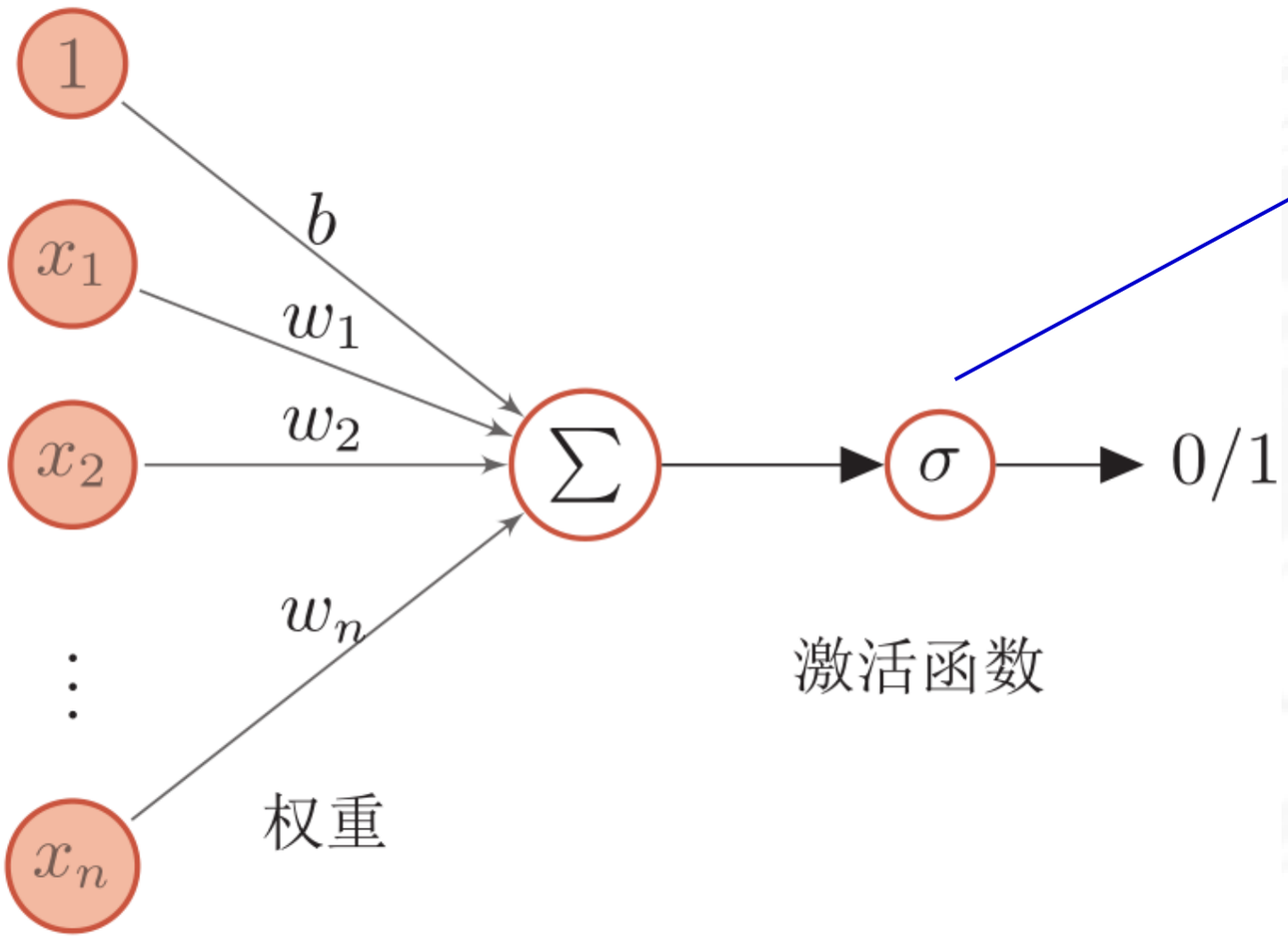
从人工神经网络到深度神经网络

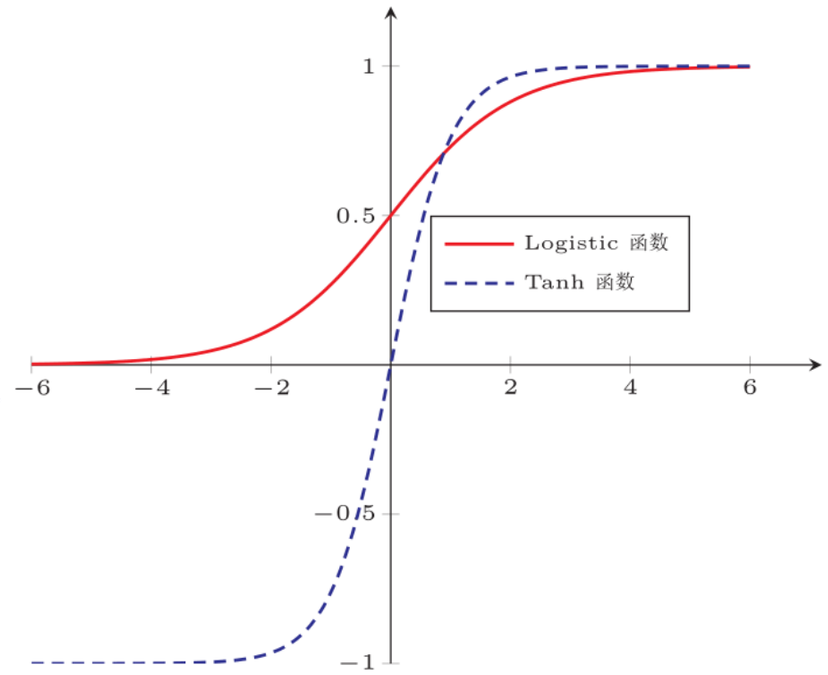
人工神经网络主要由大量的神经元以及它们之间的有向连接构成。最早的人工神经网络是单层神经网络，由输入层和输出层组成。



输入层里的输入单元只负责传输数据，不做计算。输出层里的输出单元则需要面对前面一层的输入进行计算。需要计算的层次称为计算层，并把拥有一个计算层的网络称为：单层神经网络。

从人工神经网络到深度神经网络



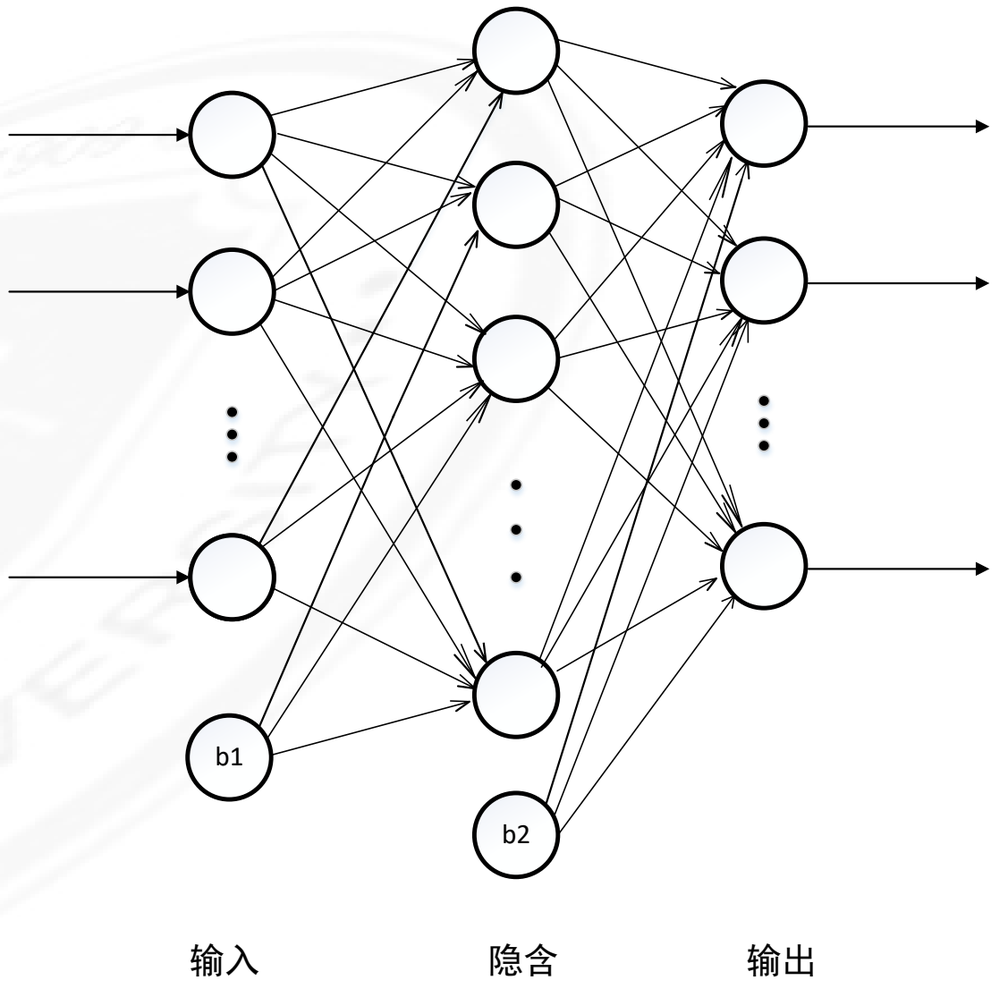


常用激活函数：阈值函数、双向阈值函数、S型函数、双曲正切函数、 高斯函数

从人工神经网络到深度神经网络

BP网络

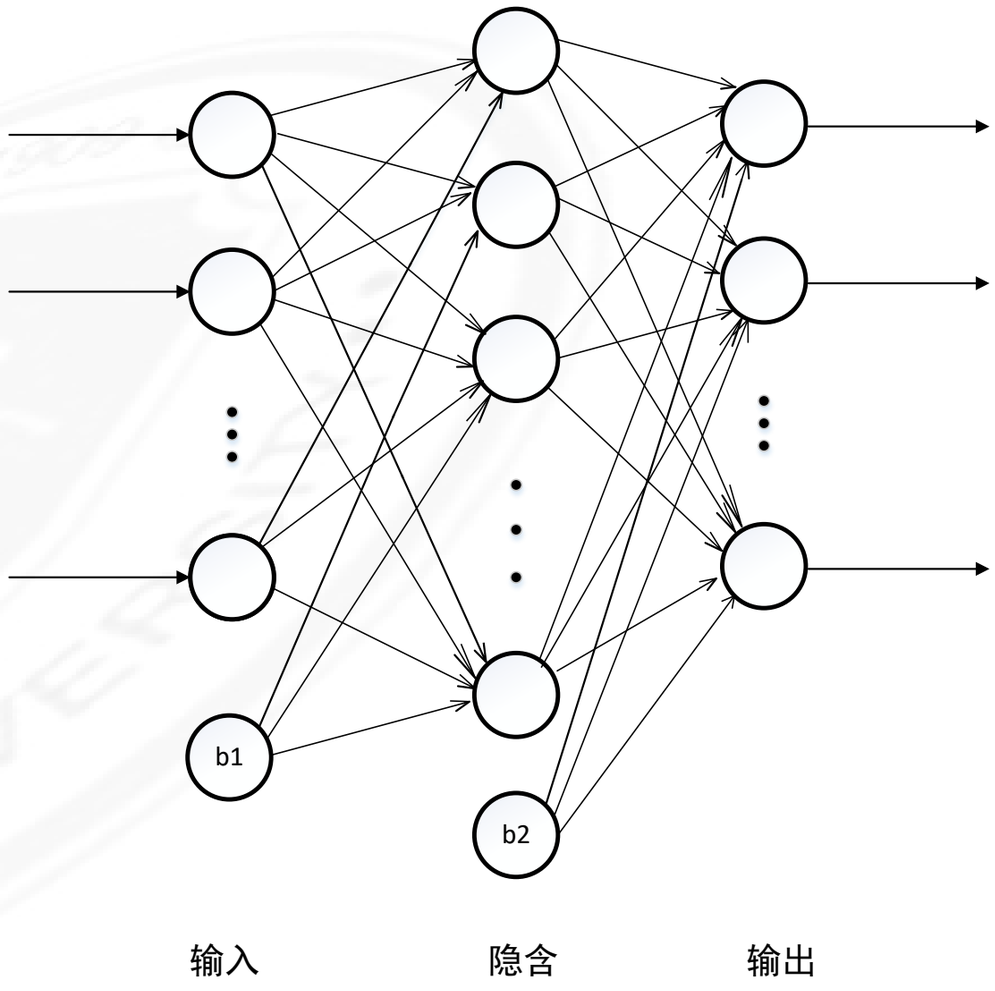
- BP神经网络由输入层、隐含层和输出层组成，其中隐含层和输出层都实现计算功能，属于两层神经网络，如右图所示。
- 输入值从输入层单元通过连接权重加权激活逐层向前传播经过隐层最后到达输出层得到输出。在信号的向前传递过程中，网络的权值是固定不变的，每一层神经元的状态只影响下一层神经元的状态。
- 神经网络的本质是通过参数与激励函数来拟合特征与目标之间的真实函数关系。
- 两层神经网络可以无限量逼近任意连续函数。



从人工神经网络到深度神经网络

BP网络——反向传播算法

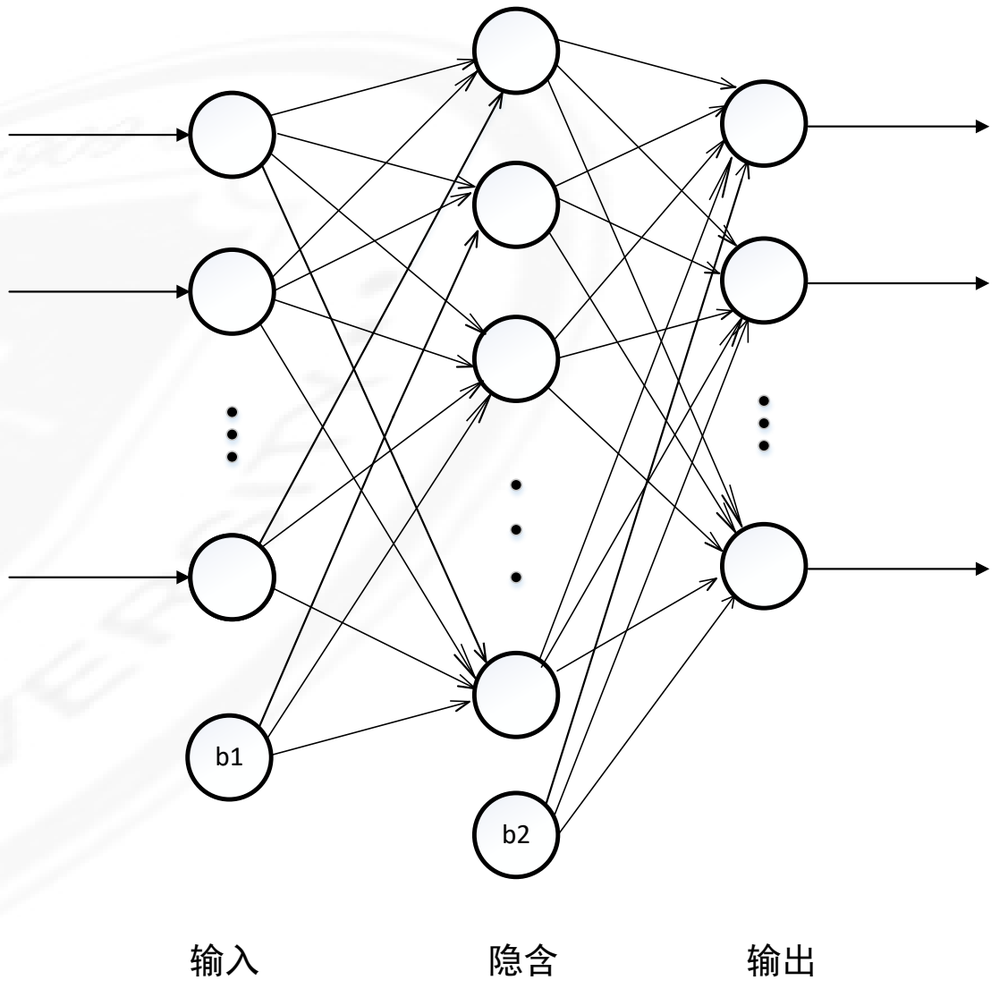
- BP网络是一种前馈神经网络，各神经元分层排列，每个神经元只与前一层神经元相连，接收前一层的输出，并输出给下一层。
- 反向传播是指从输出层开始沿着相反的方向来逐层调整参数的过程。
- 网络的实际输出与期望输出之间的差值即为误差信号。误差信号由输出端开始逐层向前传播，这是误差信号的反向传播。在误差信号反向传播的过程中，网络的权值由误差反馈进行调节，通过权值的不断修正使网络的实际输出更加接近期望输出。



从人工神经网络到深度神经网络

BP网络

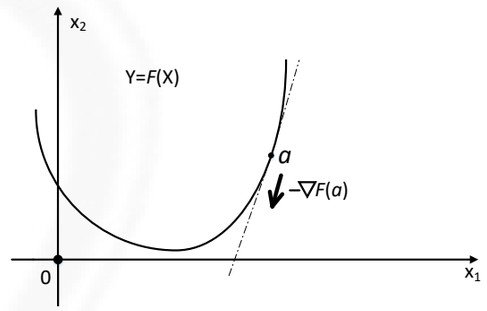
- 输入值从输入层单元通过连接权重加权激活逐层向前传播经过隐层最后到达输出层得到输出。在信号的向前传递过程中，网络的权值是固定不变的，每一层神经元的状态只影响下一层神经元的状态。
- 反向传播算法：
- 网络的实际输出与期望输出之间的差值即为误差信号。误差信号由输出端开始逐层向前传播，这是误差信号的反向传播。在误差信号反向传播的过程中，网络的权值由误差反馈进行调节，通过权值的不断修正使网络的实际输出更加接近期望输出。



从人工神经网络到深度神经网络

代价函数

- 在回归问题中，指定代价函数 $J(\theta)=\frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})^2$ 以使目标变量的真实值和预测值的距离最小
- 代价函数描述了网络输出与真实值之间的误差。
- 通过随机梯度下降的方法最小化代价函数以提高网络精度
- 可以在代价函数中引入其他约束以满足设定要求



BP算法

1. 进行前馈传导计算，利用前向传导公式，得到 $L_2, L_3, \ldots$ 直到输出层 L_{n_l} 的激活值。
2. 对输出层（第 n_l 层），计算：

$$\delta^{(n_l)} = -(y - a^{(n_l)}) \bullet f'(z^{(n_l)})$$

3. 对于 $l = n_l - 1, n_l - 2, n_l - 3, \ldots, 2$ 的各层，计算：

$$\delta^{(l)} = \left((W^{(l)})^T \delta^{(l+1)}\right) \bullet f'(z^{(l)})$$

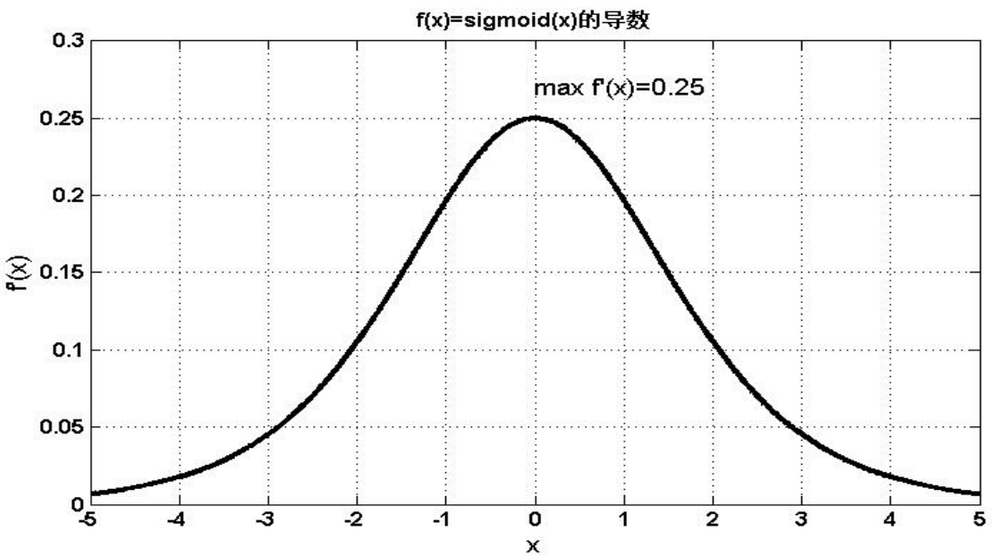
4. 计算最终需要的偏导数值：

$$\nabla_{W^{(l)}} J(W, b; x, y) = \delta^{(l+1)} (a^{(l)})^T,$$

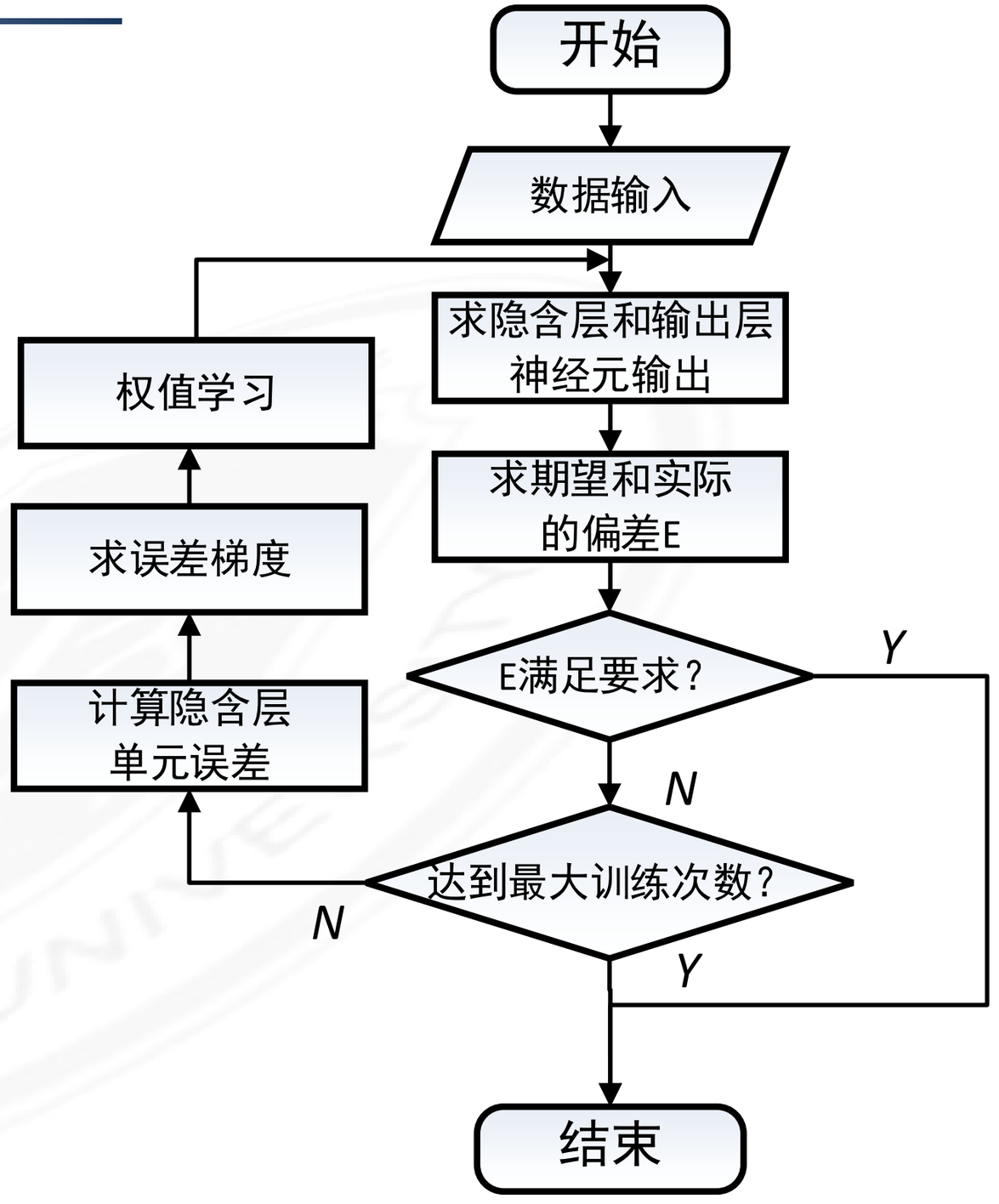
$$\nabla_{b^{(l)}} J(W, b; x, y) = \delta^{(l+1)}.$$

从人工神经网络到深度神经网络

反向传播与梯度下降



S型函数导数



从人工神经网络到深度神经网络

深度学习

深度学习算法是一类基于生物学对人脑进一步认识，将神经-中枢-人脑的工作原理设计成一个不断迭代、不断抽象的过程，以便得到最优数据特征表示的机器学习算法；该算法从原始信号开始，先做低层抽象，然后逐渐向高层抽象迭代，由此组成深度学习算法的基本框架。

一般来说，深度学习算法具有如下特点：

（1）使用链式结构非线性变换对数据进行多层抽象。

（2）以寻求更适合待解决的问题概念表示方法为目标。

（3）形成一类具有代表性的特征表示学习方法。

从人工神经网络到深度神经网络

深度学习的优点

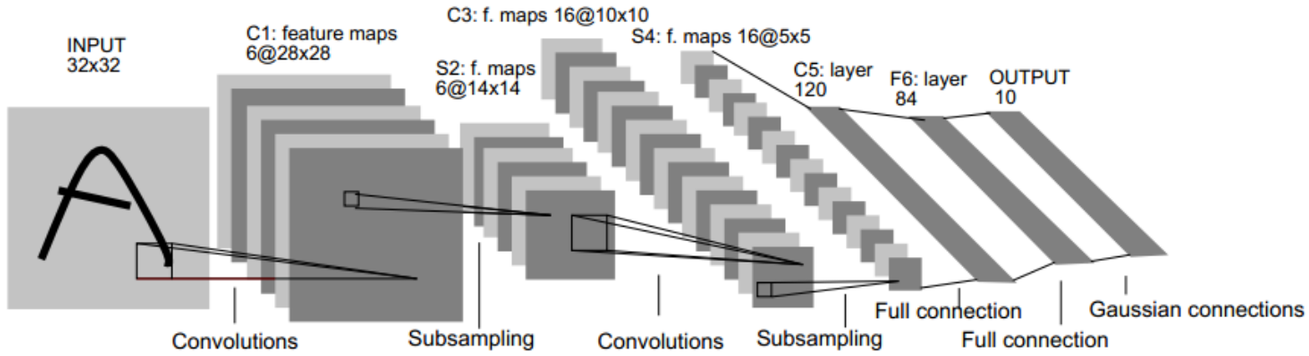
（1）概念提取可以由简单到复杂。深层下是指神经网络包含很多隐含层。

（2）每一层中非线性处理单元的构成方式取决于要解决的问题；每一层中学习模式也可按需求灵活调整为有监督或者无监督学习，有利于调整学习策略，从而提高效率。

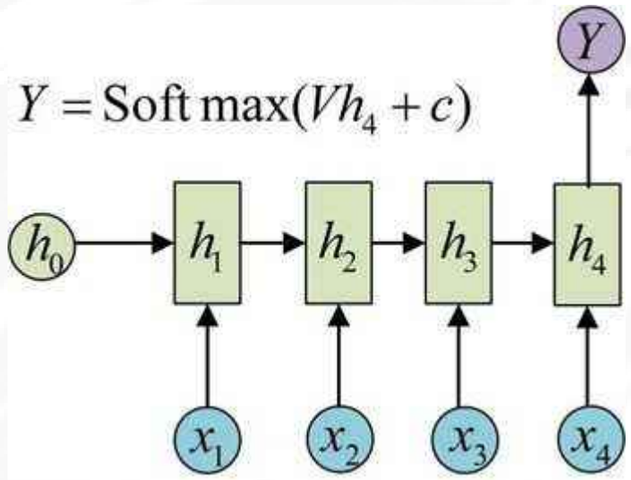
（3）学习无标签数据优势明显。

从人工神经网络到深度神经网络

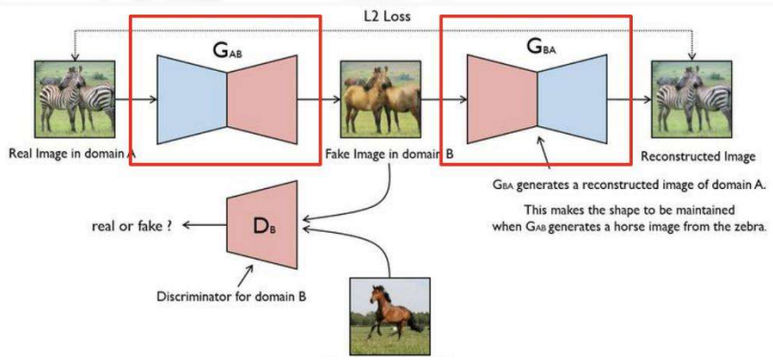
卷积神经网络



循环神经网络



生成式对抗神经网络



卷积神经网络的若干改进

常用的深度学习框架

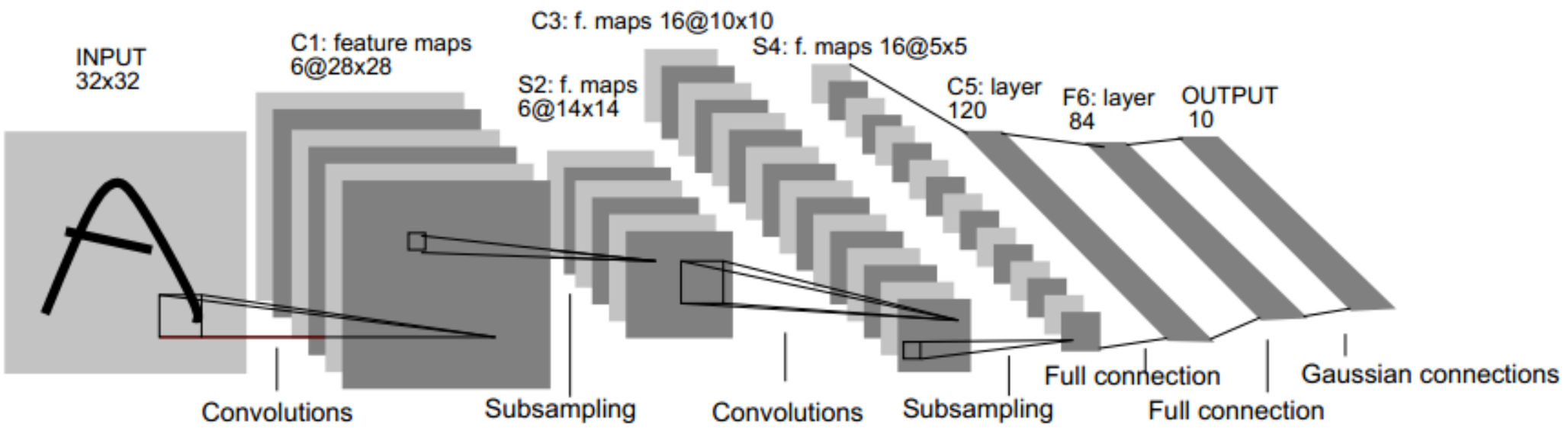
1. 简易和快速的原型设计

2. 自动梯度计算

3. 无缝CPU和GPU切换

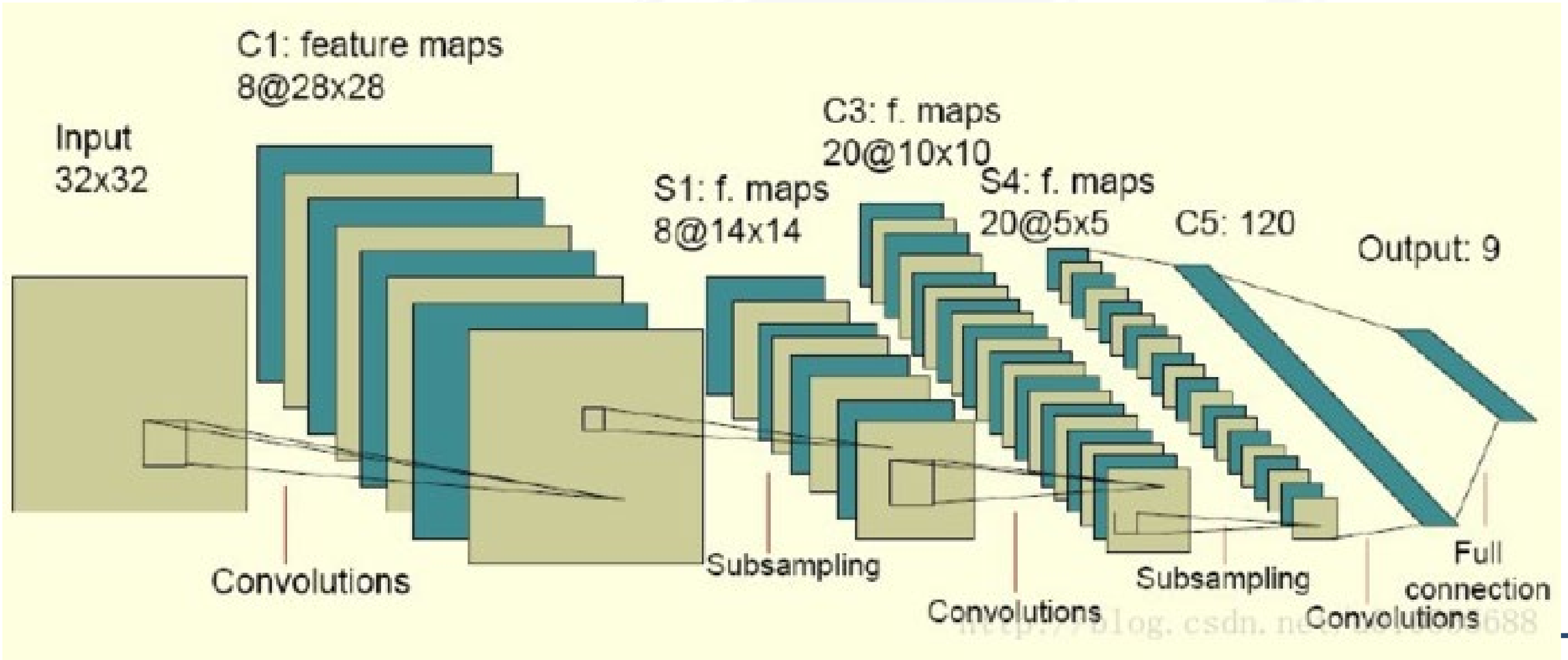
卷积神经网络的若干改进

卷积神经网络是一种前馈神经网络，已成为当前语音分析和图像识别领域的研究热点。它的权值共享网络结构使之更类似于生物神经网络，降低了网络模型的复杂度，减少了权值的数量。该优点在网络的输入是多维图像时表现的更为明显，可以使图像直接作为网络的输入，避免了传统识别算法中复杂的特征提取和数据重建过程。卷积网络是为识别二维形状而特殊设计的一个多层感知器，这种网络结构对平移、比例缩放、倾斜或者其他形式的变形具有高度不变性。



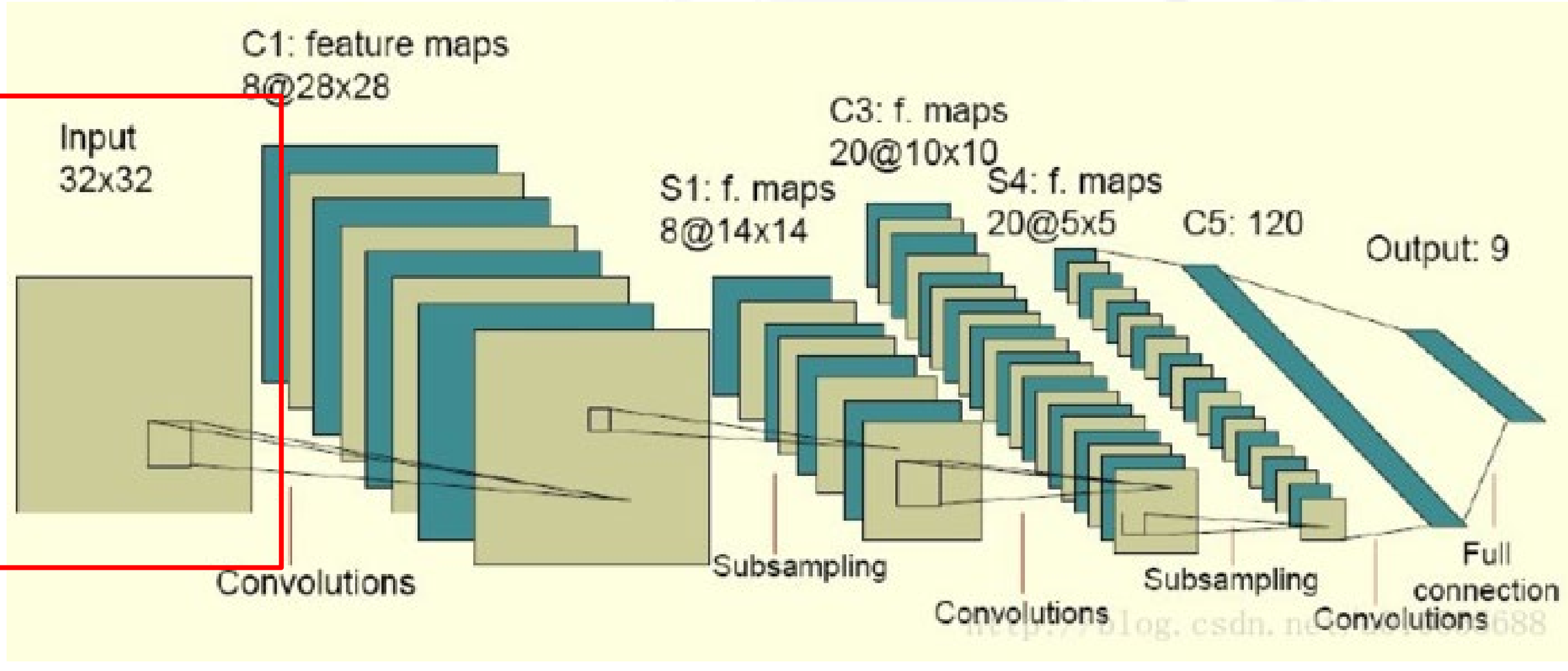
卷积神经网络的若干改进

- 卷积神经网络是由卷积层、池化层（子采样层）、全连接层输出层交叉堆叠，加上输入层和输出层构成。
  - 趋向于小卷积、大深度
  - 趋向于全卷积



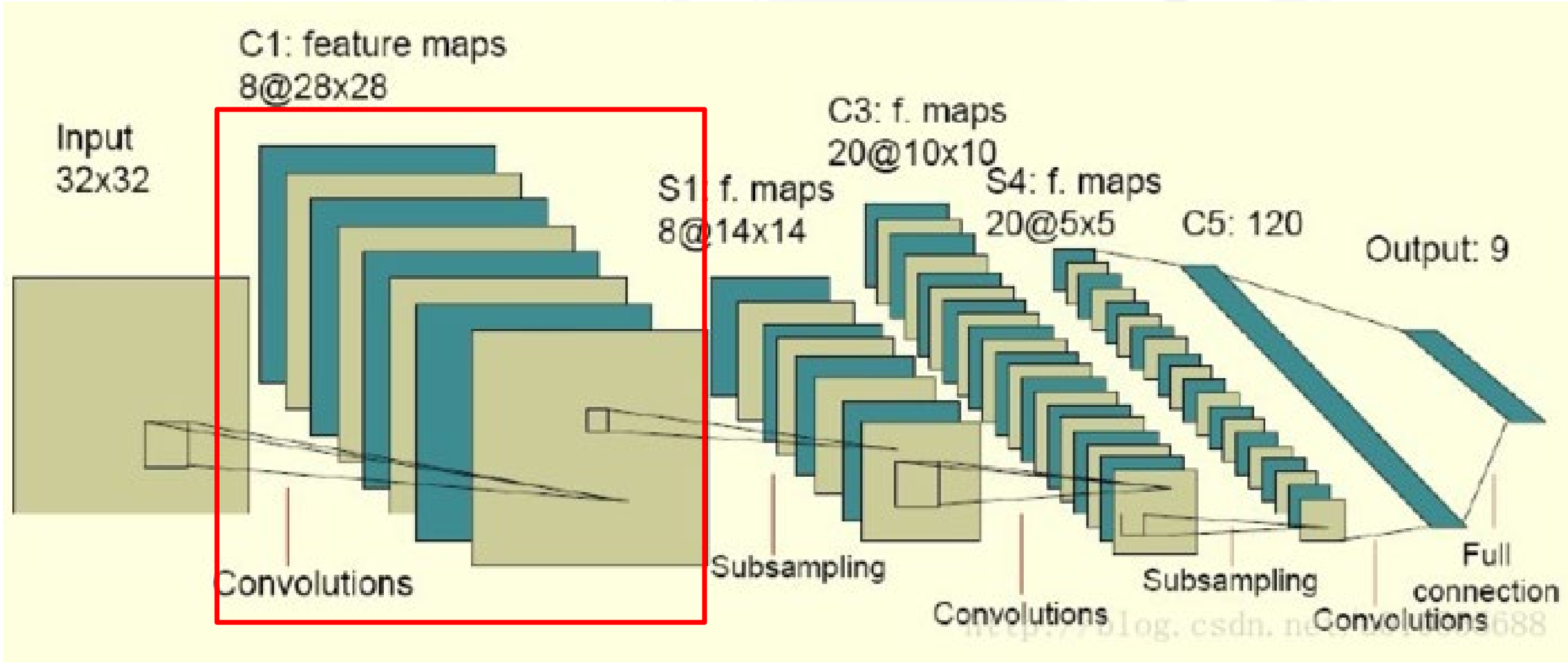
卷积神经网络的若干改进-输入层

输入层：输入层往往输入的是原始图像矩阵或者向量化的文本矩阵等形式。



卷积神经网络的若干改进-卷积层

卷积层：输入层往往输入的是原始图像矩阵或者向量化的文本矩阵等形式。



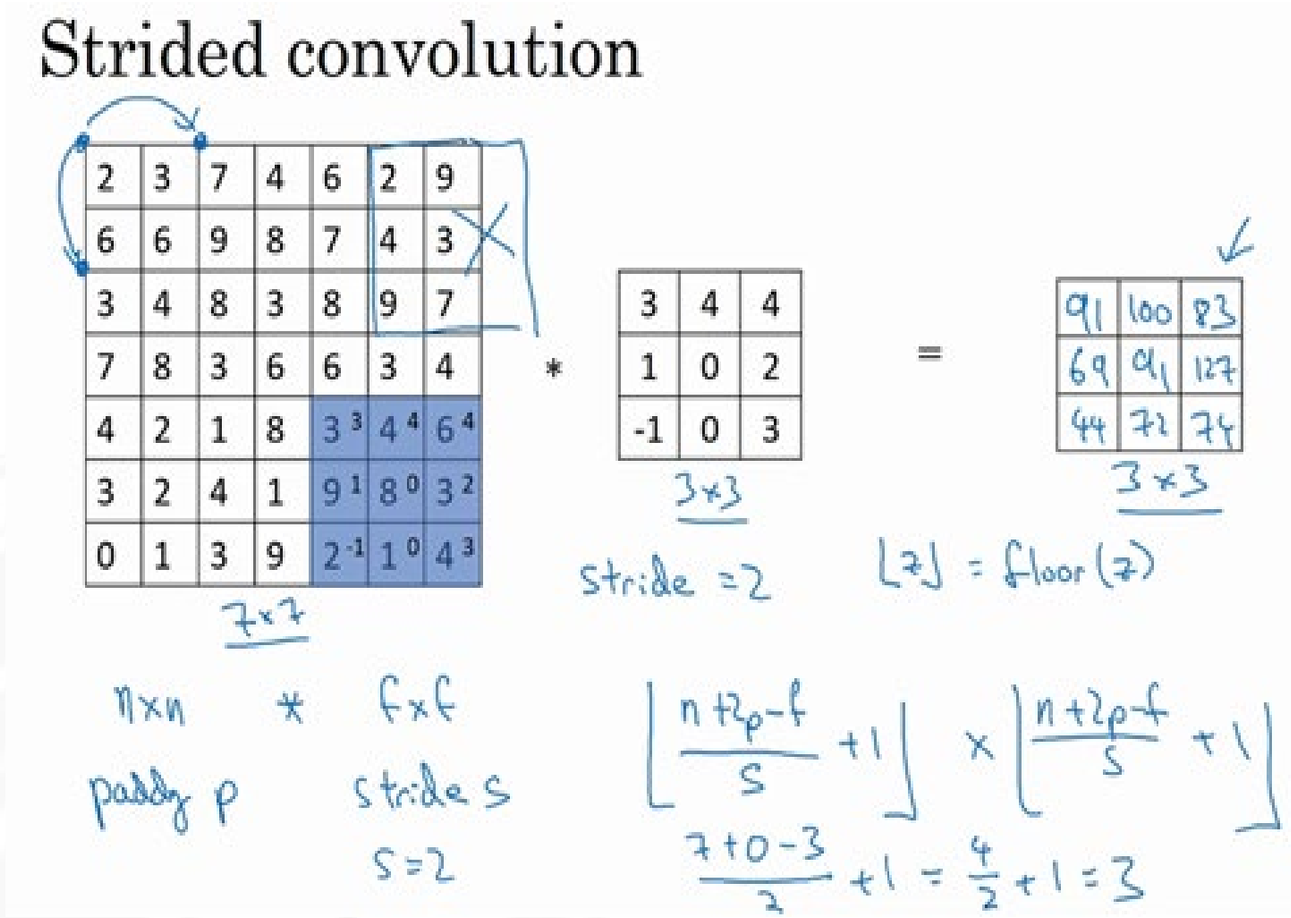
卷积神经网络的若干改进-卷积层

Padding p { **Other** **same卷积**
0 **Valid卷积**

卷积操作之前填充这幅图像

卷积步长 s

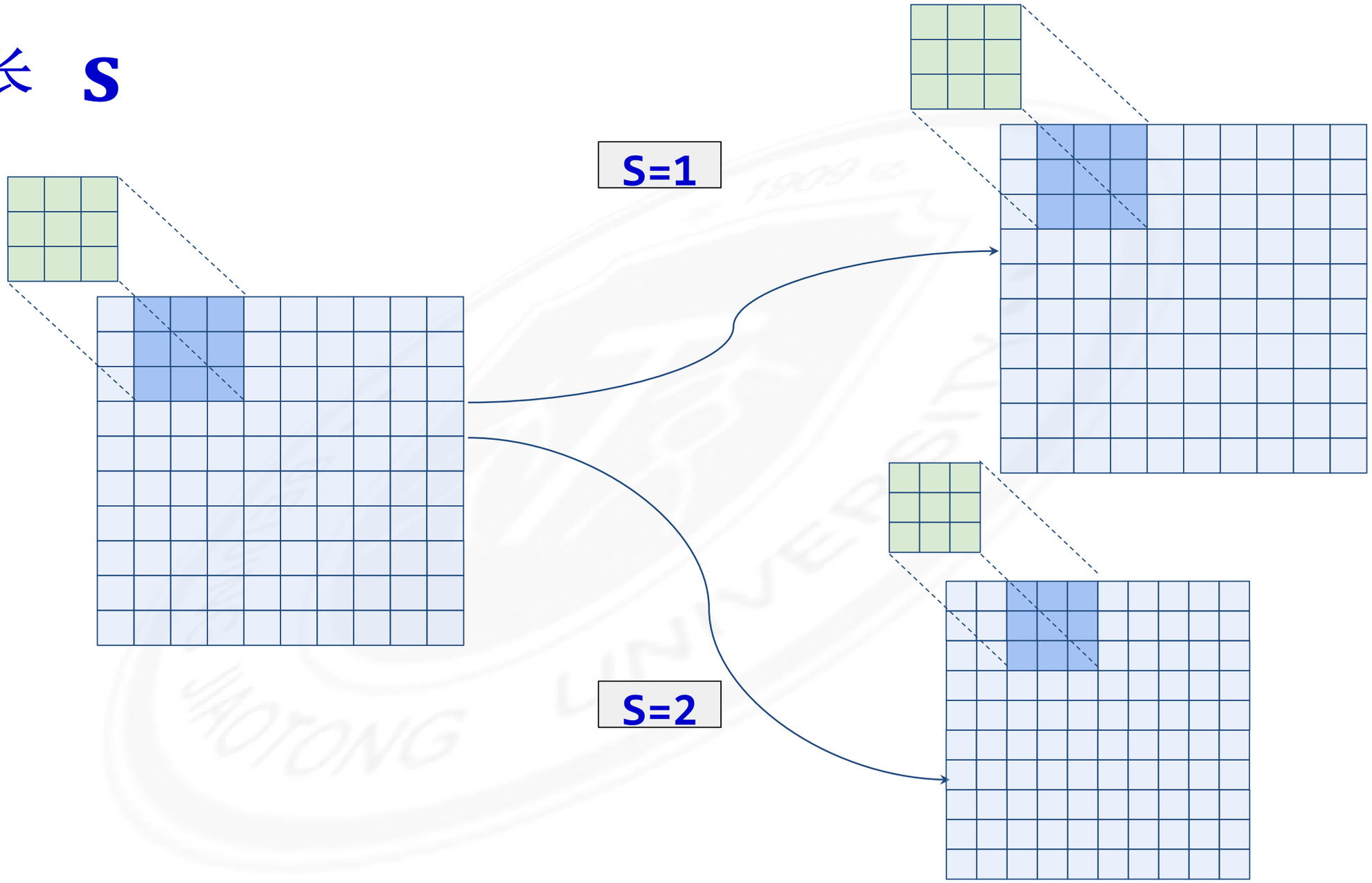
图像尺寸 n



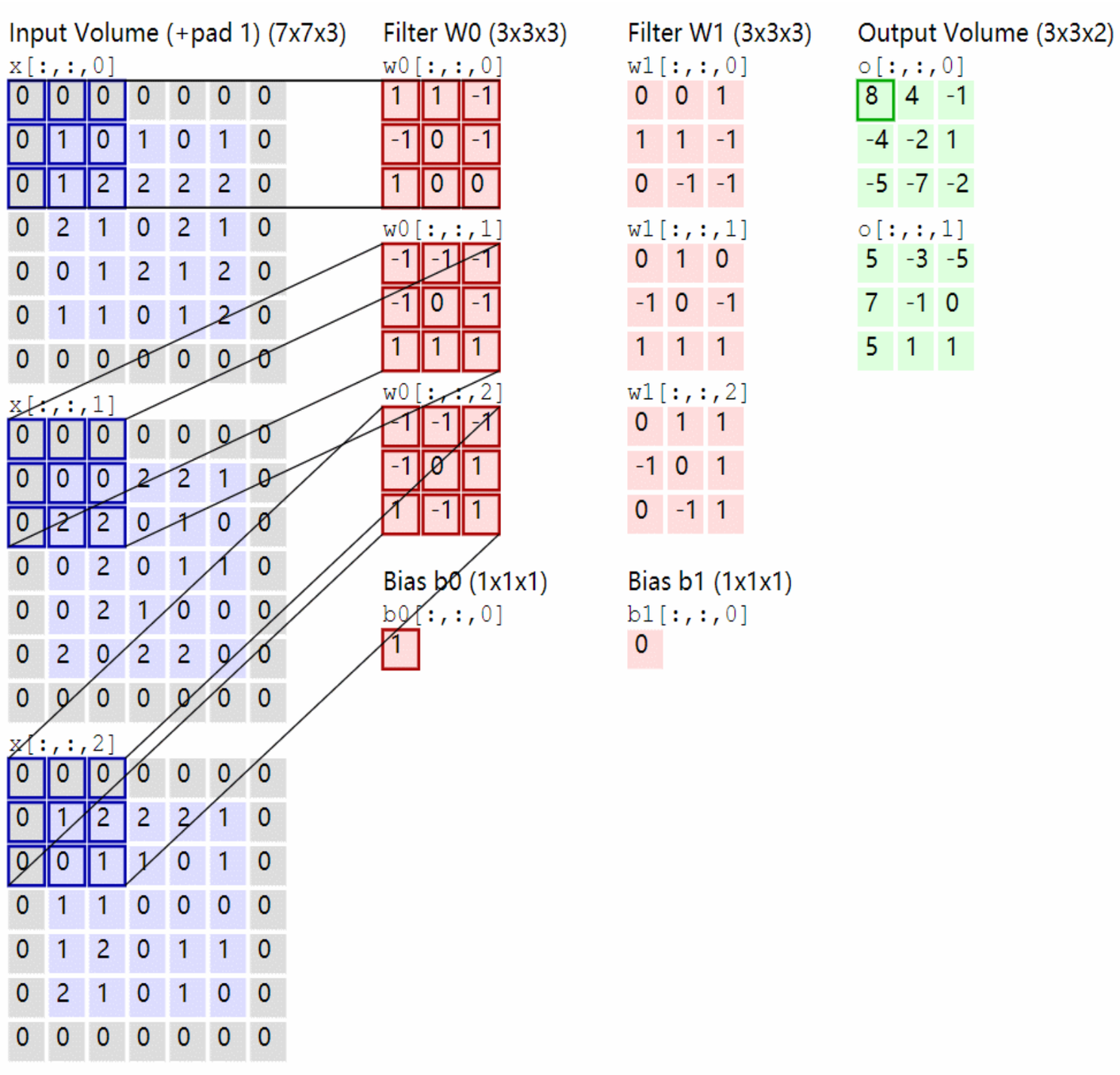
$$n = 7,\ p = 0,\ f = 3,\ s = 2,\ \frac{7+0-3}{2} + 1 = 3$$

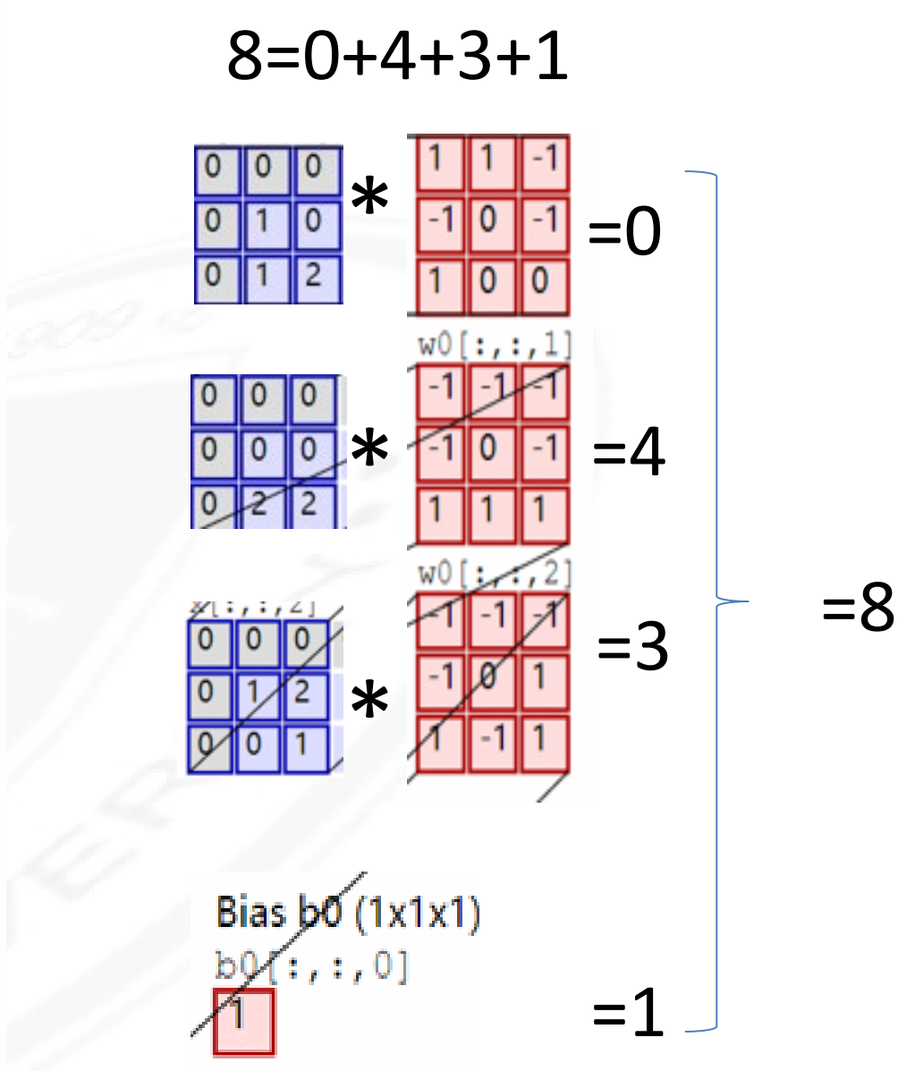
卷积神经网络的若干改进-卷积层

卷积步长 **S**



卷积神经网络的若干改进-卷积层





卷积神经网络的若干改进-参数共享

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

$*$

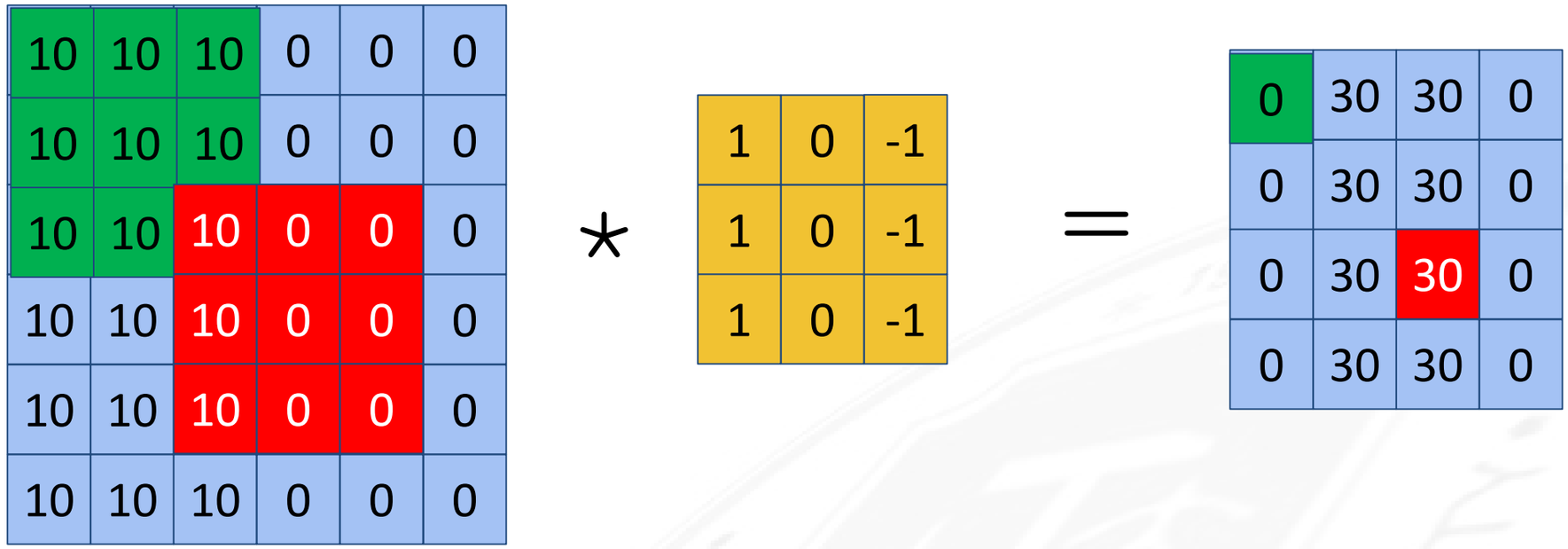
1	0	-1
1	0	-1
1	0	-1

$=$

0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0

如果你用一个3×3的过滤器检测垂直边缘，那么图片的左上角区域，以及旁边的各个区域（左边矩阵中蓝色方框标记的部分）都可以使用这个3×3的过滤器。即使减少参数个数，这9个参数同样能计算出16个输出。

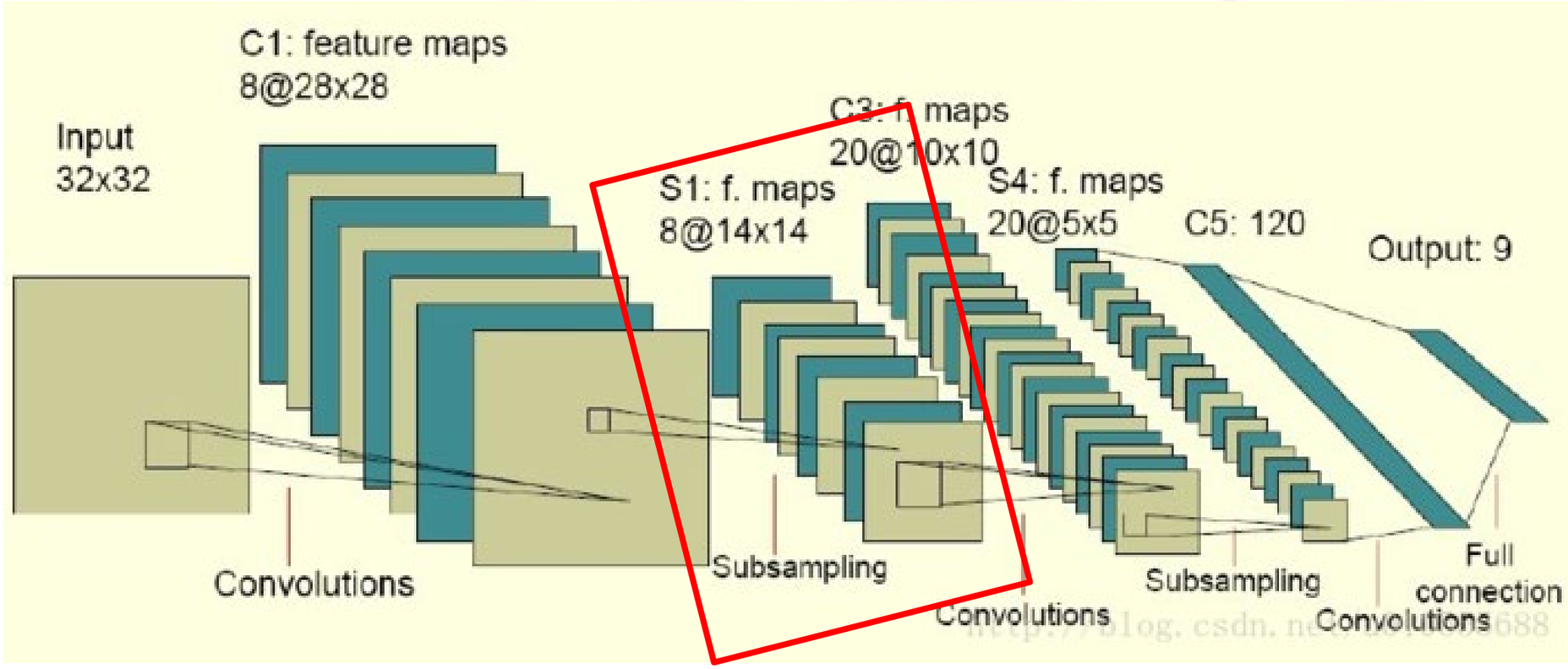
卷积神经网络的若干改进-稀疏连接



右图这个绿色格子的0是通过3×3的卷积计算得到的，它只依赖于这个3×3的输入的单元格，右边这个输出单元（元素0）仅与36个输入特征中9个相连接。而且其它像素值都不会对输出产生任影响，这就是稀疏连接的概念。

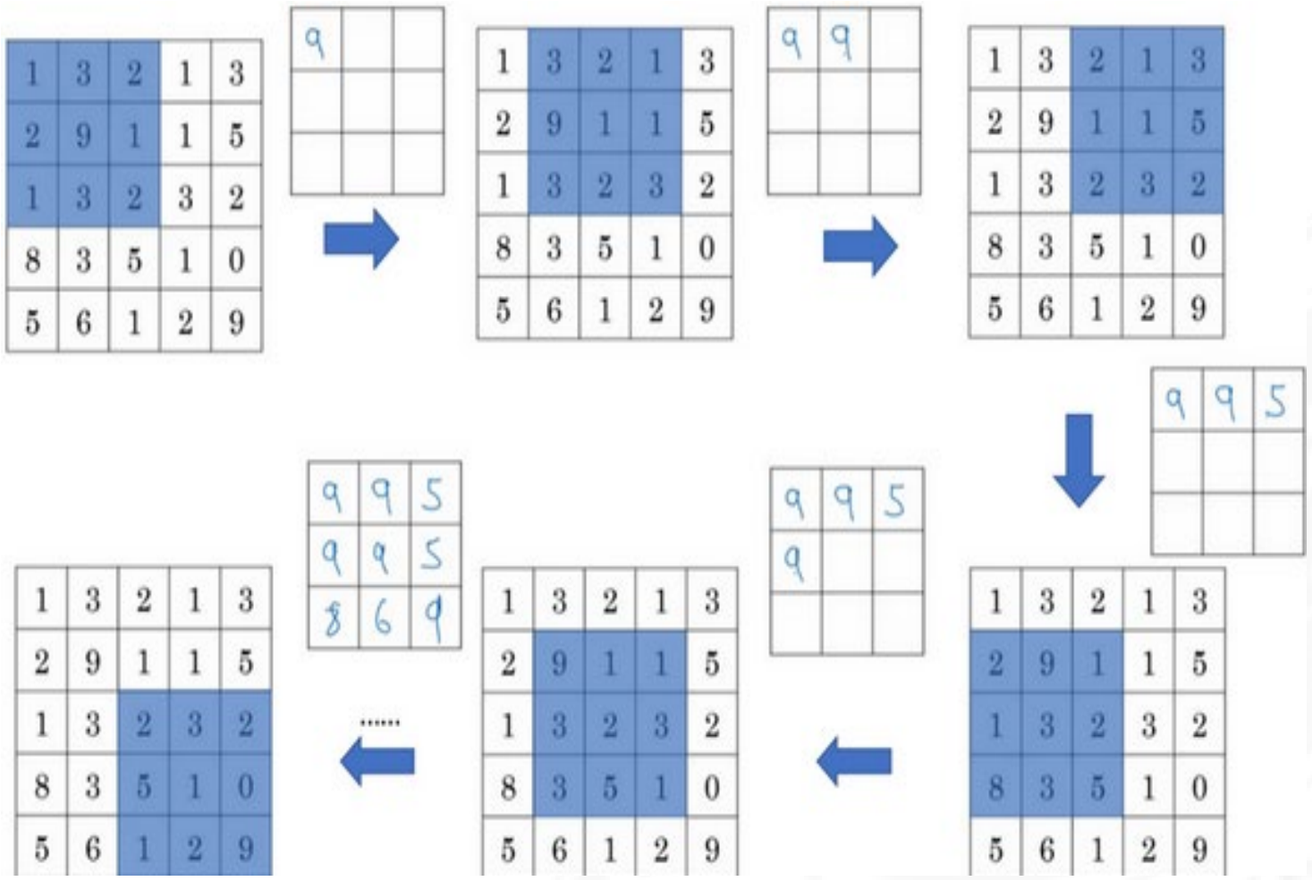
卷积神经网络的若干改进-池化层

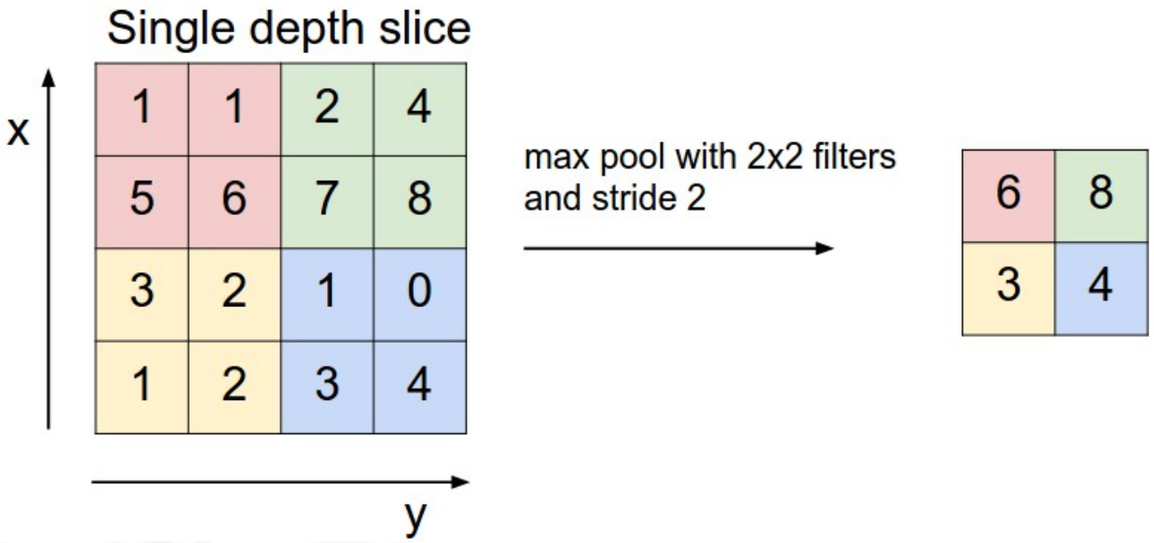
池化层(pooling layer)，通常在卷积层之后得到维度很大的特征，将特征切成几个区域，取其最大值或平均值，得到新的维度较小的特征。常用方法有：均值池化(mean pooling)、最大化池化(max pooling)、重叠采样(overlapping)、均方采样(L2 pooling)、归一化采样(local contrast normalization)，下面重点介绍最大池化实现。



卷积神经网络的若干改进-池化层

最大池化



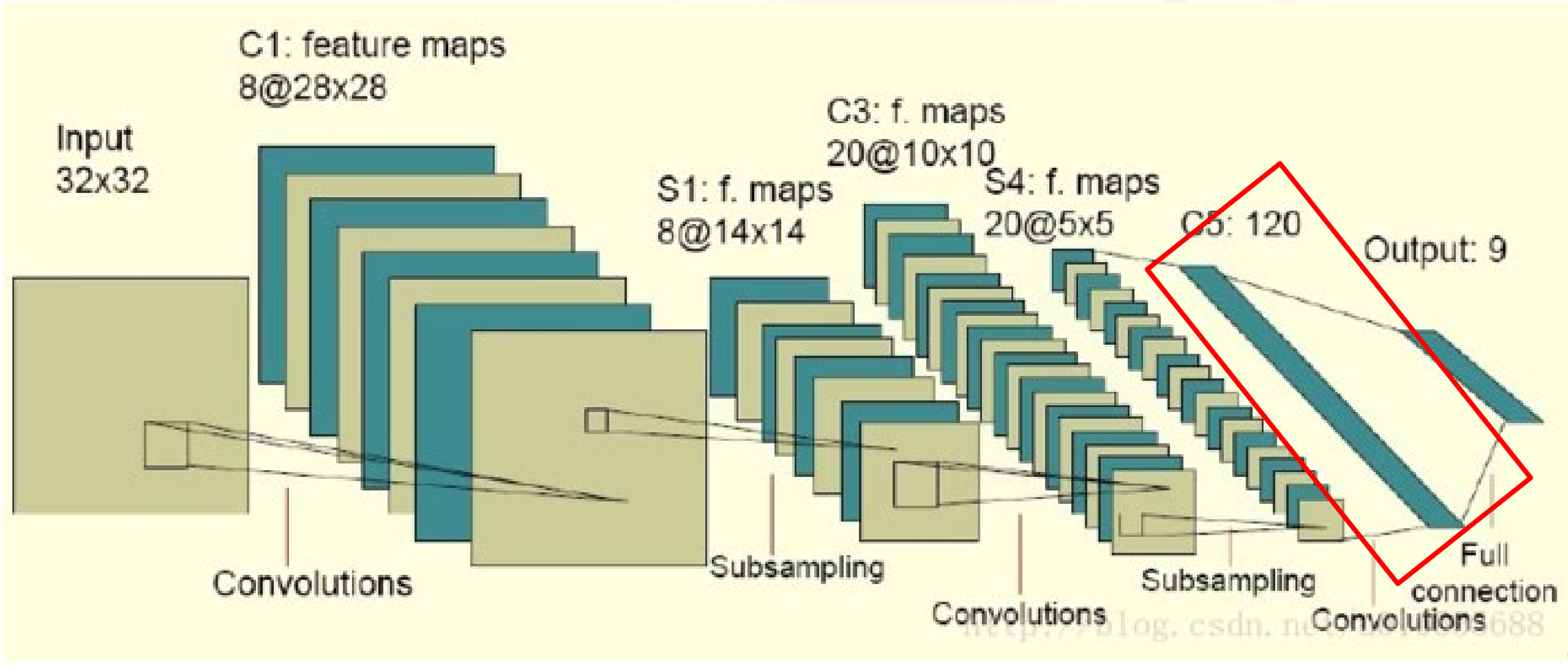


最大池化的输入就是$n_H \times n_W \times n_c$，假设没有padding，则输出

$$\left\lfloor \frac{n_H - f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n_w - f}{s} + 1 \right\rfloor \times n_c$$

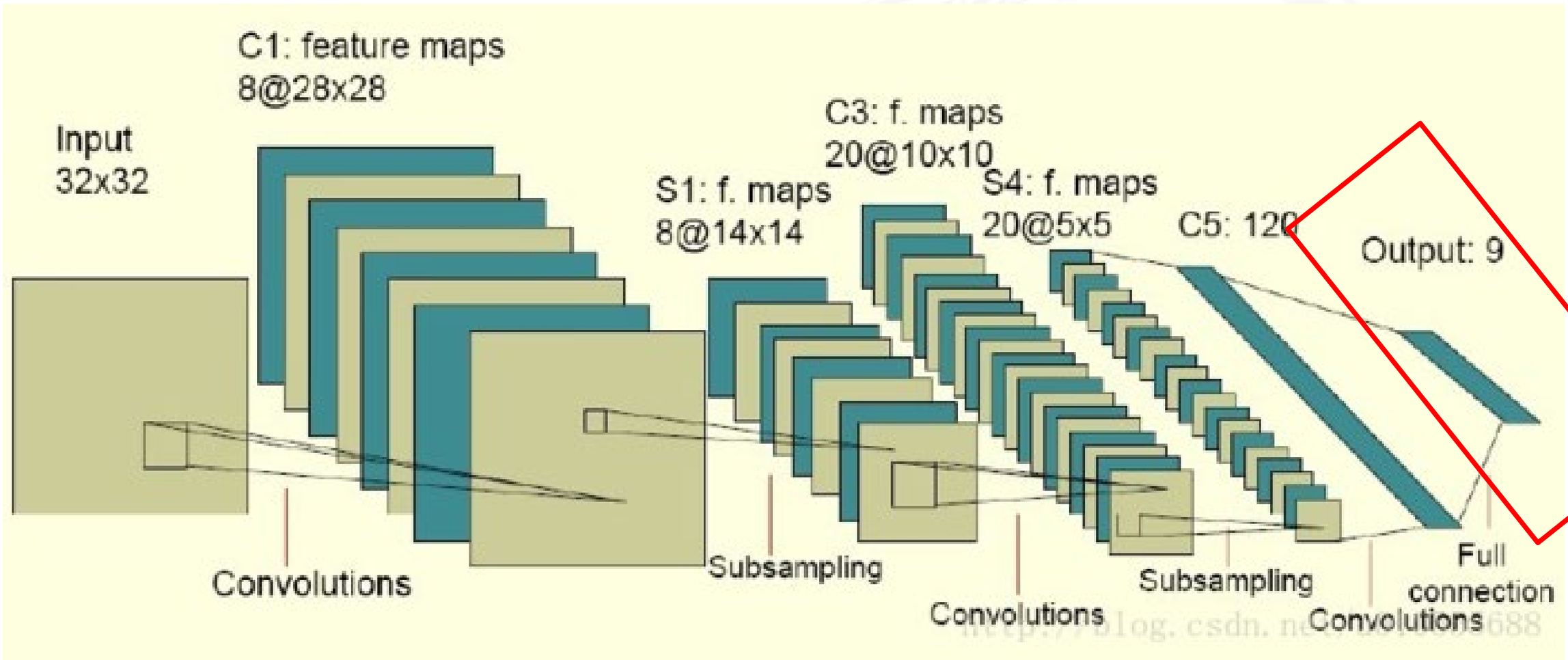
卷积神经网络的若干改进-全连接层

全连接层，把所有局部特征结合变成全局特征，用来计算最后每一类的得分。常采用softmax作为激励函数，进行多分类输出。注意，除了最后全连接层采用softmax作为激励函数外，其他卷积层往往采用RELU作为激励函数。



卷积神经网络的若干改进-输出层

输出层，输出层往往实现分类效果。在实际应用中，往往使用多层卷积，然后再使用全连接层进行训练，多层卷积的目的是一层卷积学到的特征往往是局部的，层数越高，学的的特征就越全局化。



卷积神经网络的若干改进

卷积网络的核心思想是将：局部感受野、权值共享以及降采样这三种结构思想结合起来获得了某种程度的位移、尺度、形变不变性，以达到数据降维的特征学习与分类。

- 全连接网络
  - 权重矩阵的参数非常多

- 卷积神经网络
  - 生物学上感受野

- 卷积神经网络结构上的特性：
  - 稀疏连接
  - 权重共享

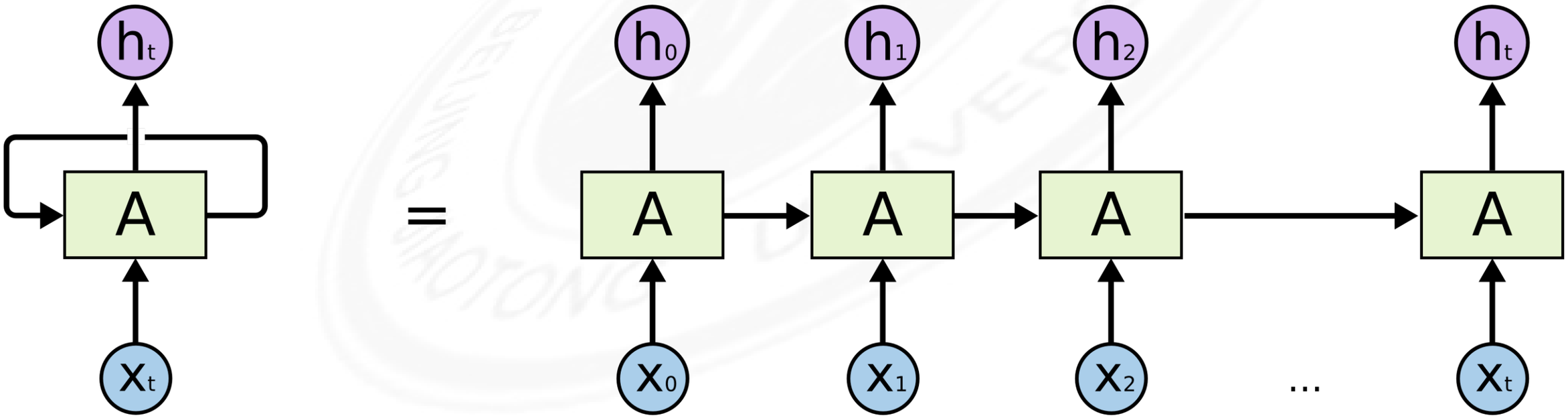
循环神经网络若干改进

循环神经网络（recurrent neural network, RNN）是一种人工神经网络，除了层间的连接以外，同层个单元之间连接构成了一个有向图序列，这种结构允许显示一个时间序列的动态时间行为。

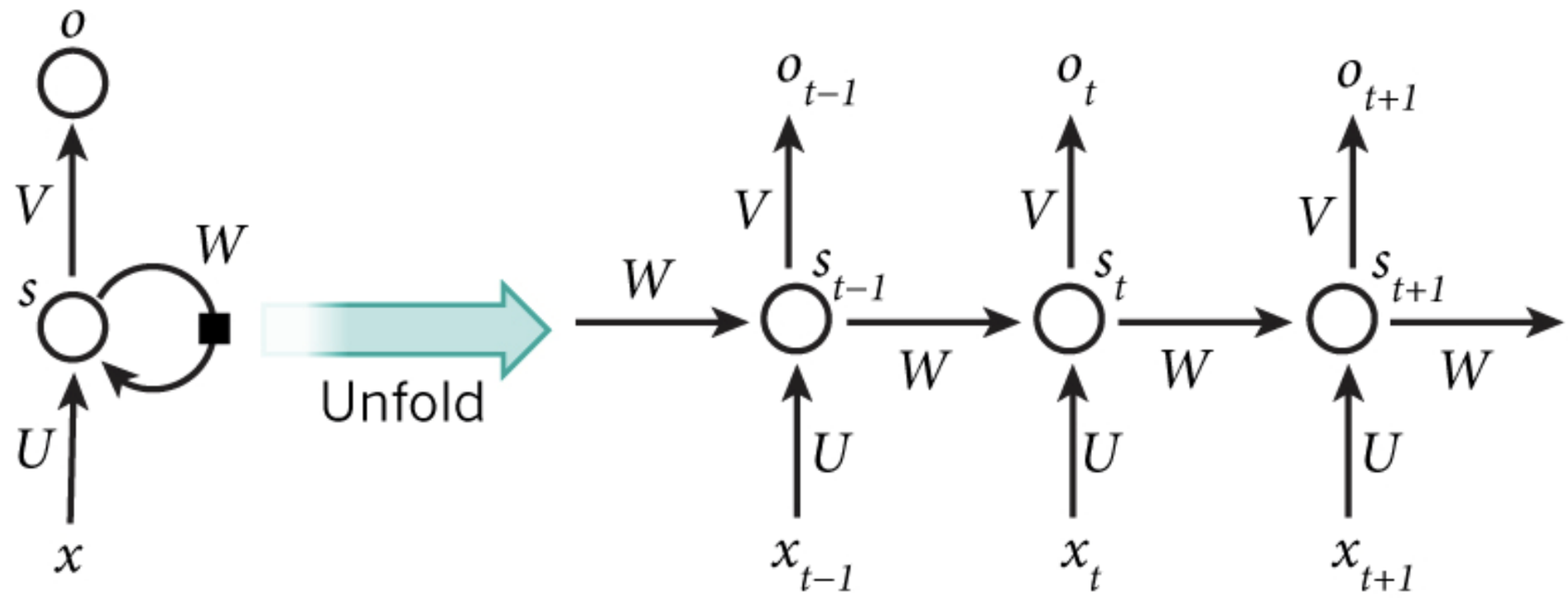
- 循环神经网络通过使用带自反馈的神经元，能够处理任意长度的序列。
- 循环神经网络比前馈神经网络更加符合生物神经网络的结构。
- 循环神经网络已经被广泛应用在语音识别、语言模型以及自然语言生成等任务上。

循环神经网络若干改进——RNN

RNN与传统神经网络的不同之处在于其允许对向量的序列进行操作，输入输出都可以是序列。由于一个序列当前的输出与前面的输出也有关，RNN需要有记忆特性，具体表现为RNN会对前面的信息进行记忆并应用于当前输出的计算中，既隐藏层之间的节点不再是无连接的而是有连接的，并且隐藏层的输入不仅包括输入层的输出还包括上一时刻隐藏层的输出。



循环神经网络若干改进——RNN

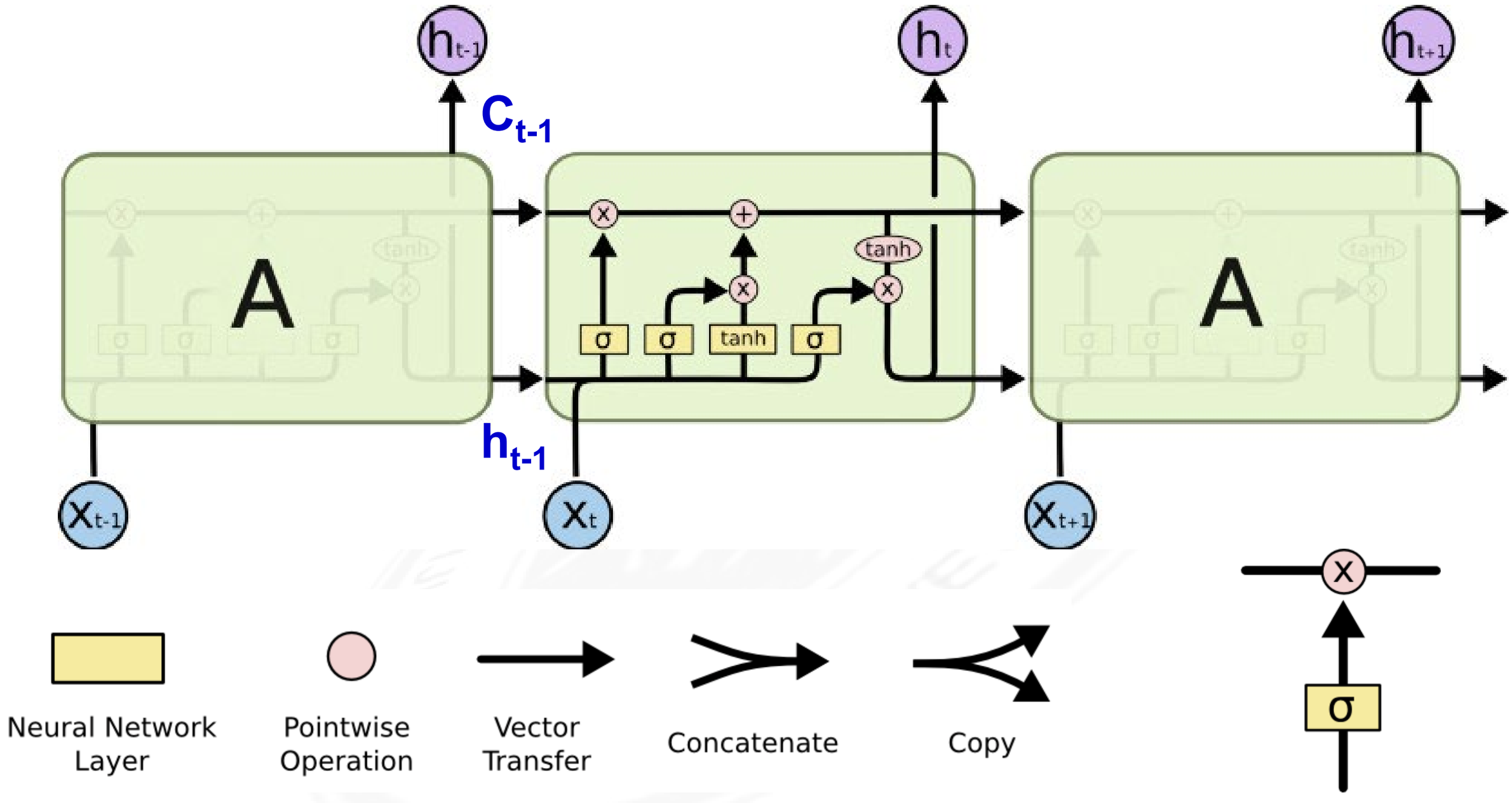


X_t 表示 t 时刻的输入，该时间序列的长度为 T；b_t 是 t 时刻的隐状态，基于上一时刻的隐状态 b_{t-1} 和当前的输入 x_t 得到，既 $b_t = f(Ux_t + Ws_{t-1})$，其中 f 一般是非线性的激活函数。O_t 表示 t 时刻的输出，$O_t = softmax(Vb_t)$

循环神经网络若干改进——LSTM

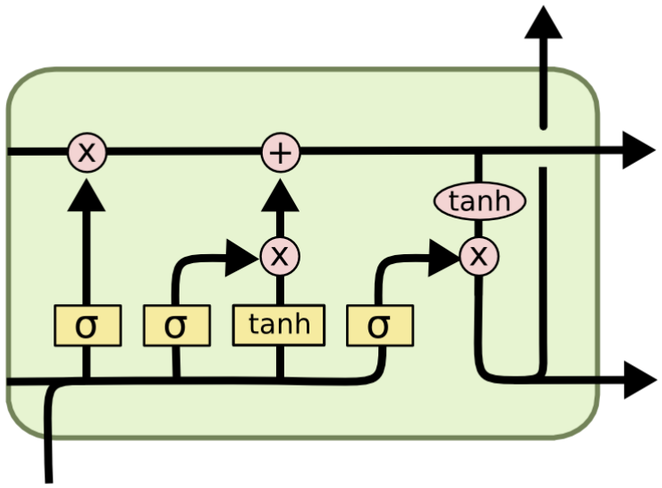
传统RNN较难训练，往往会出现梯度消失或者梯度抱着等，为了解决这一问题，RNN出现了多个扩展版本，在此介绍一种重要模型——LSTM。

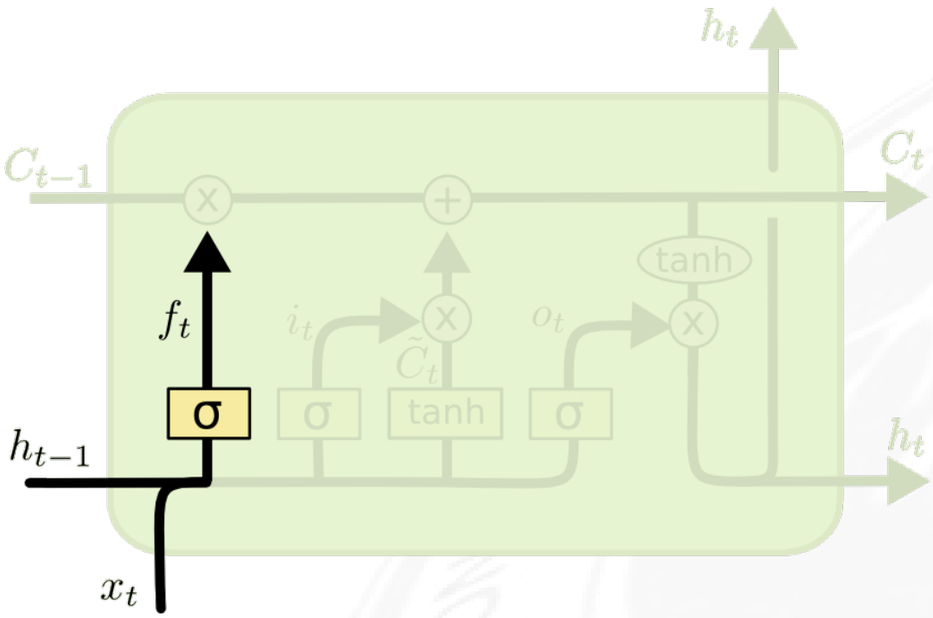
循环神经网络若干改进——LSTM



sigmoid层输出0-1之间的数字，确定每个组件应通过的量。粉色X门是逐点乘法。

长短时记忆神经网络：LSTM—忘记门

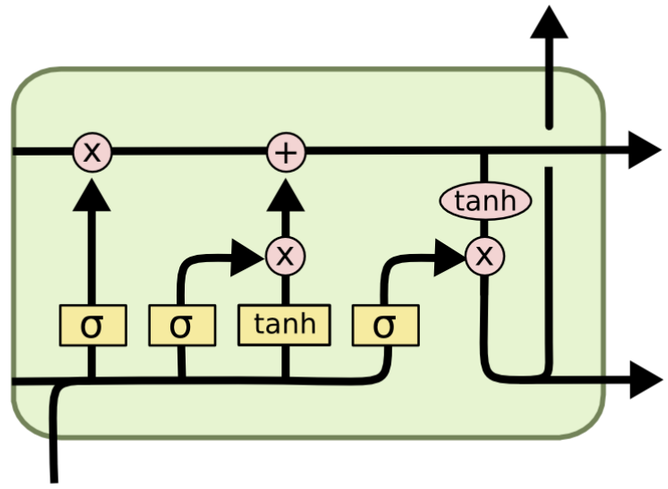


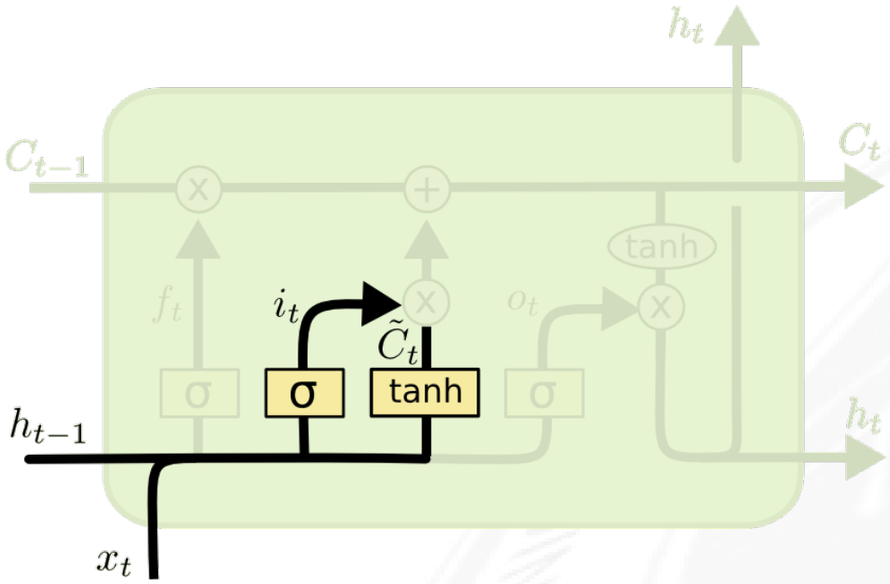


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

核心思想是这种细胞状态C_t，只有轻微的线性相互作用才能缓慢变化。信息很容易在不改变的情况下沿着它流动。

长短时记忆神经网络：LSTM—输入门





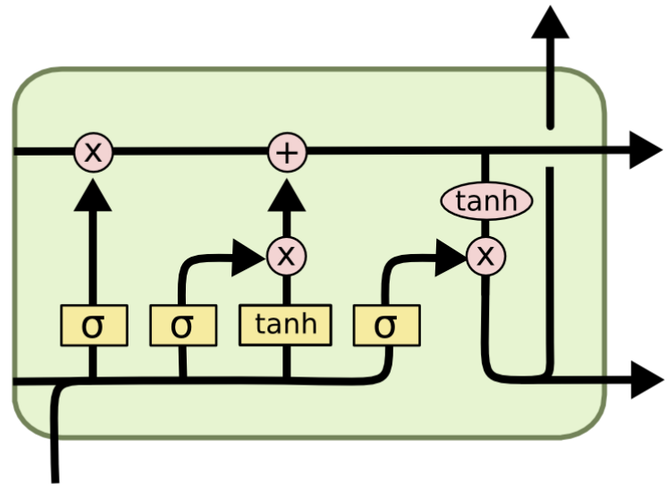
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

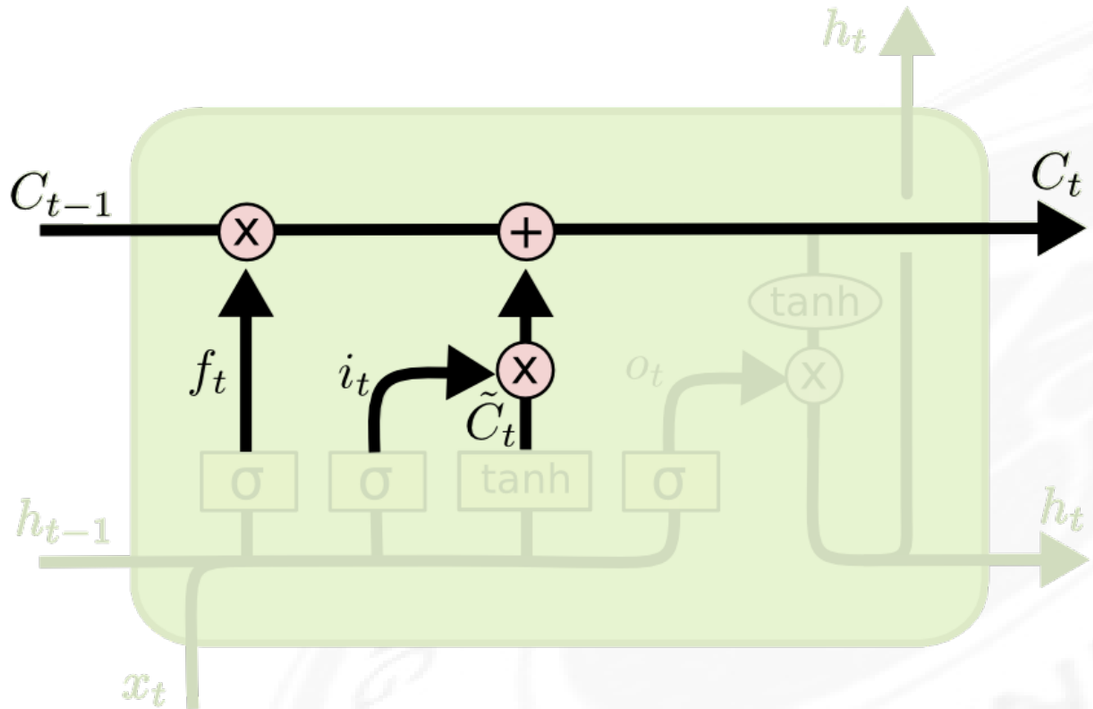
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

i_t 决定要更新的组建部分.

C'_t 提供更新内容

长短时记忆神经网络：LSTM—输出门

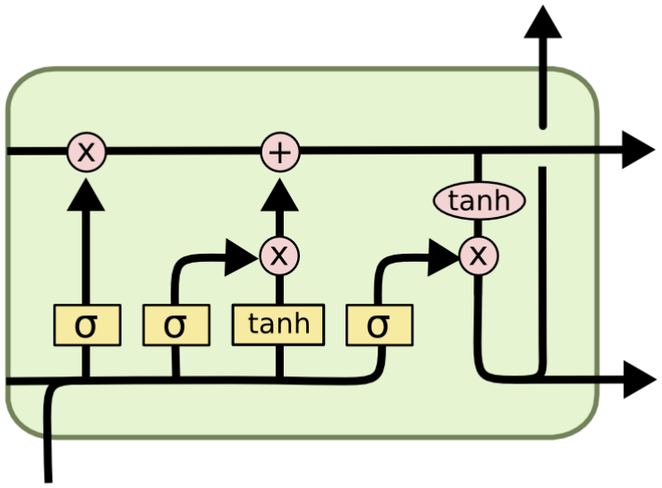


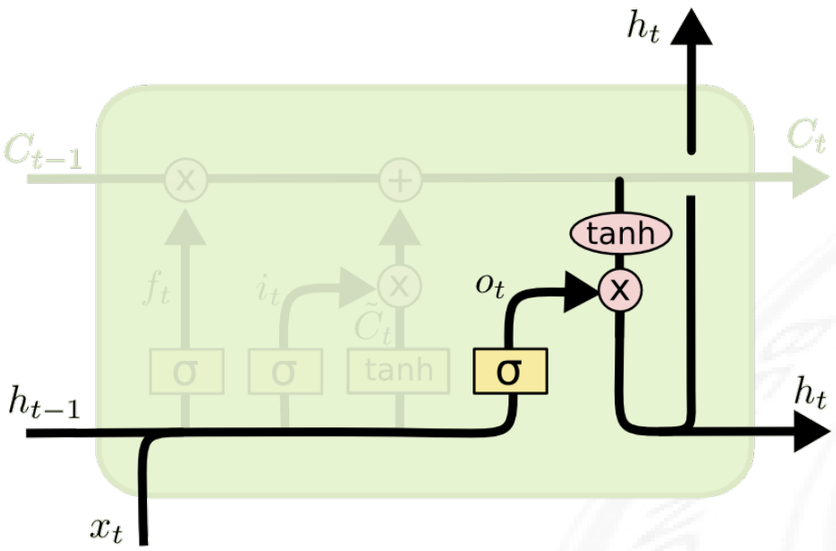


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

输出门控制进入输出的内容，更新细胞状态。

长短时记忆神经网络：LSTM—细胞更新



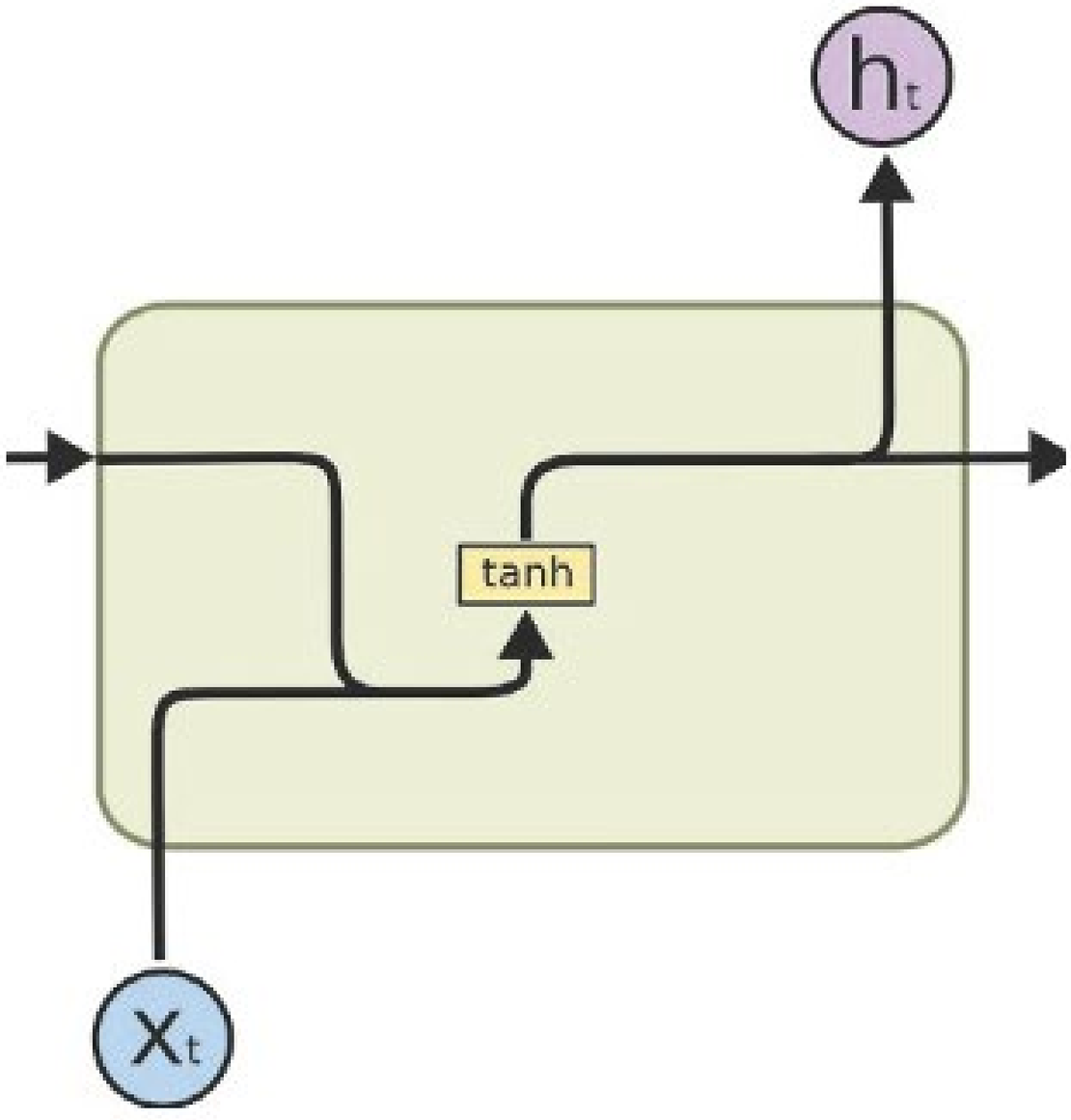


$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

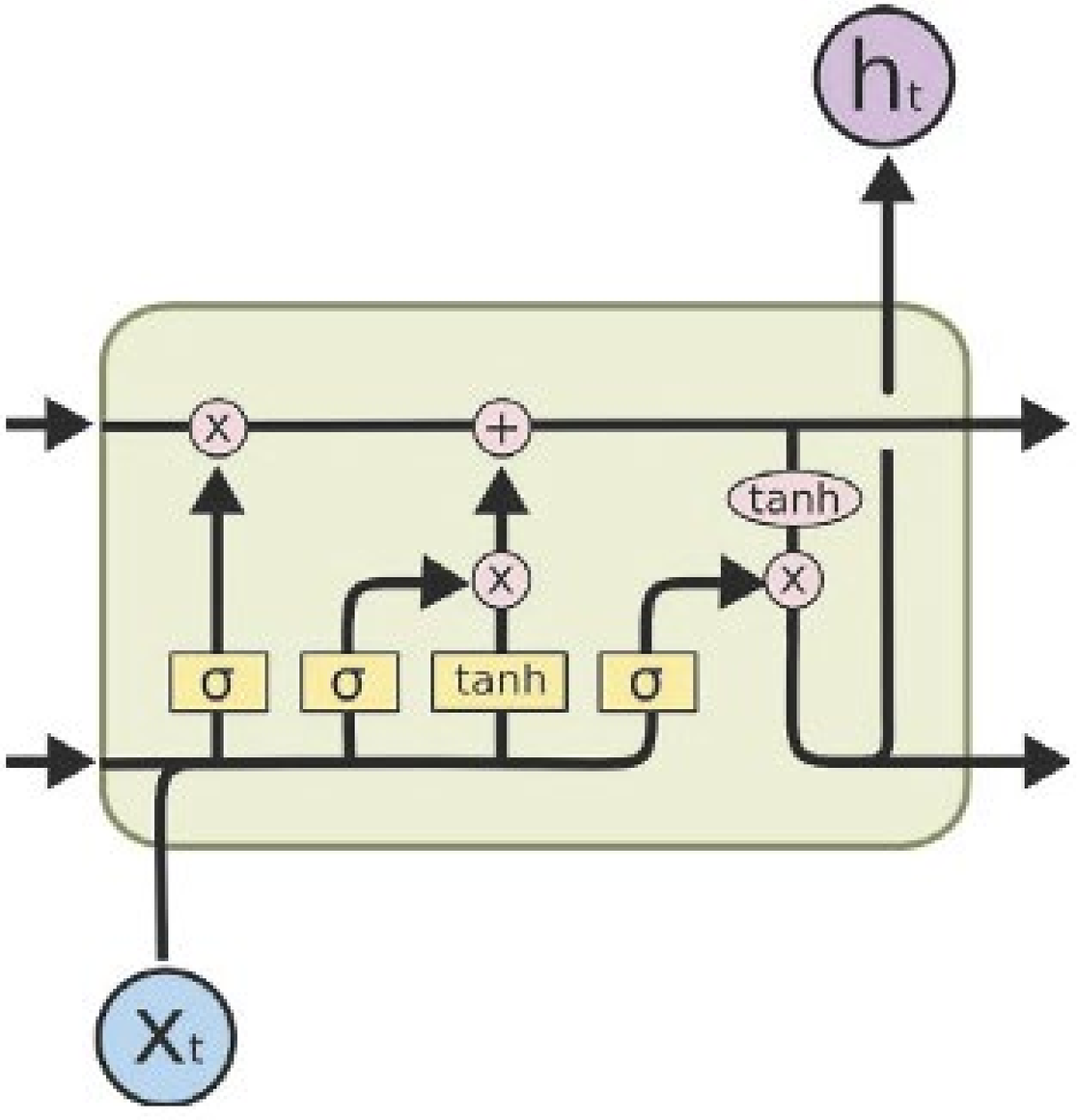
$$h_t = o_t * \tanh\left(C_t\right)$$

决定哪一部分单元格状态能够输出

RNN vs LSTM

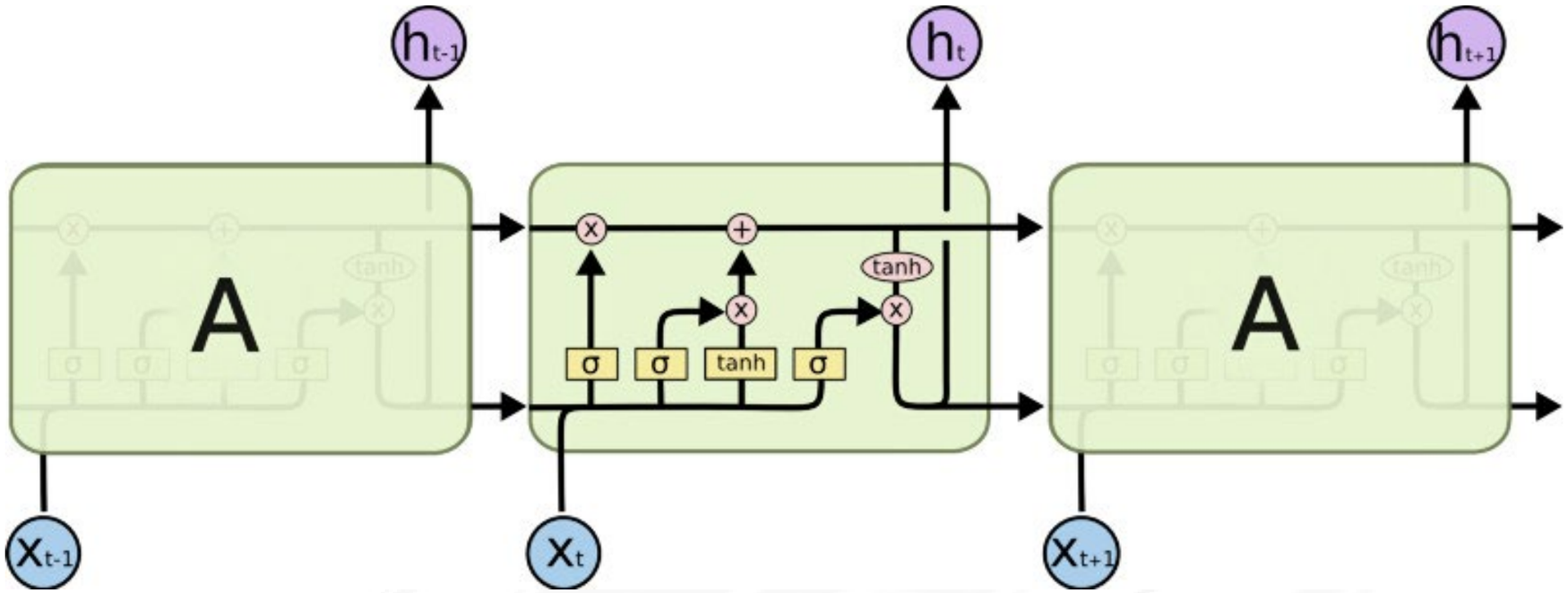


(a) RNN

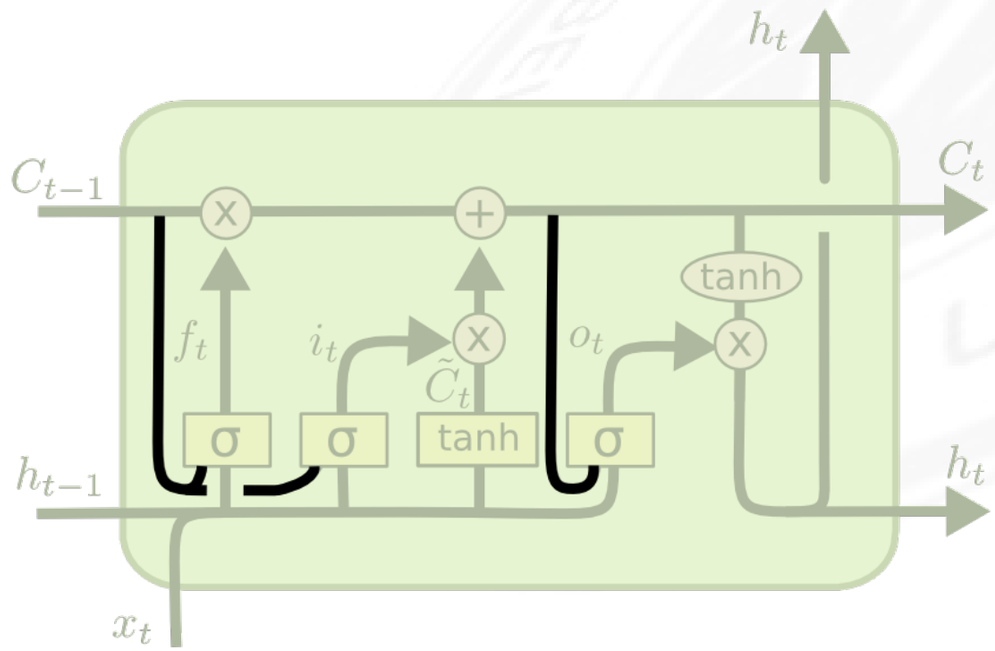


(b) LSTM

长短时记忆神经网络：LSTM



允许模型对前一状态进行窥视，既窥视记忆。



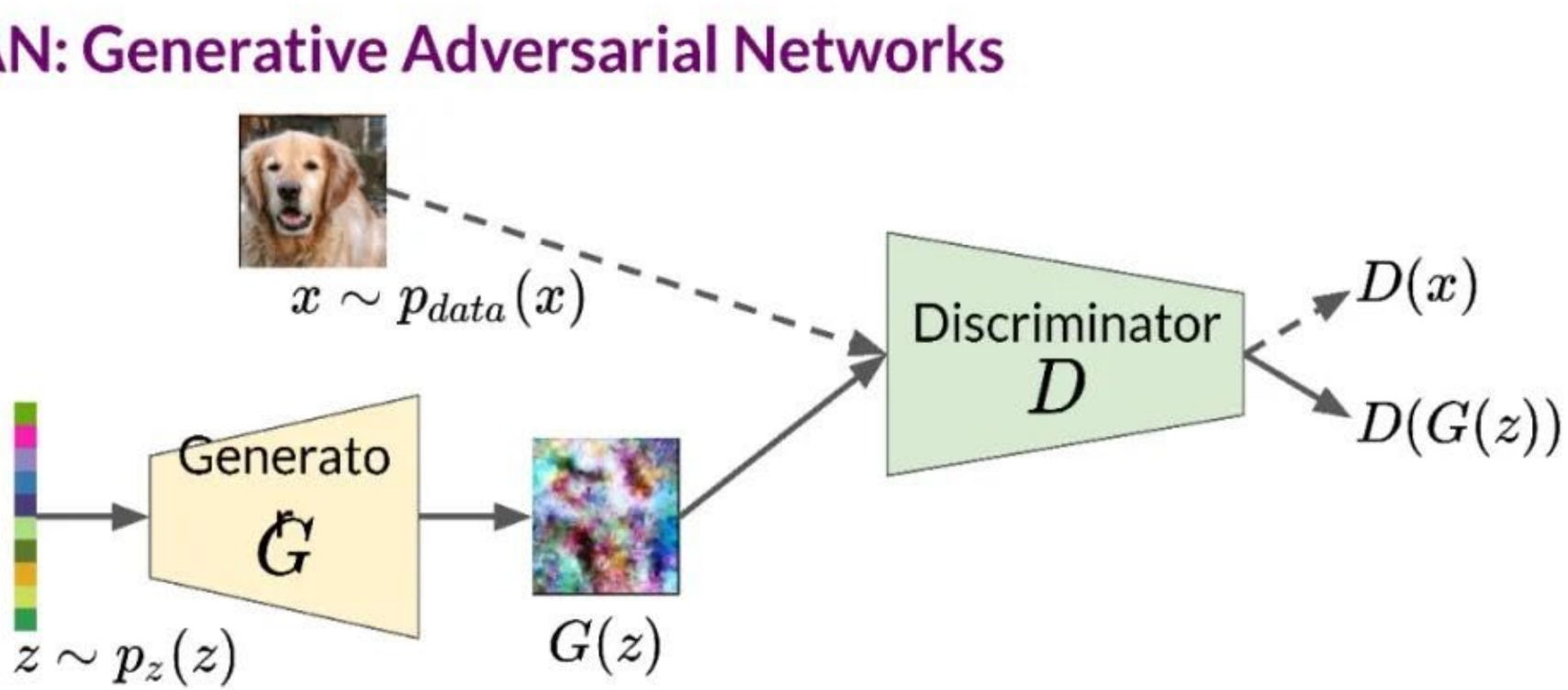
$$f_t = \sigma\left(W_f \cdot [\boldsymbol{C_{t-1}}, h_{t-1}, x_t] \; + \; b_f\right)$$

$$i_t = \sigma\left(W_i \cdot [\boldsymbol{C_{t-1}}, h_{t-1}, x_t] \; + \; b_i\right)$$

$$o_t = \sigma\left(W_o \cdot [\boldsymbol{C_t}, h_{t-1}, x_t] \; + \; b_o\right)$$

生成式对抗神经网络

- 生成对抗网络由一个生成网络与一个判别网络组成。生成网络从潜在空间（latent space）中随机采样作为输入，其输出结果需要尽量模仿训练集中的真实样本。
- 判别网络的输入则为真实样本或生成网络的输出，其目的是将生成网络的输出从真实样本中尽可能分辨出来。



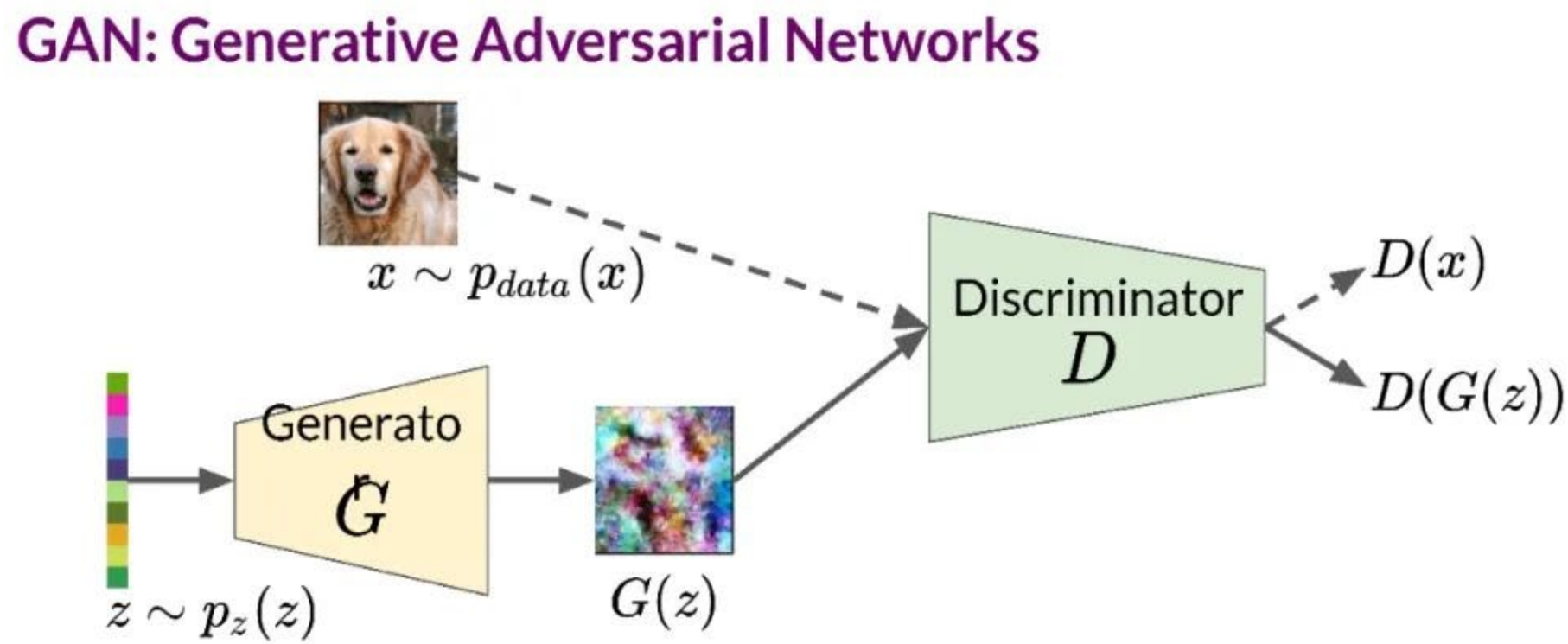
$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

生成式对抗神经网络

GANs有两个网络,分别是生成 器G和判别器D。它们的功能分别是:

- G作为一个生成图片的网络,它接收一个随机的噪声z,通过这个噪声生成图片,记做G(z)
- D作为一个判别网络,判别一张图片是不是“真实的”。它的输入参数是x,x代表一张图片,输出D(x)代表x为真实图片的概率,如果为1,就代表100%是真实的图片,而输出为0,就代表不是真实图片。
- 在训练过程中,G的目标就是尽量生成真实的图片去欺欺骗D而D的目标就是尽量把G生成的图片和真实的图片区分开。这样,G和D构成了一个动态的“博弈过程”。

生成式对抗神经网络



$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

其中,G代表牛成器。代表判别器,训练时分别对D和G进行交互迭代固定G,优化D,一段时间后,固定D再优化G,直到过程收敛。

生成式对抗神经网络—训练过程

算法 13.1: 生成对抗网络的训练过程

输入: 训练集 $\mathcal{D}$，对抗训练迭代次数 T，每次判别网络的训练迭代次数 K，小批量样本数量 M

随机初始化 θ, ϕ;

for $t \leftarrow 1$ **to** T **do**

// 训练判别网络 $D(\mathbf{x}, \phi)$

for $k \leftarrow 1$ **to** K **do**

// 采集小批量训练样本

从训练集 $\mathcal{D}$ 中采集 M 个样本 $\{\mathbf{x}^{(m)}\}, 1 \leq m \leq M$;

从分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ 中采集 M 个样本 $\{\mathbf{z}^{(m)}\}, 1 \leq m \leq M$;

使用随机梯度上升更新 ϕ，梯度为

$$\frac{\partial}{\partial \phi}\left[\frac{1}{M}\sum_{m=1}^{M}\left(\log D(\mathbf{x}^{(m)}, \phi) + \log\left(1 - D\Big(G(\mathbf{z}^{(m)}, \theta), \phi\Big)\right)\right)\right];$$

end

// 训练生成网络 $G(\mathbf{z}, \theta)$

从分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ 中采集 M 个样本 $\{\mathbf{z}^{(m)}\}, 1 \leq m \leq M$;

使用随机梯度上升更新 θ，梯度为

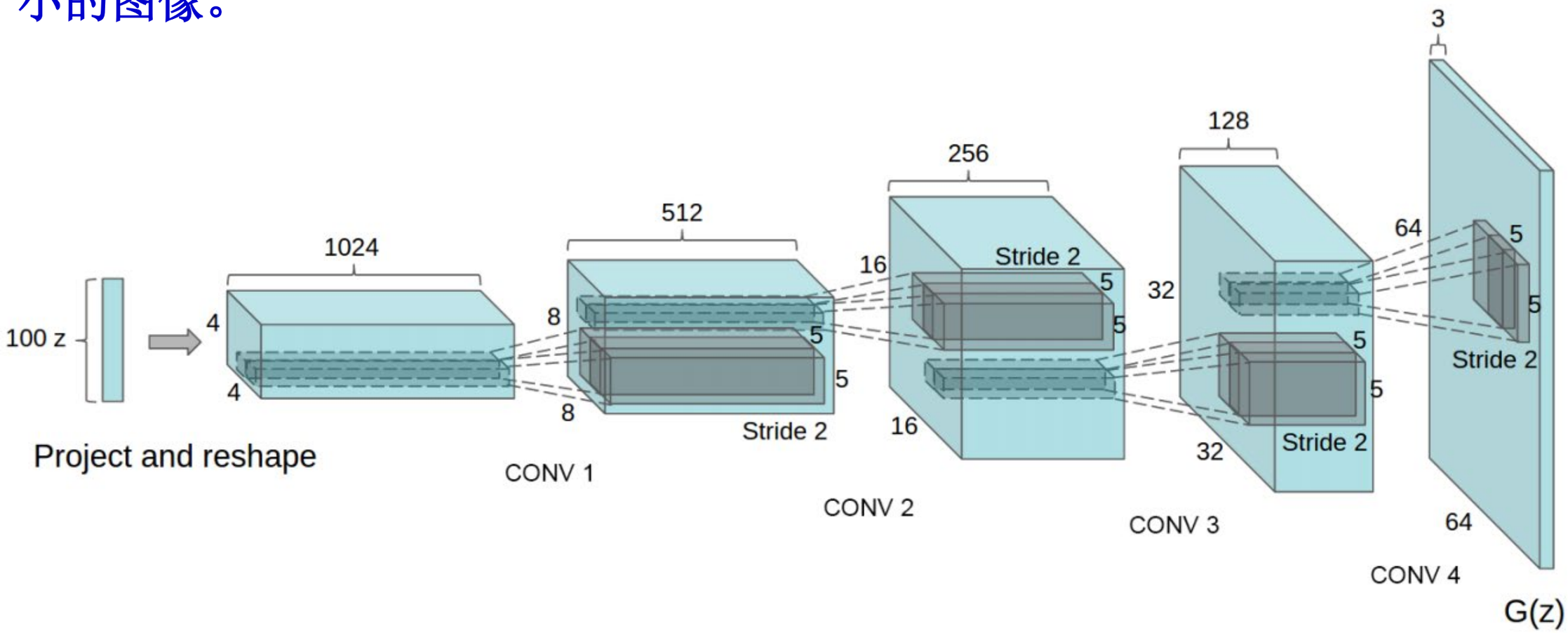
$$\frac{\partial}{\partial \theta}\left[\frac{1}{M}\sum_{m=1}^{M} D\Big(G(\mathbf{z}^{(m)}, \theta), \phi\Big)\right];$$

end

输出: 生成网络 $G(\mathbf{z}, \theta)$

生成式对抗神经网络—DCGANs

- 判别网络是一个传统的深度卷积网络，但使用了带步长的卷积来实现下采样操作，不用最大汇聚（pooling）操作。
- 生成网络使用一个特殊的深度卷积网络来实现使用微步卷积来生成64×63大小的图像。



生成式对抗神经网络—DCGANs改进

(1)使用全卷积网络。在判别模型中,使用带步长的卷积(sidedconvolutions)取代空间池化(spatial pooling)。这种形式可以让更多的前层信息传递到后层上去。另外,在生成模型中,使用反卷积机制(fractional strided),可以学习自己的空间上采样。

(2)取消了全连接层,直接用卷积层连接输入层和输出层

(3)批归一化(batch normalization)。除了生成器模型的输出层和判别器模型的输人层,在网络其他层上都使用了批归一化,使川批归化可以稳定学习,有助于处理初始化不良导致的训练问题。

(4)在生成器的输出层采用Tanh做激励函数,其他层使用ReLU做激励函数。判别器上统一使用 eaky Relu做激励函数