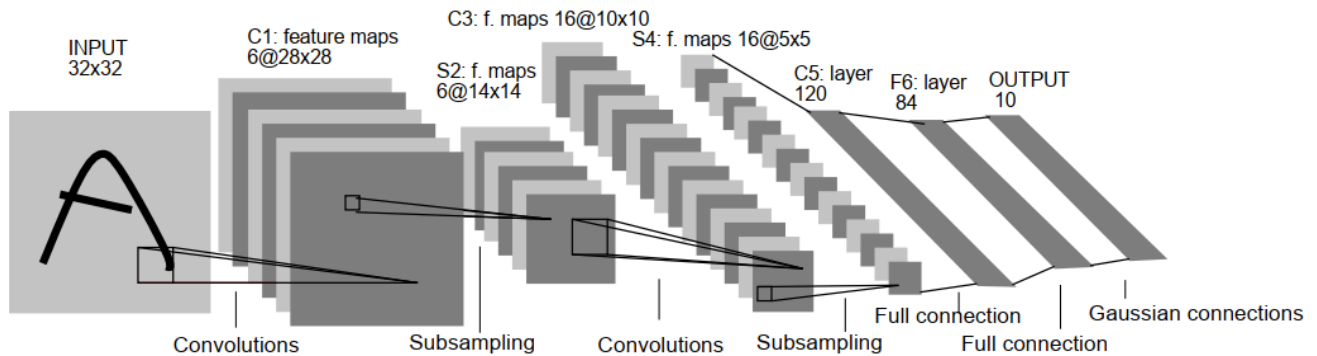


CNN

1 LeNet-5



- LeNet共分为7层

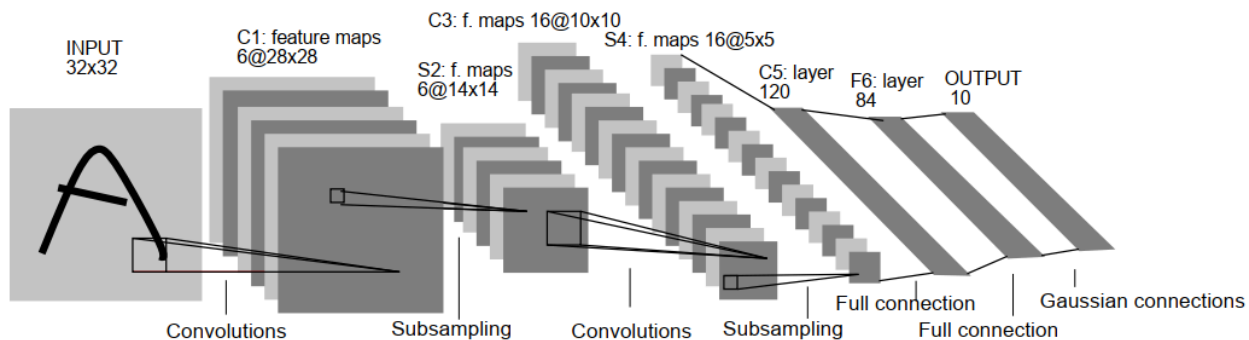
- C1 - 卷积层

- 输入：32 × 32图像
 - 卷积核种类（决定输出特征图的数量）：6
 - 卷积核大小：5 × 5（限于当时计算机资源水平，论文里使用的5×5）
 - 特征图大小：28 × 28 - 没有进行填充（padding=0）
 - 输出特征图大小公式： $32 - 5 + 1 = 28$

$$\text{输出尺寸} = \frac{\text{输入尺寸} - \text{卷积核尺寸} + 2 \times \text{填充 (padding)}}{\text{步长 (stride)}} + 1$$

- 可训练参数：(5 × 5 + 1) × 6
卷积核是一个5×5的矩阵，里面的每一个数都是要通过训练得到的。在实际运算中还要加上一个偏置bias，所以每一个卷积核需要训练5×5+1个参数，6个卷积核就需要训练(5 × 5 + 1) × 6 = 156个参数。

- S2 - 池化层



- 输入：6张 28×28 的特征图

- 采样区域： 2×2

LeNet-5 的池化层 S2 采用的是平均池化 (average pooling)

取该窗口内的 4 个像素值，计算均值后再乘以一个可训练的权重系数，再加上一个可训练的偏置，最后通过 Sigmoid 激活函数

- 输出特征图的大小： 14×14

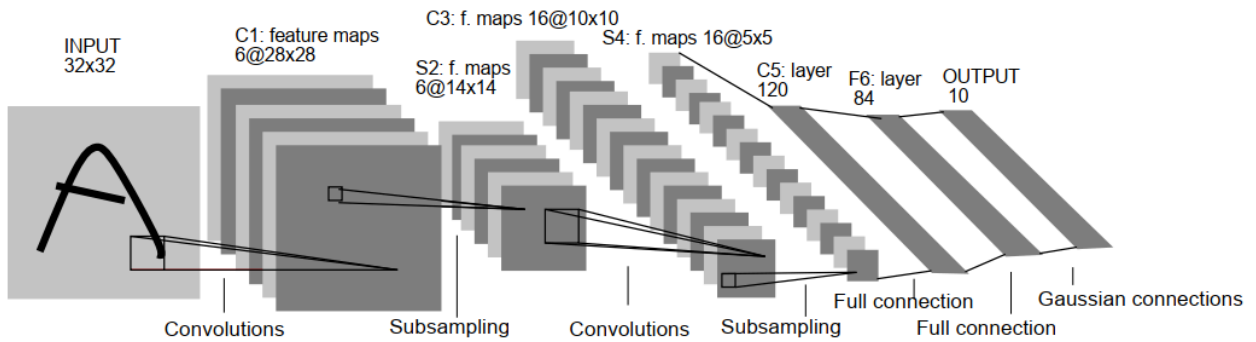
- 输出特征图数量：6

- **可训练参数： 2×6**

对于每张特征图来说，只有两个参数需要确定：用于相乘的“可训练参数 w ”与“可训练偏置 b ”，一共6张特征图，因此需要训练 2×6 个参数

- 池化基本思想其实就是使特征图尺度减小，降维，同时保留主要特征

- C3 - 卷积层



- 输入：14×14的特征图（6张）

- 卷积核种类：16

- 卷积核大小：5×5

- 输入特征图数量：6

- 输出特征图数量：16

- C3层卷积方式：

在C1卷积层中，卷积核有6种，输入图像只有一张，因此只需要将6个卷积核分别对一张图像进行卷积操作，最后得到6张特征图。

对于输入6张图像的处理方法是：“每个卷积核对多张特征图”进行处理” - 部分连接模式。

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X		X	X	X	X	X		X	X
1	X	X				X	X	X		X	X	X	X	X		X
2	X	X	X				X	X	X		X		X	X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

横轴为编号0~15的16个卷积核，纵轴为0~5张输入的特征图，X表示为连接。

- 输出特征图大小：10×10(padding=0)

输出尺寸 = $\frac{\text{输入尺寸} - \text{卷积核尺寸} + 2 \times \text{填充 (padding)}}{\text{步长 (stride)}} + 1$

输出特征图的边长为：14 - 5 + 1 = 10

- 可训练参数：1516

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X		X	X	X	X	X		X	X
1	X	X				X	X	X		X	X	X	X	X		X
2	X	X	X				X	X	X		X		X	X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

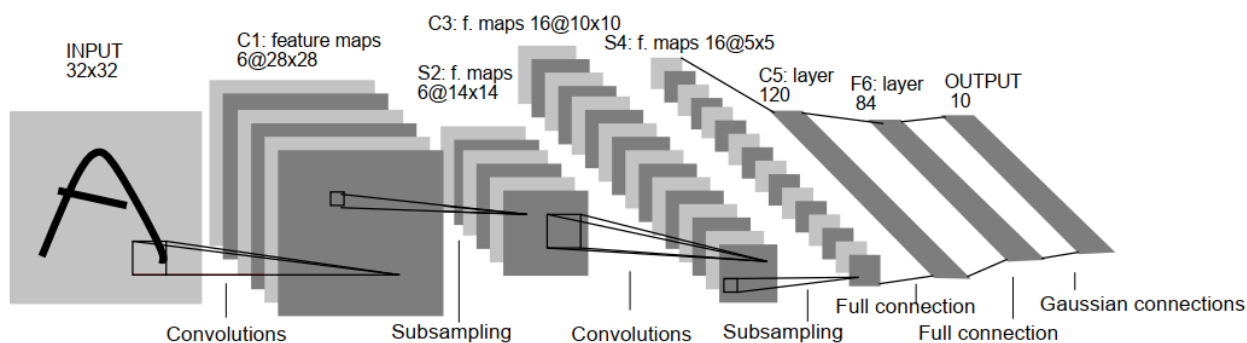
第一个红框，每个卷积核包含5×5个可训练参数；在标号为0~5的卷积核需要分别与3张特征图相连，最后还要加上一个偏置，因此需要训练3 × 5 × 5 + 1个参数。第一个红框内有6个这样的卷积核，因此共需要训练6 × (3 × 5 × 5 + 1)个参数。

同理，第二个红框，共需要训练6 × (4 × 5 × 5 + 1)个参数；对于第三个红框，其共需要训练3 × (4 × 5 × 5 + 1)个参数；对于第四个红框，其共需要训练1 × (6 × 5 × 5 + 1)个参数。

总计可训练 $6 \times (3 \times 5 \times 5 + 1) + 6 \times (4 \times 5 \times 5 + 1) + 3 \times (4 \times 5 \times 5 + 1) + 1 \times (6 \times 5 \times 5 + 1) = 1516$ 个参数。

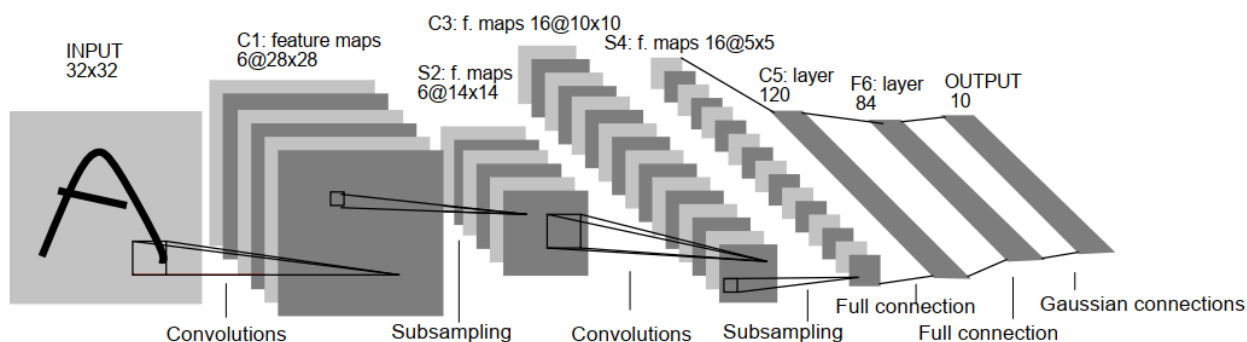
- 若采用全连接模式总参数数量：
 - 每个卷积核需要连接6个输入通道（6个输入特征图）
每个卷积核的参数数目是： $6 \times 5 \times 5 = 150$
 - 总参数量： $16 \times 150 = 2400$
- Summary - C3采用部分连接而不采用全连接
 - 计算复杂度降低（总参数量）
 - 模仿生物视觉的局部感受野（目的：不同的卷积核学习到不同的特征，增强多样性。）
LeNet-5当时设计的时候是受到生物视觉系统的启发。在视觉皮层中，不是所有神经元会连接到所有的输入信号，而是局部感受野的形式

- S4 - 池化层



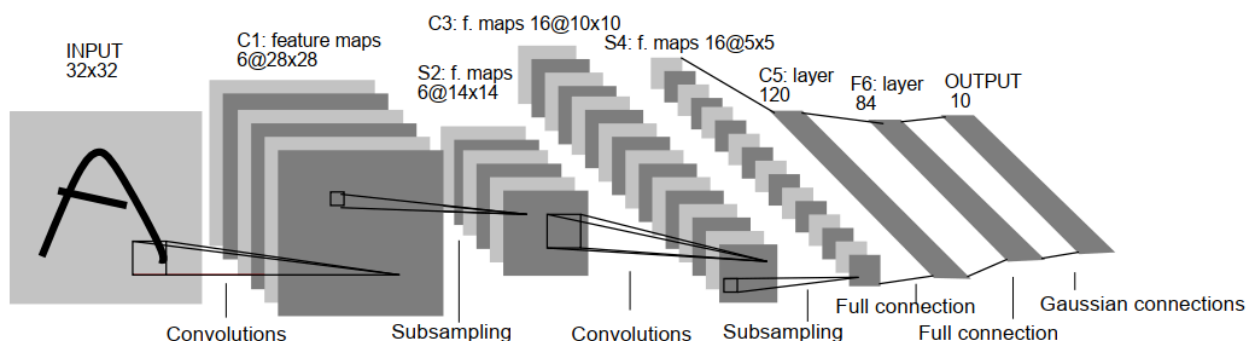
- 输入：10×10的特征图（16张）
- 采样区域：2×2
采样方式为4个输入相加，乘以一个可训练参数，再加上一个可训练偏置，并将结果通过sigmoid函数。（和S2相同）
- 输出：5×5的特征图（16张）
- 可训练参数：2×16
对于每张特征图来说，只有两个参数需要确定：用于相乘的“可训练参数 w ”与“可训练偏置 b ”，一共16张特征图，因此要训练 $16 \times 2 = 32$ 个参数。

- C5 - 卷积层（类似全连接层 - 卷积核大小和输入该层的特征图大小相等）



- 输入5×5的特征图（16张）
- 卷积核大小：5×5
- 卷积核种类120
- 输出120维向量（120个1x1的特征图）
每个卷积核与16张特征图做卷积，得到的结果求和，再加上一个偏置，结果通过sigmoid函数输出。
- 可训练参数： $(5 \times 5 \times 16 + 1) \times 120$

- F6 - 全连接层



- 输入：120维向量
- 算法：计算输入向量和权重向量之间的点积，再加上一个偏置，结果通过sigmoid函数输出。
 - 输入向量：120×1
 - 权重向量：84×120（在全连接层中，每个输出神经元都连接到所有输入神经元，因此权重矩阵的大小取决于输入神经元数量和输出神经元数量）
- 输出：84维向量（矩阵乘法规则 - 两个矩阵的行数和列数相等）
- 可训练参数：84 + 84×120
 - 权重参数：F6 层的输入是 120 维的向量，输出是 84 维的向量。这意味着 F6 层需要学习从 120 维空间到 84 维空间的映射。因此，对于每一维输出，都需要与输入的 120 维特征进行连接。换句话说，每输出一维就需要 120 个权重值。
总权重参数量 = $84 \times 120 = 10080$
 - 偏置参数：每个F6层的神经元都有一个独立的偏置参数，因此
总偏置参数 = 84
- F6层有84个节点，对应于一个7x12的比特图，-1表示白色，1表示黑色，这样每个符号的比特图的黑白色就对应于一个编码。

ASCII编码图如下：



- OUTPUT - 全连接层

- 输入：84维向量
- 输出：10维向量
OUTPUT层一共有10个节点，分别对应着数字0到9。
采用径向基函数(RBF)的连接方式，计算方式为：

$$y_i = \sum_{j=0}^{83} (x_j - w_{ij})^2$$

- y_i 计算的是输入特征向量 x 与数字 i 的比特图编码（模板）之间的距离：
- x_j ：F6层的第 j 个神经元输出值（输入特征向量）。
- w_{ij} ：值由数字 i 的比特图编码确定
- RBF输出的值（ y_i ）越接近于0，则越接近于 i ，即越接近于 i 的ASCII编码图，表示当前网络输入的识别结果是字符 i