



山东大学
SHANDONG UNIVERSITY

信息基础2 机器学习与深度学习

实验二

陈雷

山东大学信息科学与工程学院

实验内容

实验一：常规神经网络函数逼近实验

实验二：基于LeNet-5的MNIST字符识别

实验三：ResNet-18

实验四：Selective search

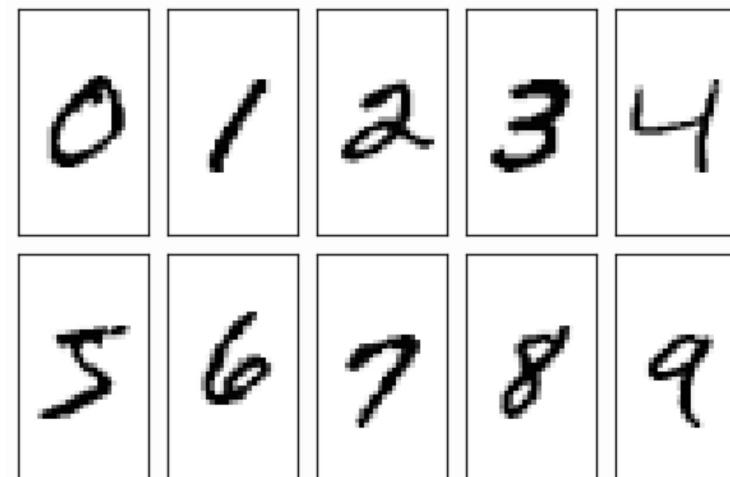
实验五：Yolo



实验二：基于LeNet-5的MNIST字符识别

LeNet-5是一种典型的非常高效的用来识别手写体数字的卷积神经网络。

LeNet-5出自论文**Gradient-Based Learning Applied to Document Recognition**，是由Yann LeCun提出的，对MNIST数据集的识别准确度可达**99.2%**。



经典卷积神经网络结构

经典卷积神经网络的结构一般满足如下表达式：

输入层 -> (卷积层+ -> 池化层?) + -> 全连接层+

上述公式中，“+”表示一个或者多个，“?”表示一个或者零个，“->”表示**forward** 方向。



经典卷积神经网络基本概念

局部感受野（Local Receptive Fields）

卷积神经网络把每一个隐藏节点只连接到图像的某个局部区域，从而减少参数训练的数量。

共享权值（Shared Weights）

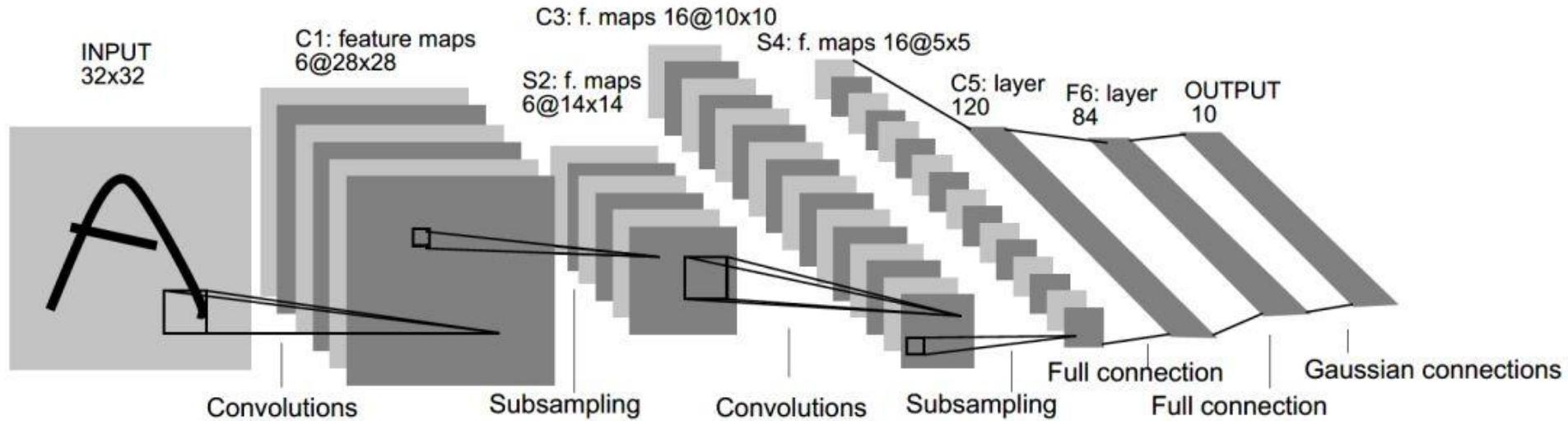
在卷积神经网络的卷积层中，神经元对应的权值是相同的，由于权值相同，因此可以减少训练的参数量。共享的权值和偏置也被称作卷积核或滤波器。

池化（Pooling）

对图像进行卷积之后，通过一个下采样过程，来调整图像的大小。



LeNet-5



LeNet-5共有7层，不包含输入，每层都包含可训练参数；每个层有多个Feature Map，每个Feature Map通过一种卷积滤波器提取输入的一种特征，然后每个Feature Map有多个神经元。



1. INPUT层-输入层

首先是数据 INPUT 层， 输入图像的尺寸统一归一化为**32*32**。

注意：本层不算LeNet-5的网络结构， 传统上， 不将输入层视为网络层次结构之一。

2. C1层-卷积层

对输入图像进行第一次卷积运算（使用 6 个大小为 $5 * 5$ 的卷积核）， 得到6个C1特征图（6个大小为 $28 * 28$ 的 feature maps, $32-5+1=28$ ）。

3. S2层-池化层（下采样层）

第一次卷积之后紧接着就是池化运算， 使用 $2 * 2$ 核进行池化， 于是得到了S2， 6个 $14 * 14$ 的特征图（ $28/2=14$ ）。

4. C3层-卷积层

第一次池化之后是第二次卷积， 第二次卷积的输出是C3， 16个 $10x10$ 的特征图， 卷积核大小是 $5 * 5$ 。



5. S4层-池化层（下采样层）

S4是pooling层，窗口大小仍然是 2×2 ，共计16个feature map，S4中每个特征图的大小是C3中特征图大小的 $1/4$ 。

6. C5层-卷积层

C5层是一个卷积层。由于S4层的16个图的大小为 5×5 ，与卷积核的大小相同，所以卷积后形成的图的大小为 1×1 。

7. F6层-全连接层

F6层是全连接层。F6层有84个节点，对应于一个 7×12 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）。

8. Output层-全连接层

Output层也是全连接层，共有10个节点，分别代表数字0到9，且如果节点i的值为0，则网络识别的结果是数字i。



通过TensorFlow实现LeNet5模型，可以增加代码可读性、提高编程效率，方便管理变量。

我们将LeNet5模型分为三部分：

- 1. 网络定义部分：**这部分是训练和验证都需要的网络结构。
- 2. 训练部分：**用于神经网络训练MNIST训练集。
- 3. 验证部分：**验证训练模型的准确率，在Tensorflow训练过程中，可以实时验证模型的正确率。

将训练部分与验证部分分开的好处在于，训练部分可以持续输出训练好的模型，验证部分可以每隔一段时间验证模型的准确率；如果模型不好，则需要及时调整网络结构的参数。



LeNet-5 的一些性质：

1. 如果输入层不算神经网络的层数，那么 LeNet-5 是一个 7 层的网络。LeNet-5 大约有 60,000 个参数。
2. 随着网络越来越深，图像的高度和宽度在缩小，与此同时，图像的 channel 数量一直在增加。
3. 现在常用的 LeNet-5 结构和 Yann LeCun 论文中提出的结构在某些地方有区别，比如激活函数的使用，以及输出层一般选择 softmax。



MNIST简介：

MNIST是一个非常有名的手写数字识别数据集，是NIST数据集的一个子集，包含了60000张图片作为训练数据，10000张图片作为测试数据。在MNIST数据集中的每一张图片都代表0~9中的一个数字。



实验要求：

1. 要求自己编程实现网络结构；
2. 采用层次化的映射模式；
3. 开发语言为**python**或者**C/C++**；
4. 要求提交报告和代码。

