

梯度学习在文档识别中的应用

(1988)—典型的神经网络leNet-5

分享人： 王嘉恒

CONTENTS

1-Motivation

2-Framework

3-Results

1.Motivation

主旨：通过更多地使用自动学习而非手动设计提取特征值的方法来实现识别分类的系统

在当时，大多数模型识别的系统是由特征提取模块和分类器模块组成的，其中，特征提取模块需要完全的手工制作，因此识别精度取决于设计者提取特征集的能力，并且遇到一个新的问题就需要重新制作。

因此，作者介绍了一种新的学习方法，GTN（图变换网络），用一种直接作用在图像像素上的算法来取代传统的手工特征提取。

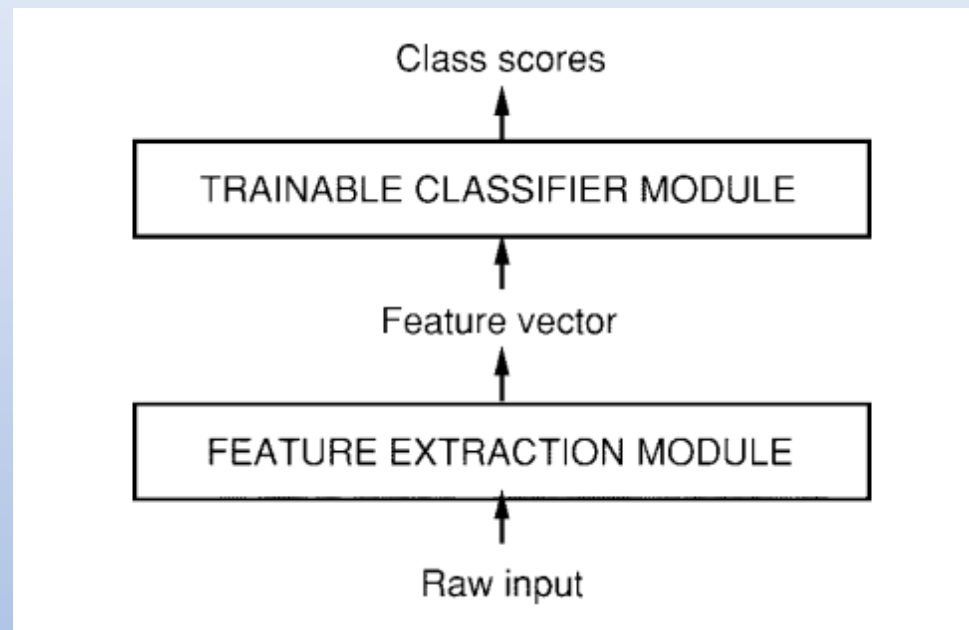


图1. 传统的模型识别系统

1.Motivation

作者在文章中还介绍了当时卷积神经网络的一些概念:

- 1、学习函数力求找到一组权重 W ，使得损失函数 $E(W)$ 达到最小
- 2、梯度下降学习： $W_K = W_{K-1} - \varepsilon \frac{\partial E(W)}{\partial W}$ 通过衡量参数权重的变化对损失函数的影响来进行优化
- 3、反向传播：通过输出到输入的传播来有效计算梯度
- 4、权重共享：使得要训练的参数数量大大减小
- 5、全局训练系统：要求保证全局损失函数可微，每个模块实现的函数必须对内部的权重参数处处可微。

2. Framework

文中作者设计了一个LeNet-5卷积神经网络，结构如下：

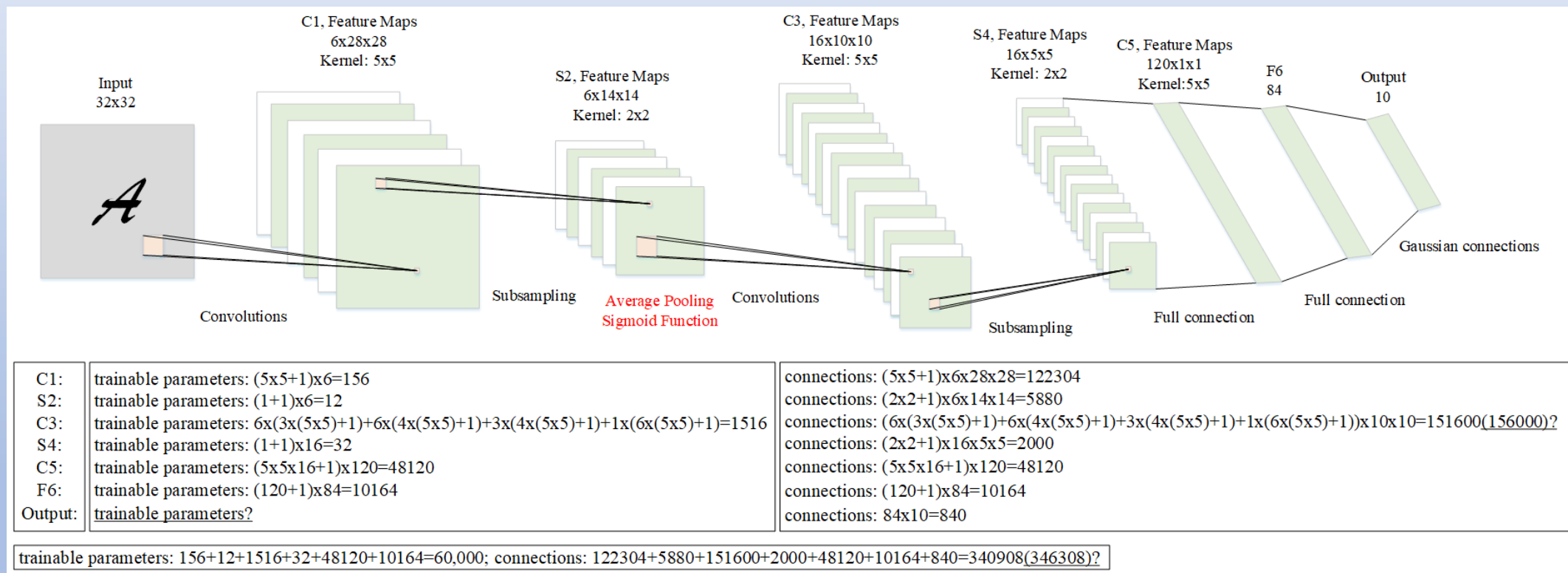


图2. Lenet-5神经网络的结构

2. Framework

整个结构分为7层：2个卷积层，2个池化层，3个全连接层。

输入：归一化成 32×32 像素的图像。背景(白色)对应于-0.1，而前景(黑色)对应于1.175。

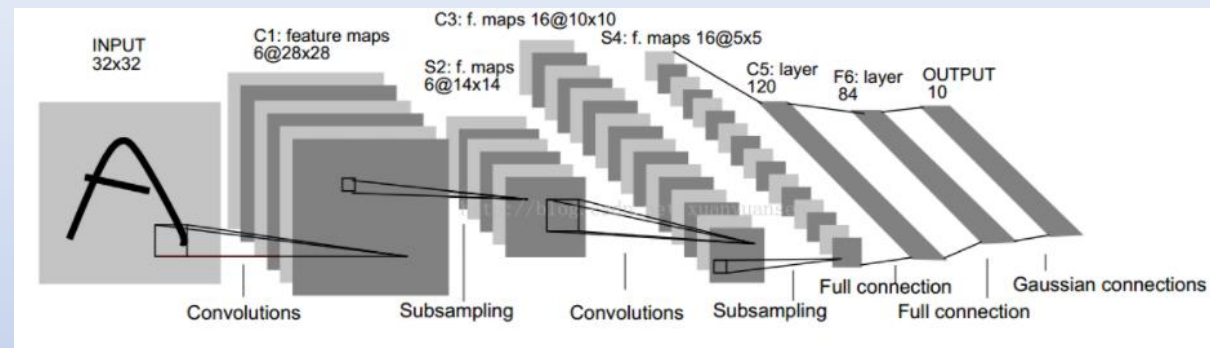


图3. Lenet-5神经网络的结构

C1层：6个特征图 大小 28×28 卷积核大小为 5×5

S2层：具有池化和激活的作用 6个特征图 大小为 14×14 感受野为 2×2 步幅为2 将4个输入相加然后乘可训练系数再加上偏置 再经过sigmoid函数激活

C3层：卷积层 16个特征图 大小 10×10 感受野为 5×5

2. Framework

C3层与S2层的连接方式与众不同，每个特征图连接到S2层的特征图不同。

这样做的目的：

- 1)破坏网络的对称性
- 2)希望提取到互补的特征

	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
1	X	X				X	X	X			X	X	X	X		X
2	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

图4. C3层与S2层的连接方式

2.Framework

S4层：16个特征图，大小为5*5，感受野为2*2，步幅为2，与S2层相同

F5层：包含120个大小1*1的特征图，感受野为5*5 C5到S4是全连接的

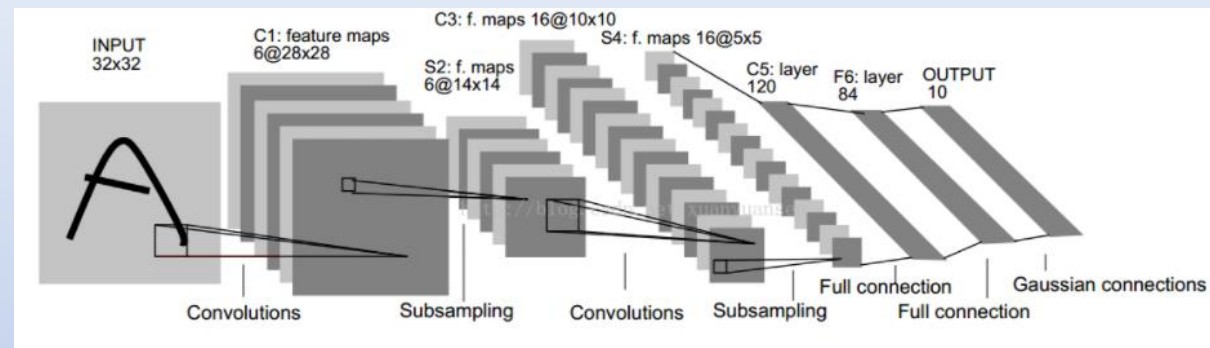


图3. Lenet-5神经网络的结构

F6层：84个大小为1*1的单元，结果通过 sigmoid 函数输出。这里84个单元对应7*12的比特图，-1代表白色，+1代表黑色，每个符号的比特图对应一个唯一的编码。

输出层：计算输入向量和参数向量的欧几里得距离， $Y_i = \sum (x_i - w_{ij})^2$ ，其中 w_{ij} 的值由比特图编码决定，即输入向量离 i 越接近，识别结果就判定为 i 。

3.Results

数据库：MNIST 500位作者写的近60000个数字图像 10000个测试样本。

进行20次迭代，学习率0.0005二次、0.0002三次、0.0001三次、0.00005四次、0.00001
剩余.

测试集的错误率在迭代10次后达到0.95%，训练集的错误率在迭代19次达到0.35%.

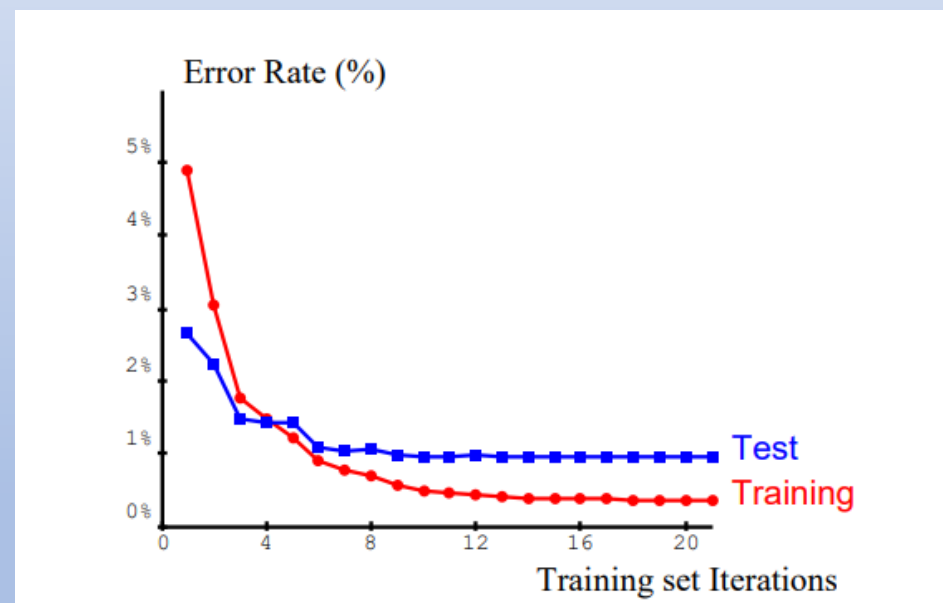


图5. 迭代次数与错误率关系图