采用卷积神经网络进行图像识别。首先将图像做0/1 normalization，将验证码设置为1，噪音和背景设置为0（这样的normalization是很有必要且很有效果的），得到30\*30的图片。

网络结构为：two convolutional layers, two fully connected layers.

两个conv layer分别拥有32和64个kernel，其中receptive field分别为5\*5和3\*3，使用ReLU作为activation。两个conv layer之后跟着一个max pooling layer，其中stride是2，filter size是3\*3，注意到这里我们使用了filter之间overlap的策略。

为了防止overfitting，fully connected layer的数目比较少（当然这和模型容量不需要太大有关系），两个fully connected layer均使用ReLU作为activation。Loss函数使用prediction和label之间（两个distribution之间）的cross entropy，即

其中p由输出层（36个neuron，分别表示0-9和a-z）做softmax得到。其中softmax函数为

并且使用SGD算法进行训练，每个batch 100张图。

一些改进：

1.经过实验后发现该模型仍然有一点overfitting，所以在两个fully connected layer之间加入了一个dropout随机丢弃一些neurons，dropout rate设为0.5。

2.在改进1的基础上，发现generalization err仍然有一些大，于是增强了数据以增强网络的泛化性能。