计算机科学与技术学院神经网络与深度学习课程实验报告

实验题目：Homework2		学号：201605130098
日期：2019.4.12	班级：16 智能	姓名： 李云开
Email：liyunkai@foxmail.com		

实验目的：

In this assignment you will practice putting together a simple image classification pipeline, based on the SVM/Softmax classifier. The goals of this assignment are as follows:

• Understand the basic Image Classification pipeline and the data-driven approach;

• Get a basic understanding of performance improvements from using higher-level representations than raw pixels (e.g. color histograms, Histogram of Gradient (HOG) features)

• Master basic neural network adjustment skills and try to improve deep neural networks

实验软件和硬件环境：

Python

实验原理和方法：

Get a basic understanding of performance improvements from using higher-level representations than raw pixels (e.g. color histograms, Histogram of Gradient (HOG) features)

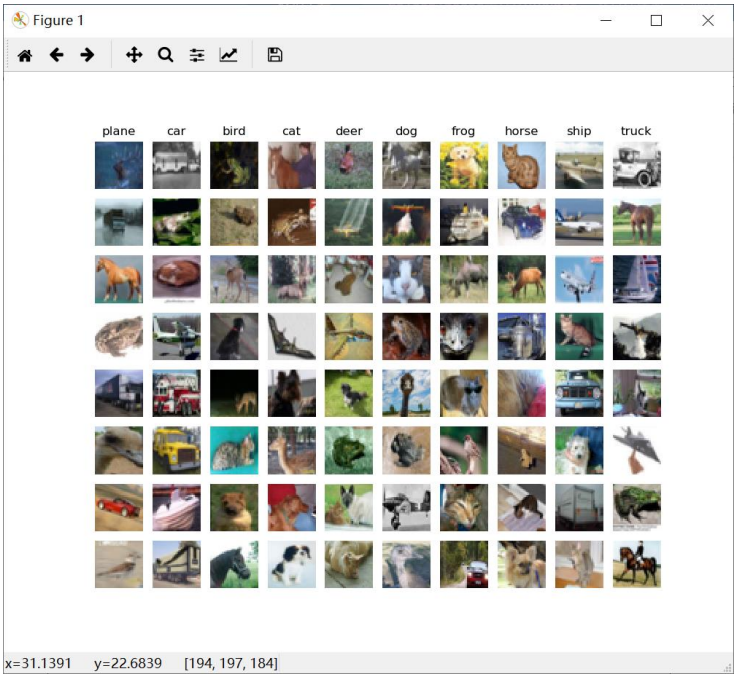
• Master basic neural network adjustment skills and try to improve deep neural networks

实验步骤：（不要求罗列完整源代码）

1 、修改 features 代码

如下，和 hw1 差不多，

程序自带的预测错误的可视化



Beastnet: Batchsize 256 learningrate0.001 learningdecay0.99 reg0.001

2、Improving Deep Neural Networks

<1> Gradient Checking:
梯度检验， 用我们bp 得到的grad 与梯度的定义得到的比较。

<2> Initialization
Different initializations lead to different results
- Random initialization is used to break symmetry and make sure different hidden units can learn different things
- Don't intialize to values that are too large
- He initialization works well for networks with ReLU activations.
<3> Optimization method

$$\begin{cases} v_{dW^{[l]}} = \beta_1 v_{dW^{[l]}} + (1-\beta_1)\frac{\partial \mathcal{J}}{\partial W^{[l]}} \\ v_{dW^{[l]}}^{corrected} = \frac{v_{dW^{[l]}}}{1-(\beta_1)^t} \\ s_{dW^{[l]}} = \beta_2 s_{dW^{[l]}} + (1-\beta_2)(\frac{\partial \mathcal{J}}{\partial W^{[l]}})^2 \\ s_{dW^{[l]}}^{corrected} = \frac{s_{dW^{[l]}}}{1-(\beta_1)^t} \\ W^{[l]} = W^{[l]} - \alpha \frac{v_{dW^{[l]}}^{corrected}}{\sqrt{s_{dW^{[l]}}^{corrected}}+\varepsilon} \end{cases}$$

贴一张 Adam 的

看起来好麻烦，但是写起来按照老师给的程序很简短
<4>regularization
L2loss and dropout.

结论分析与体会：

好的初始化与正则能得到更好的结果，Adam 能很快收敛，但最后的效果不如 SGD 效果好。

就实验过程中遇到和出现的问题，你是如何解决和处理的，自拟 1－3 道问答题：

简述 VGG?
VGGNet 全部使用 3*3 的卷积核和 2*2 的池化核，通过不断加深网络结构来提升性能。网络层数的增长并不会带来参数量上的爆炸，因为参数量主要集中在最后三个全连接层中。同时，两个 3*3 卷积层的串联相当于 1 个 5*5 的卷积层，3 个 3*3 的卷积层串联相当于 1 个 7*7 的卷积层，即 3 个 3*3 卷积层的感受野大小相当于 1 个 7*7 的卷积层。但是 3 个 3*3 的卷积层参数量只有 7*7 的一半左右，同时前者可以有 3 个非线性操作，而后者只有 1 个非线性操作，这样使得前者对于特征的学习能力更强。