



# 图片分类以及句子相似度评估

期中project

姓名： 何家栋、刘斯宇

专业： 计算机科学与技术





## 目录

---

1/ 实现思路

2/ 网络结构

3/ 结果分析

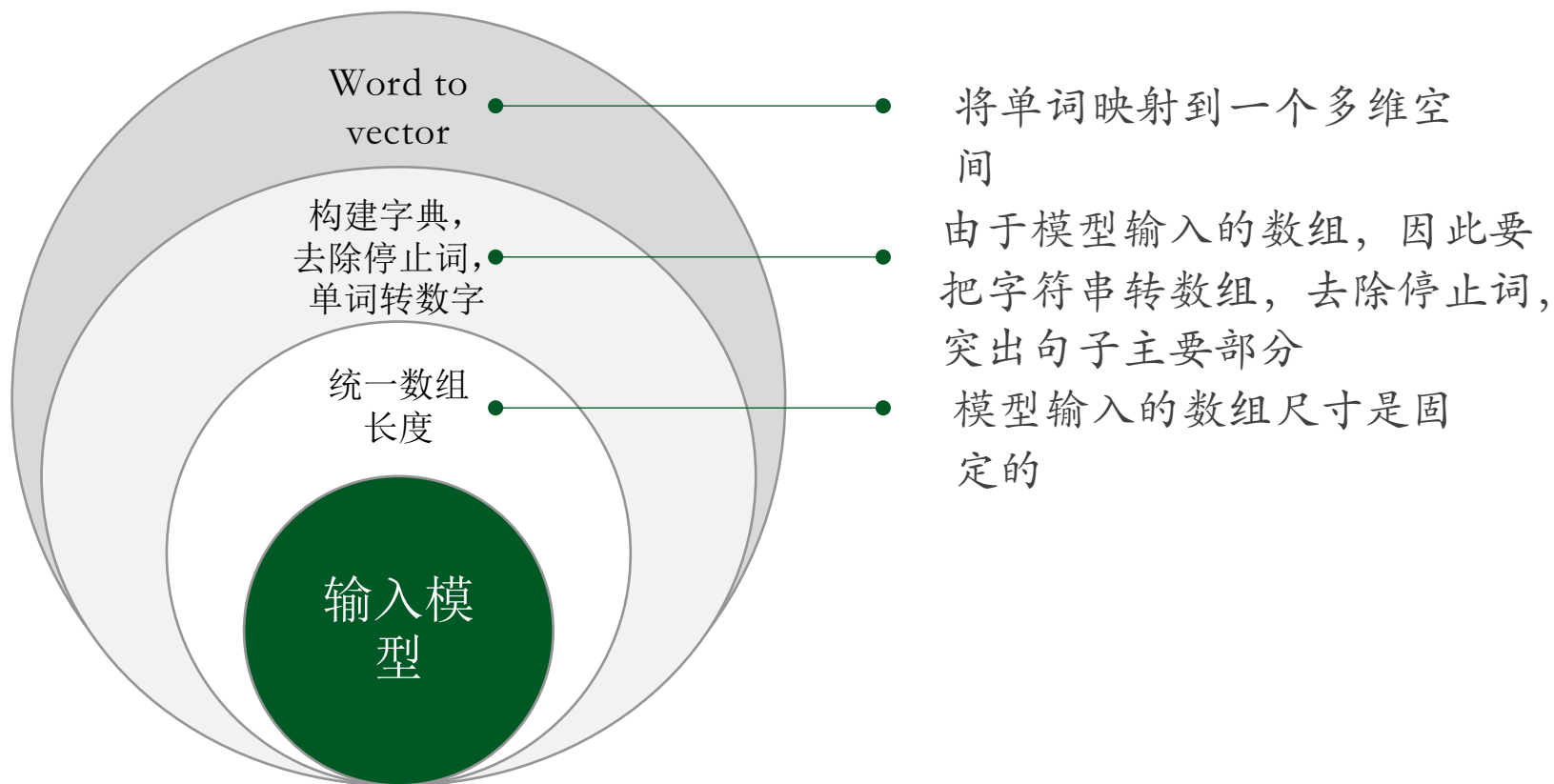
4/ 创新

---



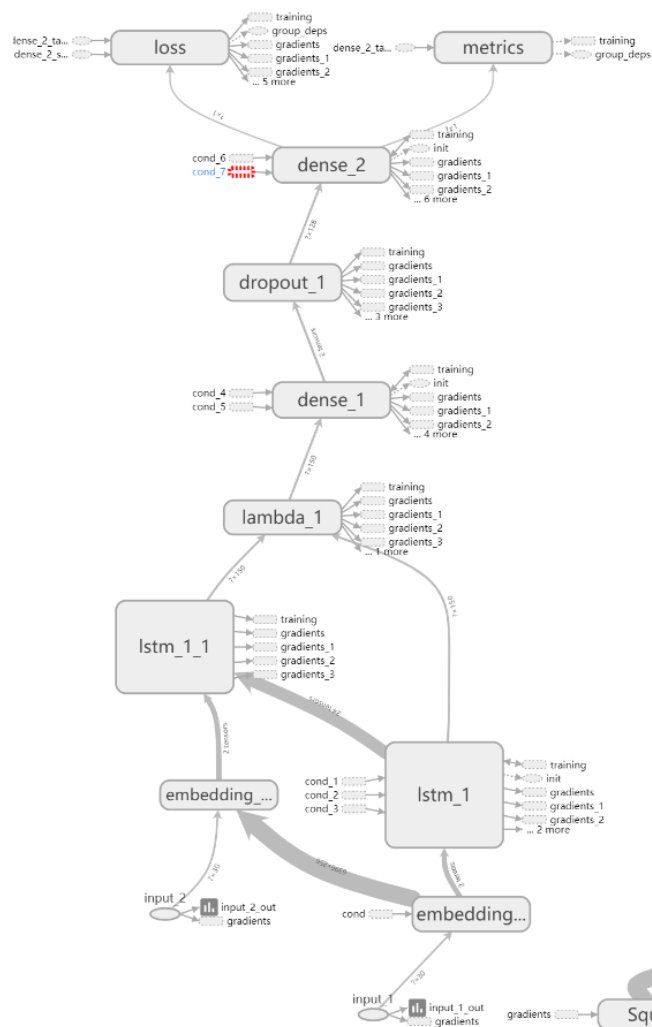
## RNN 实现思路

这里主要阐述一下整个文本的处理到模型的建立过程。





## RNN网络结构



1、这里的embedding层参数固定，用于把每一个单词转换为词向量

2、lstm层采用了参数共享的方式（lstm的原理叙述）

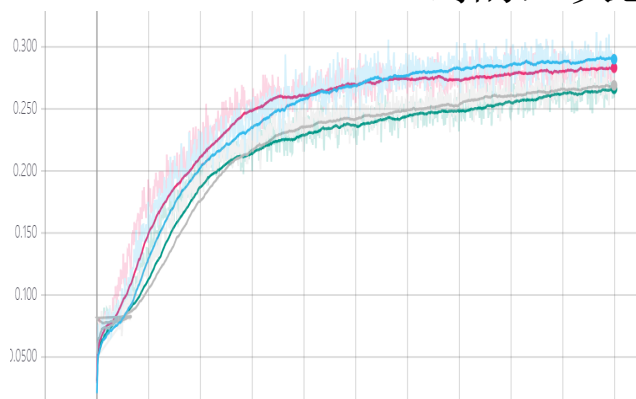
3、lambda层，用于融合两个输出，采用的融合函数是 $\exp(-|V1-V2|)$

4、dropout层，用于防止过拟合，节点的采用概率为0.5

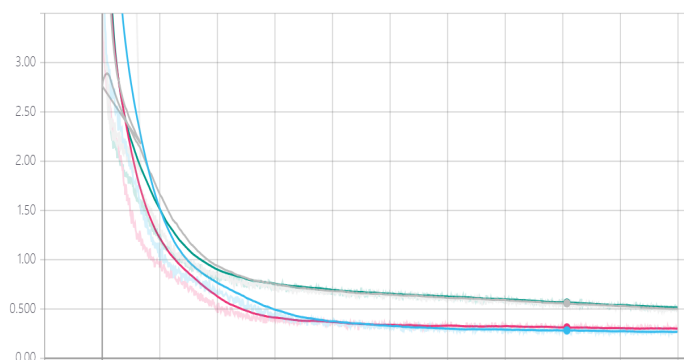


## 结果分析

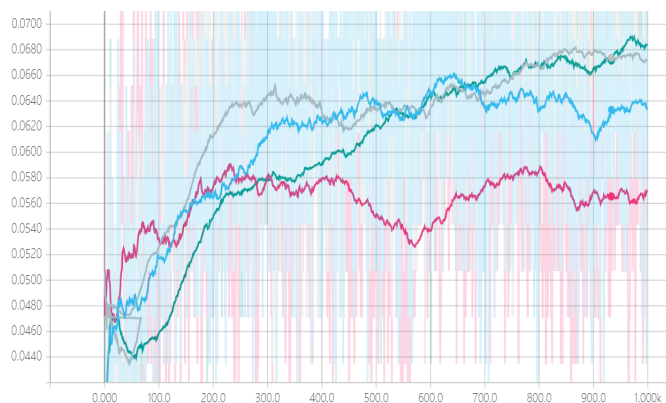
（红色的是原模型，绿色的是动态学习率，蓝色的是有停止词的，灰色的是有停止词且学习率动态的）



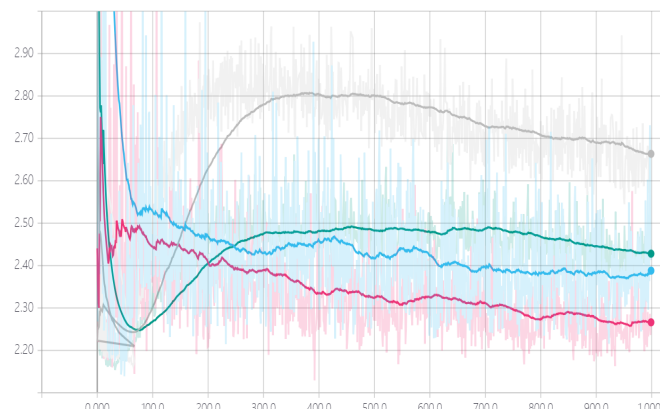
训练集准确率



训练集误差损失



验证集准确率



验证集误差损失



## 创新探索

主要是探索了一些文本预处理对模型结果的影响，以及动态学习率对模型结果的影响，这个只代表我自己的探索结果，可能会与别人相比有所出入。

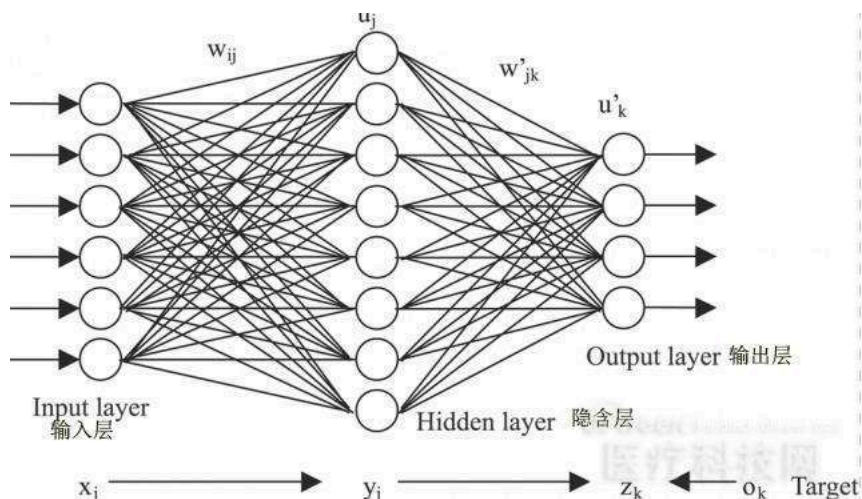
- (1) 在原模型的基础上增加了动态学习率，当loss5次不再下降后，学习率变为原来的一半
- (2) 在原模型的基础上，文本预处理时，添加其中的停止词，觉得可以使句子的句子结构更加突出，在相似度评估上应该会有更好的效果
- (3) 在增加停止词的同时，增加了动态学习率

探索的结果以及对比，都在上一部分结果分析中给出。

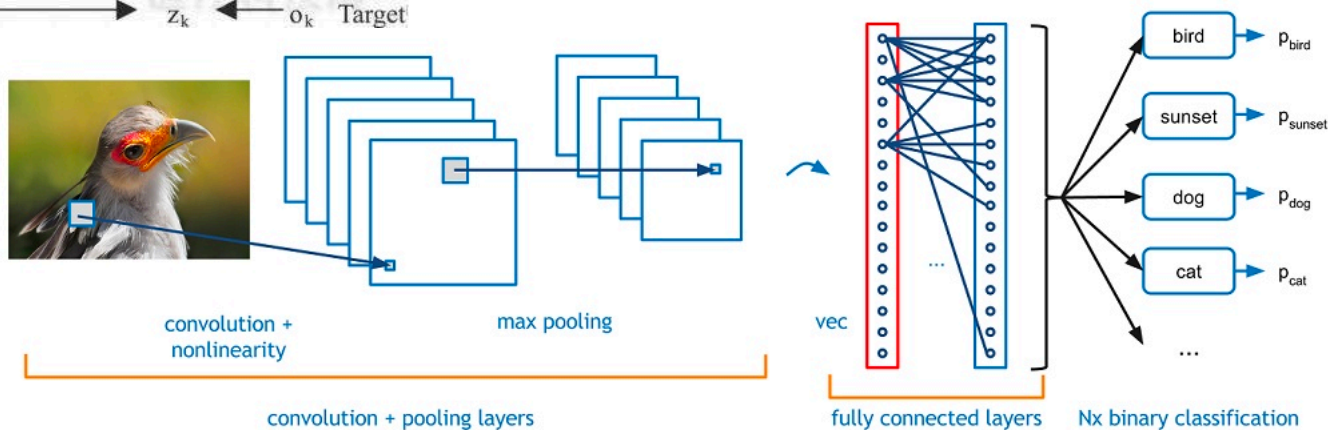


# CNN—卷积神经网络

CNN VS 全连接网络?



参数数量太多  
没用利用像素之间的位置关系  
网络层数限制

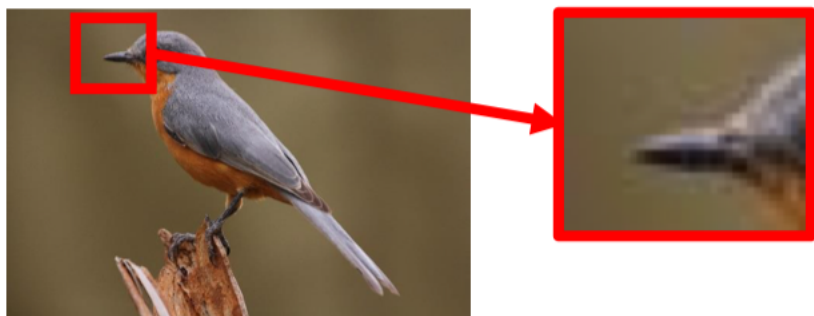




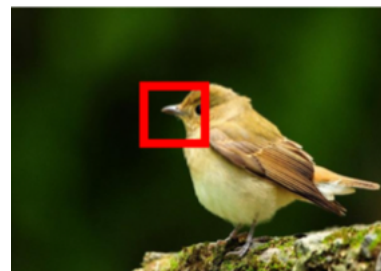


## CNN 的优点

因为很多特征要比整张图片小很多，所以一个神经元不需要看整张图片来发现这个特征。



相同的特征在不同的图片中会出现在不同的区域。



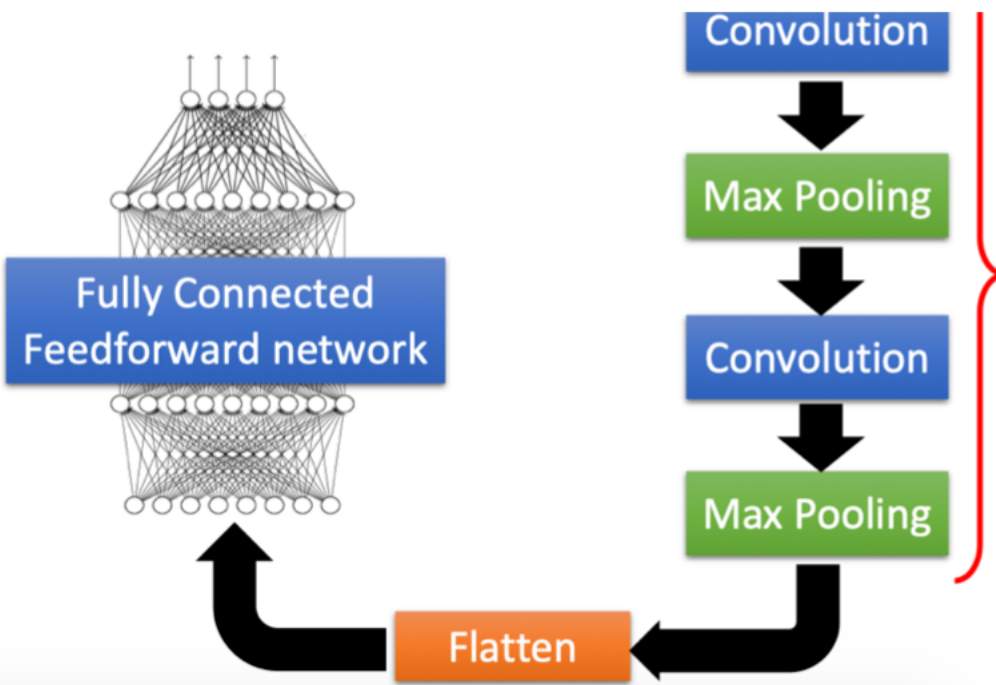
对图片进行采样不会改变图片中的物体。







## CNN的框架



Can repeat  
many times

卷积层 convolution: 每个 filter 都是需要学习出来的, 类似于 neutral, 但是比实际的 image 所需参数小

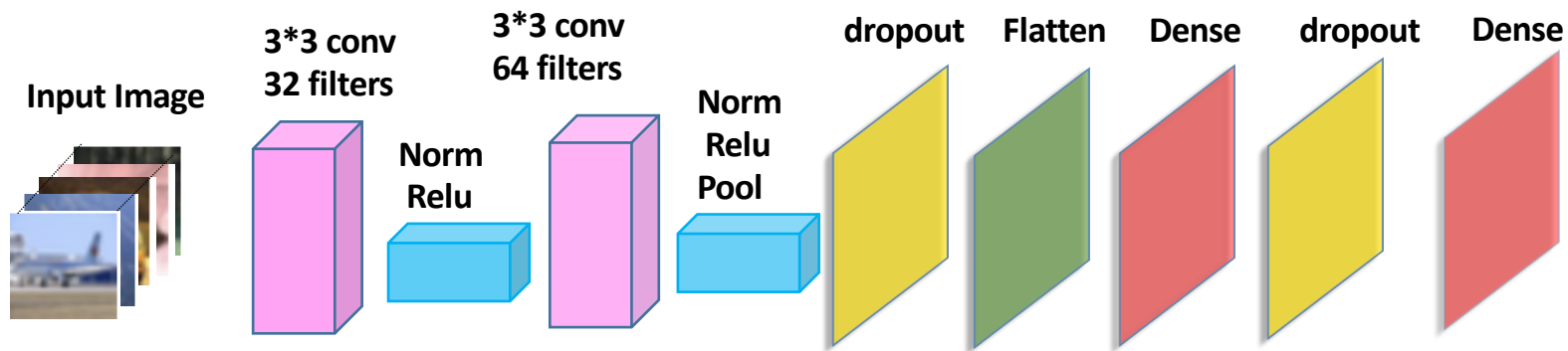
池化: 对不同位置的特征进行聚合统计, 以降低维度, 不容易过拟合。

Flatten: 来将输入“压平”, 即把多维的输入一维化

其实, CNN 的本质就是去除了一些权重的全连接网络!



## 我的简单的CNN模型





## 我的简单的CNN模型实验效果

### 训练集

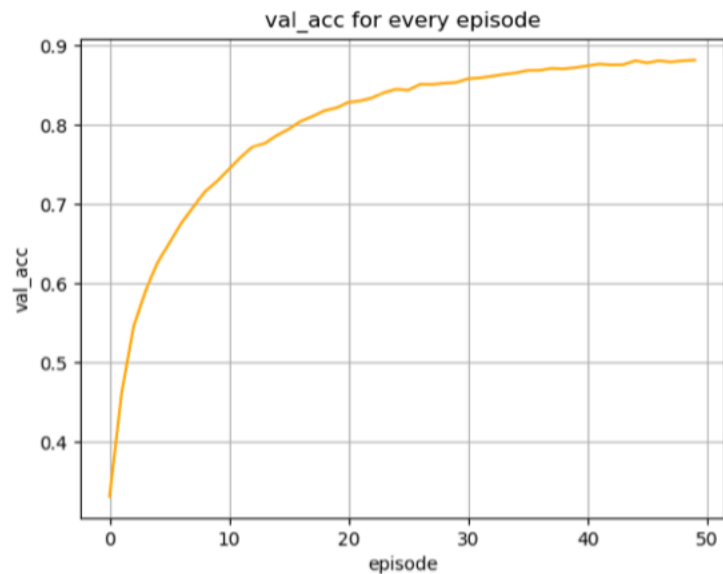


图 1: 含有 BatchNormalization

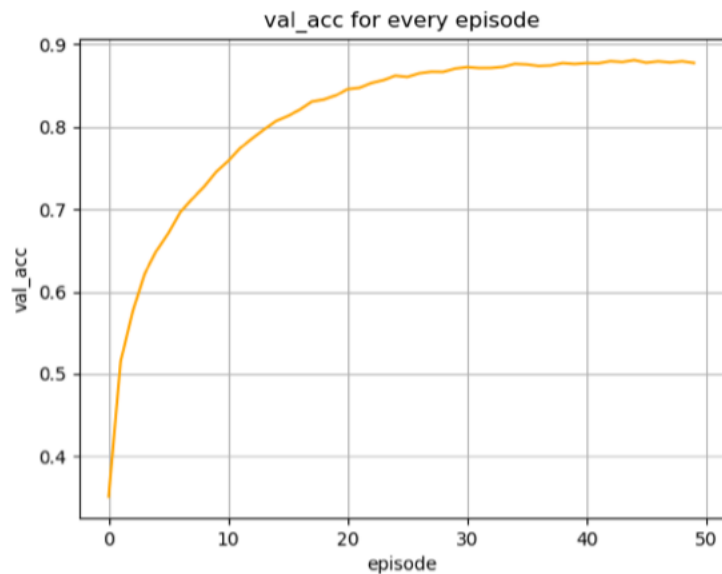


图 2: 不含有 BatchNormalization

## 我的简单的CNN模型实验效果

### 测试集



图 5: 含有 BatchNormalization

从实验结果我们可以发现其实 **BatchNormalization** 也并非是一定好，也是需要在不同的情况下特殊考虑，当然一般情况下是加上 **BatchNormalization** 的情况会更好一些的，

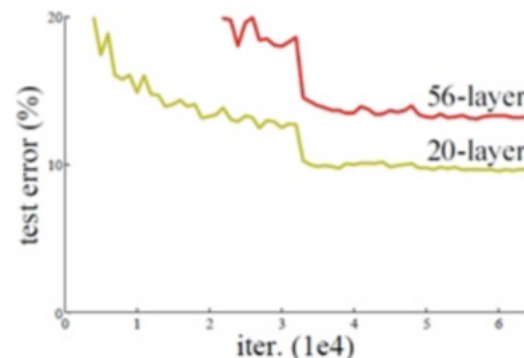
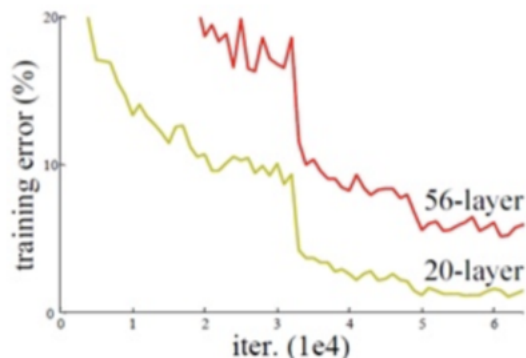


图 6: 不含有 BatchNormalization

比较有意思的是实验结果表明不加 **BatchNormalization** 的在测试集的准确率效果会更好一些。还有就是两幅图片都有比较大的波动，这说明过拟合了。

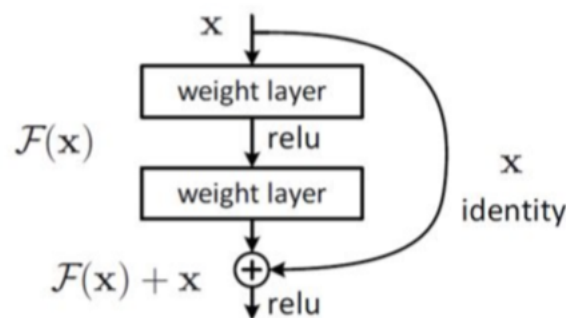
## Why Resnet?

从经验来看，网络的深度对模型的性能至关重要，当增加网络层数后，网络可以进行更加复杂的特征模式的提取，所以当模型更深时理论上可以取得更好的结果，但是更深的网络其性能一定会更好吗？实验发现深度网络出现了退化问题（**Degradation problem**）：网络深度增加时，网络准确度出现饱和，甚至出现下降。这个现象可以在图中直观看出来：56层的网络比20层网络效果还要差。这不会是过拟合问题，因为56层网络的训练误差同样高。



## Why Resnet?

由于梯度弥散等的原因，导致深度网络的实际效果要比预期中差很多，但是深度网络绝对有其自身的优点，首先我们必须承认特征的“等级”会随着网络深度的加深而变高，同时更深的网络具有更强大的表达能力，但是训练深度网络有很多的问题，比如上面所说的梯度弥散，还有一个重要的问题是退化问题。这主要是因为深度更深的卷积层不能保持其多余的卷积层的参数就是1，这是很难做到。所以何凯民博士就发明了 Resnet 深度网络。



我们要是知道只需要几层网络效果比较好的话，那么是很容易解决这个问题的，但是问题在于我们并不知道哪些网络是不需要的，所以Resnet网络正好帮我解决了这个问题！--解决了退化问题！



为什么残差学习相对更容易，从直观上看残差学习需要学习的内容少，因为残差一般会比较小，学习难度小点。不过我们可以从数学的角度来分析这个问题，首先残差单元可以表示为：

$$\begin{aligned}y_l &= h(x_l) + F(x_l, W_l) \\x_{l+1} &= f(y_l)\end{aligned}$$

其中  $x_l$  和  $x_{l+1}$  分别表示的是第  $l$  个残差单元的输入和输出，注意每个残差单元一般包含多层结构。 $F$  是残差函数，表示学习到的残差，而  $h(x_l) = x_l$  表示恒等映射， $f$  是 ReLU 激活函数。基于上式，我们求得从浅层  $l$  到深层  $L$  的学习特征为：

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i)$$

解决了梯度爆炸问题！

利用链式规则，可以求得反向过程的梯度：

$$\frac{\partial \text{loss}}{\partial x_l} = \frac{\partial \text{loss}}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_l} = \frac{\partial \text{loss}}{\partial x_L} \cdot \left(1 + \frac{\partial}{\partial x_L} \sum_{i=l}^{L-1} F(x_i, W_i)\right)$$

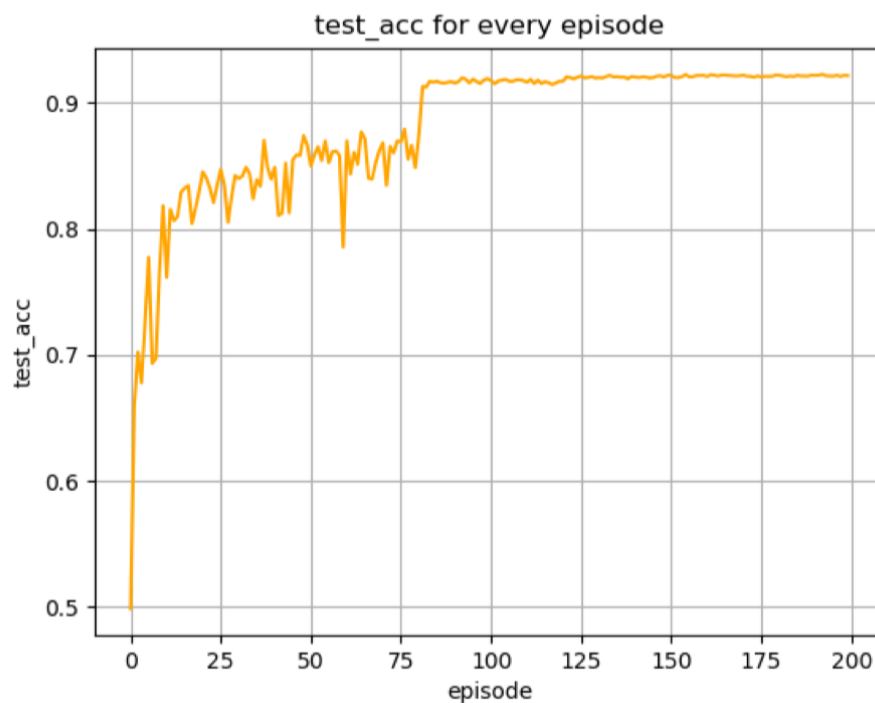
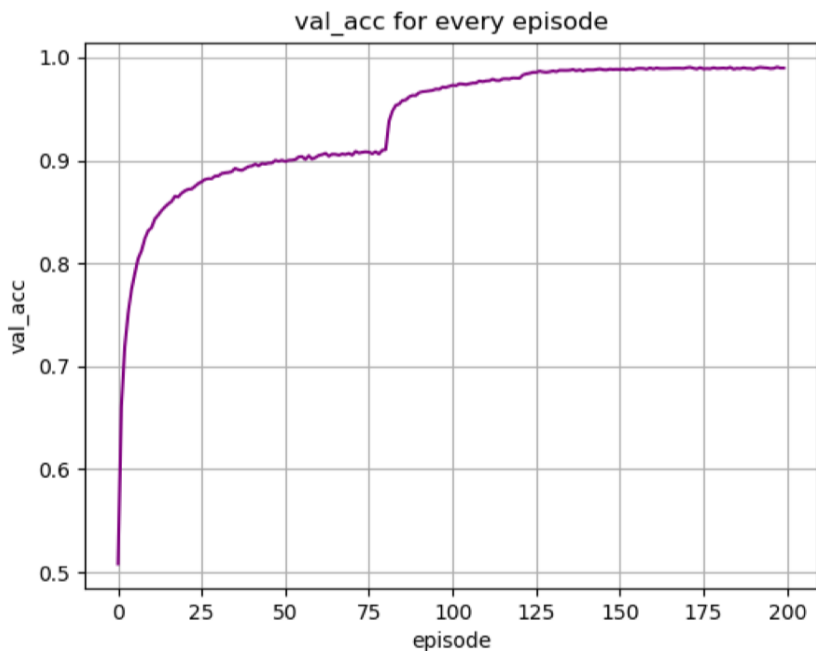
式子的第一个因子  $\frac{\partial \text{loss}}{\partial x_L}$  表示的损失函数到达  $L$  的梯度，小括号中的 1 表明短路机制可以无损地传播梯度，而另外一项残差梯度则需要经过带有 **weights** 的层，梯度不是直接传递过来的。残差梯度不会那么巧全为-1，而且就算其比较小，有 1 的存在也不会导致梯度消失。







## 实验结果



resnet 确实又很强的表达能力，它的分类准确率能够达到 92%，它所具有区别不同图片的能力更强，效果更好。



## 实验创新

### 1.探索模型中的每个部件的作用

在这次试验中，我分别探索了 **BatchNormalization** 和 **episode** 对网络的影响，发现这两个 部件都不是有一个确定的作用，也就是说并不是说加了就会一定的好，比如我在 **episode=8** 的时候我发现不加 **BN** 的实验效果会更好。还有就是 **episode** 也不是越多越好，这些东西都是需 要实际情况具体对待的。

### 2.多个模型一起分析

在这次试验中，我运用了 **resnet18** 来解决 **cifar-10** 发现效果会很好。当然这个模型为什么 会比较适合来解决图片分类的问题是值得我们思考的，因为深层的网络具有更好的表达能力，其自然要比一般简单的网络要好，当然其实 **CNN** 真的对于图片的分类问题具有较好的解决能 力，从一个简单的网络就能达到比较好的准确率就能看出。



THANKS

