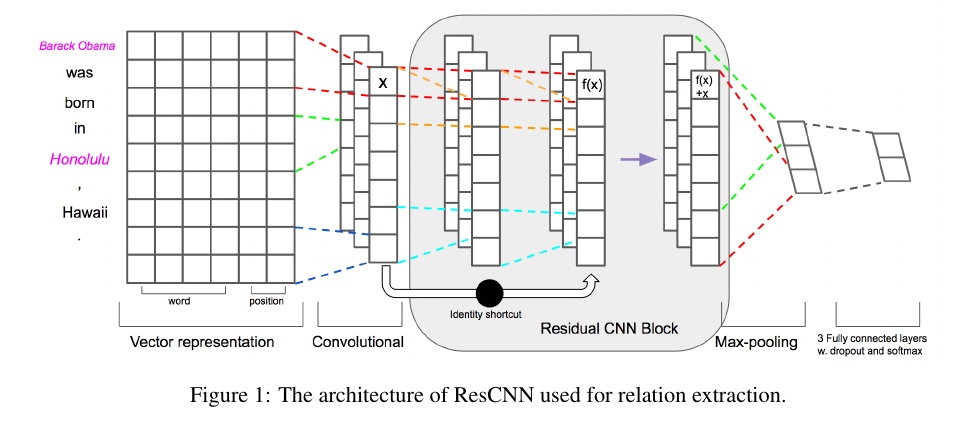
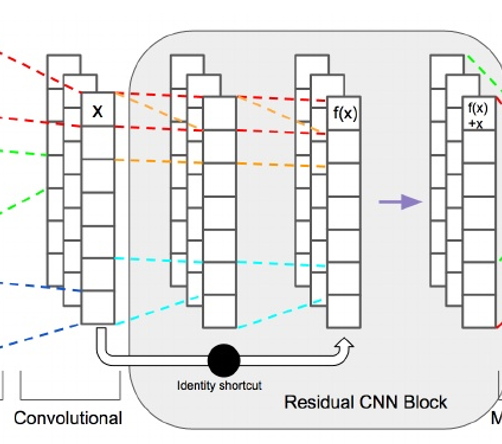
**[论文笔记]基于ResNet CNN的关系抽出**

## 1. 模型介绍

这个模型的思想主要是利用 Residual CNN 去解决 基于远程监督的关系抽出问题. 下面是模型图:

这个模型前面的输入和2015年的论文一致, 唯一不同是在后面的多层卷积上.

* **第一层的卷积层**  
  这里的公式是:

https://pic2.zhimg.com/80/v2-92baa6ef09a103c0eb51fdaed88311f5_hd.jpg这是第一层的卷积层.  
 : 代表第一层中, 根据某个卷积核计算出的 **句子的卷积特征向量**, 中的 **第 i 个元素**.  
 : 一整条句子的卷积特征向量,



https://pic3.zhimg.com/80/v2-298da530a2f5135595ff2661fdaeedfa_hd.jpghttps://pic3.zhimg.com/80/v2-10cdbf13cba5d168bfe9768aecde3c4e_hd.jpg后面的CNN Block  
这里就用到了残差网络的特点.  
CNN Block 中的第一层:



CNN Block 中的第二层:



第三层, 组成残差结构, 这里时 CNN Block 的输出:



注意, 这里每一个卷积层的映射中, 使用的卷积核的size是相同的.



* 多个Block  
  这个图中虽然只画出了一个, 但实际上是有很多的, 其中作者的实验是, 在有九个时有最好效果.



* 最后的全连接层  
  最后使用maxpool来构成了全连接层的输入.  
  在参数说明中, 看到作者使用模型中使用的卷积核共有 128 个, 意味着这里的全链接层的输入向量维度是128.



## 2. Intuition

文中给出了两个解释:

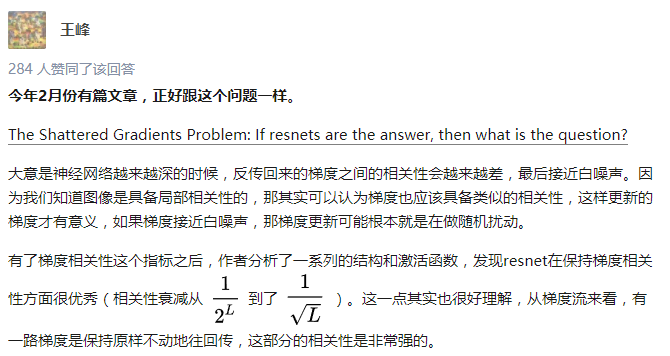
* 如果低级,中级和高级分别学习隐藏的词法,句法和语义的表示,有时它助于绕过语法直接连接词汇和语义空间.



* ResNet本身就有解决梯度消失的问题, 在这里降低了noise的影响.  
  确实是, 从结果来看, 这个模型可以和 基于attention 的模型打平甚至更优秀, 那么肯定也拥有了解决远程监督中的噪声的机制和能力, 但是为什么ResNet 就可以解决这个问题呢? 下面是自己的思考.

## 3. Why it works?

## 





## 这个问题本质上是深度网络的退化问题

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，从图2中也可以看出网络越深而效果越好的一个实践证据。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3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。我们知道深层网络存在着梯度消失或者爆炸的问题，这使得深度学习模型很难训练。但是现在已经存在一些技术手段如BatchNorm来缓解这个问题。因此，出现深度网络的退化问题是非常令人诧异的。



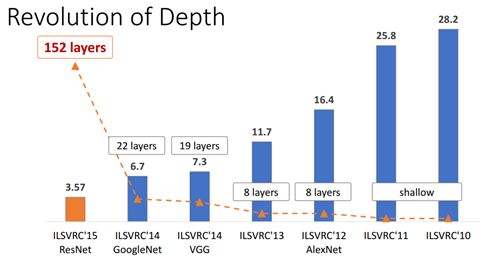
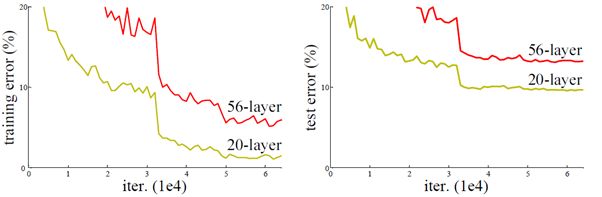
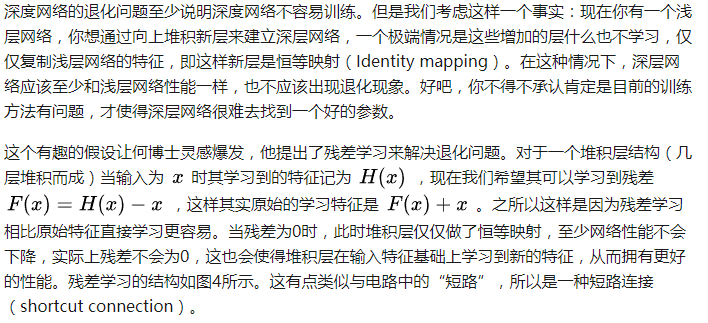
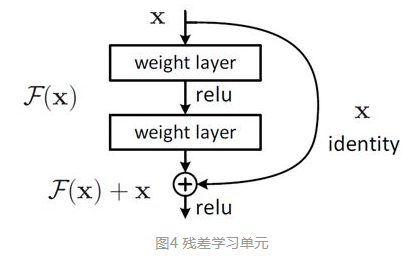


图2 ImageNet分类Top-5误差

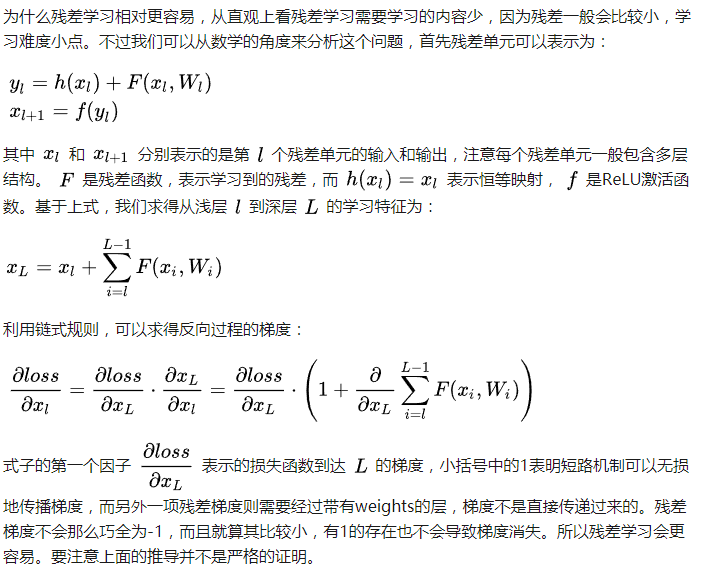


图中 20层与56层网络在CIFAR-10上的误差

## 残差学习

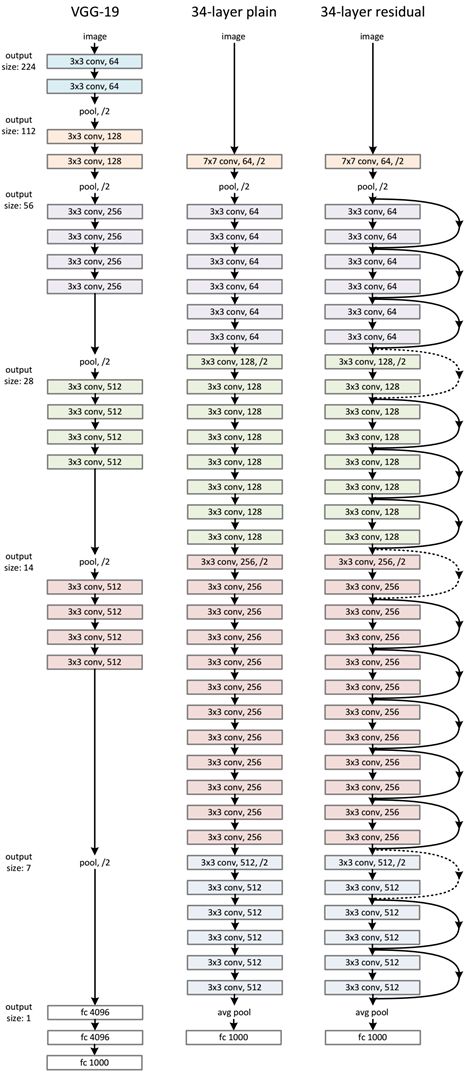








## ResNet的网络结构

ResNet网络是参考了VGG19网络，在其基础上进行了修改，并通过短路机制加入了残差单元，如图5所示。变化主要体现在ResNet直接使用stride=2的卷积做下采样，并且用global average pool层替换了全连接层。ResNet的一个重要设计原则是：当feature map大小降低一半时，feature map的数量增加一倍，这保持了网络层的复杂度。从图5中可以看到，ResNet相比普通网络每两层间增加了短路机制，这就形成了残差学习，其中虚线表示feature map数量发生了改变。图5展示的34-layer的ResNet，还可以构建更深的网络如表1所示。从表中可以看到，对于18-layer和34-layer的ResNet，其进行的两层间的残差学习，当网络更深时，其进行的是三层间的残差学习，三层卷积核分别是1x1，3x3和1x1，一个值得注意的是隐含层的feature map数量是比较小的，并且是输出feature map数量的1/4。



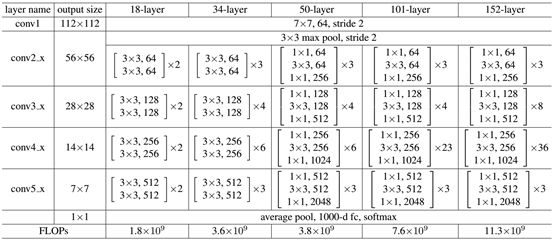


表1 不同深度的ResNet

下面我们再分析一下残差单元，ResNet使用两种残差单元，如图6所示。左图对应的是浅层网络，而右图对应的是深层网络。对于短路连接，当输入和输出维度一致时，可以直接将输入加到输出上。但是当维度不一致时（对应的是维度增加一倍），这就不能直接相加。有两种策略：（1）采用zero-padding增加维度，此时一般要先做一个downsamp，可以采用strde=2的pooling，这样不会增加参数；（2）采用新的映射（projection shortcut），一般采用1x1的卷积，这样会增加参数，也会增加计算量。短路连接除了直接使用恒等映射，当然都可以采用projection shortcut。



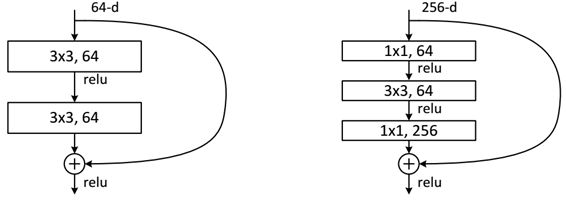
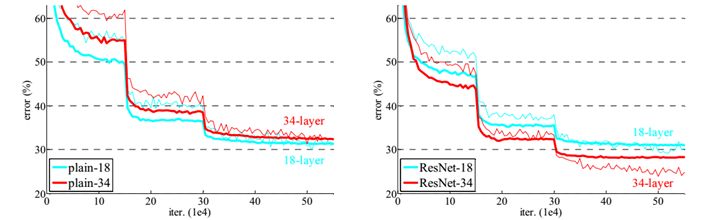


图6 不同的残差单元

作者对比18-layer和34-layer的网络效果，如图7所示。可以看到普通的网络出现退化现象，但是ResNet很好的解决了退化问题。

最后展示一下ResNet网络与其他网络在ImageNet上的对比结果，如表2所示。可以看到ResNet-152其误差降到了4.49%，当采用集成模型后，误差可以降到3.57%。

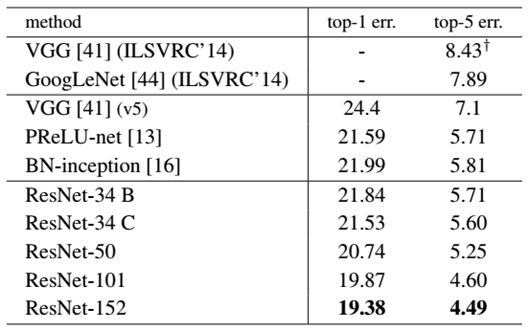


表2 ResNet与其他网络的对比结果

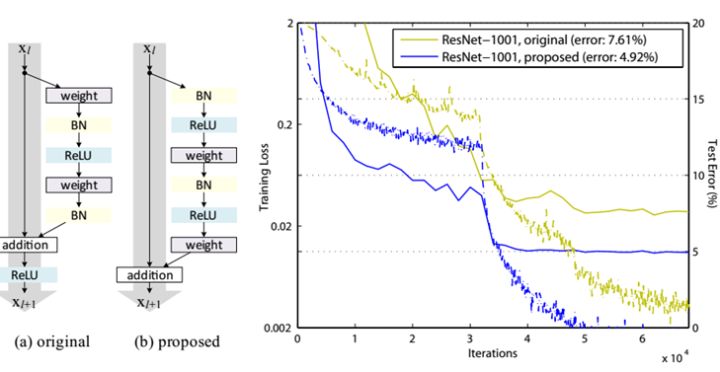
说一点关于残差单元题外话，上面我们说到了短路连接的几种处理方式，其实作者在[文献[2]](http://link.zhihu.com/?target=https%3A//arxiv.org/abs/1603.05027)中又对不同的残差单元做了细致的分析与实验，这里我们直接抛出最优的残差结构，如图8所示。改进前后一个明显的变化是采用pre-activation，BN和ReLU都提前了。而且作者推荐短路连接采用恒等变换，这样保证短路连接不会有阻碍。感兴趣的可以去读读这篇文章。



图8 改进后的残差单元及效果