# ResNet

# Deep Residual Learning for Image Recognition

在【4】VGG结论中表示网络深度是十分重要有效的

但更深的神经网络更加难以训练

本文提出残差学习框架

退化问题：随着网络深度的增加，准确率饱和，然后迅速降低

原因：不是由于过拟合，而是逐渐产生更高的训练误差

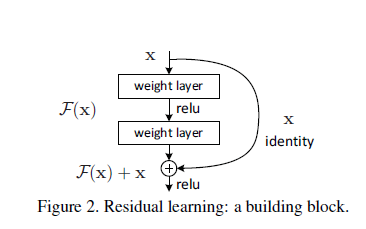
传统解决方法：数据初始化和正则化

引入深度残差学习框架：

残差：实际观察值与估计值（拟合值）之间的差

P（X）=H（X）-X  
让非线性堆叠层学习残差函数“F(x)=H(x)-x”，而不是直接学习H(x)，因此原来的映射转化为”F(x)+x“。

”F(x)+x“通过shortcut连接，shortcut 连接只是简单的执行恒等映射，再将它们的输出和堆叠层的输出叠加在一起(Fig.2)。恒等的shortcut连接并不增加额外的参数和计算复杂度。



We hypothesize that it is easier to optimize the residual mapping than to optimize the original, unreferenced mapping. To the extreme, if an identity mapping were optimal, it would be easier to push the residual to zero than to fit an identity mapping by a stack of nonlinear layers。

参考<https://www.cnblogs.com/aoru45/p/10286918.html>

***比如把5映射到5.1，那么引入残差前是F'(5)=5.1，引入残差后是H(5)=5.1, H(5)=F(5)+5, F(5)=0.1。这里的F'和F都表示网络参数映射，引入残差后的映射对输出的变化更敏感。比如s输出从5.1变到5.2，映射F'的输出增加了1/51=2%，而对于残差结构输出从5.1到5.2，映射F是从0.1到0.2，增加了100%。明显后者输出变化对权重的调整作用更大，所以效果更好。残差的思想都是去掉相同的主体部分，从而突出微小的变化。***

1. 如果同一纬度那就直接相加

IMG_256

1. 如果不同纬度我们将x卷积变换一下，将x变换为和输出结果一样的维度就可以了

IMG_256

基准网络为：基于VGGNet，采用的卷积核为3x3，其中有两个设计原则，1）对于有相同的输出feature map尺寸，filter的个数相同；2）当feature map尺寸减半时，filter的数量加倍。下采样的策略是直接用stride=2的卷积核。网络最后末尾是一个global average pooling layer（不需要参数，参考http://www.cnblogs.com/hejunlin1992/articles/7750759.html）和一个1000的全连接层（后面接softmax）。

　　残差网络为：在基准网络的基础上，插入了shortcut connections。当输入输出具有相同尺寸时，identity shortcuts可以直接使用（实线部分），就是公式1；当维度增加时（虚线部分），有以下两种选择：A)仍然采用恒等映射（identity mapping），超出部分的维度使用0填充；B) 利用1x1卷积核来匹配维度，就是公式2。对于上面两种方案，当shortcuts通过两种大小的feature map时，采取A或B方案的同时，stride=2。

