残差学习的结构下图所示，X表示这小段网络最顶层的输入，其学习到的特征记为 H ( x )，明确的让网络去拟合一个残差函数F(x):=H(x) -x.原始的方程则变成了F(x) +x.

*恒等即H(x)=x,这里所说的恒等层即图像进入网络前跟从网络出来后是一样的，也就是所增加的层中，所有的权重都是1. 这样就构成了文章中所说的恒等层*

在极端情况下，如果一个恒等映射是最优的，那么将残差置为 0 比使用一堆非线性层使它们的权重都变为1来拟合恒等映射更容易，至少网络性能不会下降。这种残差映射相对于原始的映射会更加容易进行优化的，残差不为 0 时，堆积层在输入特征基础上学习到了新的特征，从而拥有更好的性能。

F(x) + x这种形式很容易在网路中通过 ’shorcut connections‘的形式来实现，传统的 CNN 在不相邻的层间传递需要经过中间层，但深度残差网络提出在输入和输出之间添加一个线性的连接，跳过了中间层，跳过的这些层就是 short connection。在我们的示例中，’shorchut connection‘ 很容易的完成了恒等映射，同时，把这些输出添加到新堆叠层的输出中（如图2所示）。通过这种方式进行添加一方面没有增加额外的参数，另一方面也没有增加计算的复杂度。

残差学习应用到一个小的模块的示例，x和F的维度一样的公式。其中x 和y为模块的输入和输出向量。函数F(x，{wi})为网络需要学习的残差映射。没有增加新的参数，也没有增加计算复杂度。

该模块含有两层, F = W2σ(W1x)，其中σ表示ReLu, 为了简化假设，biases也被省略。进行对应位置像素的加和操作。我们在加和操作结束后，再进行第二次的非线性激活。

如果x和F的维度不一样（例如当我们改变了输入/输出的通道数），我们可以通过线性投影的方式来让他们的维度相互匹配，平方矩阵Ws。

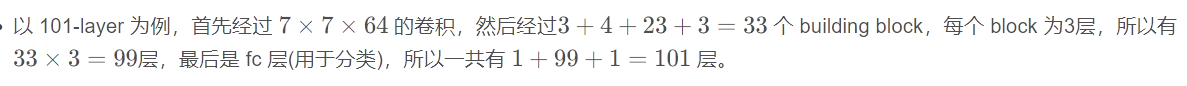
上面的所有的表述都是用的全连接层举例，它们也适用于卷积层。函数 F(x， {Wi})可以表示多个卷积层。

Bottleneck Design 通常用于较深的网络中，目的是减少计算和参数量，第一个 1 × 1 1×1 1×1 的卷积把 256 256 256 维 channel 降到 64 64 64 维，最后再通过 1 × 1 1×1 1×1 卷积恢复channel 数量。

两个在ImageNet上训练的网络架构，中:普通网络具有34层参数(36亿浮点运算次数); 右:残差网络具有34层参数(36亿浮点运算次数)，普通网络的架构(图3，中间)主要受到VGGnet的启发。其中的卷积层更多的使用3x3的过滤器，并且遵循两个简单的设计规则:(i)对于输出和输入相同的特征图，那么层具有相同数量的过滤器；(ii)如果特征图的尺寸减半，那么过滤器的数量则需要翻倍，用来保证每一层的时间复杂性。

基于上述的普通网路，我们在其中插入残差部分。当输入和输出的尺寸一致的时候，shortcut(公式1)可以被直接使用(图3中实线部分)。当维度增加的时候(图3中虚线部分)，有两种方案:（A）shortcut部分仍然使用恒等映射，但是对于新增加的维度则全部使用0来代替。这种方案不增加任何的参数。(B)通过公式2中所提供的线性投影的方式来匹配增加的维度(通过使用1x1的卷积的方式来实现)。对于这两种方案，当两个特征图的尺寸不一致时，都用过使用步长为2的卷积来让他们统一。

作者通过步长为 2 的卷积层直接执行下采样。网络以全局平均池化层和具有softmax的1000维全连接层结束。

作者一共提出了 5 种深度的 ResNet，分别是18、34、50、101和152。所有的网络都分成5部分，分别是：conv1、conv2\_x、conv3\_x、conv4\_x、conv5\_x，

综上所述，通过深度残差学习，可以显著提高图像识别和目标检测的准确性。

PestNet:一种用于大规模多类害虫检测和分类的端到端深度学习方法：

通道空间关注度、卷积神经网络、多类有害生物检测、位置敏感计分图、区域建议网络。

在这种情况下，一个潜在的后果是提取的描述符在害虫中显示出与其他描述符的强烈相似性。

PestNet由三个主要部分组成。首先，提出了一种新的模块——通道空间注意力（CSA）融合到卷积神经网络（CNN）的骨干特征提取和增强。第二个是所谓的区域建议网络（RPN），该网络利用从图像中提取的特征图提供区域建议作为潜在的害虫位置。第三个组件,位置敏感得分图（PSSM），用于取代完全连接（FC）层的害虫分类和边框回归(用于微调标记框)。

被红色粗体虚线框包围的组件将在测试阶段之前对训练图像进行前一个训练阶段。此框中的红色箭头表示训练过程中反向传播阶段的梯度下降过程。

在PestNet中，我们的害虫网包括三个阶段：害虫特征提取、害虫区域搜索和害虫预测。通过害虫图像采集设备捕获数据集中多类害虫图像，在该装置中，多光谱光阱可以发出光以吸引多类害虫，其中波长可以根据一天中害虫的习性随时间变化，以确保可以捕获不同类型的害虫。MPD2018数据集示例可视化的一部分，（MPD2018）为期 10 年的任务特定数据集多类害虫数据集 ，涵盖超过 8万 张图像，其中超过 58万种害虫由农业专家标记并分为 16 类，图像首先被输入到CNN的主干中进行特征提取，输出被称为feature maps。在这一阶段，我们提出了信道空间注意(channel - spatial Attention, CSA)模块来增强卷积块之间的信道和空间信息，用于特征增强。

然后融合RPN和PSSM分别提供害虫区域和害虫预测，预测害虫的类别，并根据区域提案微调边界框。应用区域建议网络(RPN)计算特征图中每个位置的区域建议。这些区域可以区分害虫和非害虫，指示潜在害虫的位置。在通用对象检测方法中，特征图根据裁剪为固定大小的区域建议进行池化操作。

将“（(Contextual RoIs）”添加到区域提案中，并将它们融合在一起以进行更准确的预测。候选识别区域(ROIs)【是把从RPN(Region Proposal Network)产生的候选识别框映射到Feature Map上得到的。】

CNN方法一般采用全连接（FC）层来实现预测[11]。但FC可能对空间位置不敏感，FC包含大量参数，导致计算成本增加，可能会影响实时性能。在PestNet中，为了解决这些问题，使用一个位置敏感得分图（PSSM）的模块来取代FC层，该模块可以计算出建议区域中每个位置的置信度得分。此外，在线硬样本挖掘（OHEM）[OHEM算法的核心是选择一些hard example作为训练的样本从而改善网络参数效果，hard example指的是有多样性和高损失的样本。]算法用于提高训练的有效性。整个训练和测试阶段可以自动运行，以实现害虫检测和分类，因此 PestNet 是一个端到端系统。

取大小为wl−1×Hl−1×Cl−1的特征图和过滤器组cl 在l−1层中大小为

过滤器为例，其他两个超参数填充pl 和跳格sl用于增强。， l层输出feature map的大小为wl ×Hl × Cl:

⌊⋅⌋ 表示楼层操作。请注意，过滤器的数量必须等于输入特征图的数量。更具体地说：式中，i和j分别为和范围内输入和输出特征映射的指标。Mj表示过滤器的接受域，bl j为偏置项。

PestNet使用MaxPooling它通过保留最大的接收字段并丢弃其他值来应用本地池。整个图片被不重叠的分割成若干个同样大小的小块（pooling size）。每个小块内只取最大的数字，再舍弃其他节点后，保持原有的平面结构得出 output。

损失函数被定义为交叉熵损失和盒子回归损失的总和，其中sc∗表示预测分数类c∗，而t和t∗表示边界框的{tx,ty,tw,th}。λ[c∗>0]表示我们只考虑非背景的框（如果c∗=0，则该框是背景。

这是交叉熵损失函数的公式，实际输出概率与期望输出概率之间的相似度，也就是交叉熵的值越小，两个概率分布就越接近。

边界框回归损失函数的公式，Smooth L1损失是由Girshick R在Fast R-CNN中提出的，主要用在目标检测中防止梯度爆炸。

精确率衡量在预测为正样本的样本中模型正样本的正确率，召回率衡量模型正确检测出的正样本占所有正样本的比例，

mAP目标检测中衡量识别精度的指标。