自1997年Yann LeCun提出具有卷积层的多层级连神经网络LeNet[1]以来，卷积神经网络开始进入人们的视线，并成功应用于数字预想识别应用。2012年Hintone提出AlexNet

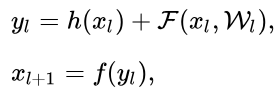
[2],这种卷积神经网络相较于LeNet更深更宽，可以说是首个现代的卷积神经网络。同时，Hintone在这个模型之中加入了ReLU非线性激活函数和GPU加速训练等技术，是的Hintone团队会的了当年ImageNet图像识别竞赛的第一名。AlexNet的成功促进了卷积神经网络接下来的大发展，接下来每年都有卷积神经网络的大量研究论文发表。总体而言，卷积神经网络作为一种特征提取的利器，近年来发展的趋势主要是三个方面：变深、变宽、变轻。

往变深方向发展出来的代表性卷积神经网络主要有VGG[3]、GoogleNet[4], ResNet[5], DenseNet[6]等。其中，被目标检测、语义分割宽泛应用的卷积神经网络ResNet中，最深的网络可以达到152层，应用该网络，Kaiming He等获得了ILCVRC2015的冠军。而往变宽方向发展的卷积神经网络的著名代表是Google公司发表的Inception v1,v2,v3,v4系列和IGCV1, IGCV2, IGCV3系列[7]。同时，由于深度学习技术的飞速发展，计算机视觉领域的一些典型场景和问题例如目标检测、人脸识别、语义分割开始成熟并达到了工业应用的精度。但是，现代生活中，我们应用计算机视觉技术的场景往往是在嵌入式芯片上计算，而非在GPU上计算。这使得设计一种参数量少、计算速度快、精度高的卷积神经网络成为十分棘手的问题。在这个背景下，各种轻量化卷积神经网络被发表提出，其中的代表性网络如MobileNet[8]、ShuffleNet[9]更是被快速的应用到移动设备中。

2.2 往变深方向发展的卷积神经网络

在一张识别任务的图像中，需要识别的物体往往占据图像大部分的空间。而在卷积神经网络之中，由于计算量的考虑，我们往往用例如3x3、5x5等尺度比较小的卷积核对其卷积。在这样的操作之中，这一层的卷积层并不能对大部分的像素或则特征矩阵中的特征点就行卷积，这样就造成卷积神经网络某一层提取的特征只能是局部特征的现象。但是由于识别需要考虑物体的全部特征而非局部特征，所以我们在卷积神经网络中往往使用MaxPooling技术对小尺度卷积核提取的特征矩阵进行尺度缩小操作。这样的操作有卒于减少计算量和增大下一层卷积核对于整张图像的感受野，也就是在减少计算量的同时还能获得更好的全局特征。那么，为了获取更全面的特征表达，MaxPooling操作就会越用越多。同时，卷积神经网络的特征提取能力越来越强。这样就造就了网络越深，性能越好的思路的出现。而VGG，GoogleNet等网络无疑是这一思路的产品，同时也产生了当时最好的图像识别效果。2014年，英国剑桥大学的VGG组提出了VGG网络。该网络包含16层卷积层和3层全连接层。其中，卷积层全部采用3x3的卷积核，作为特征提取部分，而全连接层更多的是作为一种分类器。在VGG网络中，这些卷积层的组合表现出了强大的特征提取能力。时至今日，VGG网络的卷积层结构依然被超分辨率、单目深度估计等计算机视觉任务作为特征提取网络。

卷积神经网络不断加深可以增大感受野，从而提高特征提取能力。那么是不是越深越好和能不能实现无止尽的加深呢？这个问题的答案明显是否定的。卷积神经网络越深，可能增强了性能，但也有可能造成过拟合的风险。同时，由于卷积神经网络采用的反向传播算法是SGD[11]、Adem[12]等基于梯度的优化算法。神经网络越深，那么需要计算梯度的层就越多，而梯度的计算在数学上即是求导。在这种情形下，不断地对上一次求导得来的导数求导就会导致梯度越来越小，直至最后梯度消失，使得神经网络无法优化的问题出现。这便是出现在卷积神经网络上的梯度弥散问题。虽然学者们已经研究出一些例如Relu函数、BatchNorm等方法来缓解这些问题。但是网络越深，效果反而越差的情况还是存在。为了克服以上这些问题，同时也为了能够使得卷积神经网络越来越深，性能越来越好，Kaiming He等人提出了ResNet。在ResNet中，作者采用了一种跳连结构。具体公式可以表达为：



公式1：ResNet残差结构公式

具体卷积结构可以表示为：

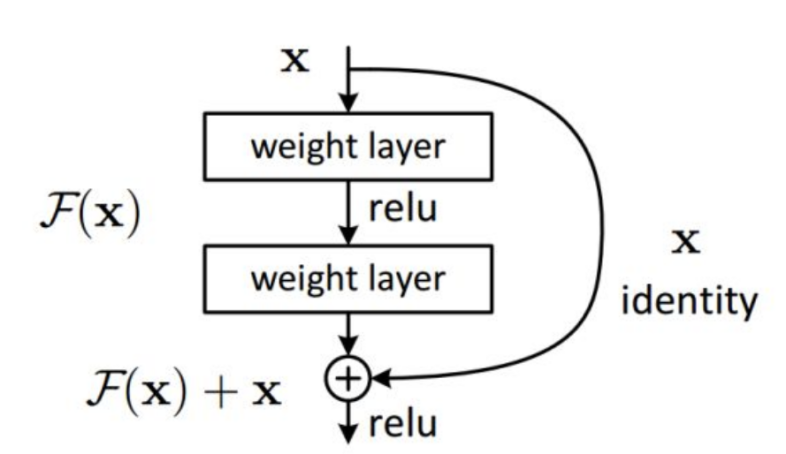


图1：ResNet残差结构示意图

在这样的一个结构之中，在极限条件下，当中间的卷积层没有学习到任何有用的特征事，残差结构相当于恒等映射。这个时候，网络的性能没有提升，但是也不会产生退化现象。当中间的卷积层能够提取到一定的特征时，残差结构就能够提升网络的性能。从结构上看，ResNet更像是不同深度的卷积神经网络组合成的集成投票系统。由于有不同深度的网络的投票集成，网络的效果自然就会更好。

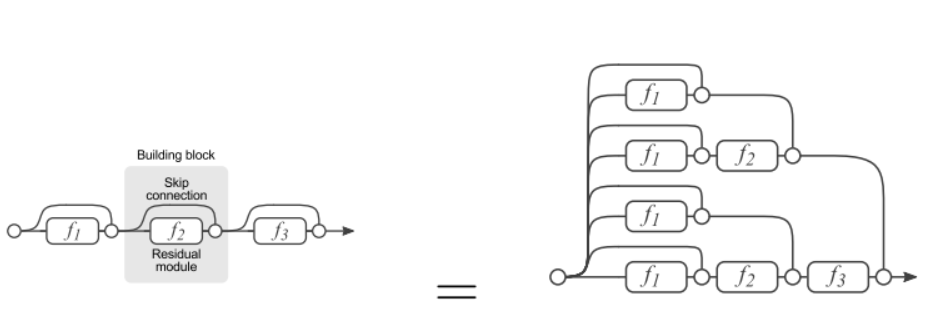
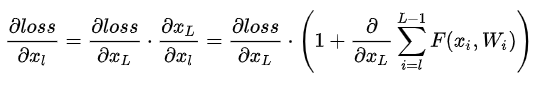


图2：残差结构相当于一个集成投票系统（Ensembling）

除此之外，在进行反向传播进行链式求导的时候，这时候可以表示为：



公式2：残差结构链式求导公式

式子的第一个因子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。所以残差学习会更容易。近年来，关于ResNet为什么有效，学者们做出了很多这方面的工作。在本篇文章中，我们将不做继续谈论。在ResNet论文中，作者提出了5种ResNet结构：

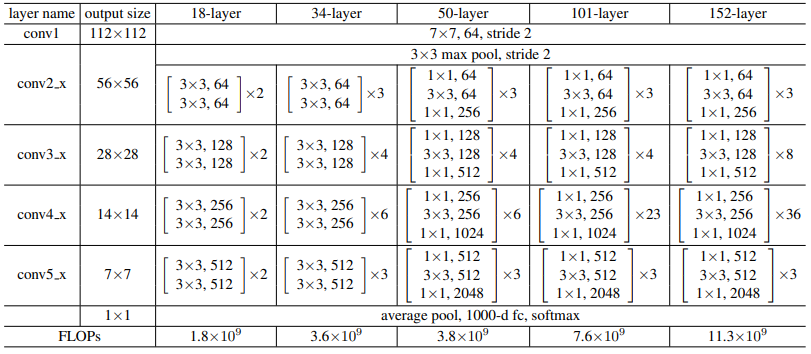


图2：ResNet结构表

在原论文中，在18层和34层的ResNet结构中，作者采用和VGG相似的两层卷积核尺寸为3X3的卷积层作为一个block。在此基础上，为了更深，同时也能减少计算量。更深的结构中，ResNet采用了用1x1卷积层先降维，再用一个卷积层进行特征提取，之后再升高维度的做法。在这样的结构的基础下，在相同的尺度下，采用不同数量的block，形成了不同深度的卷积神经网络。这种结构具有更强的特征提取能力，同时也更加的简洁，可谓是深度学习时代的卷积神经网络的代表性结构。

ResNet之后，DenseNet更加紧凑的使用了skip connection这种结构。DenseNet用skip connection把所有的特征矩阵都连接起来，是的特征得以重复利用。DenseNet也因此斩获CVPR 2017的最佳论文奖。

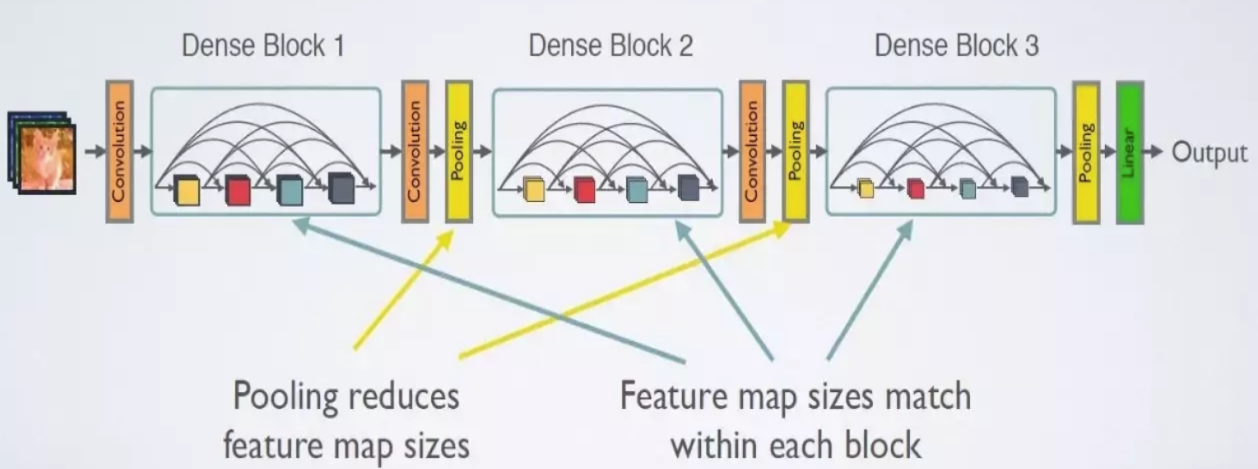
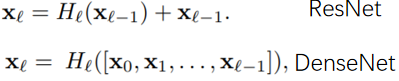


图4：DenseNet全连接结构

DenseNet的连接方式除了是所有特征矩阵全连接之外，它的连接方式也和ResNet不一样.DenseNet采用的是concat操作，可以用公式表示为：



公式3：DenseNet和ResNet连接方式异同

2.3 往变宽方向发展的卷积神经网络

在计算机视觉任务中，物体的形态大小往往各异。那么在多一层的卷积层中，我们网络的感受野都是一样的，当然对于多形态的物体的特征提取是非常不利的。这就造成了计算机视觉领域很重大的一个问题--多尺度特征提取。而卷积神经网络向宽方向发展的趋势由此开始。为了解决多尺度特征提取问题，Google公司研究出了Inception系列卷积神经网络。Inception系列网络的出发点是多尺度特征那个提取及其融合。在多尺度特征提取方面，Inception系列在同一尺度卷积层之间采用了卷积核尺度不一的卷积层进行多支路卷积，这样每一条支路都对应不同尺度的特征。之后，Inception将不同支路的特征矩阵用concat的方式融合起来，作为下一层的输入。这便是Inception系列算法的核心思想。

在Inception V1中，整体架构采用了类似于VGG的方式，并采用结构化的方式，构建一个卷积模块，并通过这个卷积模块的堆积来形成整个网络。在卷积模块层面，原始模块采用了卷积核尺寸为1x1、3x3、5x5的卷积层和尺寸为3x3的max pooling层，并通过concat将提取到特征矩阵融合起来，实现多尺度特征提取和融合，从而提升网络的效果。为了较少计算量，在3x3、5x5卷积之前，用1x1的卷积对输入特征矩阵就行降维，再进行卷积，而且在max pooling之后也就行降维操作。这就是升级版的Inception V1卷积模块，如图4。另外值得一说的是Inception V1采用了在不同尺度的特征矩阵连接分类器进行多输出训练。最近，有研究表明，不同尺度下的特征矩阵可以对应不同尺度物体，这样更加有利于识别。当然，在Inception V1原论文之中，作者更加聚焦于解决多个输出解决梯度弥散问题。

在之后的改进中，Inception V2提出用两个3x3的卷积取代5x5卷积的方案，这样做到了在同一感受野的情况下能够较少计算量的结果。之后，在Inception v3中，作者采取了将一个传统的7x7卷积改成先做一个1x7卷积后做一个7x1卷积的做法。在实验结果上，这种改变不仅减少了计算量，同时也提升了网络性能。在Inception的最后一个版本Inception v4中，受ResNet影响，Inception v4加入了残差结构，这大大加强了网络提取有效特征的能力，进一步提升了Inception 系列网络的性能。

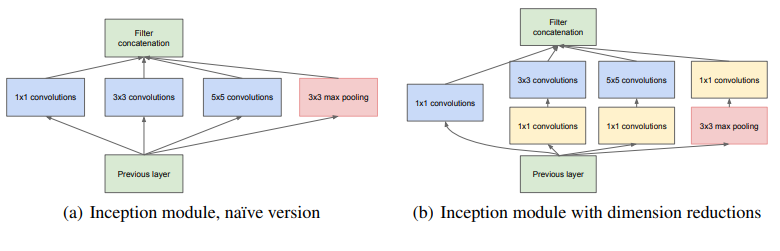
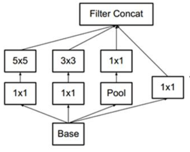
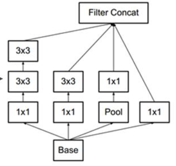
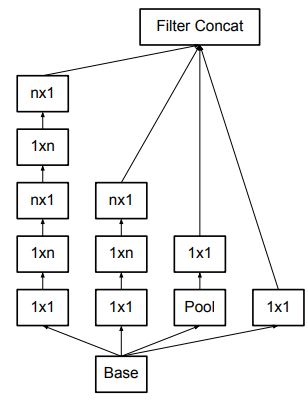
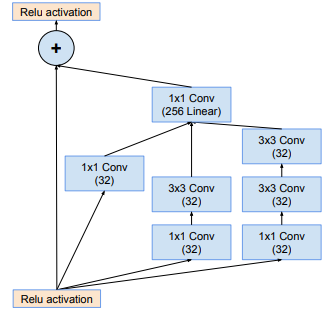


图4、Inception V1卷积模块

Inception V1 Inception V2

Inception V3 Inception V4

图5、Inception系列网络卷积模块变化

2.4 往变轻方向发展的卷积神经网络

在深度学习迅速发展的今天，各项计算机视觉因为应用了深度学习技术而得到了飞跃的突破，一些计算机视觉任务达到了工业应用的精度。但是，在移动设备越来越多的今天，如何在嵌入式设备上使用深度学习技术成为学者们急需解决的问题。所以，轻量化卷积神经网络的研究成为了时下的一个热点领域。

2.4.1 小卷积代替大卷积

在VGG中，作者使用了卷积核尺度为3x3的卷积层取代原有的卷积核尺寸为7x7、5x5的卷积层。对于卷积核尺寸为5x5的卷积层，我们可以使用两层卷积核尺寸为3x3的卷积层，这样可以做到同样的感受野，同时两层非线性的卷积层增加了网络的非线性，性能更加强大。在计算量统计上，我们可以使用浮点运算数(FLOPs)来进行比较。对于一个输入为的特征矩阵，输出为的卷积操作而言，，而采用两层3x3卷积层时，，可以看到，计算量明显较少了。从参数角度来说，对于一层5x5的卷积层，所需要的参数数量为25，而两层3x3卷积层为18。基于此，我们可以说，采用小卷积代替大卷积的做法在参数数量和计算量上都有减少的作用。在之后的研究中，类似于Inception V3，对于一个K x K的卷积层，一些学者采用了将其分解为一个1 x k和一个K x 1卷积的做法。这种采用非对称卷积取代对称卷积的做法，不仅较少了计算量和参数，还提高了网络的精度。

2.4.2 通过降维减少计算量

在一个标准卷积之中，，通过FLOPS计算公式，我们可以看到，对于一个输入输出尺寸固定的卷积层而言，以上公式中可以改变的参数是这个变量。在很多卷积神经网络之中，它们会通过构建一个卷积模块并且堆叠这个模块来构建整个网络。在这个卷积模块中，往往对于输入的特征矩阵进行多次卷积处理。那么，我们如果先对输入的特征矩阵进行降维，然后再进行卷积处理，那么计算量就会下降很多。

在Iandola 等人提出的SqueezeNet中，作者们提出了一个Fire module，如图6。在这个模块中，作者先用1x1卷积对输入特征矩阵进行降维，把通道数量减少，然后再分别使用1x1卷积层和3x3卷积层进行卷积，最后使用concat方法将两个输出融合在一起。这样的操作，在达到和AlexNe相似精度的同时，减少了大量的计算量，从而加快了前向传播速度。

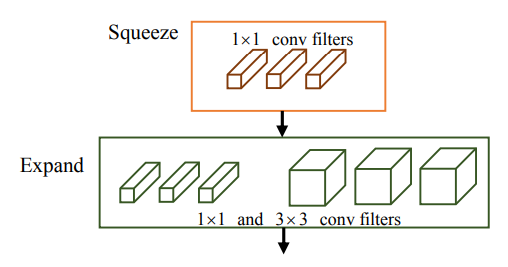


图7 SqueezeNet中的Fire module

2.4.3 分解卷积计算

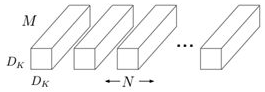
对于一个标准卷积而言，一个输入大小为的特征矩阵，输出为

的特征矩阵，标准卷积会使用个大小为的卷积核对输入特征矩阵进行卷积操作。那么，我们能不能通过改变标准卷积来减少计算量呢？在MobileNet V1中，作者提出用深度可分离卷积(Depthwise Separable Convolution)对标准卷积进行分解。深度可分离卷积有深度卷积和逐点卷积两大部分组成。深度卷积的操作具体实践为对输入特征举证的每一个维度都进行单一通道的卷积处理，之后将所有通道的输出结果进行融合。对于一个输入特征矩阵，深度可分离卷积先用深度卷积对其进行卷积处理。然后使用1x1卷积对深度卷积

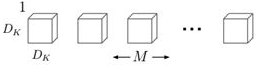
处理结果进行卷积处理，从而对每个通道的卷积结果进行融合和升高特征矩阵的

维度。对于一个输入大小为的特征矩阵，使用深度可分离卷积进行处理并得到一个特征矩阵，。

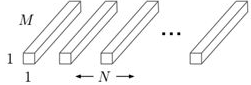
那么，深度可分离卷积计算量是标准卷积计算量的倍。一般，由于都比较大，所以深度可分离卷积相对于标准卷积计算量都大大减少。



(a)



(b)



(c)

图8 a图为标准卷积，b图为深度卷积，c图为逐点卷积。

一般情况下，在深度可分离卷积FLOP计算公式中，参数K都比较小，比较大，那么参数便是影响计算量很大的变量。这即是说，逐点卷积是影响深度可分离卷积计算量的重要因素。为了减少逐点卷积计算量，ShuffleNet V1提出了分组卷积(Group Convolution)的概念。ShuffleNet首先对输入的通道进行分组，再对每个小组进行卷积处理。如果只对每个分组进行卷积，那么，各个分组之间的信息将会局限于这一组。为了加强特征融合和每个小组的信息交流，作者使用了Shuffle操作，使得每个小组的特征进行重新的排列组合，保证信息可以进行足够的交流流通。当ResNet的残差结构使用上1x1标准卷积和深度分离卷积时，。使用分组卷积后，假设分组数量为G，那么。可以看到，分组卷积确实减少了计算量。最后，在达到和AlexNet相同精度的情况下，ShuffleNet V1速度上比AlexNet快13倍。

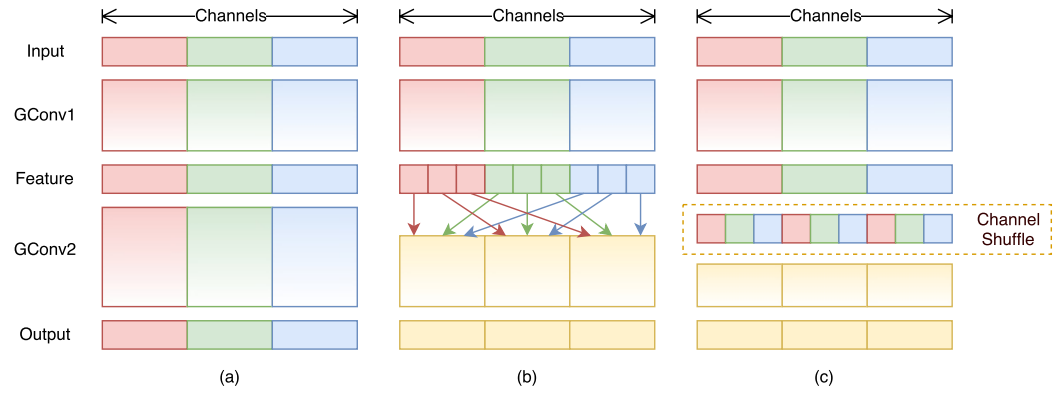


图9 (a)图为一般分组卷积，(b)(c)为加入Shuffle操作的分组卷积

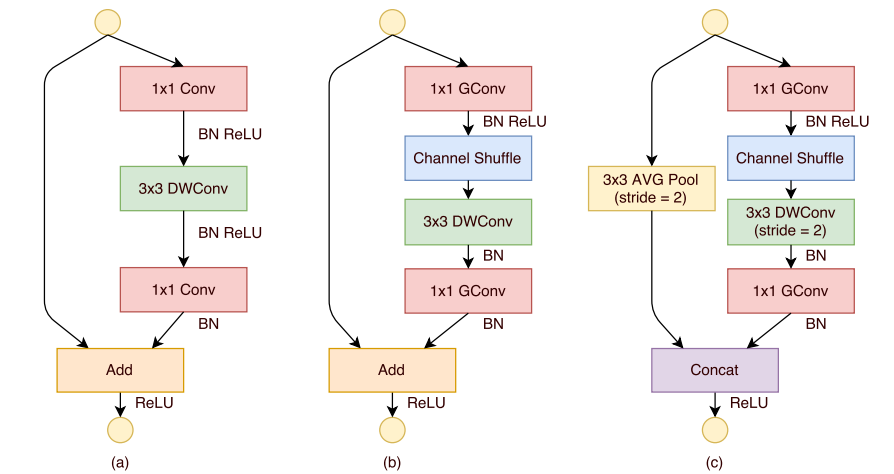


图10 (a)为使用深度卷积的残差结构，(b)为使用分组卷积和Shuffle的残差结构，(c)为当stride=2时使用分组卷积的残差结构

在ResNet的残差结构中，作者先用1x1卷积对输入特征举证进行降维，再做提取特征的卷积处理。但是当特征矩阵的通道比较少时，特征矩阵处于低维情况下，无法很好地提取特征。为了获得相当深度的卷积神经网络和更好的网络性能，MobileNet V2引入了残差结构，并改变了残差结构。在MobileNet V2中，对于输入的特征矩阵，先用1x1卷积进行升高维度，然后使用深度卷积处理，然后再降维。但是在低维情况下，使用非线性激活函数会破坏特征，所以，在这个降维过程中使用了线性激活函数。

但是，影响测试速度的因素除FLOPS外还有很多。在ShuffleNet V2中，作者提出了四种加快测试速度的方案：:1、相同的输入输出通道数 能够减少内存访问成本(MAC)；2、过多的分组卷积(Group Convolution)会增加 MAC；3、网络的碎片化程 度会减少并行化程度；4、不能忽略元素级操作。基于以上4种方案，

ShuffleNet V2在ShuffleNet V1基础上提出了新的残差模块，如图11。ShuffleNet V2 通过提出新的指标(MAC),并且在 MAC 上优化,加快了运行速度 同时也提升了准确率.新的ShuffleNet V2比MobileNet V1快58%,比ShuffleNet V1快63%,比Xception快25%.

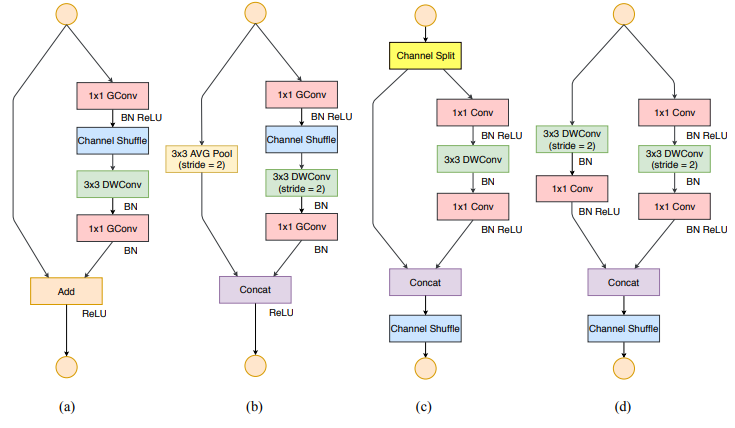


图11 (a)(b)为ShuffleNet V1残差模块，(c)(d)为ShuffleNet V2残差模块

Reference

1. LeCun, Y.; Bottou, L.; Bengio, Y. & Haffner, P. (1998). Gradient-based learning applied to document recognition.Proceedings of the IEEE. 86(11): 2278 - 2324.
2. Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.
3. Simonyan, Karen, Zisserman, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
4. Szegedy C , Liu W , Jia Y , et al. Going Deeper with Convolutions[J]. 2014.
5. He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.
6. Huang, Gao, Liu, Zhuang, Weinberger, Kilian Q. Densely Connected Convolutional Networks[J].
7. Zhang T , Qi G J , Xiao B , et al. Interleaved Group Convolutions for Deep Neural Networks[J]. 2017.Guotian Xie, Jingdong Wang, Ting Zhang, Jianhuang Lai, Richang Hong, Guo-Jun Qi; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8847-8856.Sun K . IGCV3: Interleaved Low-Rank Group Convolutions for Efficient Deep Neural Networks[J]. 2018.
8. Howard A G , Zhu M , Chen B , et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. 2017.Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmenta- tion. arXiv preprint arXiv:1801.04381
9. Zhang X , Zhou X , Lin M , et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[J]. 2017.Ma N , Zhang X , Zheng H T , et al. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design[J]. 2018.