

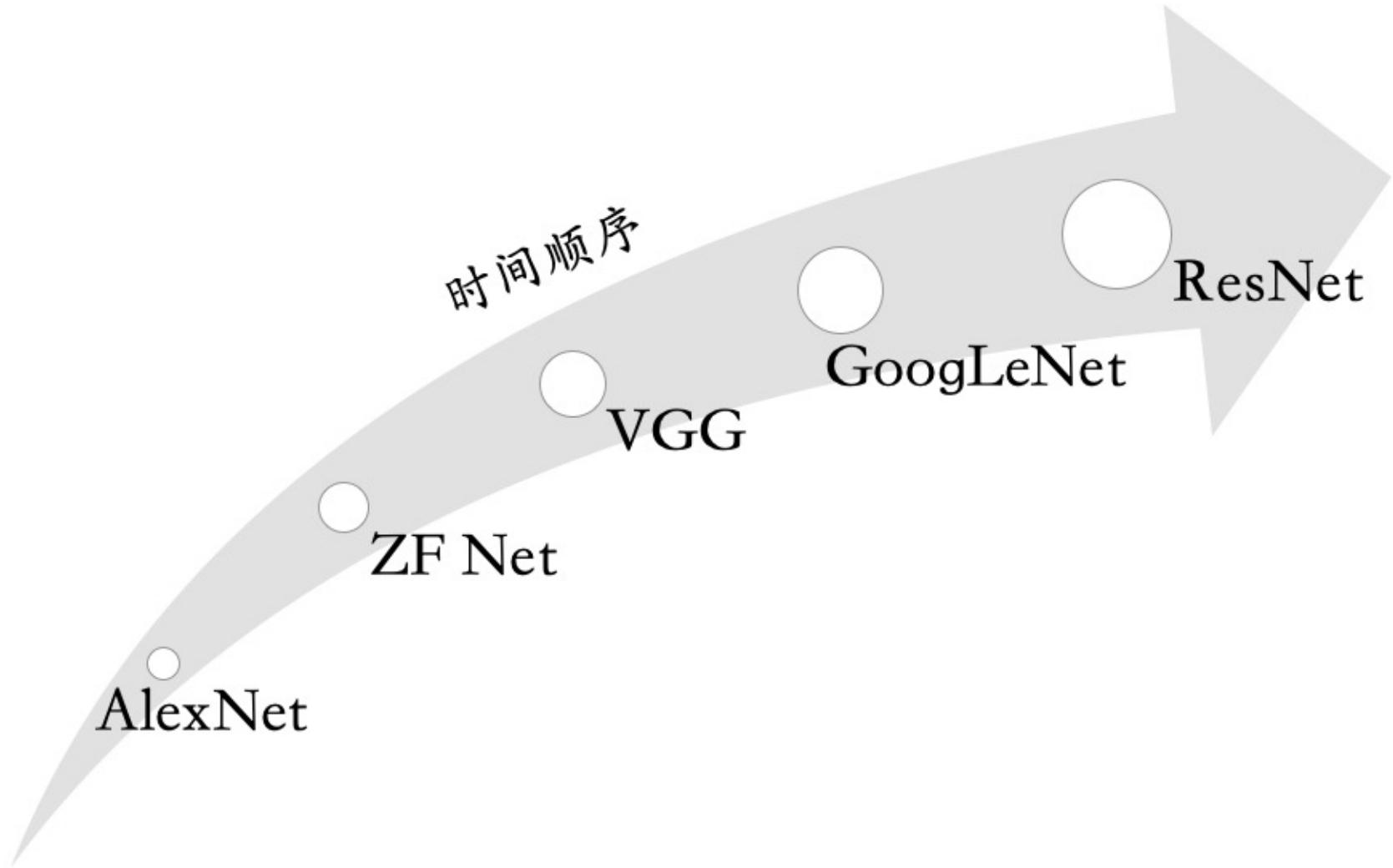


经典CNN概览 (2)

叶山 中国地质大学（北京）

yes@cugb.edu.cn

经典CNN结构



AlexNet回顾

主要的创新点包括：

- 5个卷积层+3个全连接层
- 使用ReLU
- Dropout
- 动量法
- 数据增强策略
- 使用GPU来训练

主要特点：

- 第一个卷积层中，卷积核边长较大、步长较大。
- 这导致图像的空间尺寸在第一个卷积层中就迅速收缩。
- 网络整体更加关注图片的宏观规律。

输入

卷积层+ReLU，边长11，步长4，数量96

局部响应归一化层

最大池化层，边长3，步长2

卷积层+ReLU，边长5，步长1，数量256

局部响应归一化层

最大池化层，边长3，步长2

卷积层+ReLU，边长3，步长1，数量384

卷积层+ReLU，边长3，步长1，数量384

卷积层+ReLU，边长3，步长1，数量256

最大池化层，边长3，步长2

全连接层+ReLU，神经元个数4096

全连接层+ReLU，神经元个数4096

全连接层+Softmax，神经元个数1000

输出

ZF Net (8层)

ZF Net的提出人



Matthew D. Zeiler



Rob Fergus

ZF Net的特点

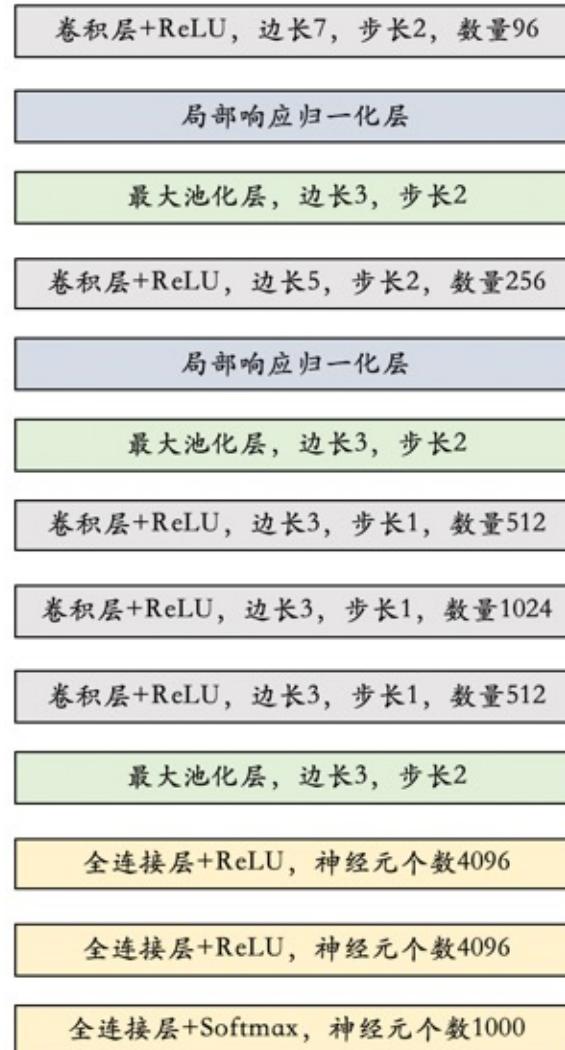
输入



输出

AlexNet

输入



输出

ZF Net

ZF Net的特点

修改版的AlexNet，改动包括：

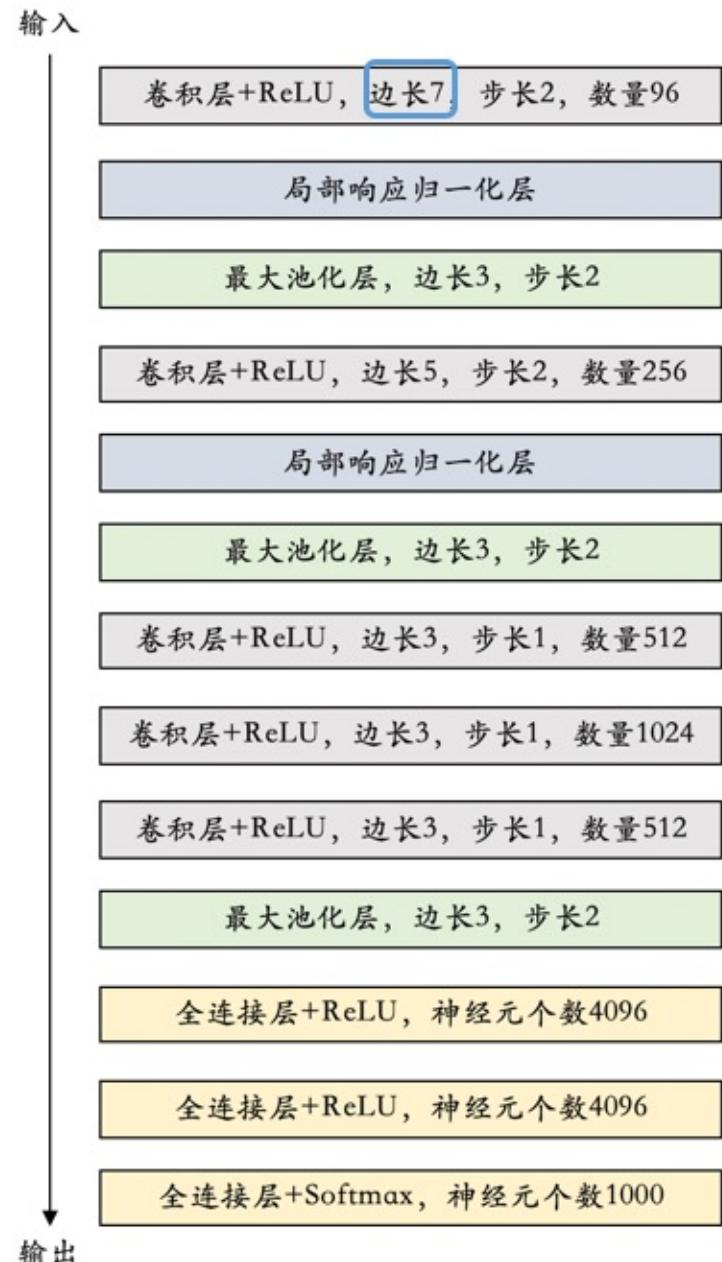
- 第一层的卷积核边长从11改为7；
- 前两层的卷积步长都改为2；
- 第三层到第五次的卷积核数量分别增加到512个、1024个和512个。



ZF Net的创新

改动1：第一层的卷积核边长从11改为7。

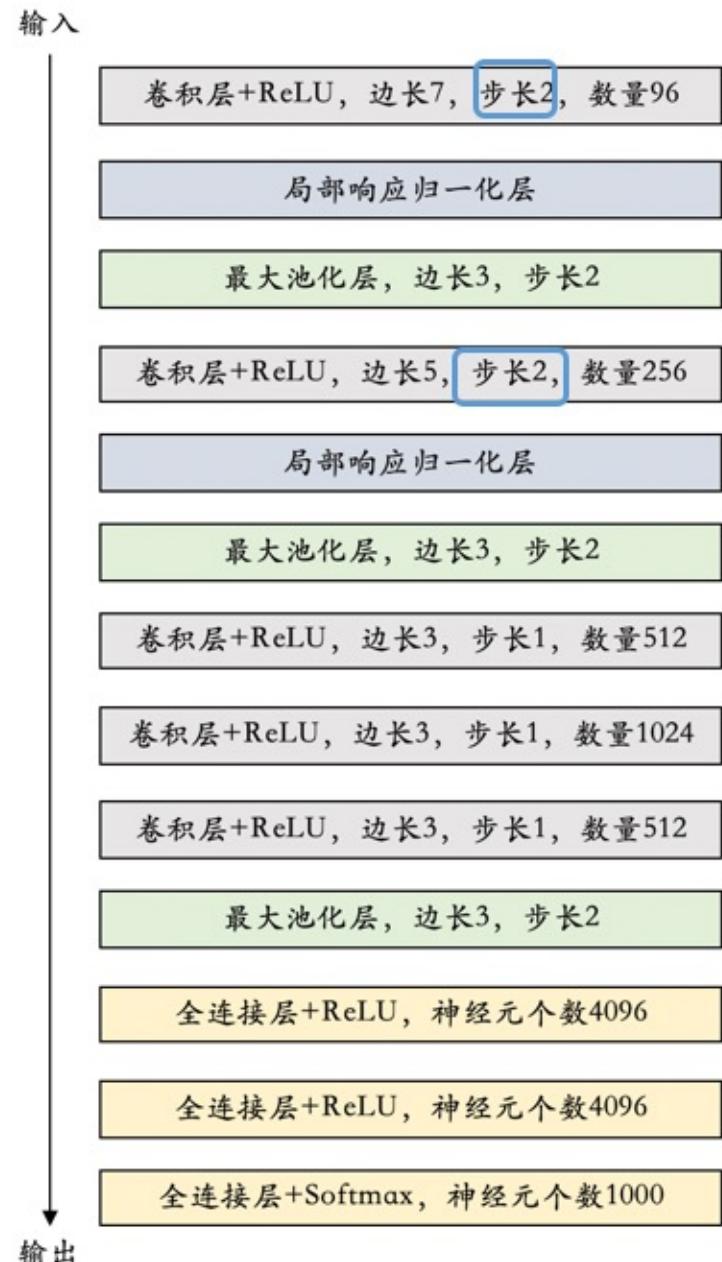
作用：让模型在第一个卷积层中，能感知到图像里粒度更细的特征。



ZF Net的创新

改动2：前两层的卷积步长都改为2。

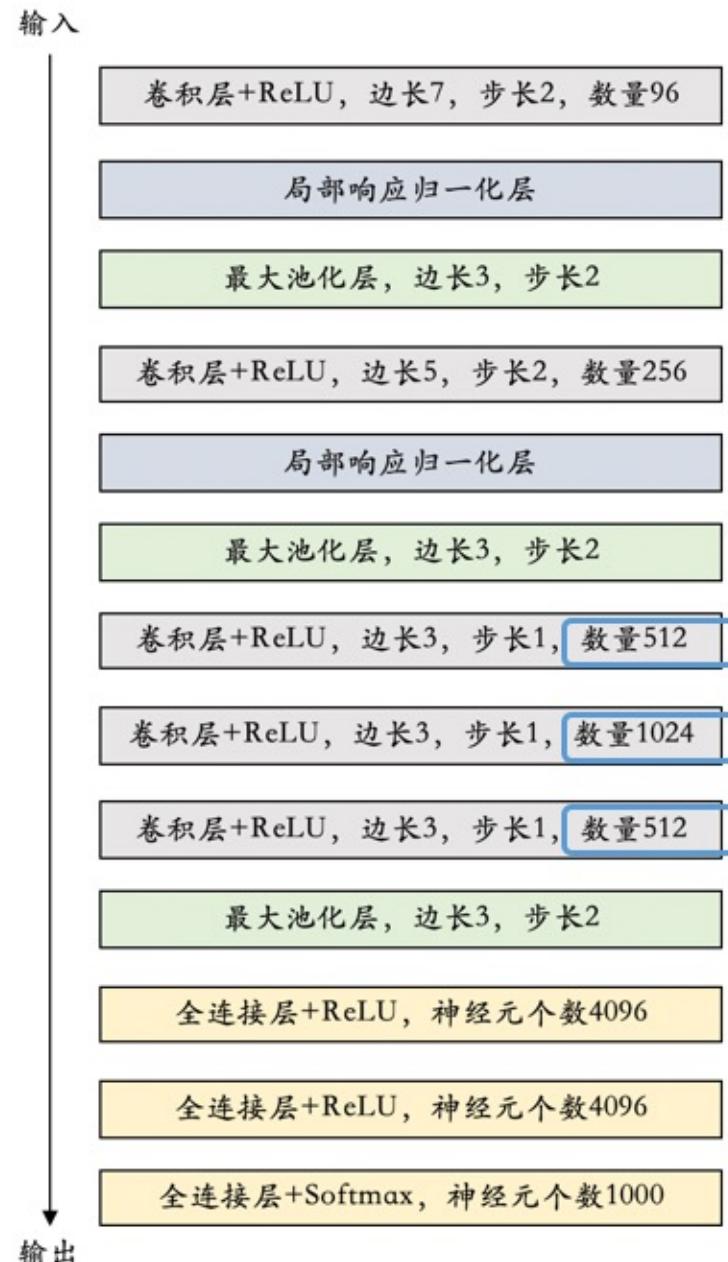
作用：减缓空间尺度的收缩速度，减少每一步的信息损失。



ZF Net的创新

改动3：增加第三层到第五层的卷积核数量。

作用：传入这几层的特征图已经具备一些抽象的“语义”信息，用更多的卷积核，可以提取到更多的语义。



VGG (16或19层)

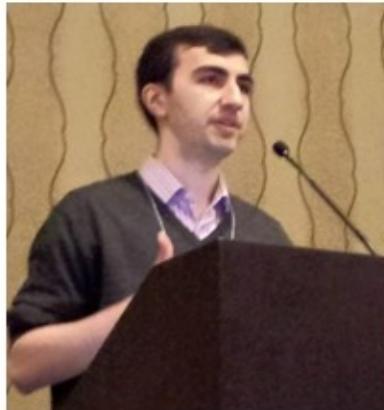
VGG的提出人



Visual Geometry Group



牛津大学视觉几何课题组



Karen Simonyan
现为微软首席科学家



Andrew Zisserman
牛津大学教授

VGG的特点

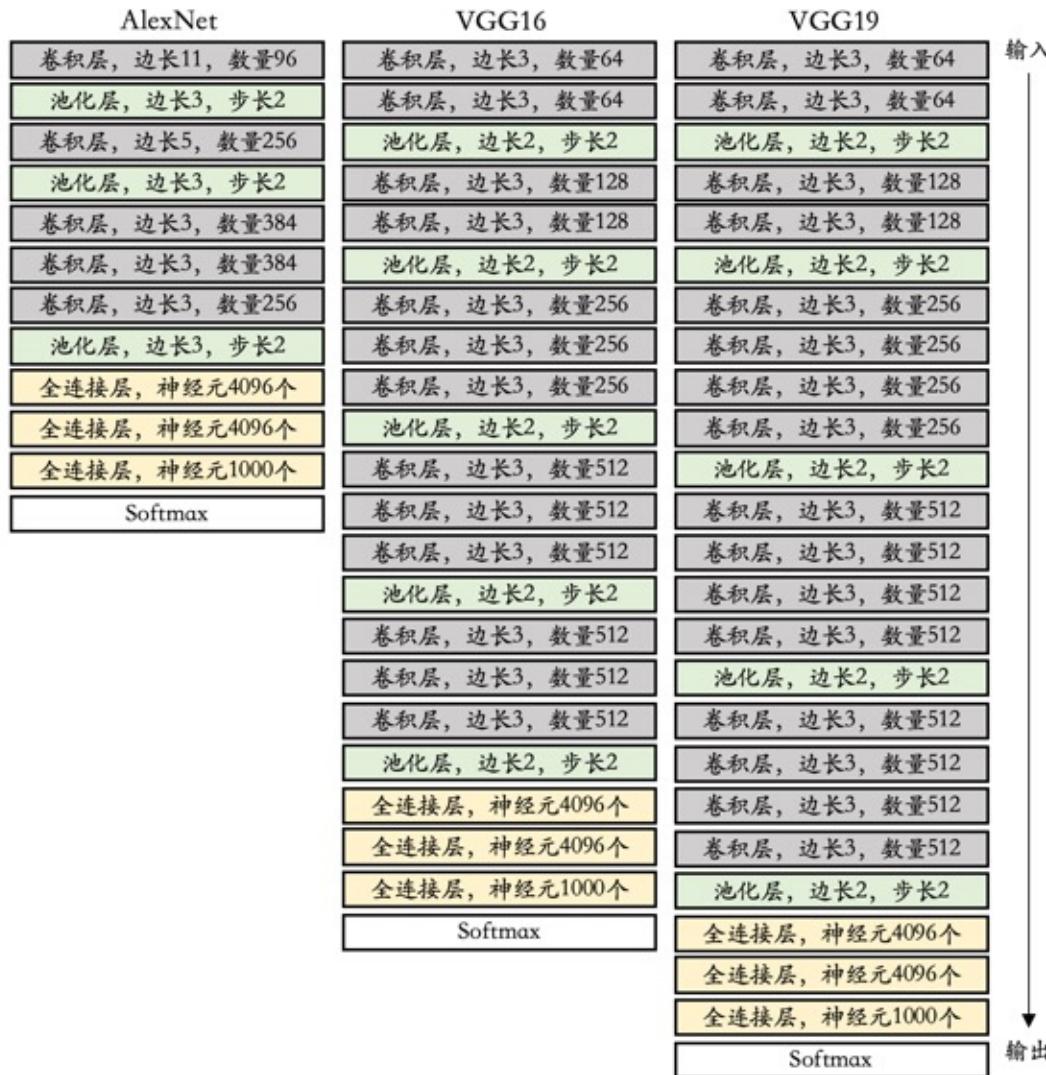
比AlexNet更深的
神经网络

AlexNet	VGG16	VGG19
卷积层，边长11，数量96	卷积层，边长3，数量64	卷积层，边长3，数量64
池化层，边长3，步长2	卷积层，边长3，数量64	卷积层，边长3，数量64
卷积层，边长5，数量256	池化层，边长2，步长2	池化层，边长2，步长2
池化层，边长3，步长2	卷积层，边长3，数量128	卷积层，边长3，数量128
卷积层，边长3，数量384	卷积层，边长3，数量128	卷积层，边长3，数量128
卷积层，边长3，数量384	池化层，边长2，步长2	池化层，边长2，步长2
卷积层，边长3，数量256	卷积层，边长3，数量256	卷积层，边长3，数量256
池化层，边长3，步长2	卷积层，边长3，数量256	卷积层，边长3，数量256
全连接层，神经元4096个	卷积层，边长3，数量256	卷积层，边长3，数量256
全连接层，神经元4096个	池化层，边长2，步长2	卷积层，边长3，数量256
全连接层，神经元1000个	卷积层，边长3，数量512	池化层，边长2，步长2
Softmax	卷积层，边长3，数量512	卷积层，边长3，数量512
	卷积层，边长3，数量512	卷积层，边长3，数量512
	池化层，边长2，步长2	卷积层，边长3，数量512
	卷积层，边长3，数量512	卷积层，边长3，数量512
	卷积层，边长3，数量512	池化层，边长2，步长2
	卷积层，边长3，数量512	卷积层，边长3，数量512
	池化层，边长2，步长2	卷积层，边长3，数量512
	卷积层，边长3，数量512	卷积层，边长3，数量512
	全连接层，神经元4096个	卷积层，边长3，数量512
	全连接层，神经元4096个	卷积层，边长3，数量512
	全连接层，神经元1000个	池化层，边长2，步长2
	Softmax	全连接层，神经元4096个
		全连接层，神经元4096个
		全连接层，神经元1000个
		Softmax

VGG的特点

创新点1：所有卷积层的卷积核边长为3、卷积步长为1、Padding为1。

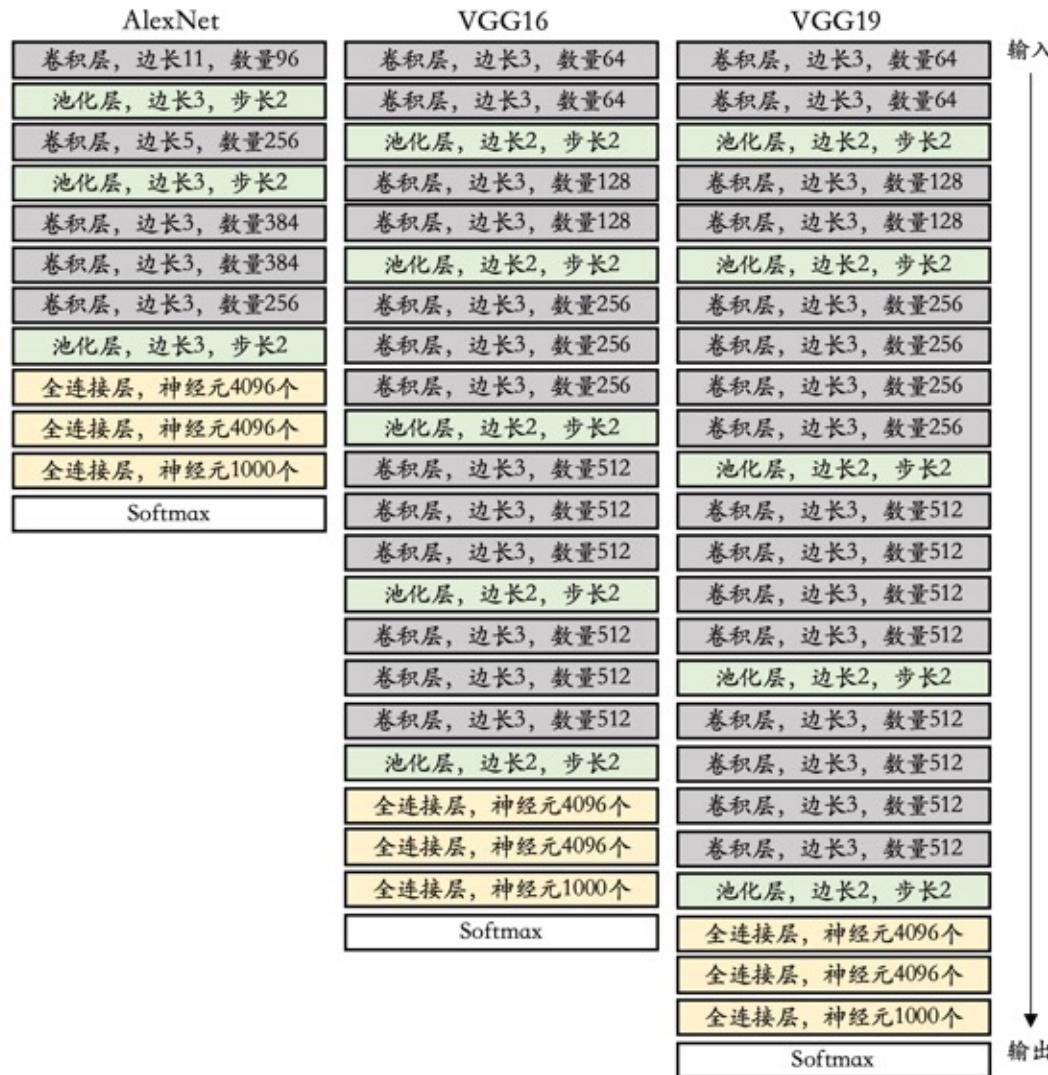
- 串联 3×3 卷积层，减小“感受野”，增加非线性操作次数；
- 减少了模型的参数量，抑制过拟合现象。



VGG的特点

创新点2：每一组“卷积层堆叠群”后面都有最大池化层。

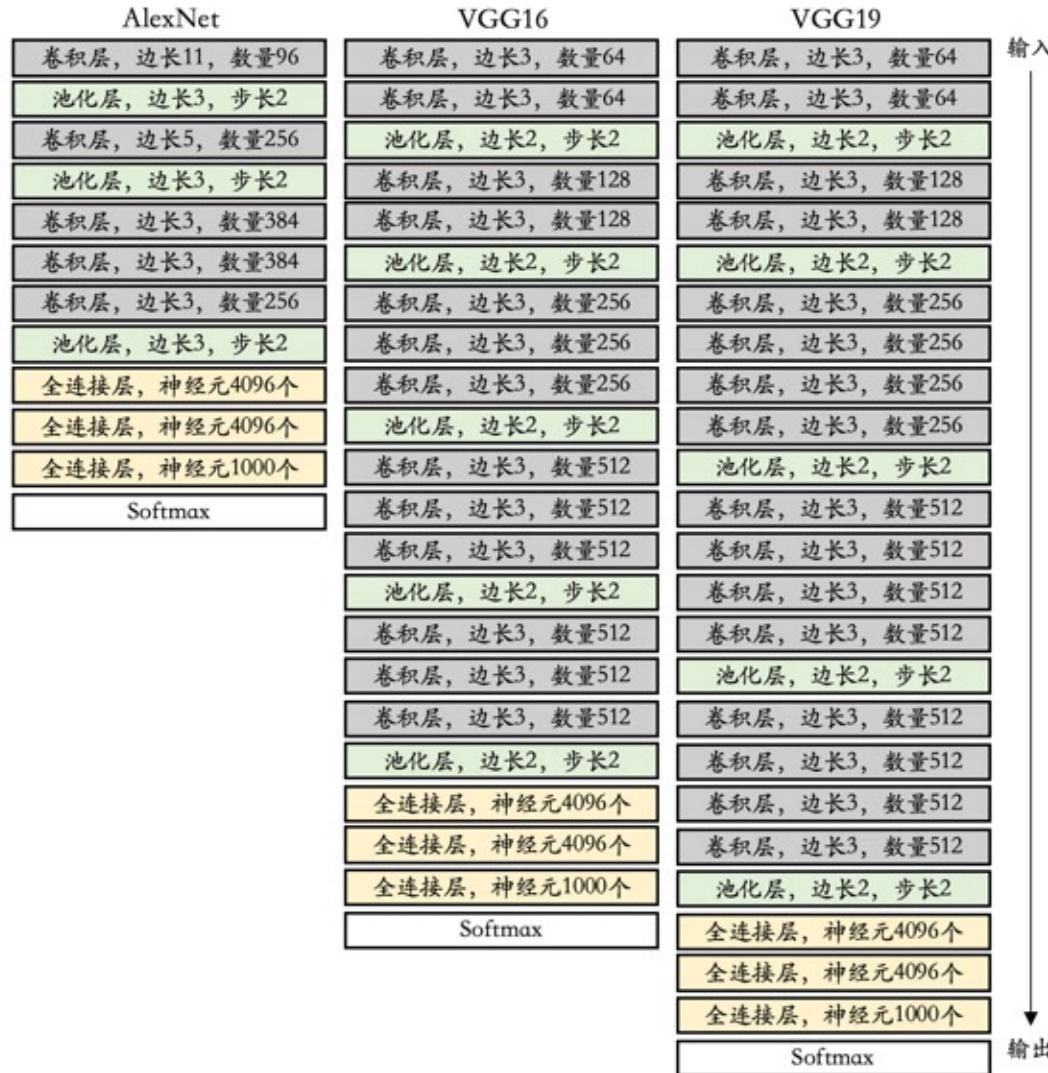
- 所有最大池化层的池化窗口边长为2、池化步长为2；
- 减少了模型的参数量，抑制过拟合现象。



VGG的特点

创新点3：前4个堆叠群，每一次池化操作之后，下一层的卷积核数量加倍。

- 前面的层主要负责学习细粒度、结构化的基本特征（点、线、环等），这种特征的种类较少；
- 靠后的层负责宏观构图的识别，因此使用更多的模板，覆盖更多的可能性。



VGG的特点

创新点3：前4个堆叠群，每一次池化操作之后，下一层的卷积核数量加倍。

- 前面的层主要负责学习细粒度、结构化的基本特征（点、线、环等），这种特征的种类较少；
- 靠后的层负责宏观构图的识别，因此使用更多的模板，覆盖更多的可能性。

类比：英语单词都是由26个字母组成的；汉字都是由横竖撇捺等基础笔画组成的。识别字母和基础笔画，不需要太多的“模板”。

AlexNet	VGG16	VGG19
卷积层，边长11，数量96	卷积层，边长3，数量64	卷积层，边长3，数量64
池化层，边长3，步长2	卷积层，边长3，数量64	卷积层，边长3，数量64
卷积层，边长5，数量256	池化层，边长2，步长2	池化层，边长2，步长2
池化层，边长3，步长2	卷积层，边长3，数量128	卷积层，边长3，数量128
卷积层，边长3，数量384	卷积层，边长3，数量128	卷积层，边长3，数量128
卷积层，边长3，数量384	池化层，边长2，步长2	池化层，边长2，步长2
卷积层，边长3，数量256	卷积层，边长3，数量256	卷积层，边长3，数量256
池化层，边长3，步长2	卷积层，边长3，数量256	卷积层，边长3，数量256
全连接层，神经元4096个	卷积层，边长3，数量256	卷积层，边长3，数量256
全连接层，神经元4096个	池化层，边长2，步长2	池化层，边长2，步长2
全连接层，神经元1000个	卷积层，边长3，数量512	卷积层，边长3，数量512
Softmax	卷积层，边长3，数量512	卷积层，边长3，数量512
	卷积层，边长3，数量512	卷积层，边长3，数量512
	池化层，边长2，步长2	卷积层，边长3，数量512
	卷积层，边长3，数量512	卷积层，边长3，数量512
	卷积层，边长3，数量512	池化层，边长2，步长2
	卷积层，边长3，数量512	卷积层，边长3，数量512
	池化层，边长2，步长2	卷积层，边长3，数量512
	卷积层，边长3，数量512	卷积层，边长3，数量512
	全连接层，神经元4096个	卷积层，边长3，数量512
	全连接层，神经元4096个	卷积层，边长3，数量512
	全连接层，神经元1000个	池化层，边长2，步长2
	Softmax	全连接层，神经元4096个
		全连接层，神经元4096个
		全连接层，神经元1000个
		Softmax

输入

输出

VGG的特点

创新点3：前4个堆叠群，每一次池化操作之后，下一层的卷积核数量加倍。

- 前面的层主要负责学习细粒度、结构化的基本特征（点、线、环等），这种特征的种类较少；
- 靠后的层负责宏观构图的识别，因此使用更多的模板，覆盖更多的可能性。

类比：由英语单词或汉字组成的文章，其结构和行文的特点各异，如需整体识别，需要更多的“模板”。

类比：英语单词都是由26个字母组成的；汉字都是由横竖撇捺等基础笔画组成的。识别字母和基础笔画，不需要太多的“模板”。

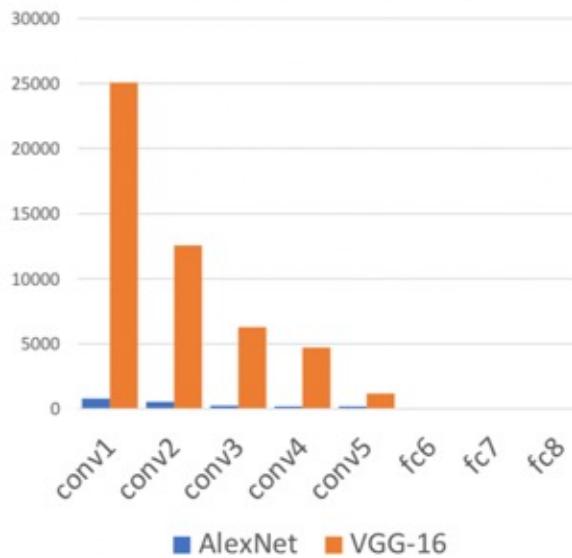
AlexNet	VGG16	VGG19
卷积层，边长11，数量96	卷积层，边长3，数量64	卷积层，边长3，数量64
池化层，边长3，步长2	卷积层，边长3，数量64	卷积层，边长3，数量64
卷积层，边长5，数量256	池化层，边长2，步长2	池化层，边长2，步长2
池化层，边长3，步长2	卷积层，边长3，数量128	卷积层，边长3，数量128
卷积层，边长3，数量384	卷积层，边长3，数量128	卷积层，边长3，数量128
卷积层，边长3，数量384	池化层，边长2，步长2	池化层，边长2，步长2
卷积层，边长3，数量256	卷积层，边长3，数量256	卷积层，边长3，数量256
池化层，边长3，步长2	卷积层，边长3，数量256	卷积层，边长3，数量256
全连接层，神经元4096个	卷积层，边长3，数量256	卷积层，边长3，数量256
全连接层，神经元4096个	池化层，边长2，步长2	池化层，边长2，步长2
全连接层，神经元1000个	卷积层，边长3，数量512	卷积层，边长3，数量512
Softmax	卷积层，边长3，数量512	卷积层，边长3，数量512
	卷积层，边长3，数量512	卷积层，边长3，数量512
	池化层，边长2，步长2	卷积层，边长3，数量512
	卷积层，边长3，数量512	卷积层，边长3，数量512
	卷积层，边长3，数量512	池化层，边长2，步长2
	卷积层，边长3，数量512	卷积层，边长3，数量512
	池化层，边长2，步长2	卷积层，边长3，数量512
	卷积层，边长3，数量512	卷积层，边长3，数量512
	全连接层，神经元4096个	卷积层，边长3，数量512
	全连接层，神经元4096个	卷积层，边长3，数量512
	全连接层，神经元1000个	池化层，边长2，步长2
	Softmax	全连接层，神经元4096个
		全连接层，神经元4096个
		全连接层，神经元1000个
		Softmax

输入

输出

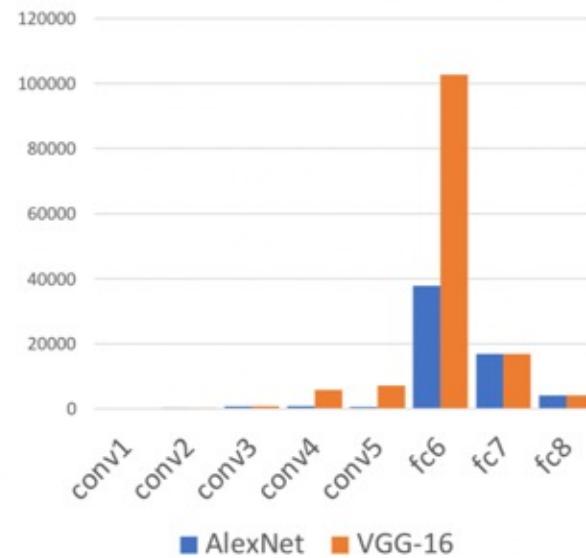
VGG和AlexNet的对比

内存用量 (KB)



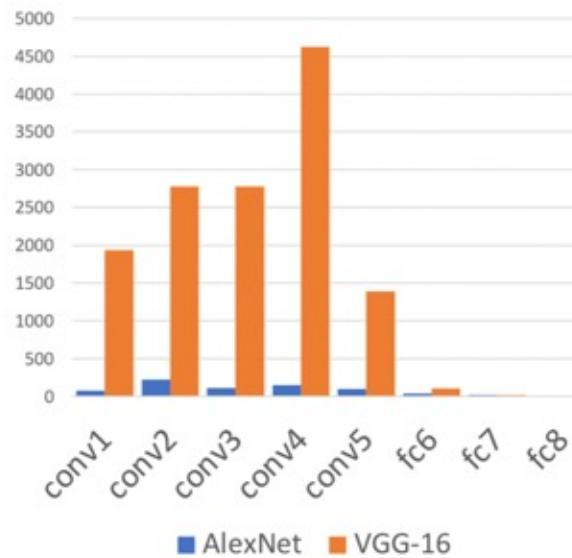
AlexNet: 1.9 MB
VGG-16: 48.6 MB (25倍)

参数总量 (百万)



AlexNet: 6.1千万
VGG-16: 1.38亿 (2.3倍)

浮点运算量 (MFLOPs)



AlexNet: 0.7 GFLOP
VGG-16: 13.6 GFLOP (19倍)

和AlexNet及ZF Net相比，VGG规模更大、更考验硬件的支持度。

GoogLeNet (22层)

GoogLeNet的设计人

Going deeper with convolutions

Christian Szegedy

Google Inc.

Wei Liu

University of North Carolina, Chapel Hill

Yangqing Jia

Google Inc.

Pierre Sermanet

Google Inc.

Scott Reed

University of Michigan

Dragomir Anguelov

Google Inc.

Dumitru Erhan

Google Inc.

Vincent Vanhoucke

Google Inc.

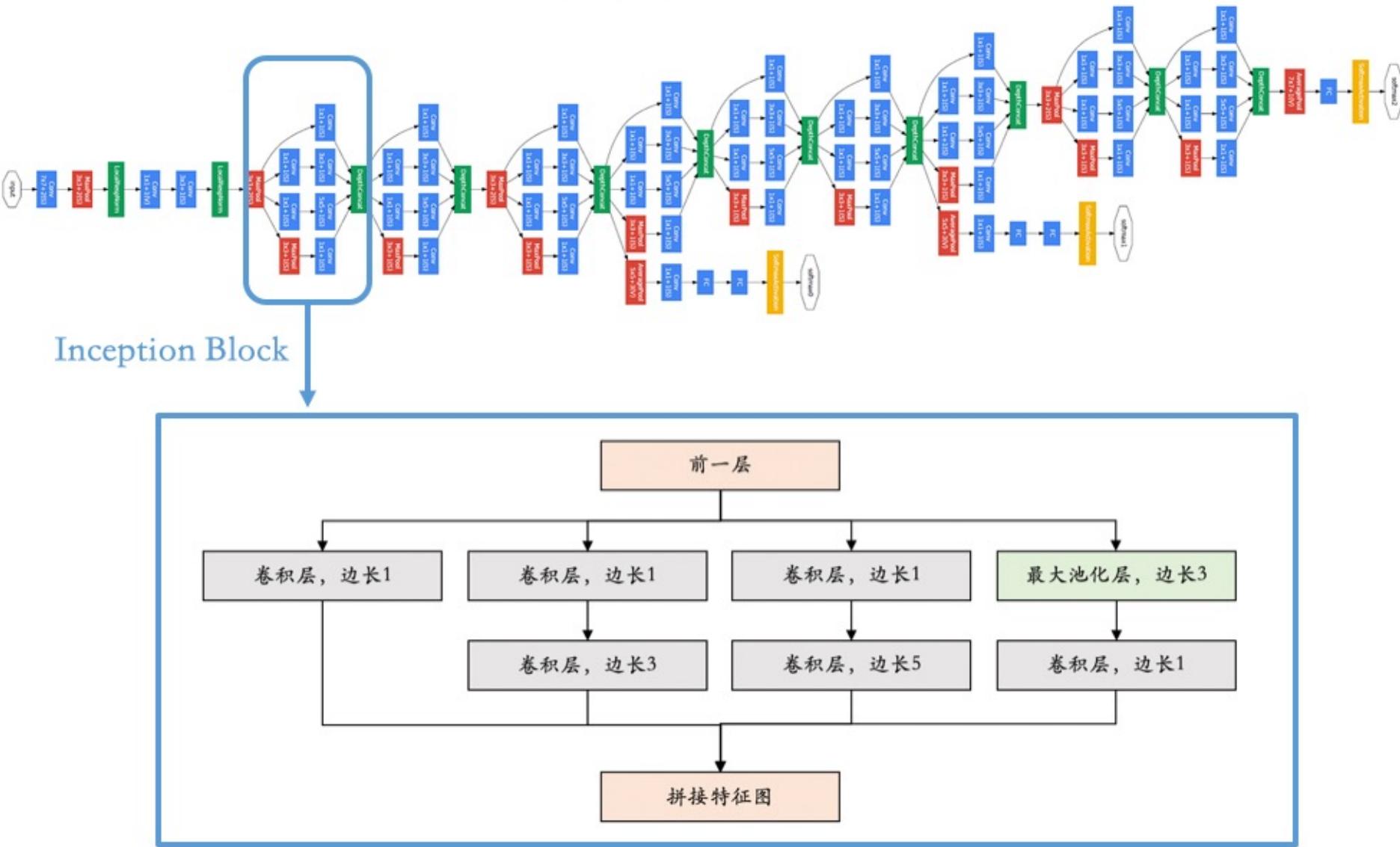
Andrew Rabinovich

Google Inc.

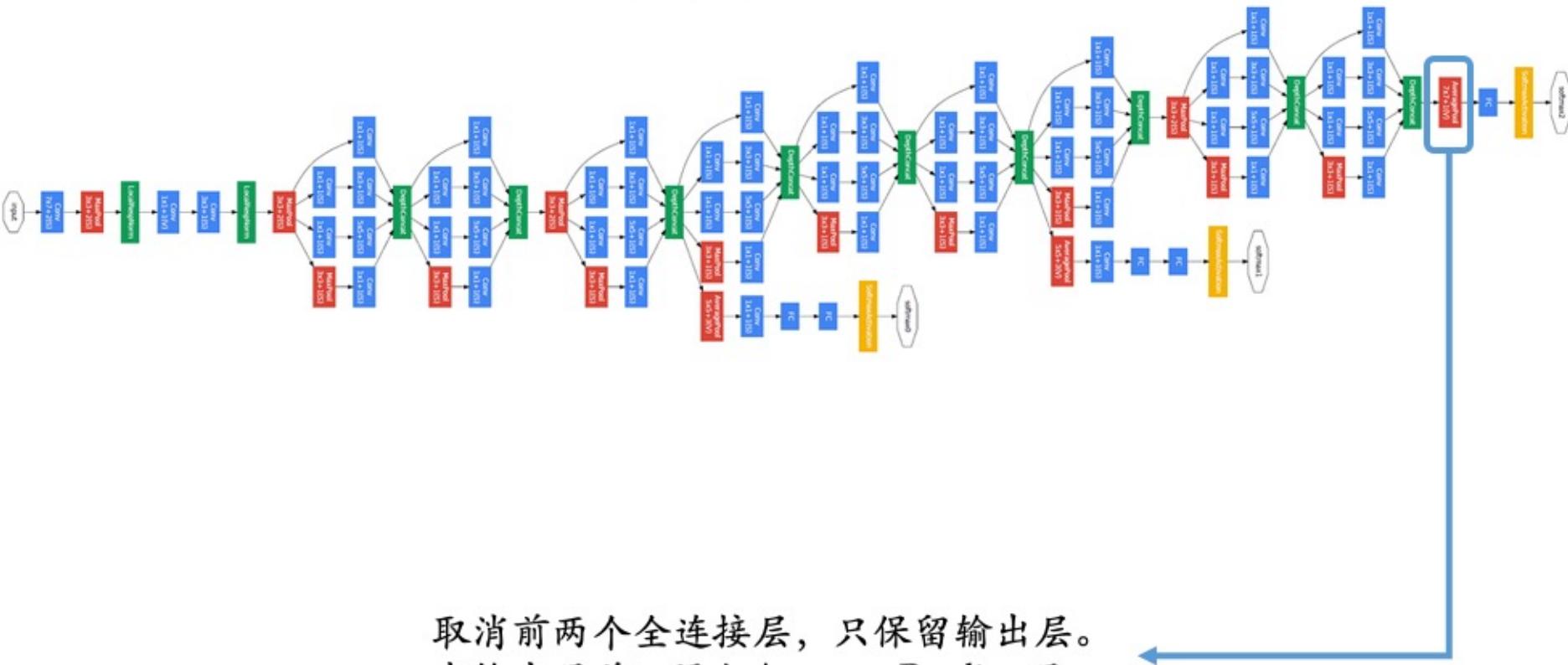


又名：Inception Net
灵感来自电影《盗梦空间》

GoogLeNet的特点



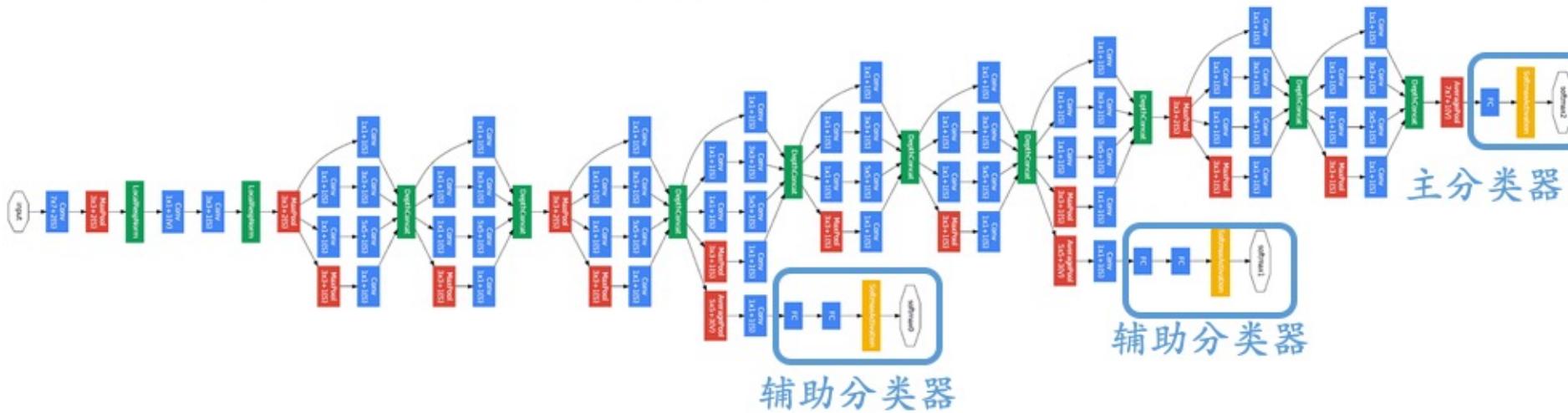
GoogLeNet的特点



取消前两个全连接层，只保留输出层。
在输出层前，添加Average Pooling层。

减少两个全连接层后，参数量较少，只有大约500万，
比AlexNet缩小约12倍！

GoogLeNet的特点



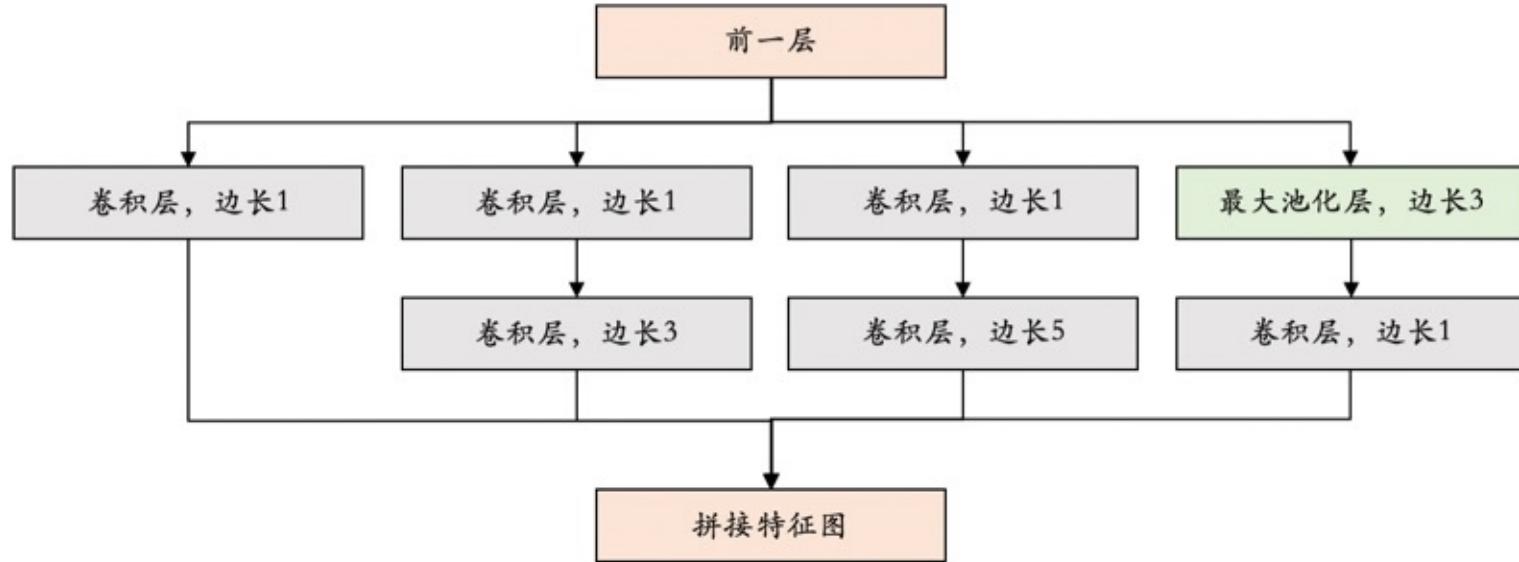
添加了两个辅助分类器，从网络的中间位置输出当前分类结果。

训练阶段，最后的分类结果由主分类区和两个辅助分类器的结果加权求得，权重为主分类器0.4，两个辅助分类器各0.3。

辅助分类器也会计算并回传损失，按权重参与训练。
作用：防止梯度消失、抑制过拟合。

实战阶段，辅助分类器会被关闭，只由主分类器输出结果。

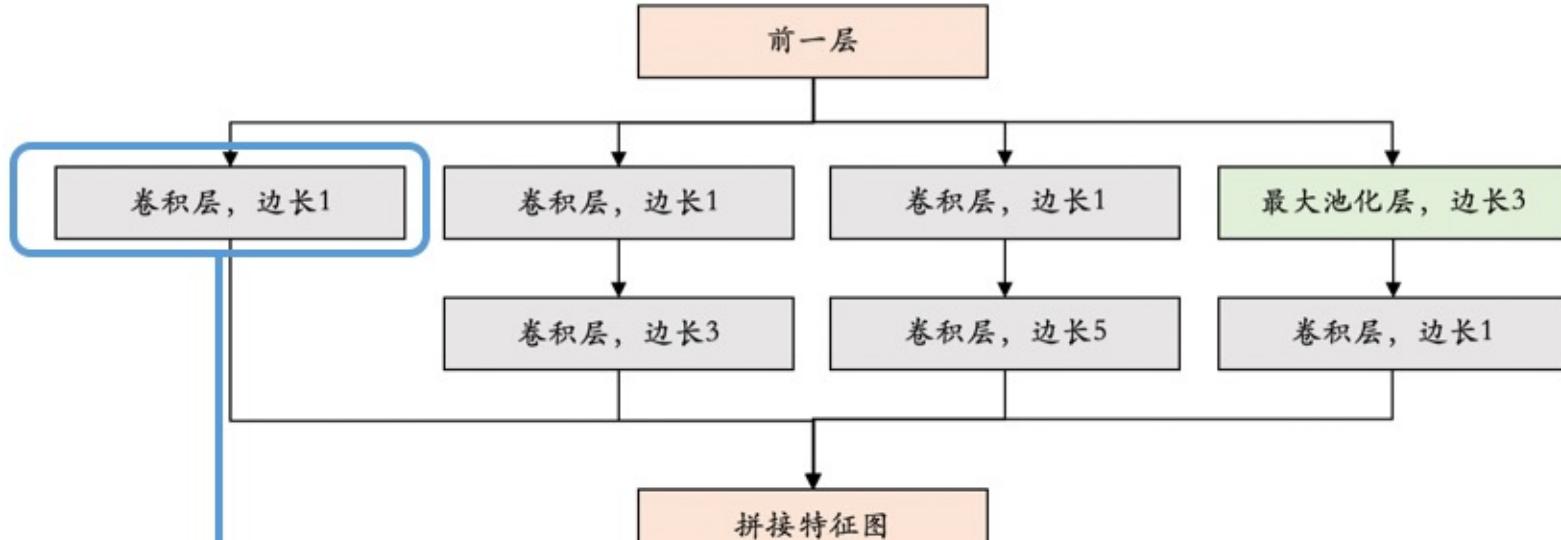
Inception Block



作用：把串联的卷积层改为并联模式。

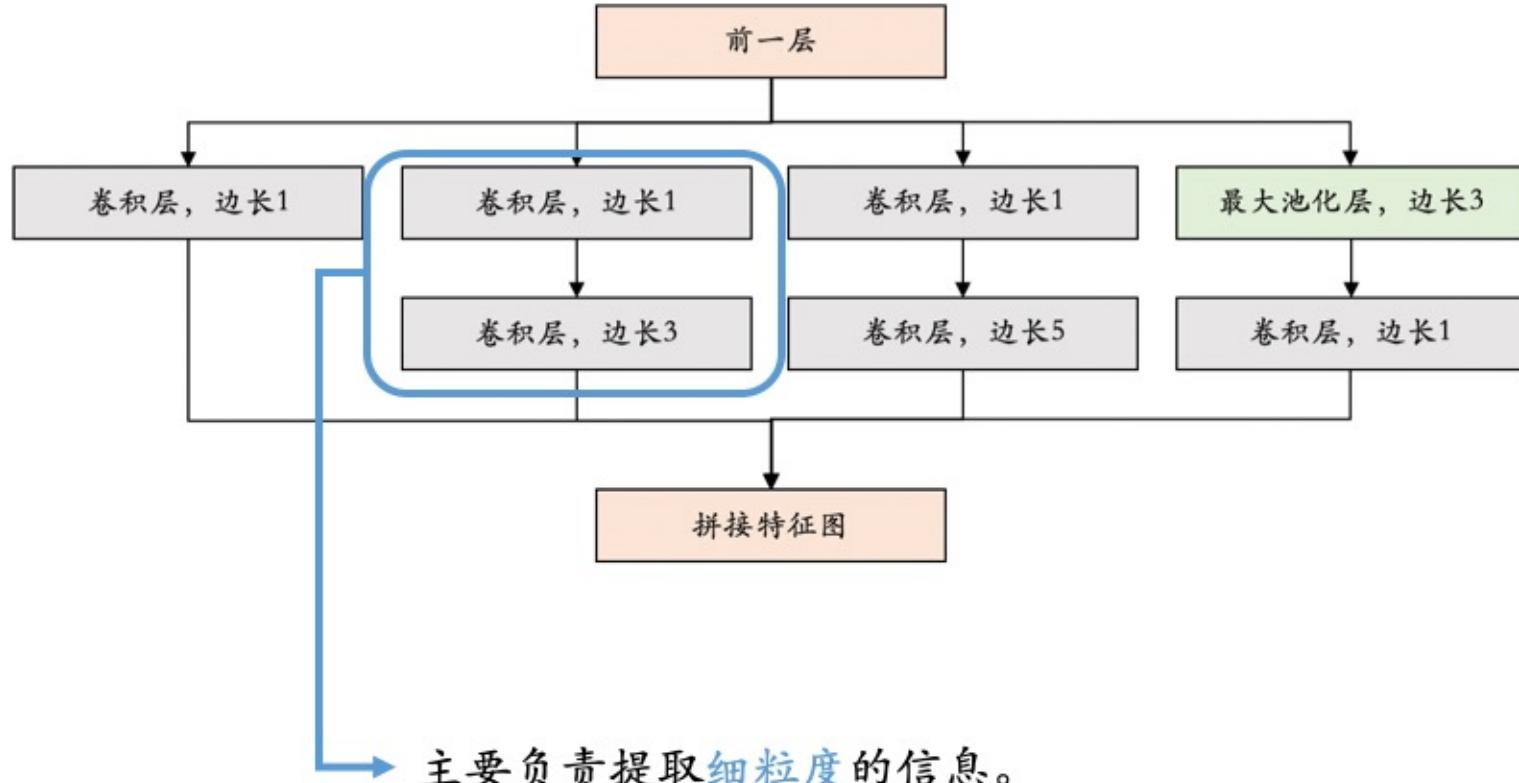
- 串联模式的缺点：前一层卷积核可能丢掉了后一层卷积核擅长提取的关键特征，导致信息损失。
- 并联模式的改良：对于同一个输入，并列进行4个不 同尺度的卷积操作，同时获得不同尺度的特征响应。

Inception Block

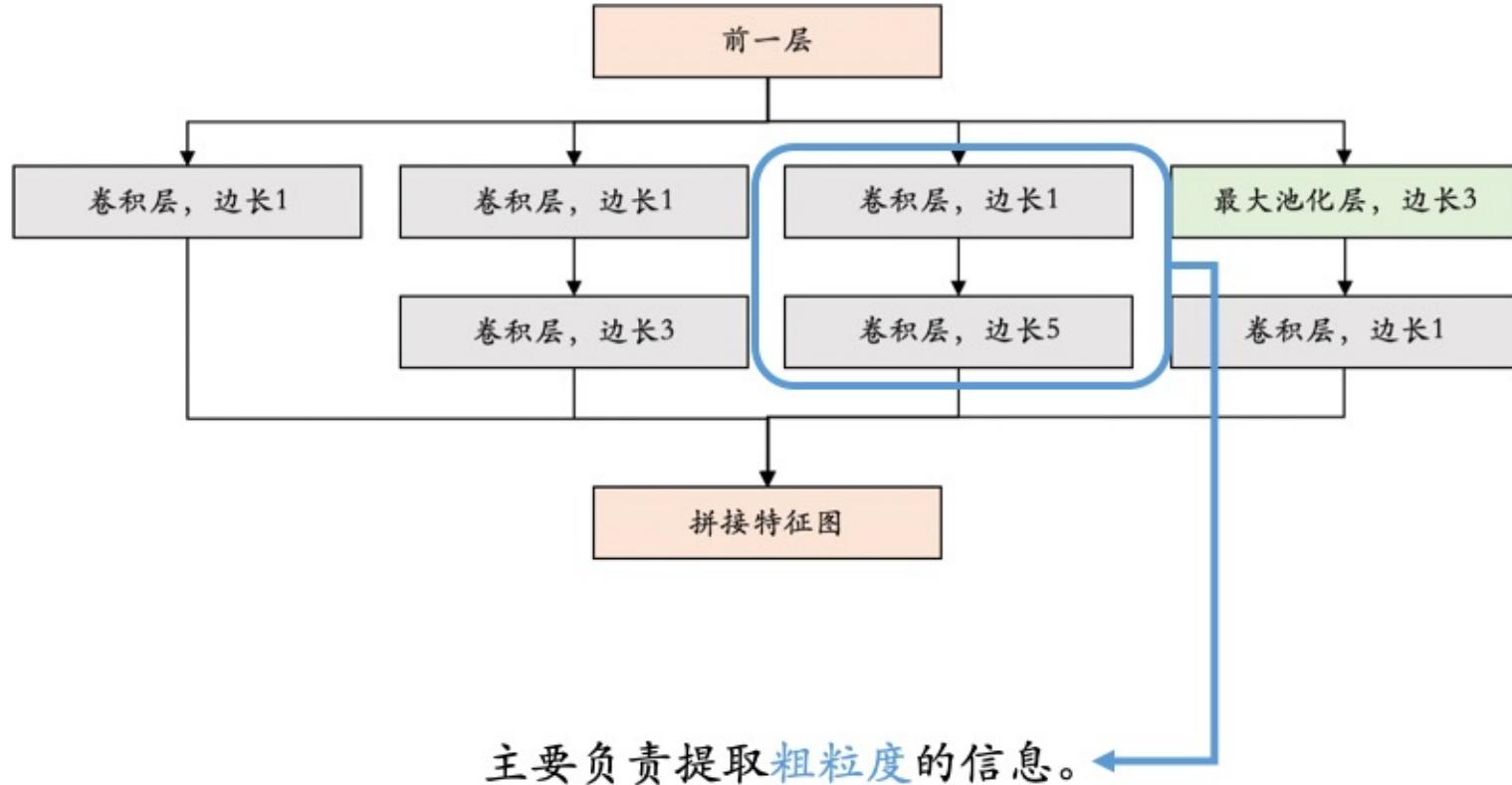


→ 不改变空间尺寸，主要负责处理输入
数据的深度通道上的信息，对其进行
压缩保留。

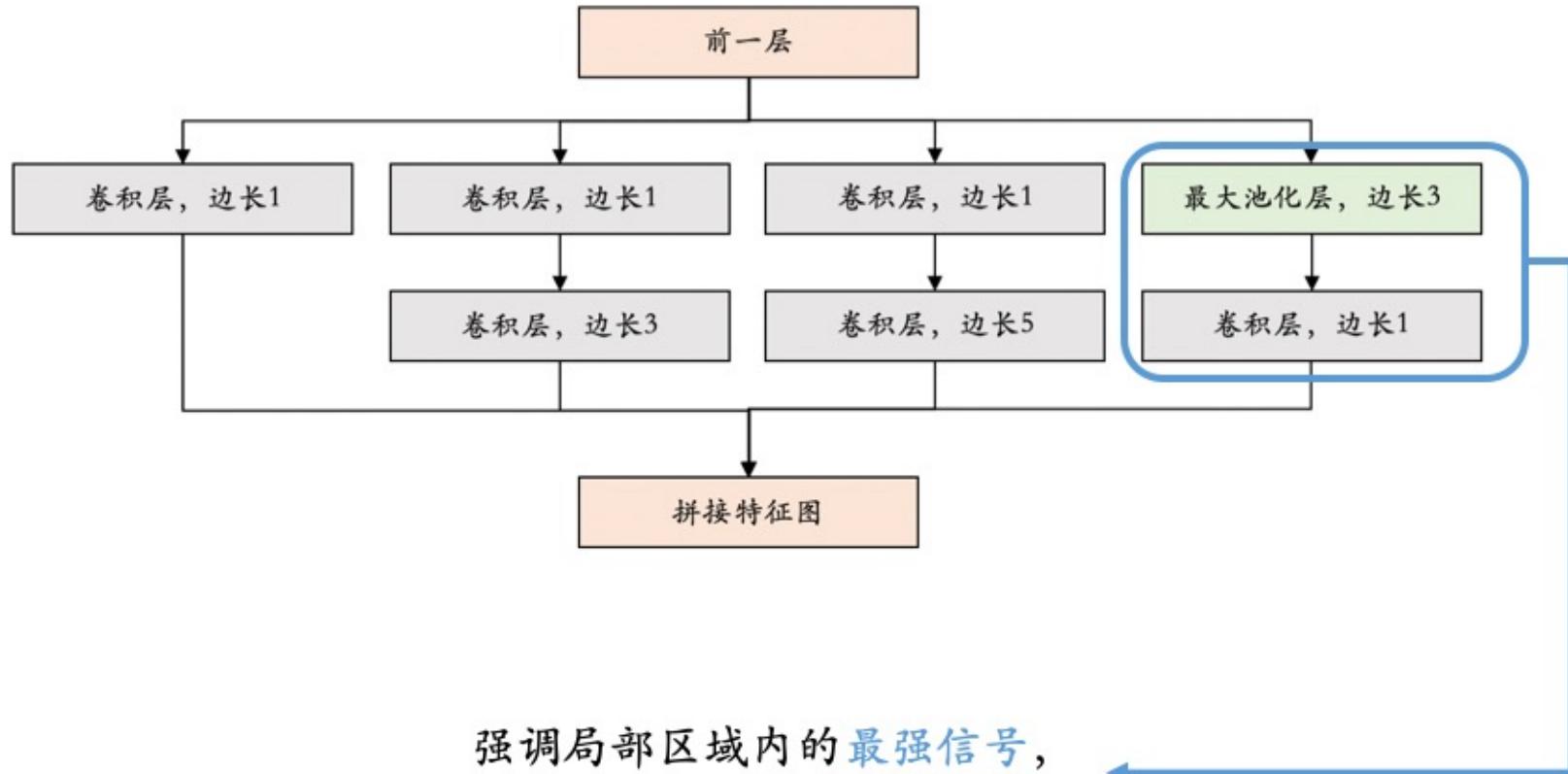
Inception Block



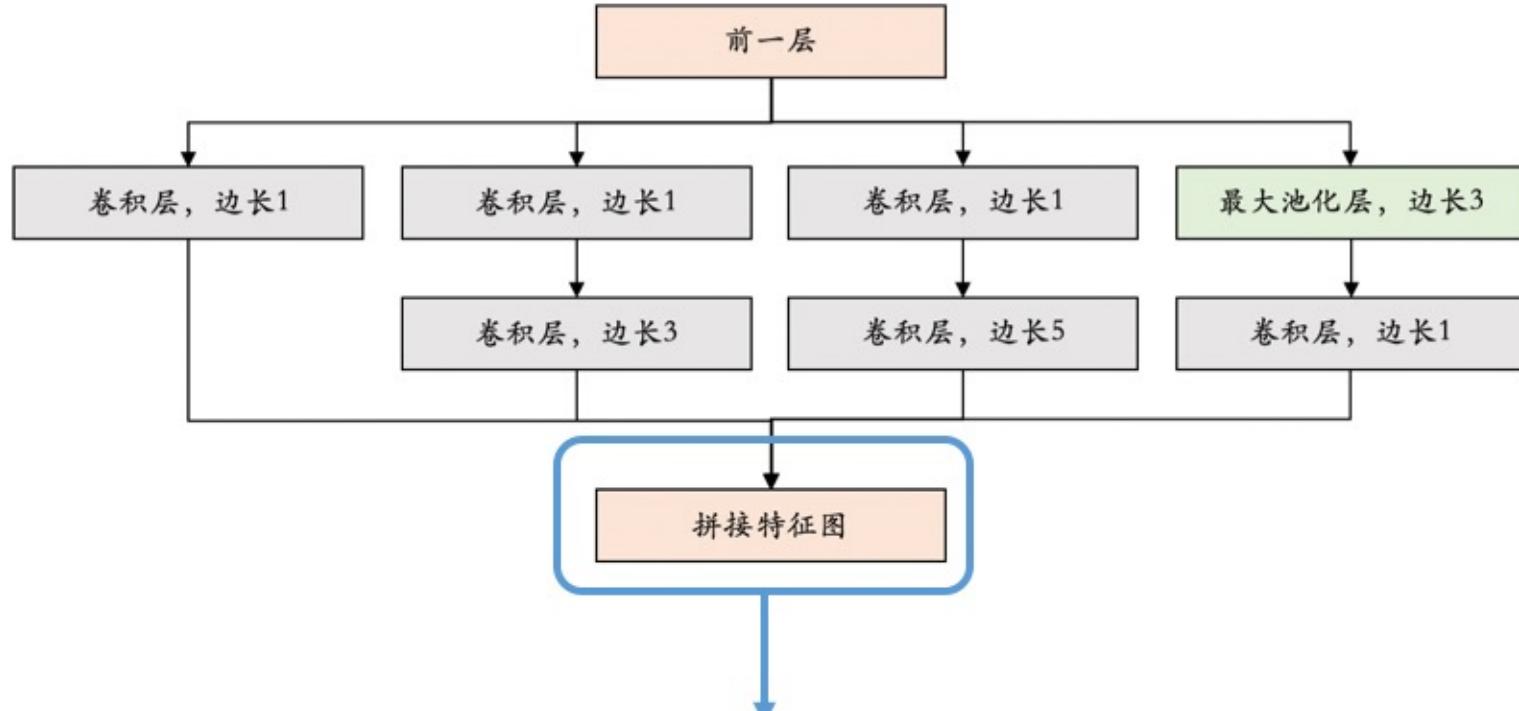
Inception Block



Inception Block



Inception Block



把四个特征图拼接起来，作为
下一层的输入。

Inception Block的优势

提高准
确率

提高计
算效率

缓解过
拟合

模块化
易拓展

适用性
广泛

Pop Quiz

1. AlexNet有多少个最大池化层?

- A. 1个 | B. 2个 | C. 3个 | D. 4个

2. AlexNet中的什么结构在后来已经被淘汰?

- A. 卷积层 | B. 池化层 | C. 局部响应归一化层 | D. 全连接层

3.以下哪一项不是设计卷积核组时要重点考虑的需求?

- A. 特征类型 | B. 尺度 | C. 方向 | D. 特征重复次数

4.在卷积神经网络中，非第一层的卷积核深度由什么决定?

- A. 下一层的卷积核边长 | B. 下一层的卷积核数量
- C. 上一层的卷积核边长 | D. 上一层的卷积核数量

5. 期末考试在哪一天?

- A. 6月23日 | B. 6月24日 | C. 6月25日 | D. 6月26日



ResNet (56层)

ResNet的设计人

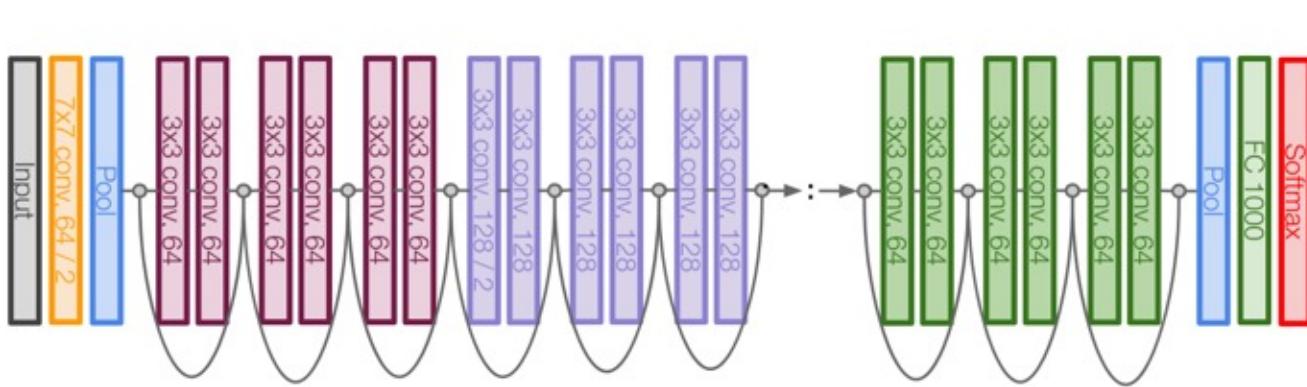
[Submitted on 10 Dec 2015]

Deep Residual Learning for Image Recognition

Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreference functions. We provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth. On the ImageNet dataset we evaluate residual nets with a depth of up to 152 layers---8x deeper than VGG nets but still having lower complexity. An ensemble of these residual nets achieves 3.57% error on the ImageNet test set. This result won the 1st place on the ILSVRC 2015 classification task. We also present analysis on CIFAR-10 with 100 and 1000 layers.

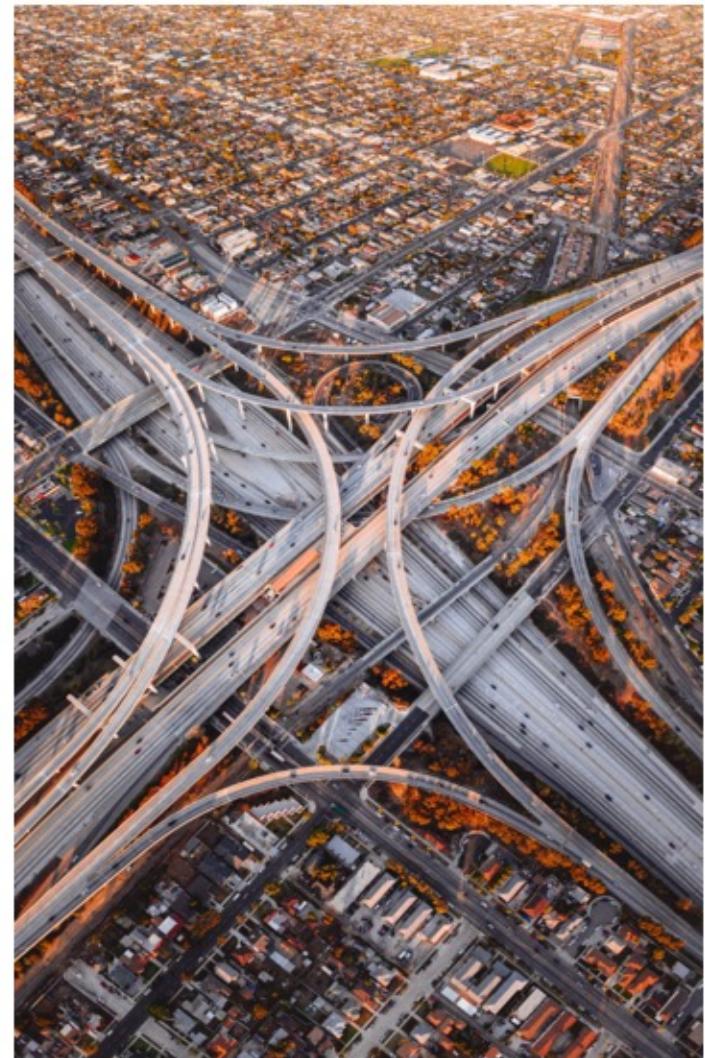
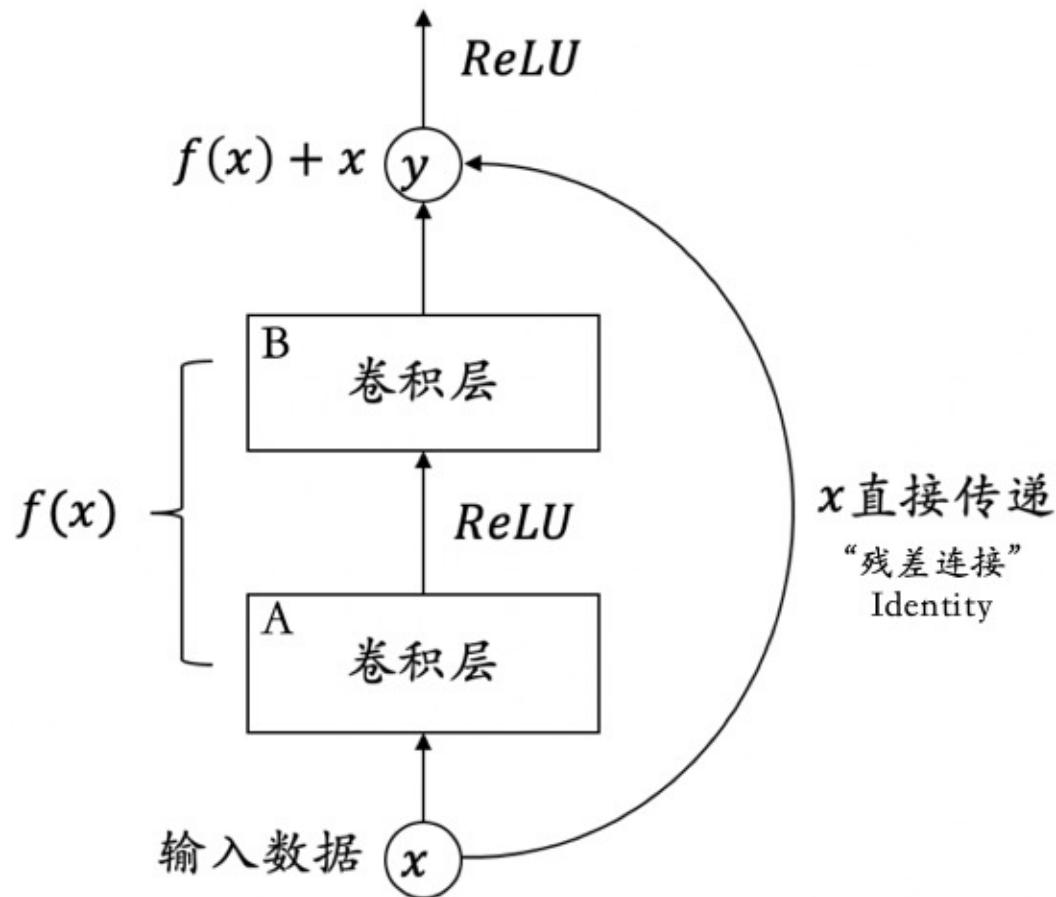
The depth of representations is of central importance for many visual recognition tasks. Solely due to our extremely deep representations, we obtain a 28% relative improvement on the COCO object detection dataset. Deep residual nets are foundations of our submissions to ILSVRC & COCO 2015 competitions, where we also won the 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation.



全名：残差神经网络 (residual neural network)

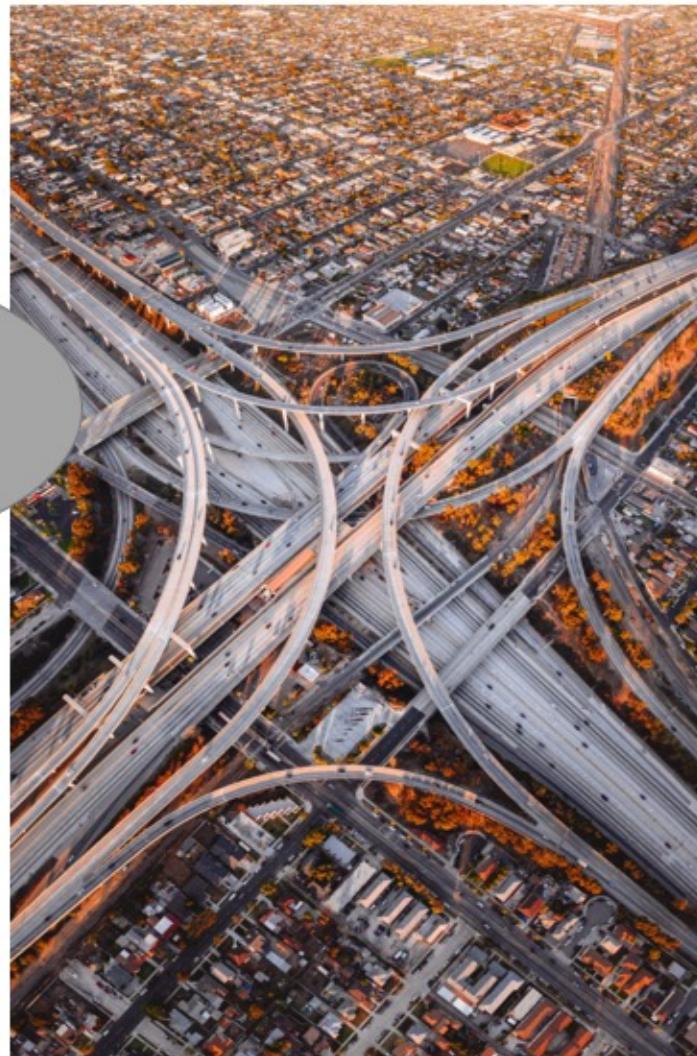
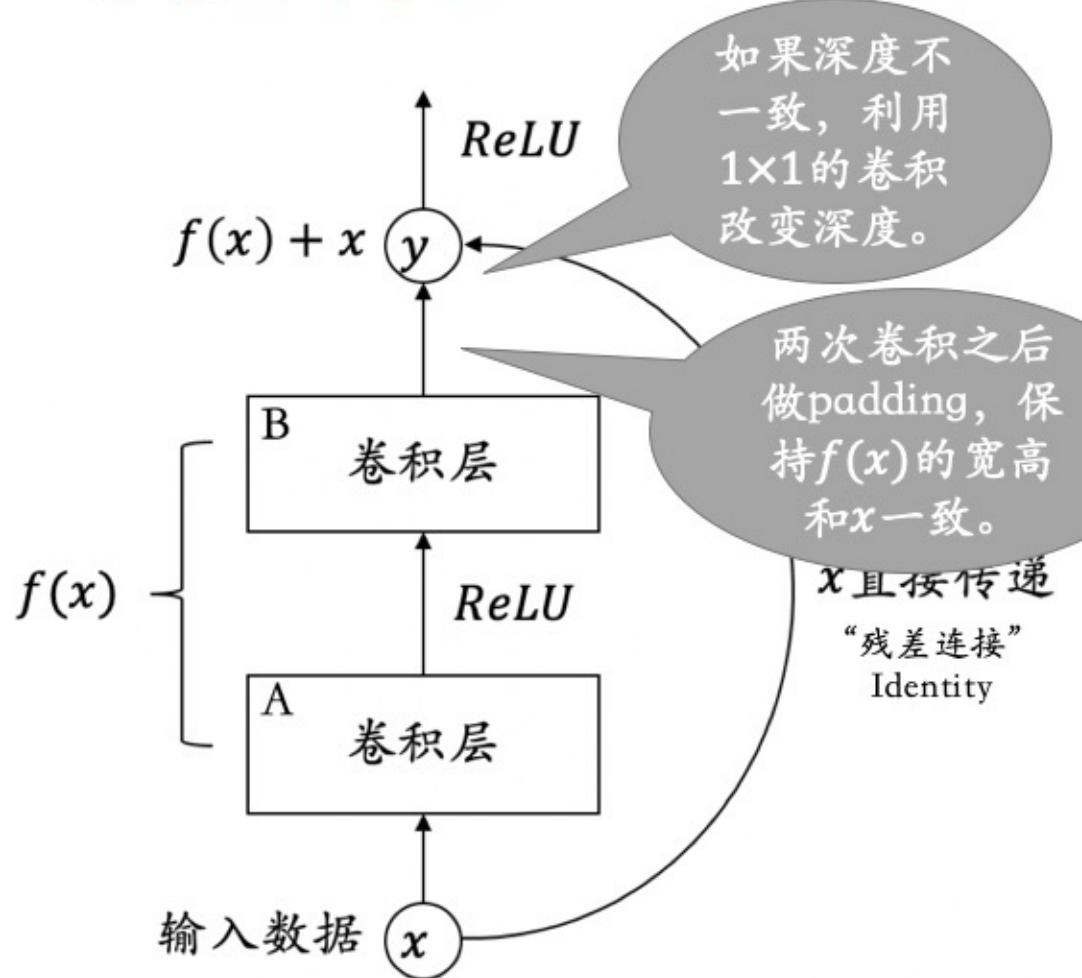
何恺明

残差单元



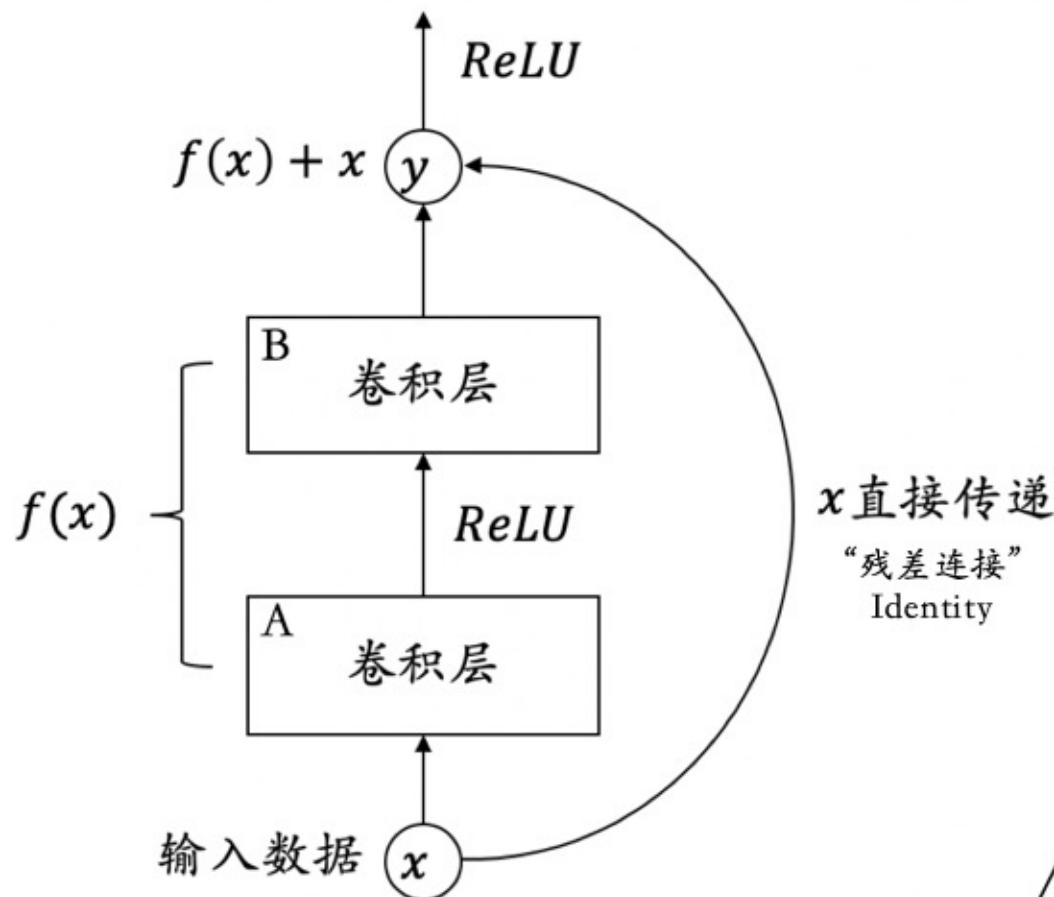
“城际直达高速公路”

残差单元



“城际直达高速公路”

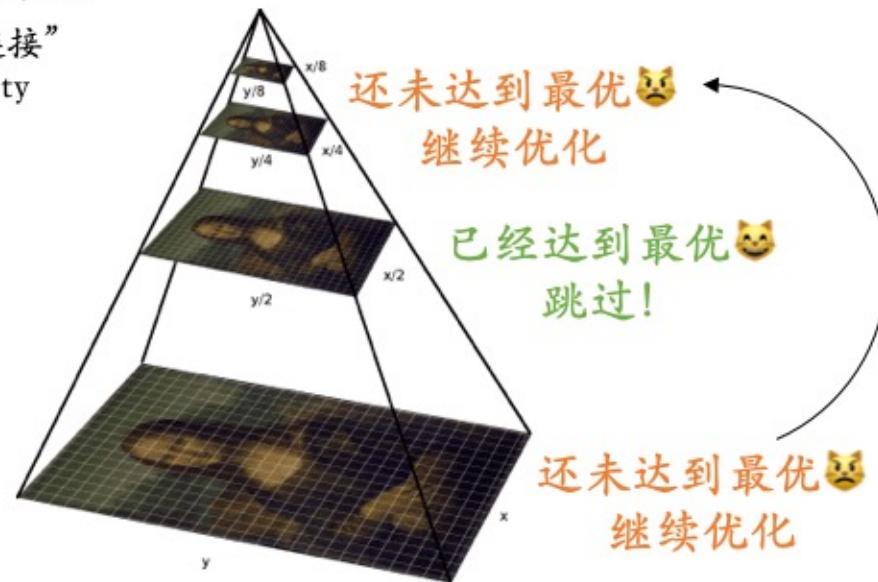
ResNet的意义



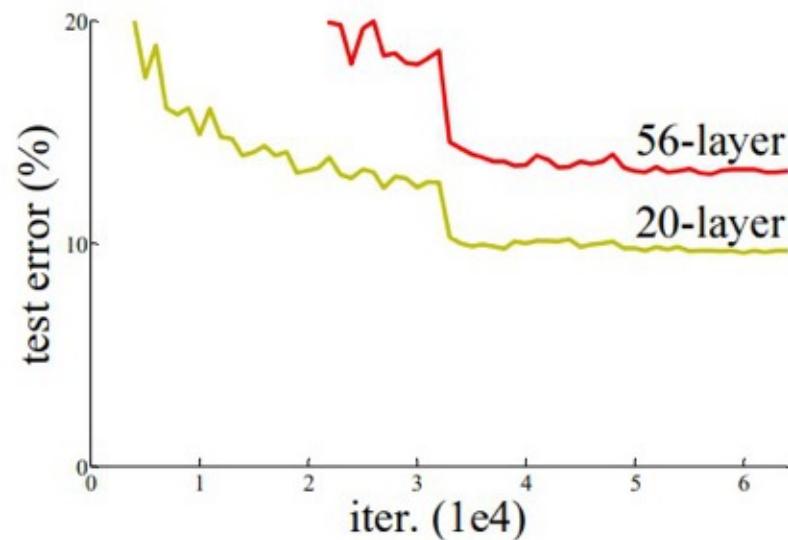
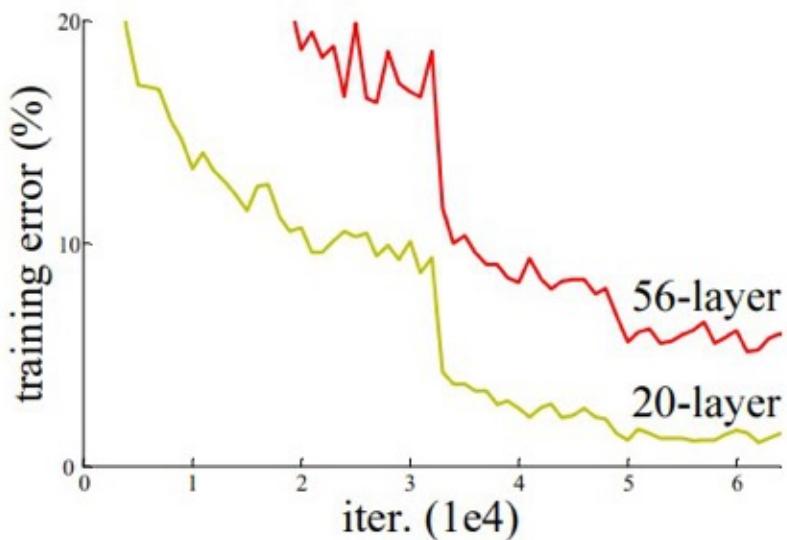
主要作用：缓解梯度消失问题。

当在图像的某一个颗粒度上，模型已经达到最优，而其他颗粒度尚未达到最优时，达到最优的颗粒度可以停止更新参数。

或者，当图像在某一个颗粒度上没有任何规律时，可以跳过该颗粒度，继续在其他颗粒度层面进行训练。

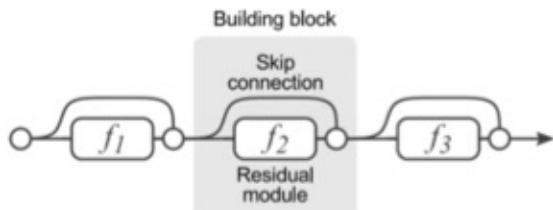


ResNet的提出：发现“退化现象”

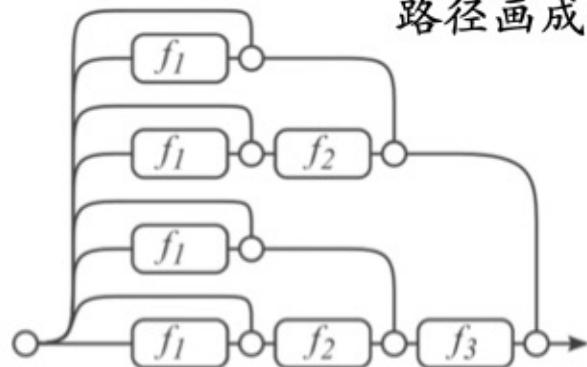


在不用残差单元的前提下，56层的网络错误率反而比20层的网络高（训练集和测试集都是如此，因此并非过拟合）。

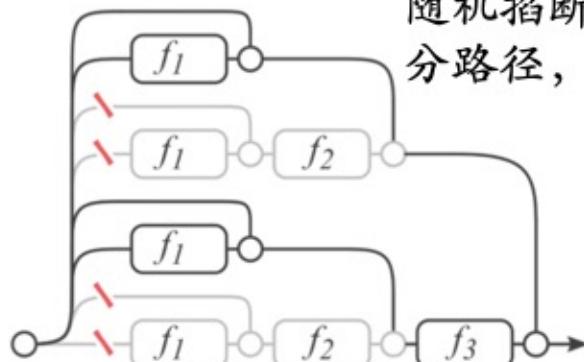
ResNet的容错率



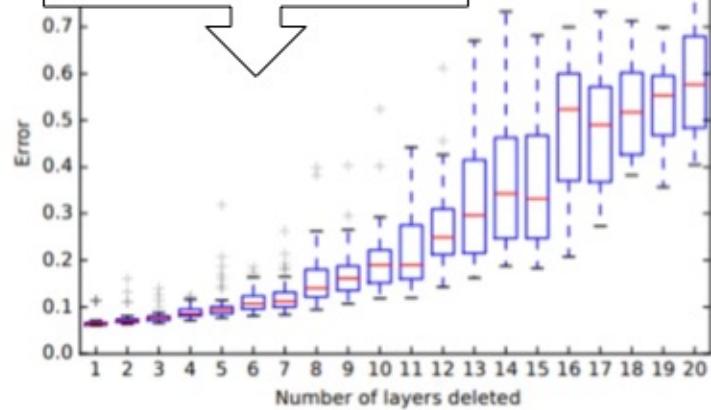
将所有可能的传播路径画成树形图



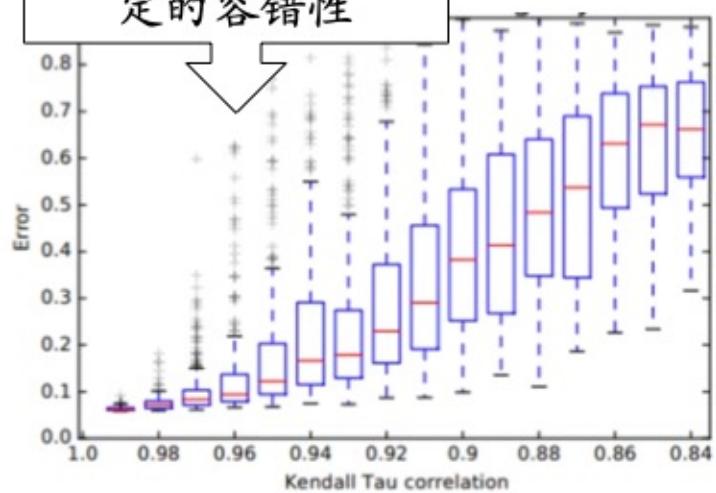
随机掐断其中的部分路径，查看性能



删除线路小于8时，
性能下降不明显

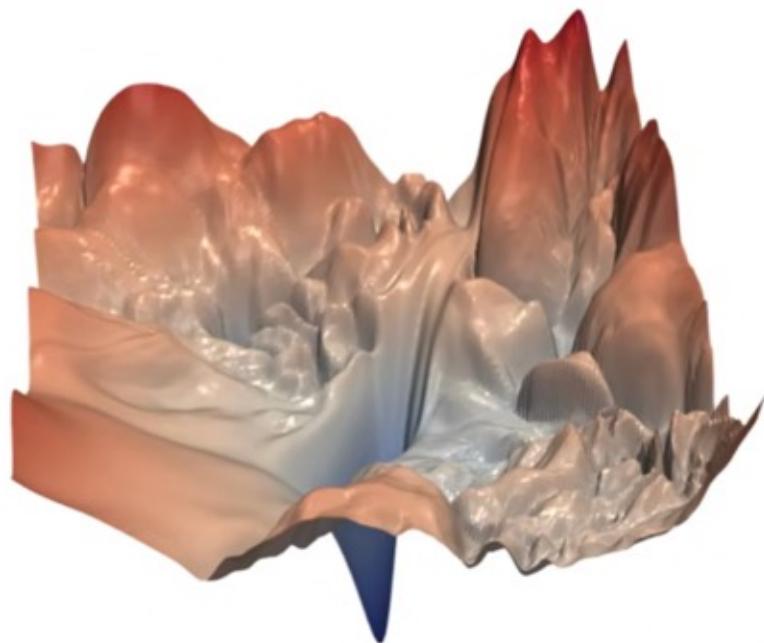


对打乱线路也有一定的容错性

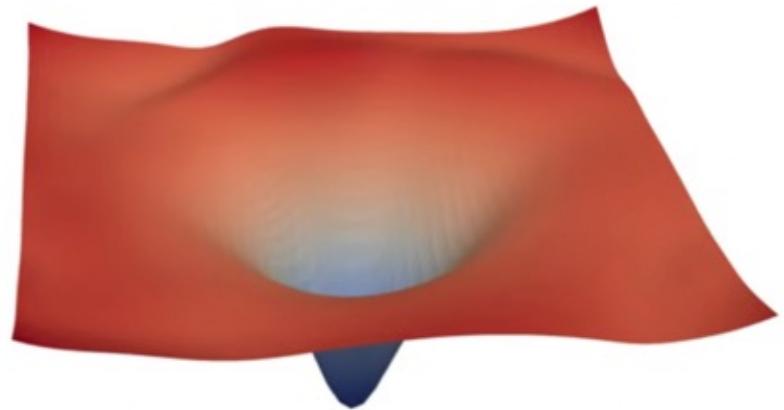


ResNet：损失函数视角

无“直连通道”的损失函数

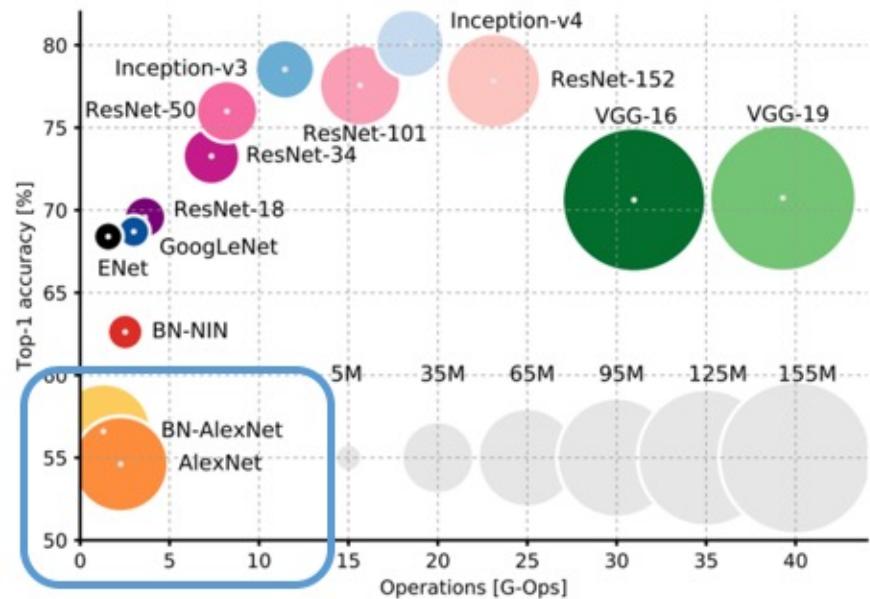
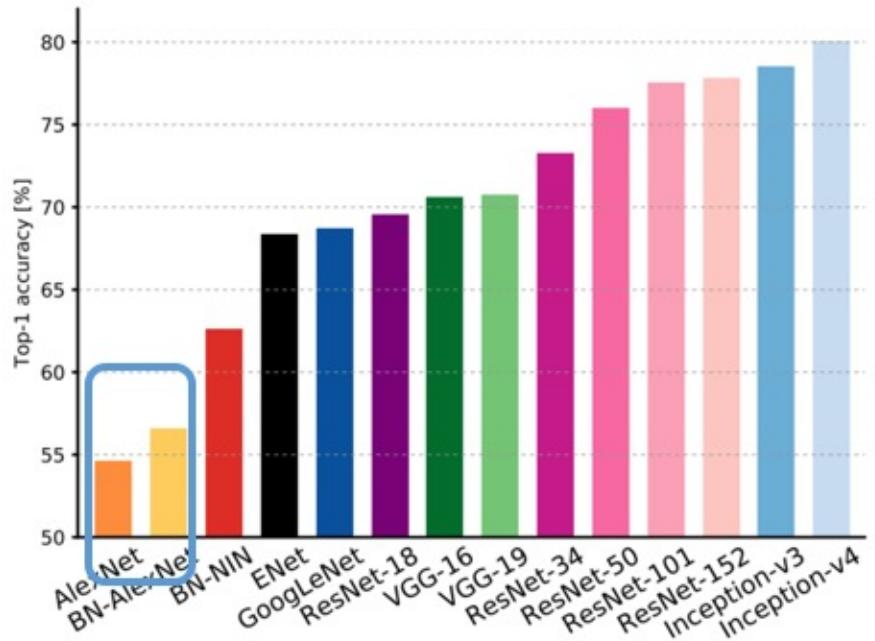


有“直连通道”的损失函数



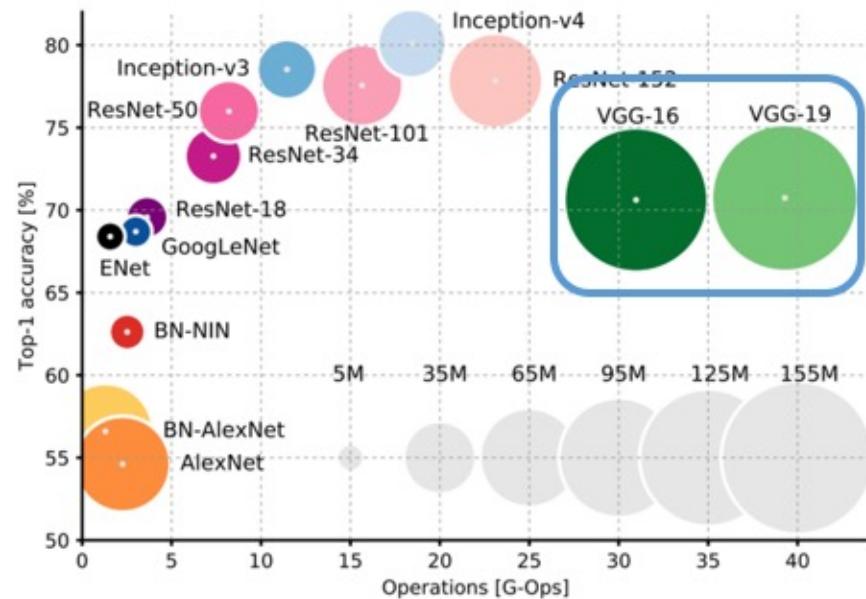
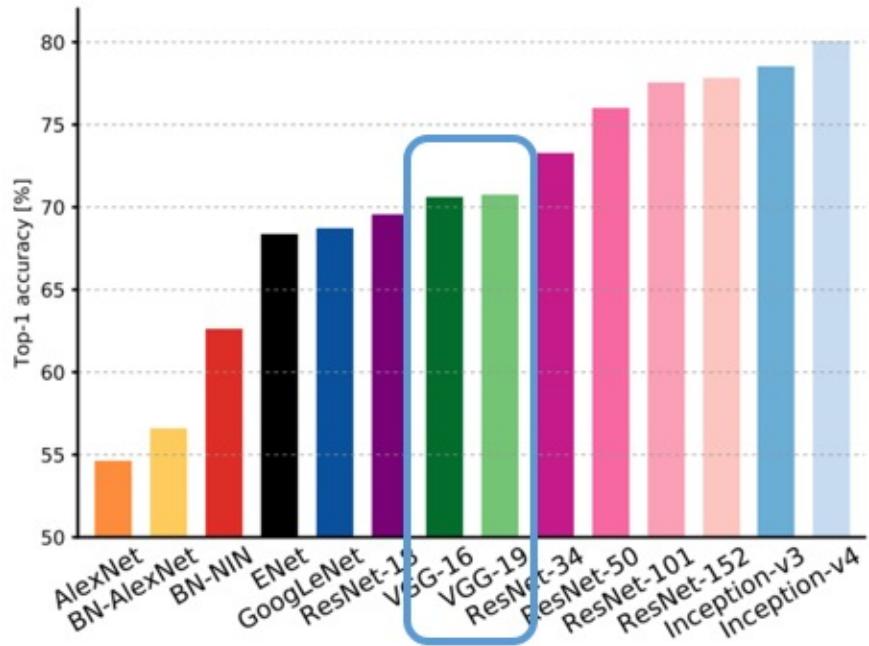
经典网络架构对比

经典网络的对比



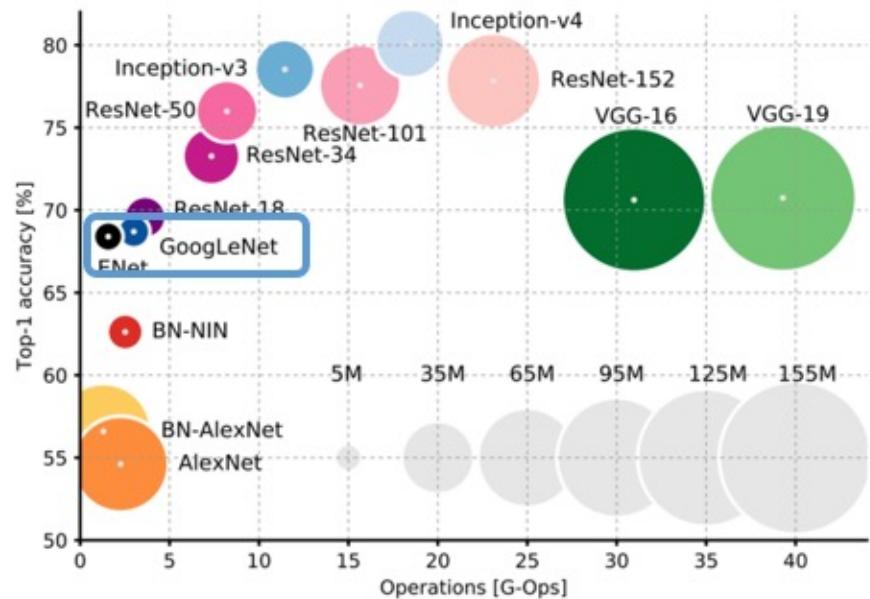
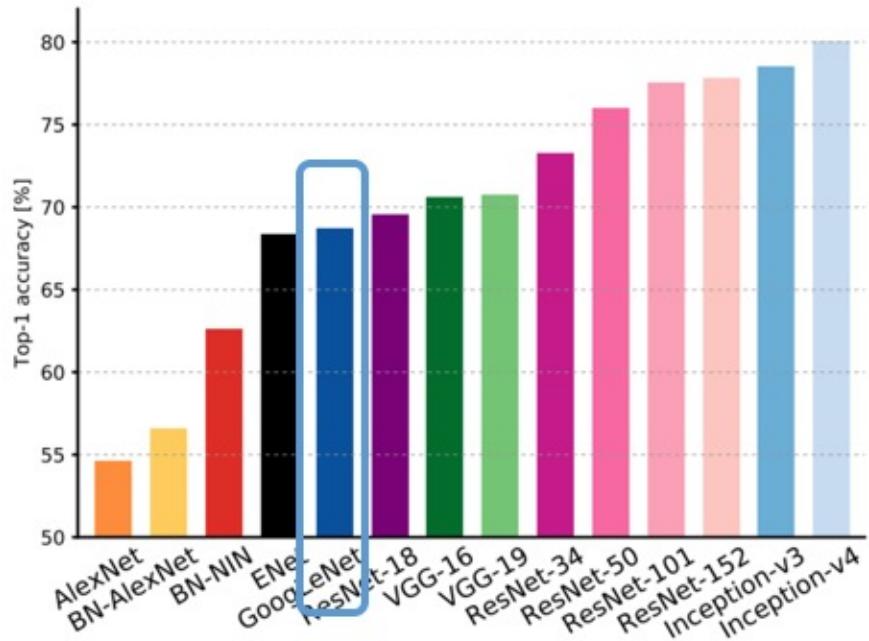
AlexNet: 运算量小，但分类性能已经落后。

经典网络的对比



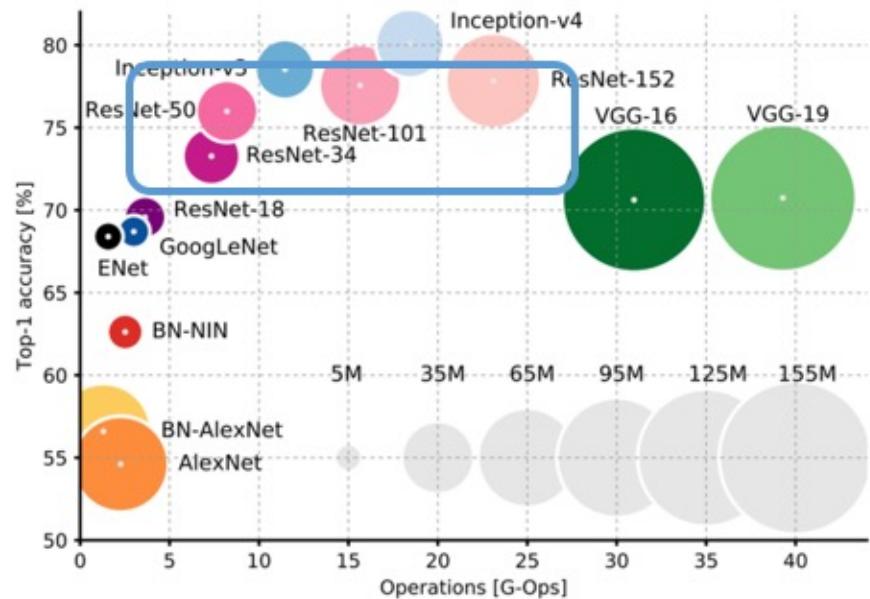
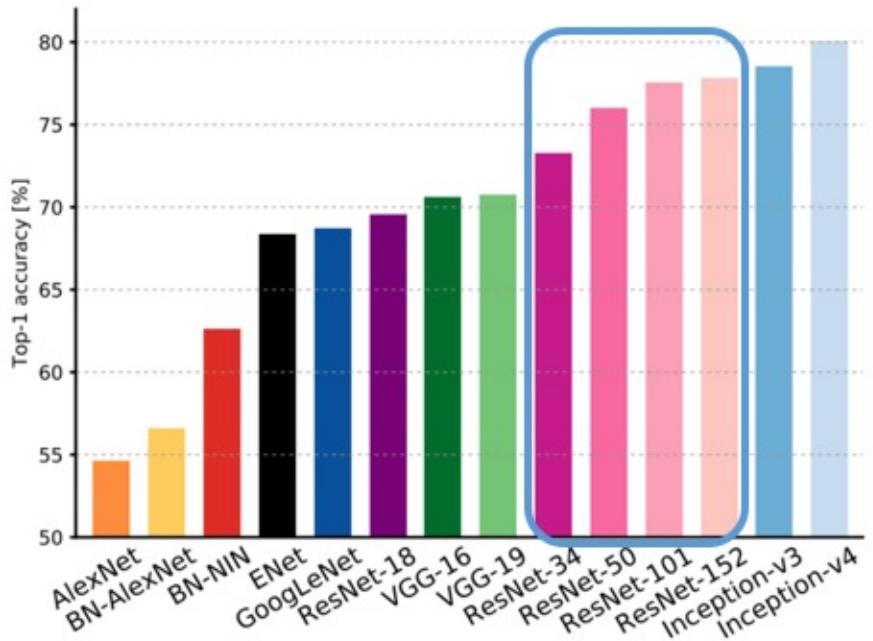
VGG: 性能较好，但运算量和内存需求大。

经典网络的对比



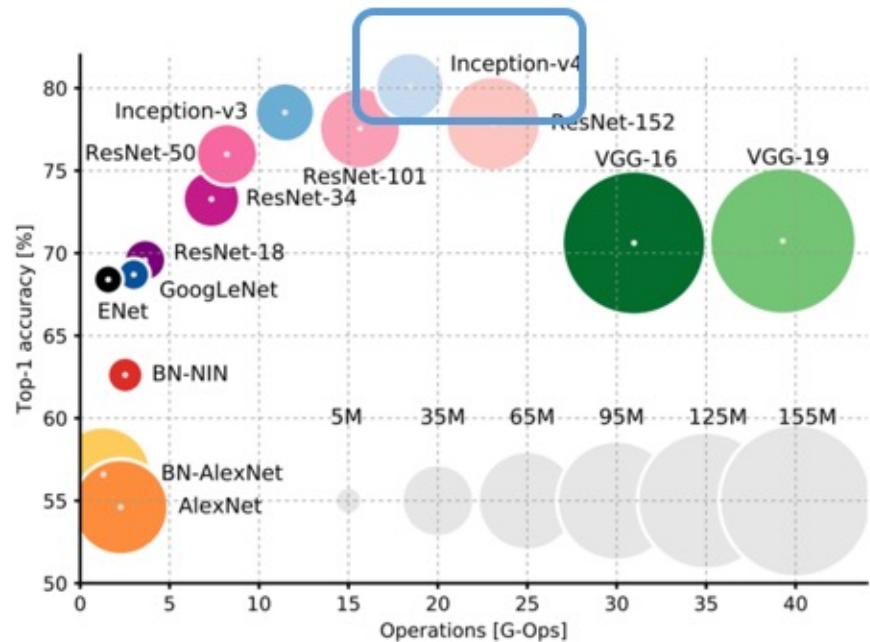
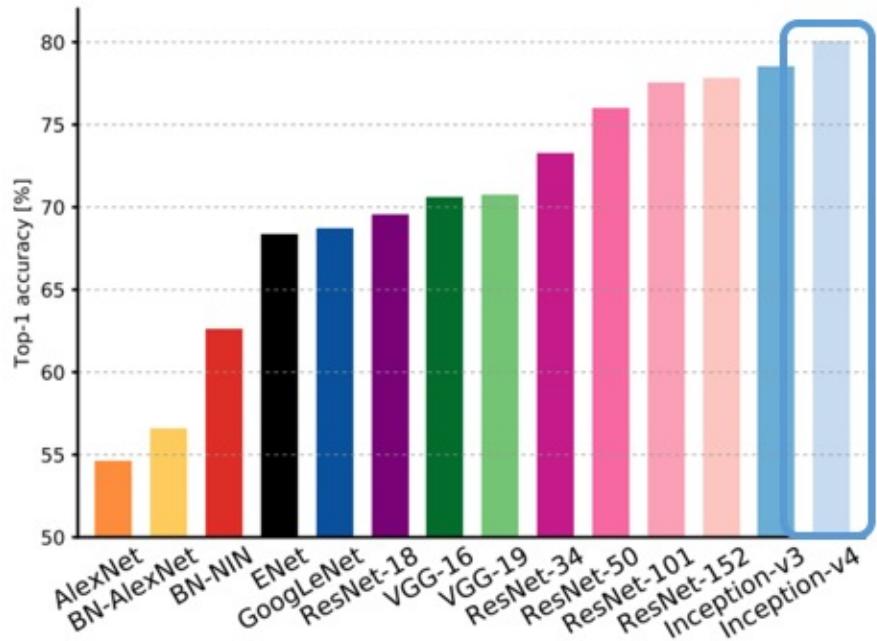
GoogLeNet (初版) : top-1正确率略弱于 VGG，但运算量和内存需求很小。

经典网络的对比



ResNet及其变种：分类性能好，计算量和内存需求为中等水平。

经典网络的对比



Inception-v4: GoogLeNet + ResNet