人工智能基础及应用

《人工智能基础及应用》第1版.
王方石，李翔宇，杨煜清，飞桨教材编写组.
北京：清华大学出版社, 2023年11月出版.
ISBN ： 9 787 302 644244

第6章 典型卷积神经网络

本章学习目标

- 掌握 LeNet-5 模型的结构及特点。
- 掌握 AlexNet 模型的结构及特点。
- 了解 VGGNet、GoogLeNet、ResNet、DenseNet的结构及特点。

6.1 LeNet

1. Lecun, Y. Generalization and network design strategies [EB/OL]. Technical Report CRG-TR-89-4, University of Toronto. (1989-06) [2023-05-06]. http://yann.lecun.com/exdb/publis/pdf/lecun-89.pdf.
2. LeCun Y, Boser B, Denker J, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1(4): 541-551.
3. LeCun Y, Boser B, Denker J, et al. Handwritten digit recognition with a back-propagation Network[C]// Advances in neural information processing systems, 1990:396-404.
4. LeCun Y, Jackel L, Bottou L, et al. Comparison of Learning Algorithms for Handwritten Digit Recognition[C]//International Conference on Artificial Neural Nervorks, 1995:53-60.
5. LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proc. of the IEEE, 1998, 86 (11): 2278-2324.

6.1.1 LeNet模型的发展历程

- 1987—1988年，杨乐昆在辛顿教授的实验室做博士后，开始研究CNN。
- 构建了Net-1、Net-2、Net-3、Net-4、Net-5五种不同结构的模型。
- 当时还没有 MNIST数据集，他用鼠标画了一些数字，用数据增强技术扩充了数据量，形成480张大小为16×16二值图像。
- 然后用这个数据集训练和测试5种模型识别手写体数字的效果，其中Net-5是局部连接且共享参数的网络，即第一代卷积神经网络，包含1个输入层、2个卷积层和1个输出层。
- 1988年10 月，杨乐昆加入AT&T贝尔实验室，那里的USPS数据集包括5000 个训练样本。他用三个月扩大了模型规模，训练了一个CNN，其效果比AT&T实验室或外部人员尝试过的所有方法都好。

6.1.1 LeNet 模型的发展历程

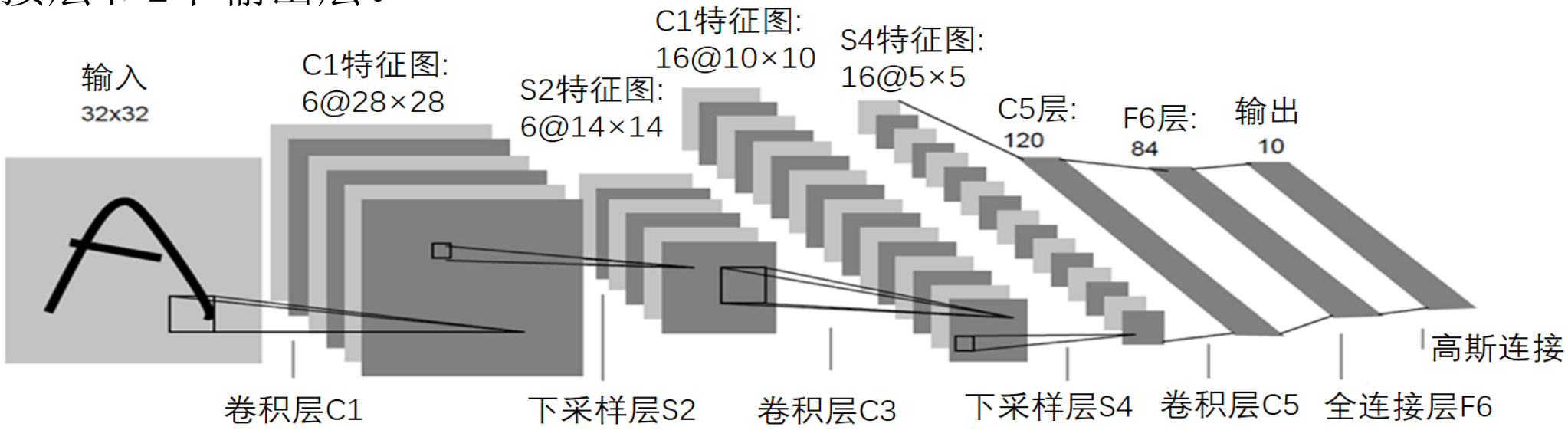
- 1989年，杨乐昆及同事合作发表了论文《反向传播在手写邮政编码中的应用》，第一次使用了“卷积”（convoluation）和“核”（kernel）的术语。
- 且明确说明：本文设计的网络是**在Net-5的基础上增加了1个全连接层，采用了带有步长移动的卷积运算**，但没有单独的池化层，每个卷积直接进行下采样。
- 这样设计的原因是因为当时的计算机无法承担每个点都有一个卷积的计算量。采用BP学习算法和随机梯度下降（stochastic gradient descent）法训练模型。**这就是第一个版本的卷积神经网络**。
- 1990年，杨乐昆等人在NIPS上又发表了《利用反向传播网络完成手写数字识别》。该文中构建的CNN包含1个输入层、2个卷积层、2个平均池化层和1个输出层。这就是**第二个版本的卷积神经网络，即LeNet-1**。

6.1.1 LeNet 模型的发展历程

- 后来，杨乐昆等人又在LeNet-1的基础上分别增加了1个全连接层和2个全连接层，形成了LeNet-4和LeNet-5。
- 当时这项技术仅在 AT&T 内部应用，几乎没有在外部使用。
- 直到1995年，被AT&T的一支产品团队将LeNet模型嵌入到能读取支票的ATM 机等设备中，用于识别银行支票上的手写体数字，随后被部署到美国的一家大型银行1。基于这个成功的商业应用案例。
- 1998年，杨乐昆等人发表了《应用于文档识别的基于梯度学习》，该文给出了著名的**LeNet-5模型**的结构图，使得LeNet-5模型广为人知，即现在引为经典的LeNet卷积神经网络。

6.1.2 LeNet-5模型的结构

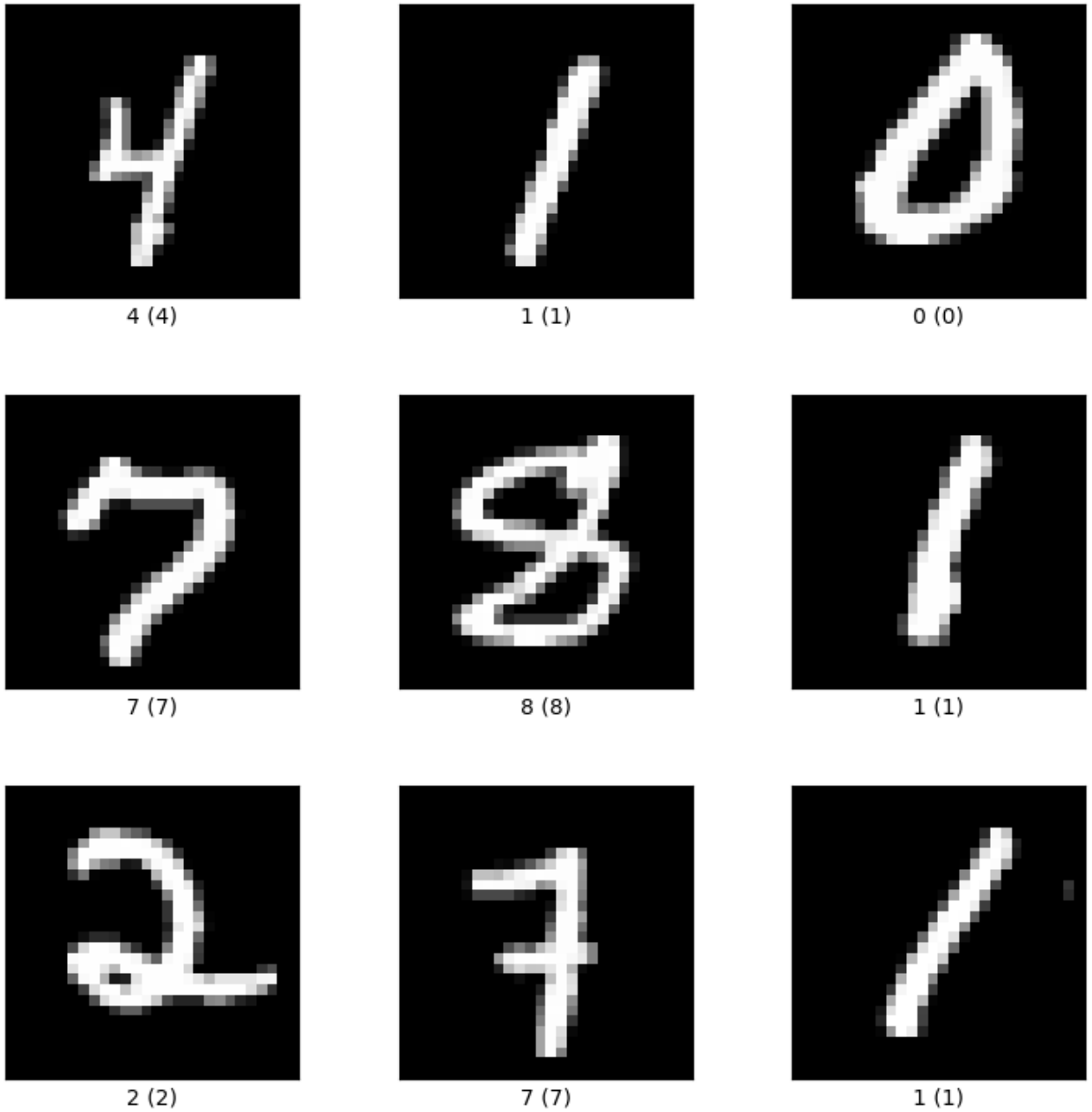
- LeNet-5采用MNIST 数据集，其中共有70000张**28×28**的灰度图像。
- 为了使图像边缘的笔画出现在卷积核感受野的中心，将原图像填充两圈零值像素，使之成为大小为 **32×32**的灰度图像。
- LeNet-5 模型的**学习目标**：从给定的32×32灰度图像中识别出手写体数字的类别。
- LeNet-5的网络结构如下图所示，包括1个输入层、2个卷积层、2个池化层、2个全连接层和1个输出层。



MNIST 数据集

- 美国国家标准技术研究所，National Institute of Standards and Technology，简称NIST
- MNIST数据集（**Mixed** National Institute of Standards and Technology database）是从NIST的两个手写数字数据集：Special Database 3 和Special Database 1中分别取出部分图像，并经过一些图像处理后得到的。
- MNIST数据集共有70000张已标注的手写体数字图像，所有图像都是**28×28**的灰度图像，**每张图像只包含一个手写数字** (0~9)，共10个类别。
- 训练集包括60000张图像，其中30000张来自NIST的Special Database 3，30000张来自NIST的Special Database 1。
- 测试集包括10000张图像，其中5000张来自NIST的Special Database 3，5000张来自NIST的Special Database 1。

MNIST数据集—示例



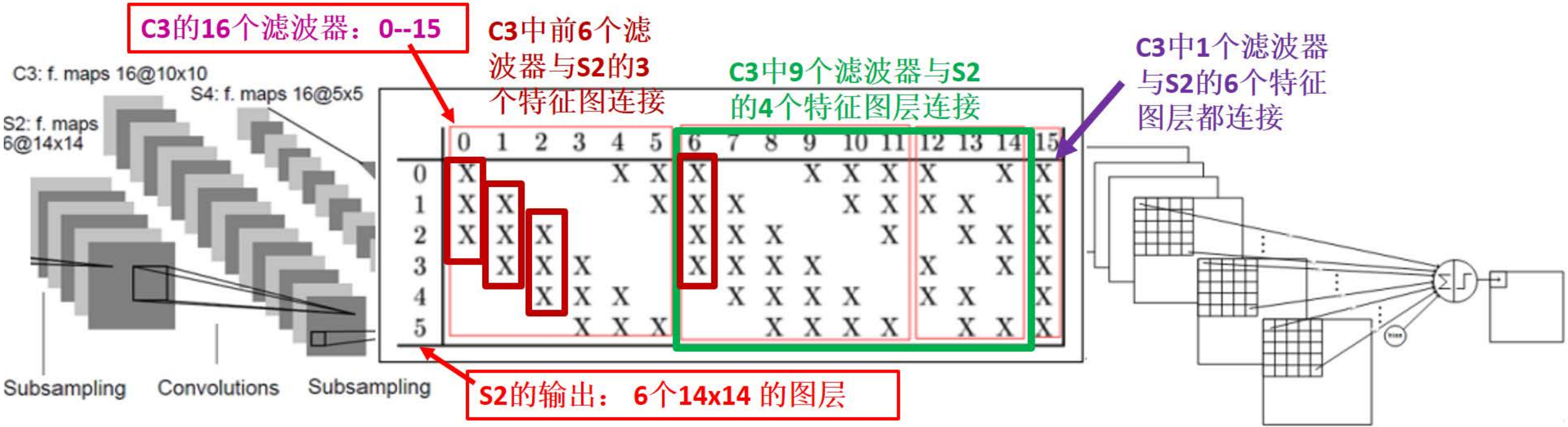
- 将**28*28** 的图像填充为**32*32**，即pad=2。
- 将32×32×1灰度图像输入到LeNet-5，识别 0-9。

LeNet-5的结构

- **C1卷积层：** 输入32×32×1的灰度图
  - 采用**6个5×5卷积核**（**深度为1的滤波器**），步长为1
  - **输出：** 28×28×6【(32-5)/1+1=28】。
  - **参数量：（5×5+1偏置）*6=156**
  - **连接数：**(5×5+1) ×28×28×6=**122304**
- **S2池化层：** 输入28×28×6的特征图
  - 采用大小为2×2、步长为2的不重叠的滑动窗口做平均池化
  - **输出：**14×14×6【(28-2)/2+1=14】。
  - **参数量：** (1+1偏置)×6=**12个参数（平均池化）**
  - **连接数：**(2×2+1个偏置) ×14×14×6=**5880**

◆ **C3卷积层：** 输入 14×14×6（即S2的输出）；**输出：10×10×16** 【14-5+1=10】

- C3层采用**16个5*5的滤波器**，C3的每个滤波器与S2中的多个（不是所有）特征图相连，即**S2与C3不是全连接**。
- **这16个滤波器的深度各不相同，有6个深度为3，9个深度为4，1个深度为6，**连接的方式**如下表** 所示。
- **参数量：** C3层该层有(5x5x3+1)x6 + (5x5x4 + 1) x 9 + (5x5x6+1)x1 = **1516个训练参数**
- **连接数**：共有1516x10x10=**151600**



◆ **S4池化层：** 输入**10x10x16**特征图

- 采用大小为2x2、步长为2的不重叠的滑动窗口做平均池化
- **输出5x5x16**。
- **参数量**：(1+1)*16=**32**
- **连接数：** (2x2+1个偏置)x5x5x16=**2000**。连接的方式与S2层类似，如下图所示。

◆**C5卷积操层，也是全连接层（但不写作F5）：** 输入5 × 5 × 16

- 采用120个 5 × 5 × 16的滤波器；
- **输出**：1 × 1 × 120，即120个分量的1维向量。
- 每个滤波器都与S4层的16个图层全部相连，所以是全连接。
- **参数量：** (5 × 5 ×16+1) x120 = **48120**
- **连接数：同参数量，即48120。**

◆**F6全连接层：输入**1 × 1 × 120。

- 包含 84个 1 × 1 × 120的神经元
- **输出**： 1 × 1 × 84，形成一个7x12的比特图（可识别ASCII集中的字符）
- 该层的**参数量**和**连接数**都是**(120 + 1)x84=10164**.

◆**输出层**（全连接层）：**输入** 1 × 1 × 84

- 共有10个节点，即10个(1×1×84)的滤波器，分别代表数字0到9，
- 若节点 i 的输出值为0，则所输入的字符被网络识别为数字i。
- **采用径向基函数（RBF）作为激活函数**。

$$y_i = \sum_{j=0}^{83} \left(x_j - w_{ij}\right)^2 \quad , \quad i = 0..9$$

此值不是概率，选择最小值所对应的类别

其中x_j是F6层的输出，y_i是激活函数的输出，参数w_{ij}是F6层和输出层之间的权重。

- **参数量**和**连接数**均为 (84+1)×10=850

◆可见，y_i越接近于0，则表明y_i的输入越接近数字i的比特图编码{w_{ij}，j=0…83}，当前输入的字符应该被识别为数字i。

◆在现在的深度学习框架中，通常采用 softmax函数取代 RBF函数，用以输出分类的概率。

◆LeNet-5模型中**共有60,850个训练参数，340,918个连接**。

LeNet-5模型中的参数量和连接数

- C1层参数个数：(5×5+1)×6=**156**，有(5x5+1)x28x28x6=**122304个连接**
- S2层的可训练参数是每一个池化都有一个权重w和一个偏置b，总共6个卷积核，故有 (1+1)×6=**12个参数**， **5880个连接**
- C3层参数个数: **1516**， 1516x10x10=**151600个连接**
- S4层总共16个池化卷积核，故有 (1+1)×16=**32个参数**，**2000个连接**
- C5层参数个数: (5×5×16+1)×120 =**48120**， **48120个连接**
- F6层参数个数:(120+1)×84=**10164**，**10164个连接**
- F7层参数个数:（84+1)×10=**850**，**850个连接**

LeNet模型中共计**60850个参数**，**340,918个连接**。

6.2 AlexNet

- **2011年，ILSVRC比赛中图像分类的最好成绩：top-5错误率为25.8%。**
- **2012 年AlexNet夺得图像分类任务的冠军，top-5错误率为15.3%，比同年第二名的26.2%低了近10个百分点**
- Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systens, 2012: 1097-1105.

ImageNet数据集

◆ImageNet是一个用于视觉对象识别研究的大型图像数据库，是由斯坦福大学李飞飞教授带领其研究团队于2007年起开始构建的。

◆他们从互联网上下载图片，手工分类、注释了超过1400万张图像，并且在至少一百万张图像中提供了对象的边界框。

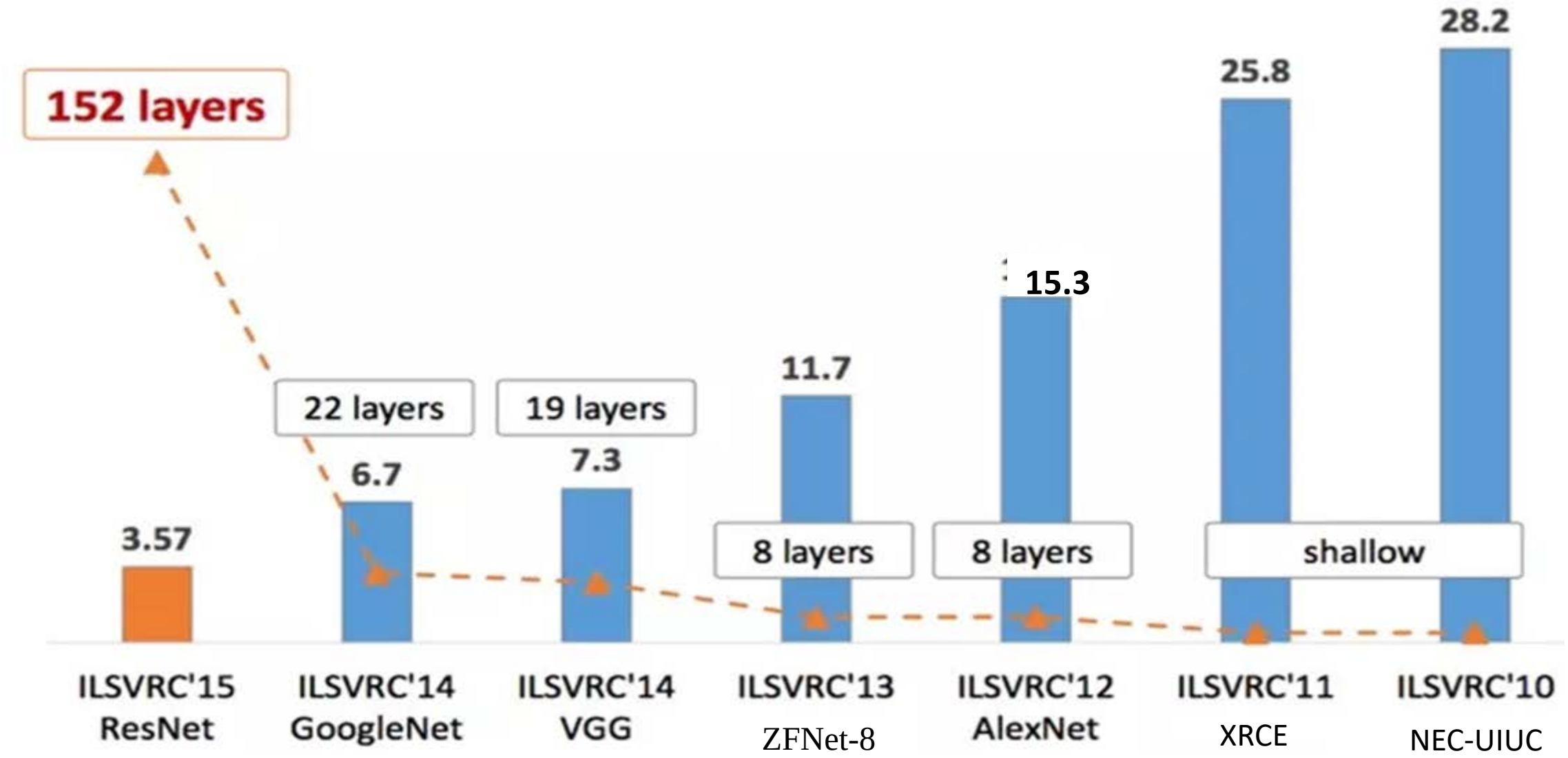
◆其中包括大约**22000个类别**，如“气球”、“草莓”等。

◆**ILSVRC**是国际上视觉领域最具权威的学术竞赛之一，代表了图像领域的最高水平，始于2010年，2017年举办了最后一届。

◆**图像分类比赛**使用ImageNet数据集的一个**子集**，总共包括**1000类图像**。

◆ImageNet 图像分类比赛以**top-5错误率**作为评价指标，即为每幅图像预测5个标签类别，只要其中一个标签与人工标注的类别相同，则视为预测正确，否则视为预测错误。

ImageNet 图像分类大赛历年的top-5错误率

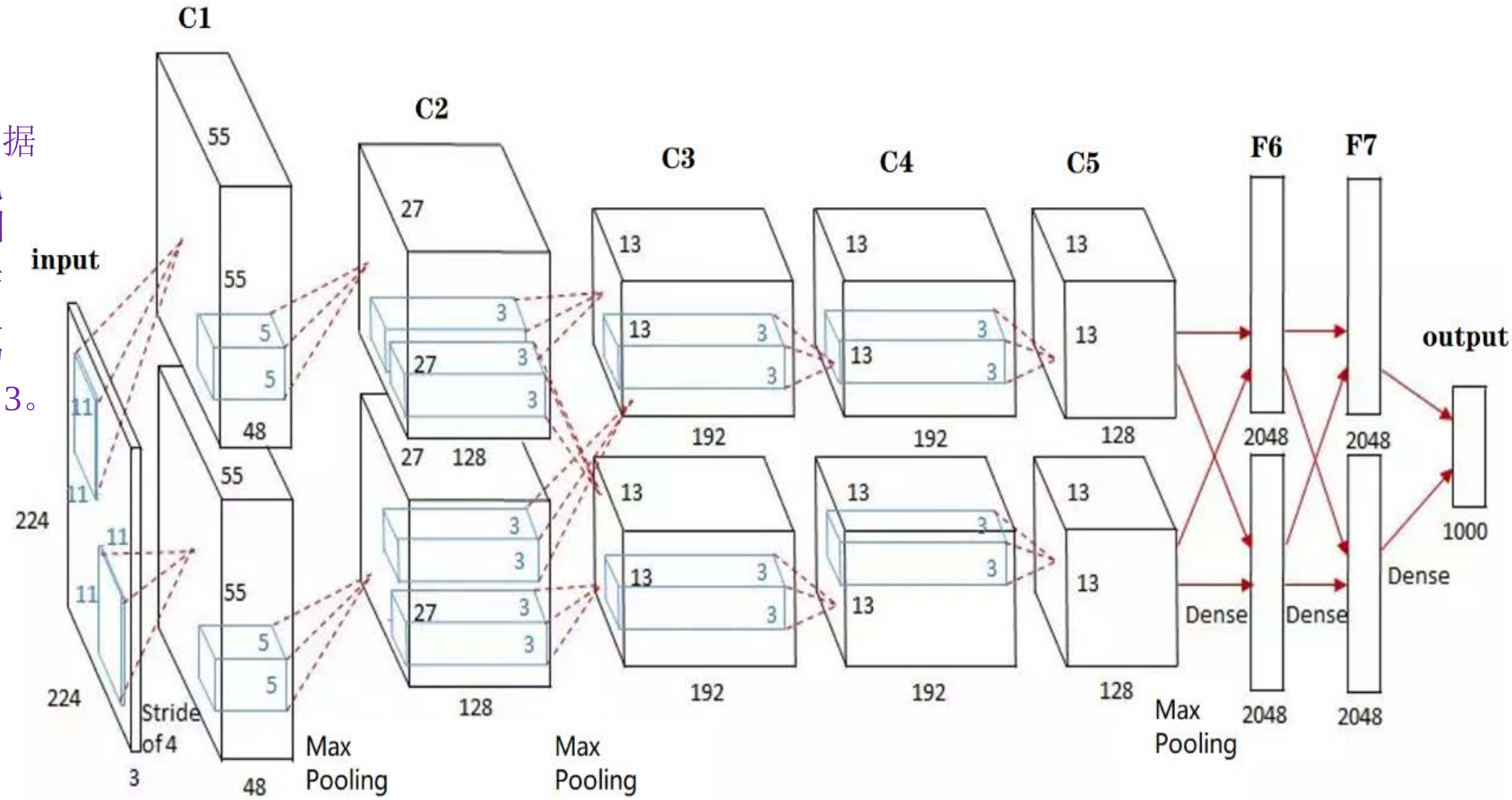


6.2.1 AlexNet 模型的结构

- AlexNet与LeNet-5在网络结构设计上的差别并不大，
- 但GPU的出现和ImageNet庞大数据量的助力，促使AlexNet表现出了卓越的性能。
- AlexNet是一个8层的深度神经网络。
- 池化层和局部响应归一化层（Local Response Normalization，LRN）不计入层数。
- 前5层为卷积层，后3层为全连接层，最后一个全连接层也是输出层。
- 分别在第一、二、五层这3个卷积层后面增加了最大池化层。
- 受当时 GPU 算力的限制，整个网络模型被一分为二，分别在两块显存为3GB的NVIDIA GTX580 GPU 上实现了快速卷积运算，两个GPU只在某些特定的网络层（C3、F6和F7）上通信。

6.2.1 AlexNet 模型的结构

输入是ImageNet数据集中归一化后的RGB图像样本，每张图像的尺寸被裁切为227×227×3。



AlexNet 模型的结构（1）

◆ **C1层：** 输入224×224×3的彩色图

- 96个11×11×3、s=4的滤波器**分为2组**，无填充（即**p=0**）
- 分别在两个独立的GPU上进行卷积运算，**输出2组**：55×55×48【(**227**-11)/4+1=55】
- 然后，采用ReLU作为激活函数，又进行局部响应归一化操作，输出2组：55×55×48。

◆ 执行**重叠的最大池化**(3×3,s=2)操作，得到**2组：27×27×48**

◆ **C2层：** 输入2组 27×27×48的特征图

- 256个5×5×48、s=1的滤波器**分为2组**，有填充（**p=2**）
- 分别在两个独立的GPU上进行卷积运算，**输出2组**： 27×27×128
- 然后，采用ReLU作为激活函数，又进行局部响应归一化操作，输出2组：27×27×128。

◆ 执行**重叠的最大池化**(3×3,s=2)操作，得到**2组：13×13×128**

AlexNet 模型的结构（2）

◆ **C3层：** 输入2组13×13×192的特征图

- C2层输出的两组13×13×128特征图（共计256层特征）共同参与每个GPU上的卷积运算，故C3层上滤波器的深度为256.
- 384个3×3×256 、s=1的滤波器**分为2组**，有填充（**p=1**）
- 分别在两个独立的GPU上进行卷积运算，**输出2组**：13×13×192
- 然后，采用ReLU作为激活函数，输出2组： 13×13×192

◆ **C4层：** 输入2组 13×13×192的特征图

- 384个3×3×192、s=1的滤波器**分为2组**，有填充（**p=1**）
- 分别在两个独立的GPU上进行卷积运算，**输出2组**： 13×13×192
- 然后，采用ReLU作为激活函数，输出2组：13×13×192

AlexNet 模型的结构（3）

- **C5层：** 输入2组13×13×192的特征图
  - 256个3×3×192 、s=1的滤波器**分为2组**，有填充（即**p=1**）
  - 分别在两个独立的GPU上进行卷积运算，**输出2组**：13×13×128
  - 然后，采用ReLU作为激活函数，输出2组： 13×13×128
- 执行**重叠的最大池化**(3×3,s=2)操作，得到**2组：6×6×128**

- **F6是第1个全连接层：** 输入2组 6×6×128的特征图
  - C5层输出2组6×6×128的特征图，共同参与F6层每个GPU上的全连接操作，故F6层上滤波器的深度为256。
  - 4096个6×6×256、s=1的滤波器**分为2组**，无填充
  - 分别在两个独立的GPU上进行卷积运算，**输出2组**： 1×1×2048
  - 然后，采用ReLU作为激活函数，输出2组：1×1×2048，共计4096个神经元。
  - 再以0.5的概率对这4096个神经元采用**随机失活（dropout）技术**，目的是使某些神经元不参与训练，以避免发生过拟合。

AlexNet 模型的结构（4）

◆ **F7是第2个全连接层：** 输入2组 1×1×2048 的特征图

- F7层包含4096个1×1神经元
- F6层的4096个神经元与F7层的4096个神经元进行全连接。
- F7层采用ReLU作为激活函数，仍以0.5的概率对F7层的4096个神经元采用dropout技术，输出4096个数值。

◆ **输出层是第3个全连接层：** 输入4096个数值

- 包含1000个神经元，分别对应于1000个图像类别。
- F7层的4096个神经元与输出层的1000个神经元进行全连接。
- 输出层采用softmax函数作为激活函数，输出1000个在[0,1]的数值，分别表示属于所对应类别的概率。输入图像被归入最大概率值所对应的类别。

AlexNet模型中共计60 965 128个参数，约是LeNet-5模型参数量（60 850个）的1000倍。

6.2.2 AlexNet 模型的创新性

(1) 采用ReLU 函数作激活函数，可提高网络的收敛速度。

(2) 采用重叠池化，可提高精度，防止过拟合。

(3) 训练网络时采用dropout技术，用以减少过拟合。

(4) 采用LRN技术，可防止过拟合，增强模型的泛化能力。

(5) 在两个GPU上同时训练，提高训练速度。

(6) 利用数据增强，扩充数据集，以减少过拟合，提升泛化能力。

另外，AlexNet的成功还**首次证明了机器学习的特征可以取代手工设计的特征**，研究者们无需花费大量精力和时间去设计各种特征。

AlexNet标志着从浅层网络跨越到深层网络的里程碑。

3. VGG

牛津大学的视觉几何组

（Visual Geometry Group，VGG）

◆2014 年ILSVRC竞赛图像分类任务的亚军，Top5 错误率为 7.32%

◆2014 年ILSVRC竞赛目标定位任务的冠军

6.3.1 VGG 模型的结构

- VGGNet模型在LeNet-5和AlexNet模型结构的基础上引入了“**模块化**”的设计思想：
  - 将若干个相同的网络层组合成一个模块，
  - 再用模块组装成完整的网络，
  - 而**不再是以“层”为单元组装网络**。
- VGGNet模型的研究人员给出了5 种不同的VGGNet配置，如表6.1所示，分别用A~E来表示。

VGG的五种配置

- 模型E就是著名的VGG-19网络。
- 模型D则是广为人知的VGG-16。
- 这里的层数是指卷积层与全连接层的层数之和，不包括池化层。

模型编号	A	A-LRN	B	C	D	E
层数	11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
	input (224 × 224 RGB image)					
卷积模块1	conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
池化层1	maxpool					
卷积模块2	conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
池化层2	maxpool					
卷积模块3	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
池化层3	maxpool					
卷积模块4	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
池化层4	maxpool					
卷积模块5	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
池化层5	maxpool					
全连接模块	FC-4096					
	FC-4096					
	FC-1000					
	soft-max					

五种VGGNet模型配置的共同点

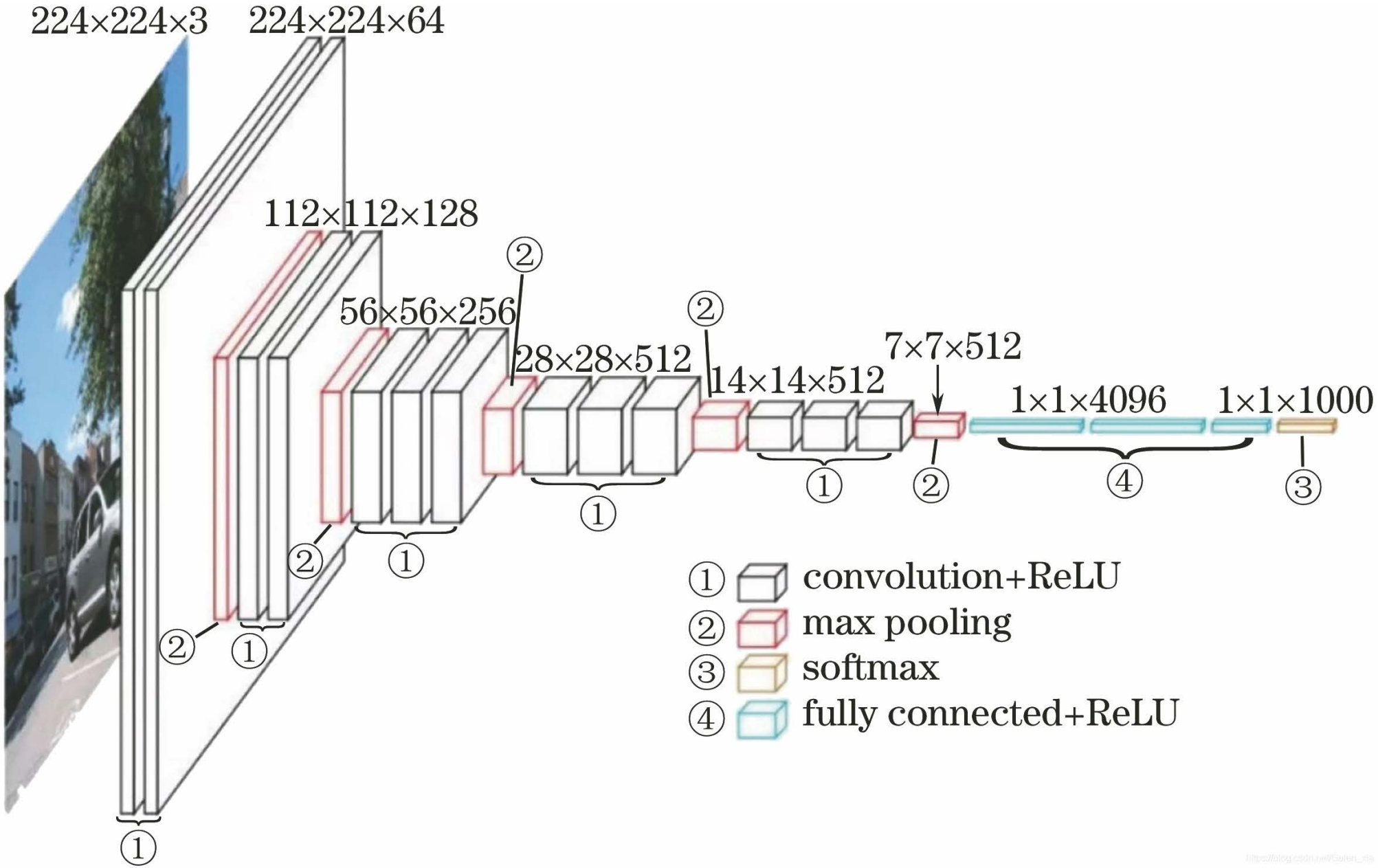
1. 输入ImageNet数据集中的RGB图像样本，每张图像的尺寸被裁剪为 224×224，并对图像进行了**零均值化**的预处理操作，即图像中每个像素均减去所有像素的均值。
2. 由5个卷积模块和1个全连接模块组成，每个卷积模块由1~4个卷积层构成，全连接模块由3个全连接层构成。
3. 几乎所有卷积核都是大小为3×3、步长为1，只有模型C用了3个1×1的卷积核，其目的是为了增加非线性表达能力和减少模型的参数量。做卷积运算时，采用“相同填充”（即padding=same）方式，保证卷积运算前、后的特征图的大小相同，即输入特征图与输出特征图的尺寸相同。
4. 在每个卷积层后面都采用 ReLU作为激活函数。

五种VGGNet模型配置的共同点

5. 每个卷积模块的最后一层之后都会有一个最大池化层，用以缩小特征图的尺寸。池化层均采用大小为2×2、步长为2的不重叠方式，使得特征图的宽和高是原来的一半。
6. 特征图的尺寸在卷积模块内不是变的，但每经过一次池化，特征图的高度和宽度减少一半，为了弥补特征量的减少，其通道数增加一倍，分别为64、128、512。
7. 全连接模块的前两层均包含4096个神经元，使用ReLU作激活函数，且使用 Dropout技术，用以防止过拟合；第三个全连接层是输出层，包含1000个神经元，采用Softmax作为激活函数，输出1000个[0,1]区间的概率值，分别对应于1000个图像类别。

- VGG-19网络的19层是指卷积层与全连接层的层数之和，不包括池化层。
- 模型D则是广为人知的VGG-16。
- VGG-16 和 VGG-19 并没有本质上的区别，只是网络深度不同，前者有 16 层（13 层卷积、3 层全连接），后者有 19 层（16 层卷积、3 层全连接）。

VGG-16


224×224×3
224×224×64
112×112×128
56×56×256
28×28×512
14×14×512
7×7×512
1×1×4096
1×1×1000
① convolution+ReLU
② max pooling
③ softmax
④ fully connected+ReLU

6.3.2 VGGNet模型的优势

- 人们的直观感受是：卷积核的感受野越大，看到的图像信息就越多，获得的特征就越丰富，则模型的效果就越好，但参数量和计算量也越大。
- 如何平衡卷积核感受野大小与参数量/计算量大小之间的关系呢？
- 值得庆幸的是：VGGNet研究人员发现：
  - 两个级联的3×3卷积核的感受野相当于一个5×5卷积核的感受野，
  - 三个级联的3×3卷积核的感受野相当于一个7×7卷积核的感受野。

选择3×3卷积核的两个好处

VGGNet选择了3×3的卷积核，这样做有以下两个好处。

（1）若干个小尺寸卷积核的参数量要远远小于一个相同感受野的大尺寸卷积核的参数量。

例如，为方便计算，假设每个卷积层的输入特征图与输出特征图的通道数相同，均为C，则三个级联的3×3卷积核的参数量为(3×3×C×C)×3个，而一个7×7卷积核中的参数量为(7×7×C×C)个，显然，前者的参数量只是后者的55.6%。

（2）每个卷积层后面都有一个非线性激活层，三个连续的卷积层就有三个非线性激活函数，可增加网络的非线性表达能力，提高网络的识别能力。

这也正是VGGNet引入模块化设计思想的初衷，从此之后，3×3卷积核被广泛应用在各种CNN模型中。

VGG的研究结论

VGGNet的研究工作发现：

- 用多个尺寸较小的卷积核代替一个大尺寸卷积核，既可保证相同的感受野，又可减少参数量；
- 证明了在大规模图像识别任务中，增加卷积神经网络的深度可有效提升模型的精确度。
- 实验数据表明：LRN层的作用不大，所以在B~E型网络结构中不再使用。

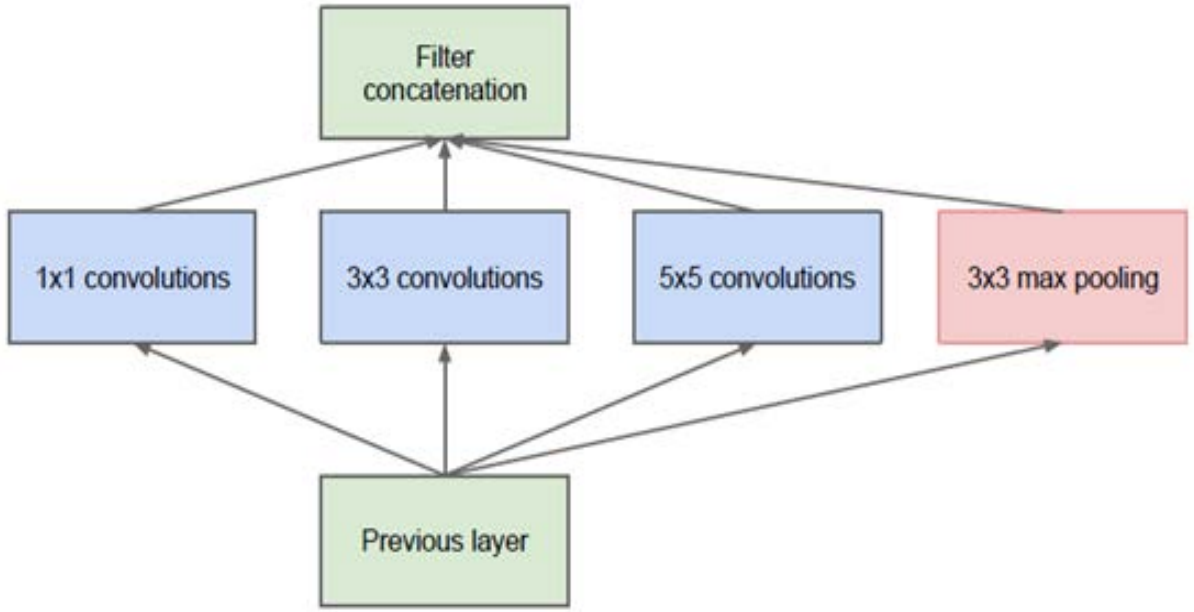
VGGNet模型的结构简单，泛化能力强，因而受到研究人员的青睐，并被广泛使用，至今仍经常被用于图像的特征提取。

6.4 Inception/GoogLeNet

- 谷歌研究团队提出的
- 为了向发明LeNet网络的杨乐昆教授致敬，参赛团队及Inception架构均命名为GoogLeNet。
- GoogLeNet是2014年ImageNet图像分类任务的冠军
- 图像分类的top-5错误率仅为6.67%。

6.4.1 GoogLeNet模型的研究思路

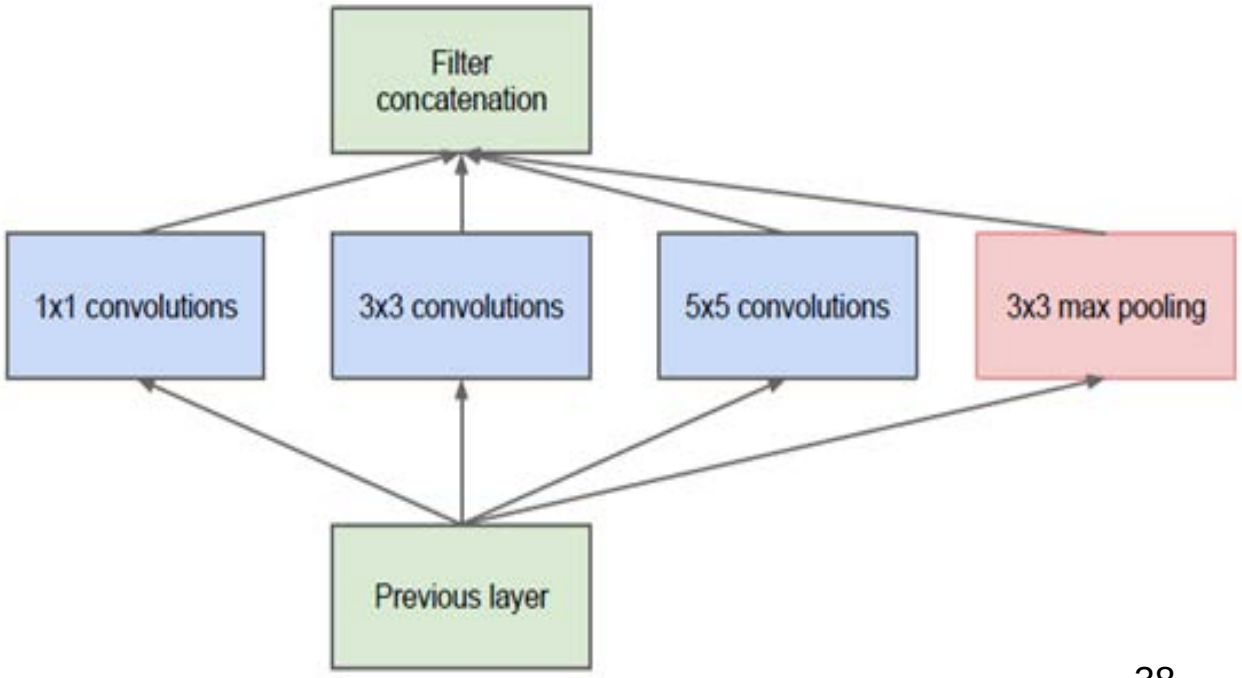
- GoogLeNet没有继承LeNet模型或AlexNet模型的框架结构，而是做了创新性的尝试。
- Inception是Google研究团队首创设计的一种网络模块，用来代替之前的“卷积+激活函数”的经典组件。
- 最初设计的Inception模块称为**Inception初级模块**（Inception Module，Naive Version）。
- Inception初级模型是基于**赫布理论**设计的一种具有优良局部拓扑结构的网络，并结合了多尺度处理的思路。



6.4.1 GoogLeNet模型的研究思路

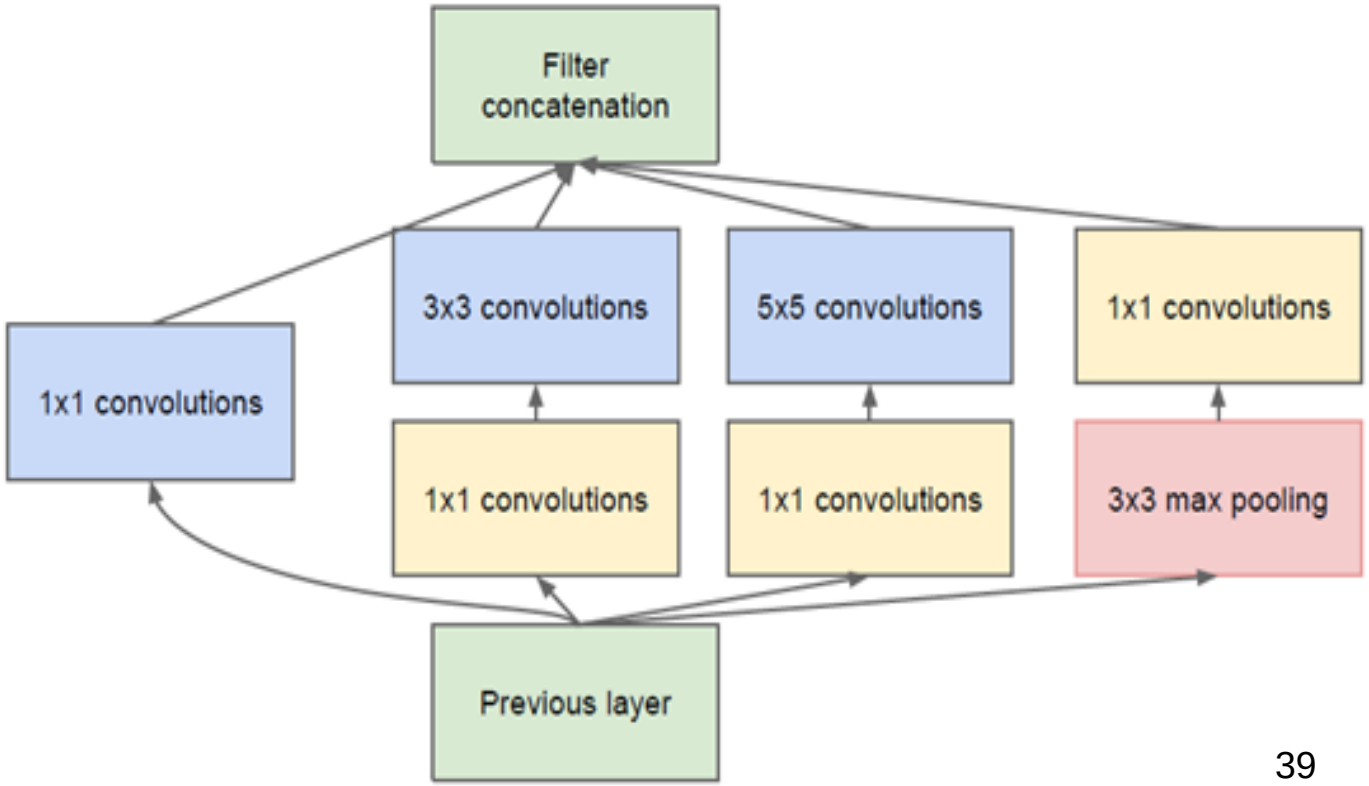
◆研究人员设计了**多条并行的、有高度相关性的分支**，分别是3个不同大小（1×1、3×3、5×5）的卷积运算和1个最大池化操作，用于模拟若干个不同的、关联性很强的神经元，对上一层的输出进行特征提取；然后将所有分支的运算结果**拼接**（concatenate）起来，作为下一层的输入。

◆这种高度关联的神经元的集成就形成了Inception初级模块，它**不仅增加了网络的宽度，还可提取不同尺度的特征**，增加了网络对多尺度的适应性。



6.4.1 GoogLeNet模型的研究思路

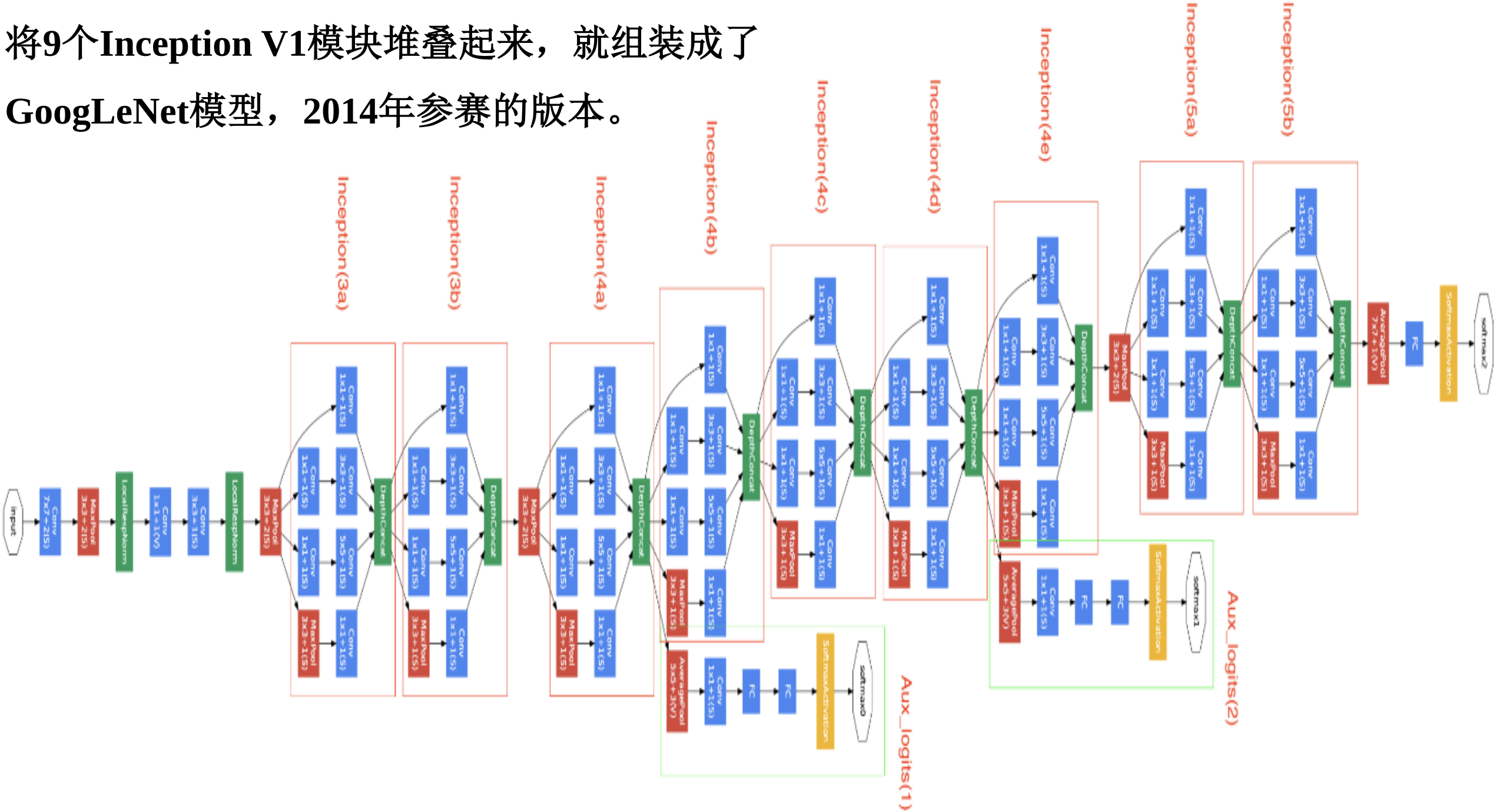
- 一个Inception初级模块的滤波器参数量是其所有分支上参数量的总和。
- 模块中包含的层数越多，则模型的参数量就会越大。
- 为了减少算力成本，在Inception初级模型内置的3×3和5×5卷积层之前，分别增加了1×1卷积层，称为**降维层**。
- 在最大池化层之后也增加了1×1卷积层，称为**投影层**。
- 如此改进后的模块称为**Inception V1模块**。



6.4.1 GoogLeNet模型的研究思路

- 在Inception V1模块中**增加1×1卷积层**，有**两个好处**：
  - 一是可以减少输入特征图的通道数，即减少卷积运算量，以便降低计算成本；
  - 二是既能增加网络深度，又能增加一层跨通道的特征变换和一次非线性函数，提取不同尺度的特征，以提高网络的表达能力。
- 将多个Inception V1模块堆叠起来，就组装成了**GoogLeNet模型**， 故GoogLeNet模型又称为**InceptionNet 模型** 或 **Inception V1 模型**。

将9个Inception V1模块堆叠起来，就组装成了GoogLeNet模型，2014年参赛的版本。



6.4.2 GoogLeNet模型结构的总体说明

- GoogLeNet模型由9个Inception V1模块线性堆叠而成，
- 其中包含22个带可学习参数的网络层，
- 并且在最后一个Inception V1模块处使用了全局平均池化，减少了全接连层的参数量，也可防止过拟合。

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

(a) “#3×3 reduce”、“#5×5 reduce”表示在 3×3、5×5 卷积操作之前的1×1 滤波器的数量。

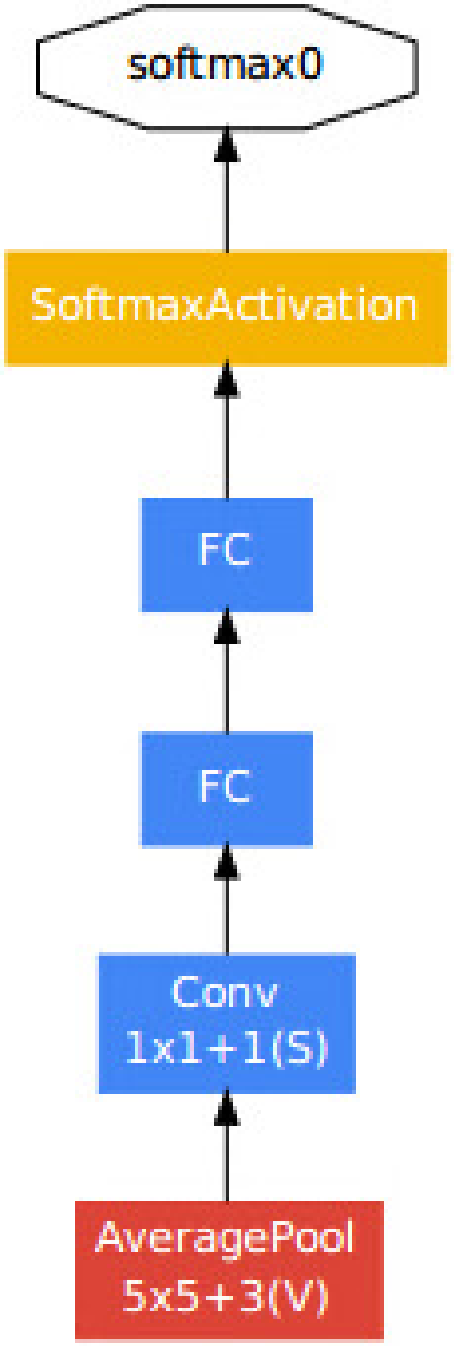
(b) “pool proj”列表示：Inception模块中内置的最大池化层后面的1×1滤波器的数量。

GoogLeNet结构的总体设计

（1）GoogLeNet采用模块化结构，浅层部分仍采用传统的卷积形式，只在较深层部分采用Inception模块堆叠的形式，可方便增添和修改模块结构。

（2）所有卷积层，包括Inception模块内部的卷积层，其后均使用ReLU激活函数，模块内用于降维和投影（Reduction/Projection）的1×1卷积层之后，也都采用了ReLU激活函数。

（3）网络尾部只保留了一个全连接层，用平均池化来代替其他的全连接层，以0.7的概率使用了dropout技术，实验证明：这样可以将准确率提高0.6%。保留一个全连接层，是为了方便对输出进行灵活调整。



（4）为防止梯度消失，增强网络的泛化能力，在网络中间部分设置了**两个辅助分类器**，这些分类器采用规模较小的卷积网络形式，依次由**一个平均池化层、一个1×1卷积层、两个全连接层和一个Softmax函数层**组成。

- 大量实验表明：处于网络模型中间层的特征往往具有很强的判别能力，故这两个辅助分类器被分别放置在Inception (4a) 和Inception (4d) 模块之后，即分别将这两个模块的输出作为输入，进行分类。
- 在网络训练阶段，辅助分类器的**损失函数以一个较小的权重（取值为0.3）加到总损失函数中**，为网络增加了反向传播的梯度信号，也起到了一定的正则化作用。但**在推断（即预测）过程中，去掉这两个辅助分类器。**

6.4.3 GoogLeNet模型结构解析（图6.7）

以5个独立的池化层（即不包含在Inception模块内）作为分界线，可将GoogLeNet模型划分为6部分：

1. 输入层。
2. **第一部分**只包含1个卷积层。
3. 第一个**独立的最大池化层**。
4. **第二部分**包含2个卷积层。
5. 第二个**独立的最大池化层**。
6. **第三部分**由2个Inception模块组成，分别是Inception (3a)和(3b)模块。
7. 第三个**独立的最大池化层**。

6.4.3 GoogLeNet模型结构解析（图6.7）

8. 第四部分由5个Inception模块组成，分别是Inception (4a)、(4b)、(4c)、(4d)和(4e)模块。

9. 第一个辅助分类器放置于Inception (4a)模块输出（14×14×512的特征图）之后，由一个较小的卷积网络构成。

10.第四个独立的最大池化层。

11.第五部分由2个Inception模块组成，分别是Inception (5a)和(5b)模块。

12.第五个独立的平均池化层。

13.第六部分由1个全连接层和softmax激活函数组成。

6.4.4 GoogLeNet模型的特点

(1) 在Inception模块中采用多分支并行处理数据的特征图，拼接多分支的输出结果。

(2) 采用 1×1卷积减少特征图的通道数，以降低计算量和参数量。

(3) 采用全局平均池化层代替全连接层，使模型参数大幅度减少。

(4) 通过精心设计的Inception模块，在增加了网络深度和宽度的同时，还能保持计算量不变。

在随后的几年里，研究人员对GoogLeNet又进行了数次改进，形成了更深的神经网络Inception V2、V3、V4等版本。

6.5 ResNet 残差网络（Residual Networks）

◆ResNet是由微软研究院提出的，是2015年CVPR年的最佳论文。

◆在2015年ILSVRC比赛的ImageNet数据集上**图像分类、目标检测、目标定位以及MS COCO数据集上目标检测、图像语义分割**等**5项任务中全部获得冠军**。

◆ResNet是在计算机视觉的深度学习领域中继AlexNet之后最具开创性的工作，因为它使得训练**成百甚至上千层的深度神经网络成为可能**。

◆ResNet在ILSVRC-2015比赛ImageNet图像分类任务中的top-5错误率仅为3.57%，比2014年冠军GoogLeNet的错误率下降了3.1%，**首次超过了人眼识别能力**（人的错误率为5.1%）。

6.5.1 ResNet模型的研究动机 (1)

◆ResNet出现之前的研究结果表明：网络的性能会随着层数的加深而增加，从AlexNet的8层到VGG的19层，再到GoogLeNet的22层，都验证了这一结论。

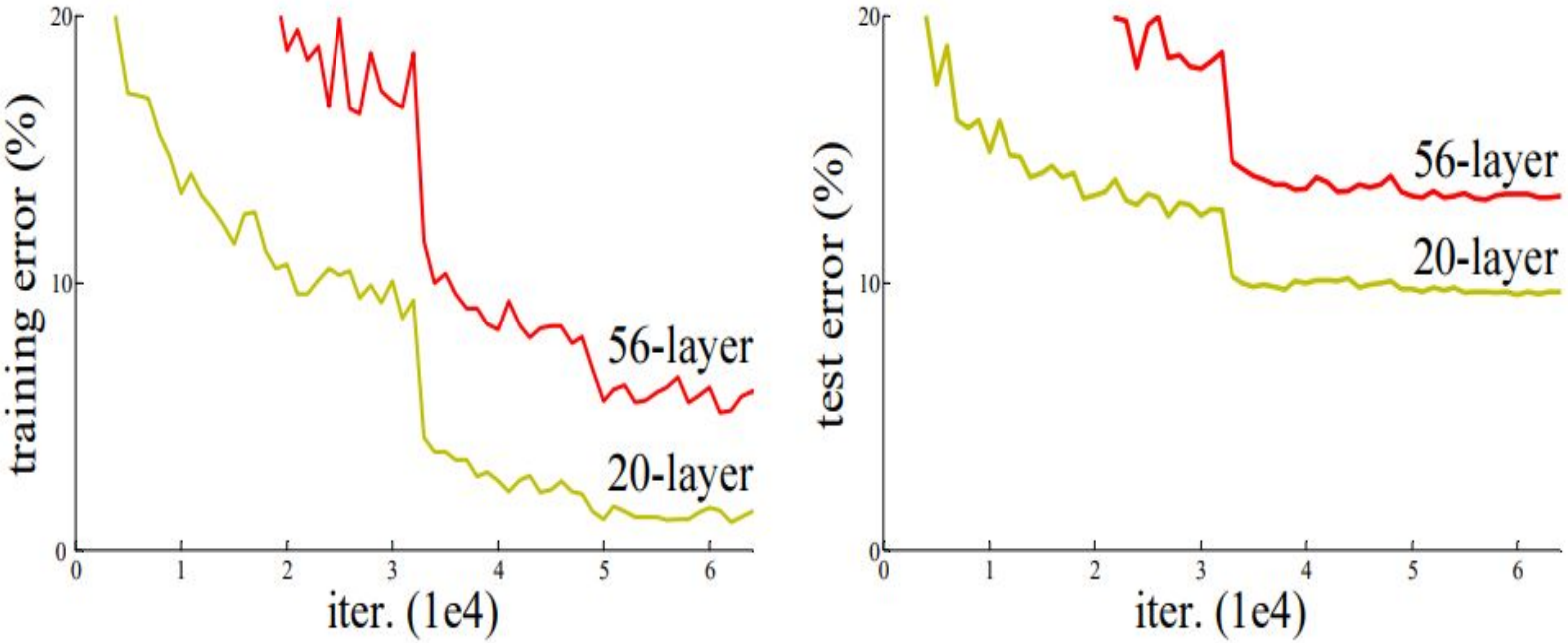
◆虽然深度网络中常出现梯度消失或梯度爆炸的问题，会导致训练过程不收敛，但这一问题在很大程度上已采用初始归一化（Normalized Initialization，即将输入数据映射到[0,1]或[-1，1]区间内）和中间层归一化（Intermediate Normalization，即将中间层的数据映射到[0,1]或[-1，1]区间内）解决了，这使得采用反向传播和随机梯度下降（Stochastic Gradient Descent, SGD）方法的几十层网络都能收敛。

6.5.1 ResNet模型的研究动机 (2)

◆但在解决了收敛问题后，又出现了**网络退化问题**：在深度神经网络达到一定深度后，随着网络层数的加深，分类的准确率反而下降了，即神经网络的训练误差随着网络层数的加深而变大。

◆可见，引起网络退化问题的原因既不是不收敛，也不是过拟合。

◆研究者们发现：无论是在训练过程中，还是在测试过程中，一个56层网络的性能还不如一个20层网络的性能。



6.5.1 ResNet模型的研究动机 (3)

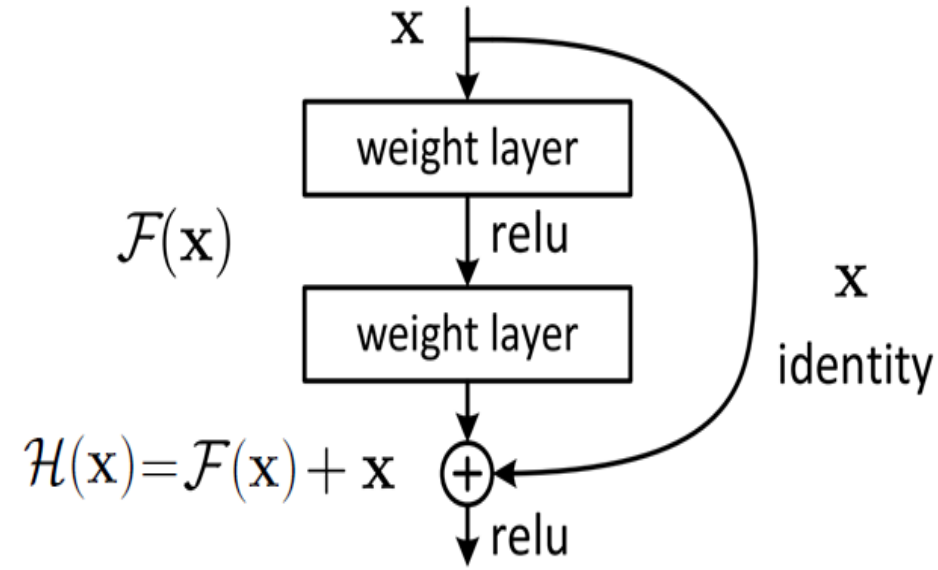
◆假设如此构建一个深层网络：先训练得到一个已达到一定准确度的浅层网络结构，然后复制上述浅层网络，在其基础上增加一些**恒等映射**（Identity Mapping）层，即**前面的某一层或某些层**，得到深层网络。

◆按常理推测，深层网络至少可以达到与浅层网络相同的准确度，不会比浅层网络的错误率高。

◆但实验结果表明：这样得到的深层网络却表现得更差。

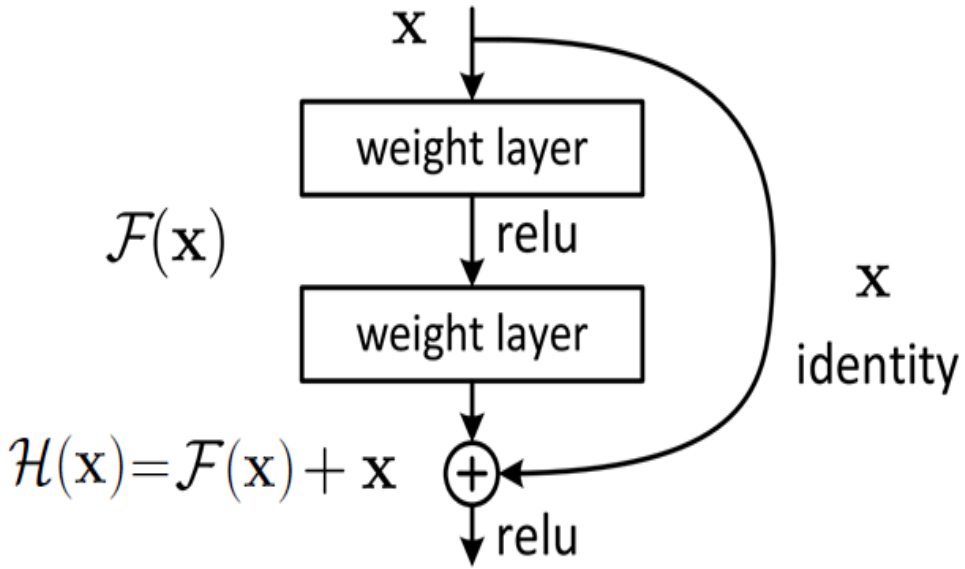
◆**网络退化问题说明：采用多个非线性层去逼近恒等映射是有困难的**。

6.5.1 ResNet 模型的研究动机 (4)



- 为了解决网络性能退化的问题，ResNet的研究人员提出了残差模块的结构。
- 其思想是：假设x为输入，令**H(x)**为需要学习得到的**基础映射**，采用堆叠的非线性层拟合另一个映射，称为**残差映射**，记为**F(x)**，令F(x)= H(x)-x，则原本需要学习的基础映射H(x) =F(x)+x（**按元素叠加起来**）。
- F(x)在数学上称为残差，所以该网络称为残差网络。
- 研究人员假设：**优化残差映射比优化原来的基础映射容易**。所以，**不直接**用若干个堆叠的网络层去**拟合基础映射H(x)**，而是**先拟合残差映射F(x)**，然后用F(x)+x得到基础映射H(x)。

6.5.2 ResNet模型的结构



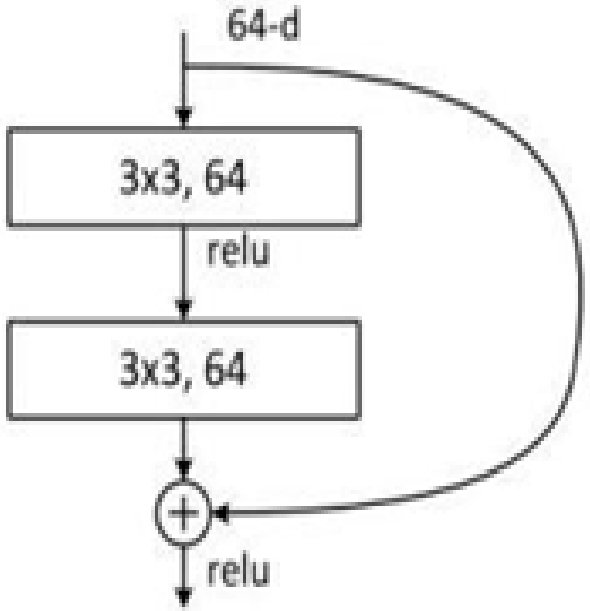
- ResNet的每个残差模块都是由**一个主分支**和**一个捷径分支**（Shortcut）并行组成的，
- **主分支**是由若干个前馈神经网络层组成的残差映射；
- **捷径分支**则是跳过一层或多层，直接将该模块的输入特征图和输出特征图连接在一起，可看作是**恒等映射**。
- **恒等映射**操作既不增加额外的参数，也不增加计算复杂度。
- 将学习得到的残差映射F(x)与恒等映射x按元素叠加起来，就得到了基础映射H(x)
- 在做叠加操作时，要使得输入 x 的特征形状与 F(x) 输出的**特征形状一致**，否则需要对 x 做**线性投影（可以下采样）**，使之与F(x)输出的维度匹配。

◆有两种形式的残差模块，“瓶颈”设计的残差模块的目的是为了降低参数个数，

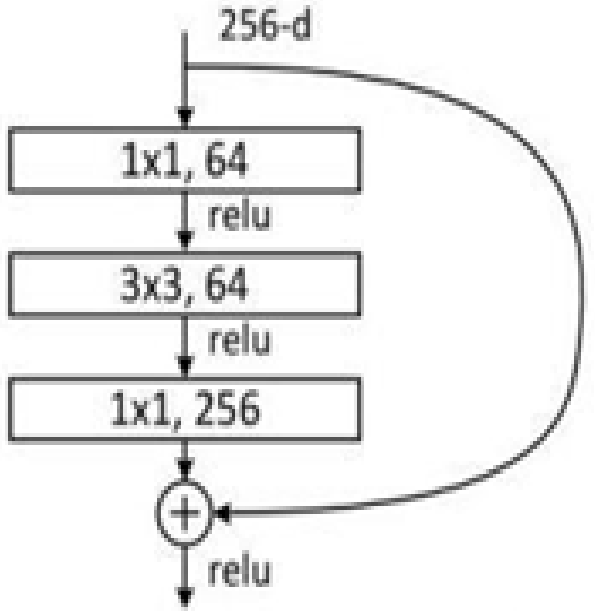
◆(b)中第1个1×1卷积将通道数由256维降到64维，第2个1×1卷积将通道数恢复为256维，总参数量为(256+1)×64 + (3×3×64+1)×64 + (1+64)×256 = 70016。

◆若不采用“瓶颈”设计，只使用两个3×3×256的滤波器，如图 (a)所示，则总参数量为(3×3×256+1)×256×2 = 1180160，是(b)中参数量的16.86倍。

◆可见，采用“瓶颈”设计的残差模块来构造深度网络模型，在训练过程中可有效减少参数量和计算量。



(a) 一般残差模块



(b) “瓶颈”残差模块

- 采用残差模块构建的ResNet网络不仅能有效地解决网络退化问题，还可极大地减缓梯度消失和梯度爆炸的问题，
- 因为ResNet的梯度能直接通过捷径分支跳跃地传回到较浅的层，避免了梯度在反向传播经过多层时过大或过小，而导致无法收敛。
- 埃明·奥尔汉（Emin Orhan）等人对深度神经网络的退化问题进行了更深入的研究，认为：**深度神经网络的退化才是深度网络难以训练的根本原因，而不是梯度消失**。

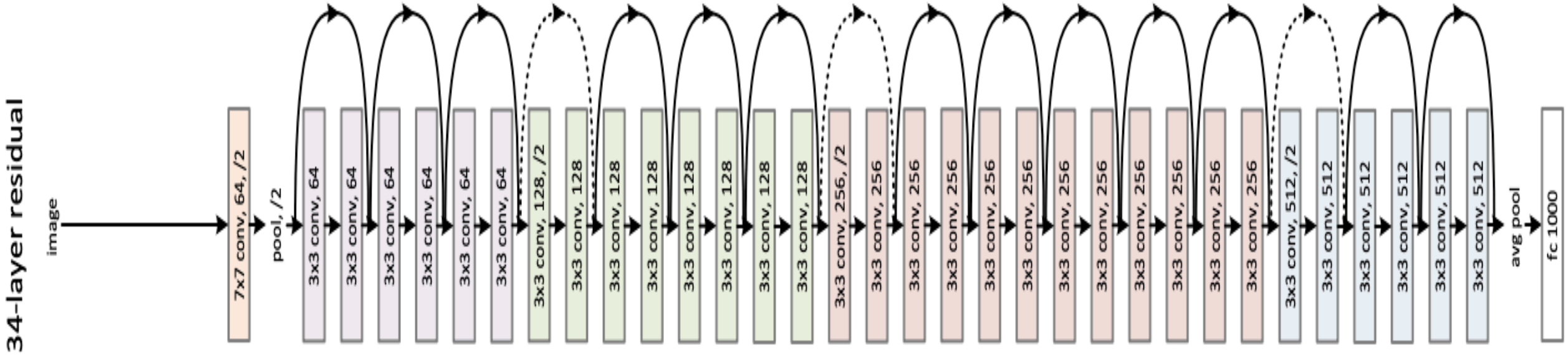
- 采用不同形式的残差模块，可以组装出不同深度的神经网络。
- 微软研究人员给出了5种ResNet网络配置。
- 其中ResNet-18、ResNet-34采用的是图 (a)形式的残差模块，
  ResNet-50、ResNet-101和ResNet-152采用的是图 (b)形式的残差模块。

表6.3 五种ResNet网络配置

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

注：“×n”表示该残差模块连续堆叠n次。

ResNet-34网络结构示意图



- 5种ResNet网络的输入信息均为224×224的RGB图像，
- 第一个卷积层都包含64个大小为7×7、步长为2 、padding=same（p=3）的滤波器，输出112×112×64的特征图，$\left\lfloor \frac{224-7+2*3}{2} \right\rfloor + 1 = 112$；
- 第一个池化层都是执行大小为3×3、步长为2的重叠最大池化操作，输出56×56×64的特征图。
- 经过不同深度（3，4，6，3）的4组残差模块后，输出大小为7×7的特征图；
- 然后，采用AvePooling (7×7) 的平均池化操作，得到大小为1×1的特征图；
- 最后是包含1000个神经元的全连接层，以Softmax作为激活函数，输出图像属于各个类别的概率值。

◆微软在ILSVRC 2015比赛中赢得冠军的网络是由6个不同深度模型集成的。

◆训练时，数据批的大小（Batch Size）设置为256个样本，在ResNet模型中的**每个卷积层和激活函数之间均执行批归一化**（Batch Normalization，BN）操作，**其作用有三：**

（1）在使用梯度下降法求最优解时，防止梯度爆炸或弥散，加快收敛速度；

（2）可以提高训练时模型对于不同超参（如学习率、初始数据）的鲁棒性；

（3）可以使大部分激活函数能够远离其饱和区域。

◆ResNet具有强大的表征能力，能训练数百层甚至上千层的神经网络。

◆在诸如图像分类、目标检测、语义分割和人脸识别等计算机视觉应用领域，取得了很大进展。ResNet也因其简单的结构与优异的性能成为计算机视觉任务中最受欢迎的网络结构之一。

6.6 DenseNet（Dense Convolutional Network）

◆DenseNet由康奈尔大学、清华和 Facebook合作提出，获得了2017年CVPR最佳论文奖。

◆研究者发现，ResNet模型的核心是在不相邻的前、后层之间建立直接的捷径（“短路连接”或“跳跃连接”），这样做有助于在训练过程中反向传播梯度值，从而能训练出更深的CNN网络。

◆基于相同的思路，他们提出了DenseNet模型，直接将前面所有的特征图与后面的特征图连接起来（前后特征图的尺寸必须匹配），构造一种具有密集连接（Dense Connection）的卷积神经网络，以确保最大信息量在网络各层之间传播，由此将模型命名为DenseNet。

◆与ResNet模型的不同之处在于：**DenseNet模型采用的是一种更密集的短路连接机制**。

1. Dense Block模块

◆一个Dense Block模块包含若干个网络层，每个网络层的特征图大小都相同，每层都与同模块中其前面的所有层相互连接，即同模块中任意两层之间都有直接的连接。

◆一个L层（包括输入层）的Dense Block模块，一共包含L(L+1)/2个连接。

◆第l层的输入特征图 $[X_0, X_1, \ldots, X_{l-1}]$ 经过映射H_l后，输出$X_l = H_l([X_0, X_1, \ldots, X_{l-1}])$，其中每个$H_l$都是由“BN-ReLU-Conv(3×3)”操作序列组成的，其中BN（Batch Normalization）是将一批数据的特征进行归一化，其作用是加快收敛过程；ReLU是激活函数，以增加非线性表达能力；

◆DenseNet的研究者规定：在一个Dense Block中，每个映射H_l都提取k个特征，则第l层的特征图的通道数为$k_0 + (l-1)k$，其中k_0是该Dense Block模块输入层的通道数。**k是一个超参数**，称为**增长率**。

◆增长率是控制Dense Block 中网络层的宽度的，它规定了每层上包含的滤波器的数量。

Dense Block模块示例

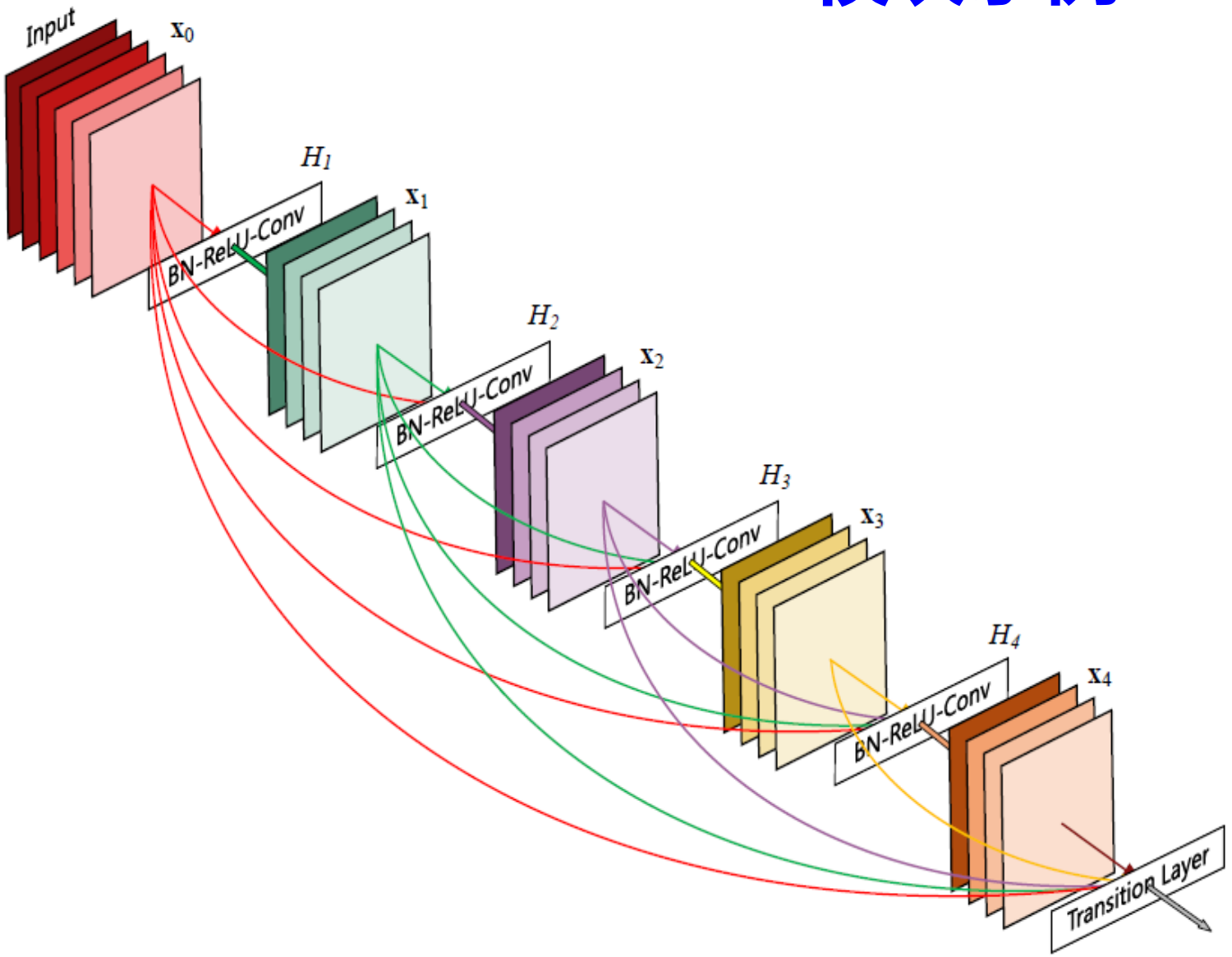


图6.13 一个5层密集连接模块（Dense Block)，增长率k=4，一共有15个连接

2. 转换层

◆为了控制特征图通道数的快速增长，在每两个相邻的Dense Block模块之间都设置一个转换层。

◆转换层的作用：降低特征图的维度，去掉冗余的特征，保证训练的高效性。

◆每个转换层均先采用瓶颈层，即Conv(1×1)，减少特征图的通道数，然后再采用池化操作缩减特征图的大小。

◆因此，每个转换层都是由BN-ReLU-Conv(1×1)-AvePooling(2×2)操作序列组成的，其中

- BN和ReLU的作用与Hl中的相同；
- Conv(1×1)是大小为1×1的卷积，以减少特征图的通道数；
- AvePooling(2x2) 是大小为2×2的不重叠平均池化层，将特征图的高和宽缩小一半。

3. DenseNet-BC模型

实现时，为了控制参数量和计算量，DenseNet的研究者还作了如下设置。

（1）在每个Dense Block模块的每个3×3卷积层之前均增加一个1×1的卷积层，以减少3×3卷积层的输入特征图的通道数，可以大大减少计算量，提高计算效率。

- 在实验中，每个Dense Block模块中每层映射H_l的操作序列为BN-ReLU-Conv(1×1)-BN-ReLU-Conv(3×3)，其中Conv(1×1) 滤波器的个数设置为$4k$。
- 仅作此设置的网络记为**DenseNet-B模型**。

3. DenseNet-BC模型

（2）如果一个Dense Block模块输出m个特征图，则使得紧随其后的转换层产生$[\theta m]$个特征图，其中θ $(0<\theta\leqslant 1)$ 称为压缩因子，以控制向下一个Dense Block模块输入的特征图的通道数。

◆转换层的瓶颈层中的Conv(1×1) 滤波器的深度为m，个数为$[\theta m]$。

- 当θ=1时，经过转换层，特征图的通道数保持不变，即无压缩；
- 当θ<1时，转换层减少特征图的通道数。

◆在实验中，设置θ=0.5，每个转换层的操作序列为BN-ReLU-Conv(1×1)-AvePooling(2×2)，其中Conv(1×1)滤波器的个数为$[m/2]$个。

◆仅作此设置的网络记为**DenseNet-C模型**。

（3）同时采用上述两种设置，构造出来的DenseNet网络结构记为**DenseNet-BC模型**。

3. DenseNet-BC模型

（2）如果一个Dense Block模块输出m个特征图，则使得紧随其后的转换层产生$[\theta m]$个特征图，其中θ $(0<\theta\leqslant 1)$ 称为压缩因子，以控制向下一个Dense Block模块输入的特征图的通道数。

◆转换层的瓶颈层中的Conv(1×1) 滤波器的深度为m，个数为$[\theta m]$。

- 当θ=1时，经过转换层，特征图的通道数保持不变，即无压缩；
- 当θ<1时，转换层减少特征图的通道数。

◆在实验中，设置θ=0.5，每个转换层的操作序列为BN-ReLU-Conv(1×1)-AvePooling(2×2)，其中Conv(1×1)滤波器的个数为$[m/2]$个。

◆仅作此设置的网络记为**DenseNet-C模型**。

（3）同时采用上述两种设置，构造出来的DenseNet网络结构记为**DenseNet-BC模型**。

4. ResNet与DenseNet中短路连接机制的不同

（1）两个网络中短路连接的密集程度不同。

- ResNet中一个残差模块一般只包含2~3个卷积层，故每层只跨越2~3层与其前面的某一层直接连接，形成短路（或捷径）。
- DenseNet中一个Dense Block模块一般包含6~64个卷积层，同一模块中的所有层均互相连接。显然，DenseNet网络中的短路连接比ResNet中的更密集，更好地实现了特征重用，增强了特征在各个层之间的传播。

（2）两个网络中短路连接的方式不同。

- 在ResNet网络的同一残差模块中，对于残差映射特征图与恒等映射特征图，执行对应位置上的元素级相加操作。
- 在DenseNet网络的同一Dense Block模块中，每层都会与其前面所有层在通道维度上作拼接操作，并作为下一层的输入。

5. DenseNet模型的优点

DenseNet网络在Dense Block模块中所有特征图之间采用密集的短路连接机制，具有如下显著优势。

（1）缓解了梯度消失问题，因为每一层都能从损失函数层直接访问梯度信息，也使得网络易于训练。

（2）实现了特征重用，因为每一层都能通过捷径接收到前面所有层的特征图作为输入。

（3）增强了特征在各个层之间的传播，因为每两层之间都存在直接的连接。

（4）极大地减少了参数量，提高了训练效率。

6. ImageNet图像分类任务上的网络配置

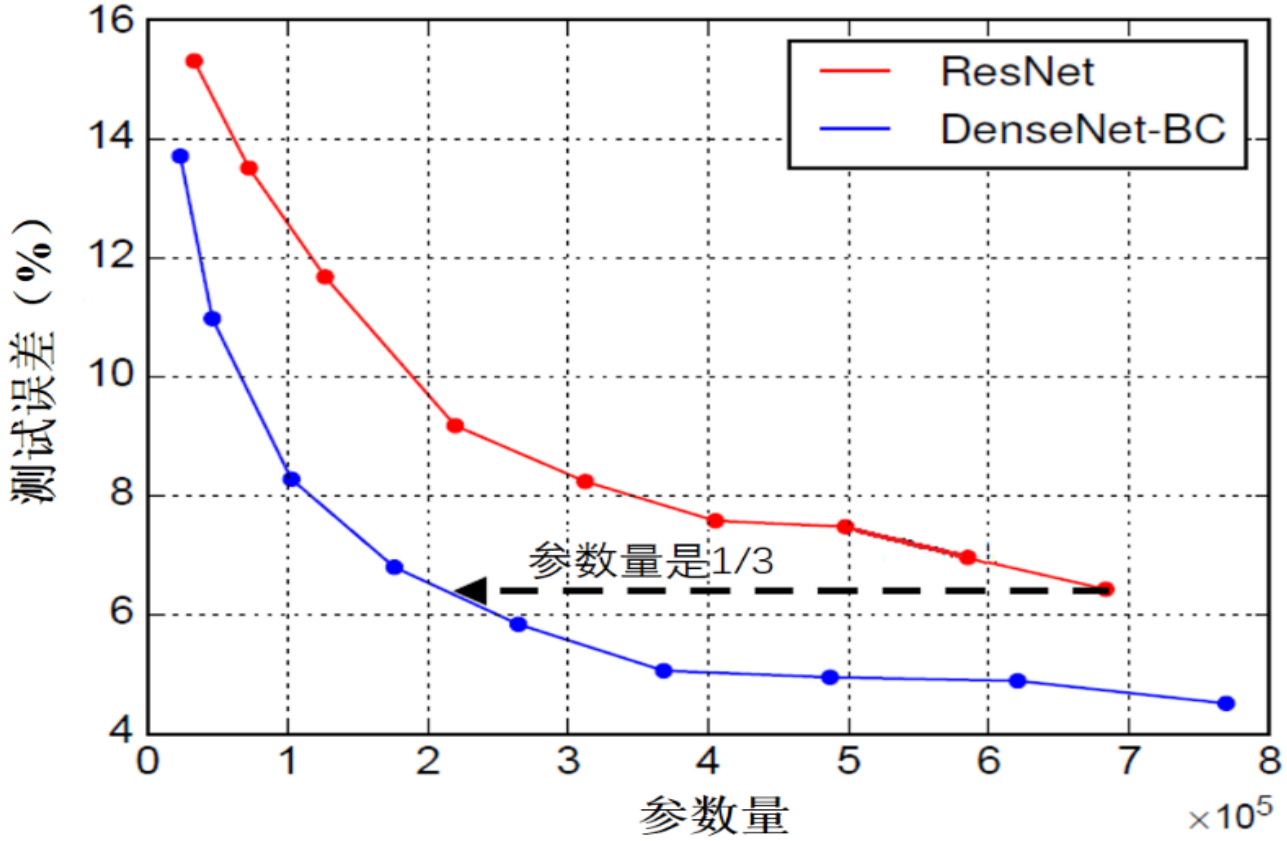
表 6.4 ImageNet 分类任务上的 4 种 DenseNet-BC 网络配置，k=32

层	输出的尺寸	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv，步长为2			
Pooling	56 × 56	3 × 3 max pool，步长为2			
Dense Block (1)	56 × 56	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool，步长为2			
Dense Block (2)	28 × 28	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool，步长为2			
Dense Block (3)	14 × 14	[1 × 1 conv; 3 × 3 conv] × 24	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 48	[1 × 1 conv; 3 × 3 conv] × 64
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool，步长为2			
Dense Block (4)	7 × 7	[1 × 1 conv; 3 × 3 conv] × 16	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 48
分类层	1 × 1	7 × 7 全局平均池化			
		1000D 全连接，softmax			

注：表中每个“conv”层都表示 BN-ReLU-Conv 操作序列，“×n”表示该 Dense Block 模块包含 n 个“Conv(1×1)+ Conv(3×3)”组合。

6. ImageNet图像分类任务上的网络配置

- DenseNet研究团队的实验数据显示：在ImageNet图像分类任务中取得相近准确率的情况下，DenseNet需学习的参数量只是ResNet-152模型参数量的约1/3。
- 由此可见，密集连接方式可极大地减少参数量和计算量。
- DenseNet在参数量和计算成本更少的情形下实现了比ResNet更优的性能。



6.7 本章小结（1）

1. LeNet

1989年杨乐昆提出了LeNet模型，奠定了现代卷积神经网络的基本结构。LeNet-5成为第一个大规模成功商用的卷积神经网络模型。

2. AlexNet

2012年辛顿研究团队提出的AlexNet模型在ILSVRC比赛中以15.3%的top-5错误率夺得图像分类任务的冠军，准确率高出亚军近10%，成为卷积神经网络研究史上一个非常重要的里程碑。从此，深度学习成为人工智能研究领域的主流方法。

3. VGGNet

由牛津大学VGG研究组提出的VGGNet模型在2014年的ILSVRC大赛图像分类任务中以7.32%的top-5错误率夺得亚军。VGGNet模型是在AlexNet结构的基础上引入了“模块化”的设计思想，网络深度可达19层。该项工作证明了：

①用多个尺寸较小的卷积核代替一个大尺寸卷积核，既可保证相同的感受野，又可减少参数量；

②在大规模图像识别任务中，增加卷积神经网络的深度可有效提升模型的精确度。

6.7 本章小结（2）

4. GoogLeNet

- GoogLeNet模型在2014年的ILSVRC大赛图像分类任务中以6.67%的top-5错误率夺得冠军。
- 提出了Inception模块。每个Inception模块采用多分支并行处理数据的特征图，拼接多分支的输出结果，同时采用 1×1卷积减少特征图的通道数，以到达降低计算量和参数量的目的。
- GoogLeNet模型是由多个Inception模块堆叠组装而成的，网络深度可达22层，是当时最深的网络模型。

5. ResNet

- 在2015年的ILSVRC比赛上，由微软亚洲研究院提出的ResNet模型荣获5项任务的冠军。其中，ImageNet图像分类的top-5错误率仅为3.57%，首次超过了人眼识别能力。
- ResNet是继AlexNet之后最具开创性的工作，提出了残差结构，不仅有效地解决了网络退化问题，还极大地减缓了梯度消失和梯度爆炸问题。
- 采用“瓶颈”残差模块更能有效地减少参数量和计算量，使得训练成百甚至上千层的深度神经网络成为可能。

6.7 本章小结（3）

6. DenseNet

- 康奈尔大学提出的DenseNet模型，获得了2017年CVPR的最佳论文奖。
- 一个DenseNet网络由若干个Dense Block模块和转换层组装而成。
- 与ResNet相比，DenseNet模型采用了一种更为密集的短路连接机制。
- ResNet采用了“瓶颈”设计，实现了特征重用，增强了特征在各个层之间的传播，缓解了梯度消失问题，极大地减少了参数量，提高了训练效率。