GoogLeNet & ResNet

——ILSVRC：2014/2015 Champion

学生创新中心：肖雄子彦

SJTU

目录 / CONTENTS

01

GoogLeNet 与 Inception模块

电影《盗梦空间》(inception)

卷积核与感受野?

信息分布更全局性的图像偏好较大的卷积核，
信息分布比较局部的图像偏好较小的卷积核。

较小卷积核

稍大卷积核

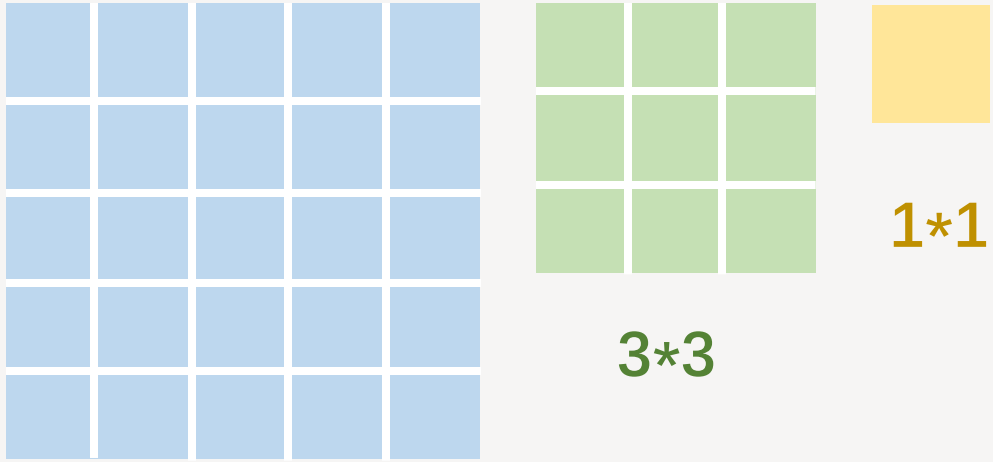
大卷积核

Inception：deeper or wider?

Inception 模块

一般提升网络性能最直接的办法是增加深度，但存在以下问题：

- 参数多，简单堆叠卷积层非常消耗计算资源，梯度更新较为困难
- 训练数据集有限，非常深的网络容易过拟合。



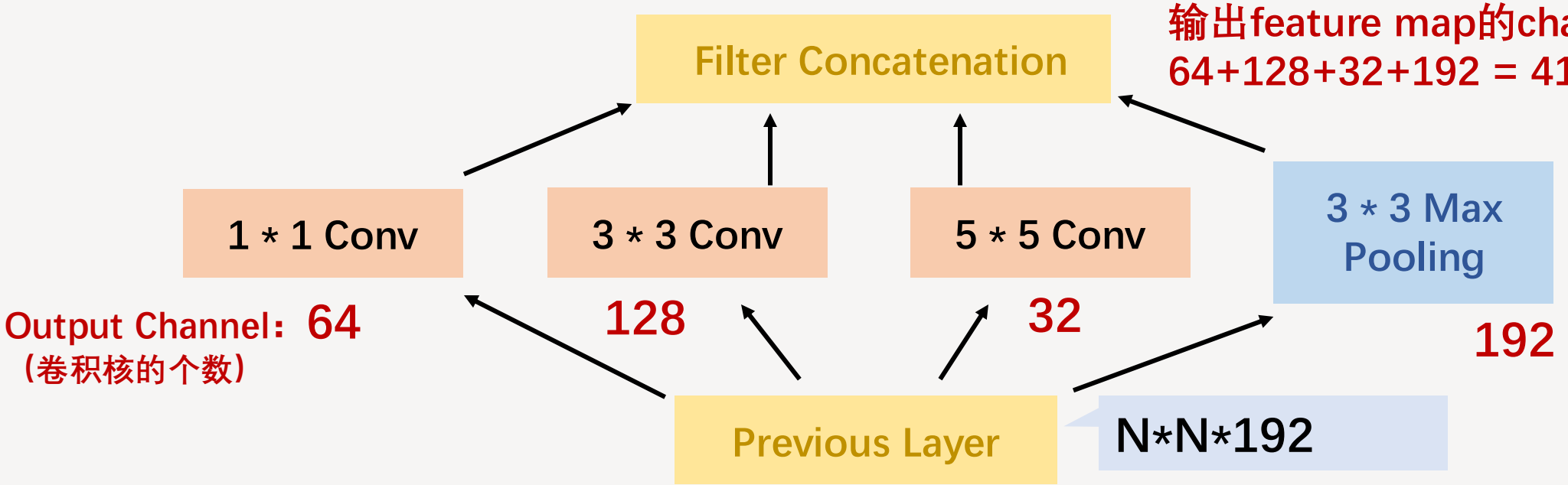
inception在**同一层级上**运行多个不同尺寸的**filter**
网络本质上会变得稍微「**宽一些**」，而不是「更深」

CNN 发展史上一个重要的里程碑

Inception V1

并行执行多个卷积或池化操作，
将输出结果拼接为一个非常深的特征图。

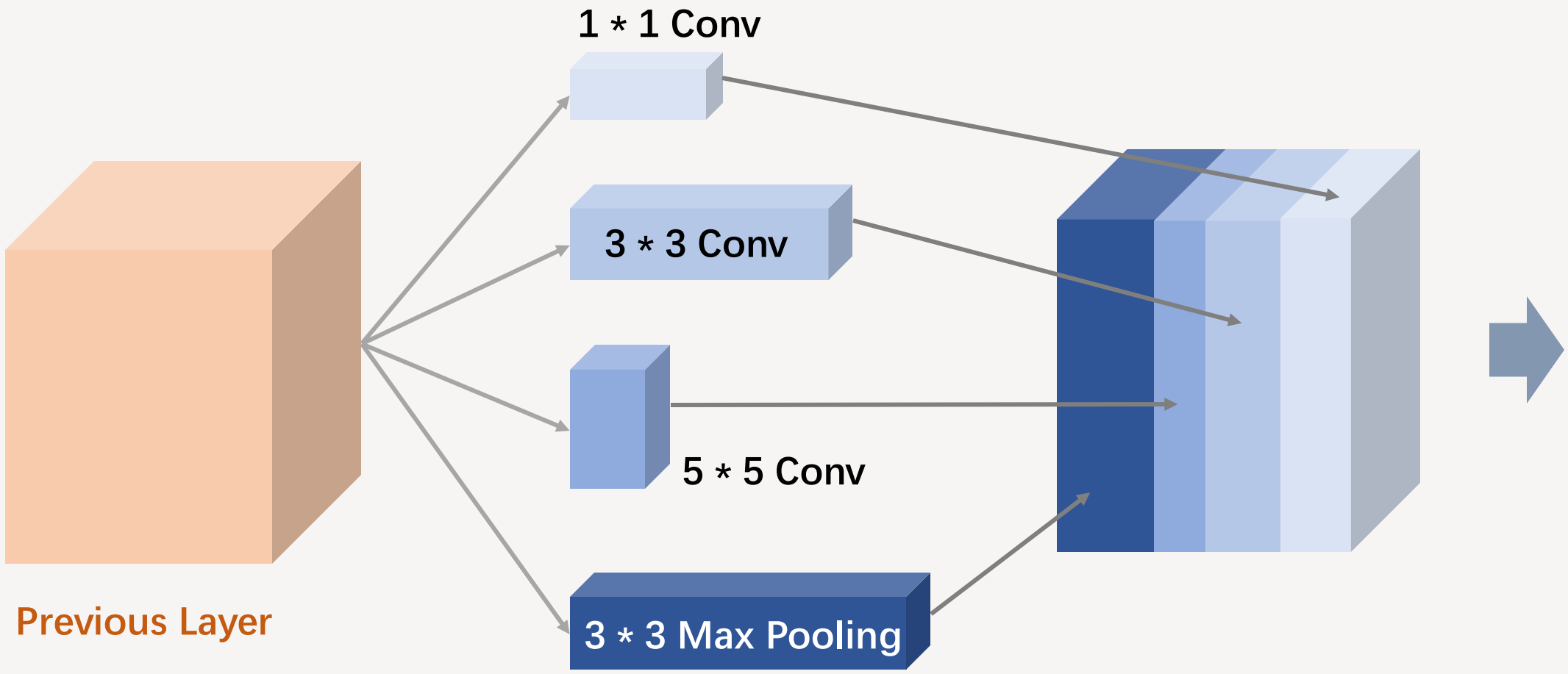
使用 3 个不同大小的kernel（1x1、3x3、5x5）对input 进行卷积操作，同时执行max-pooling
所有子层的输出最后会被级联拼接起来，并作为下一个 Inception 模块的输入。



参数：(1×1×192×64) + (3×3×192×128) + (5×5×192×32) = 387072

Inception模块

Inception Module

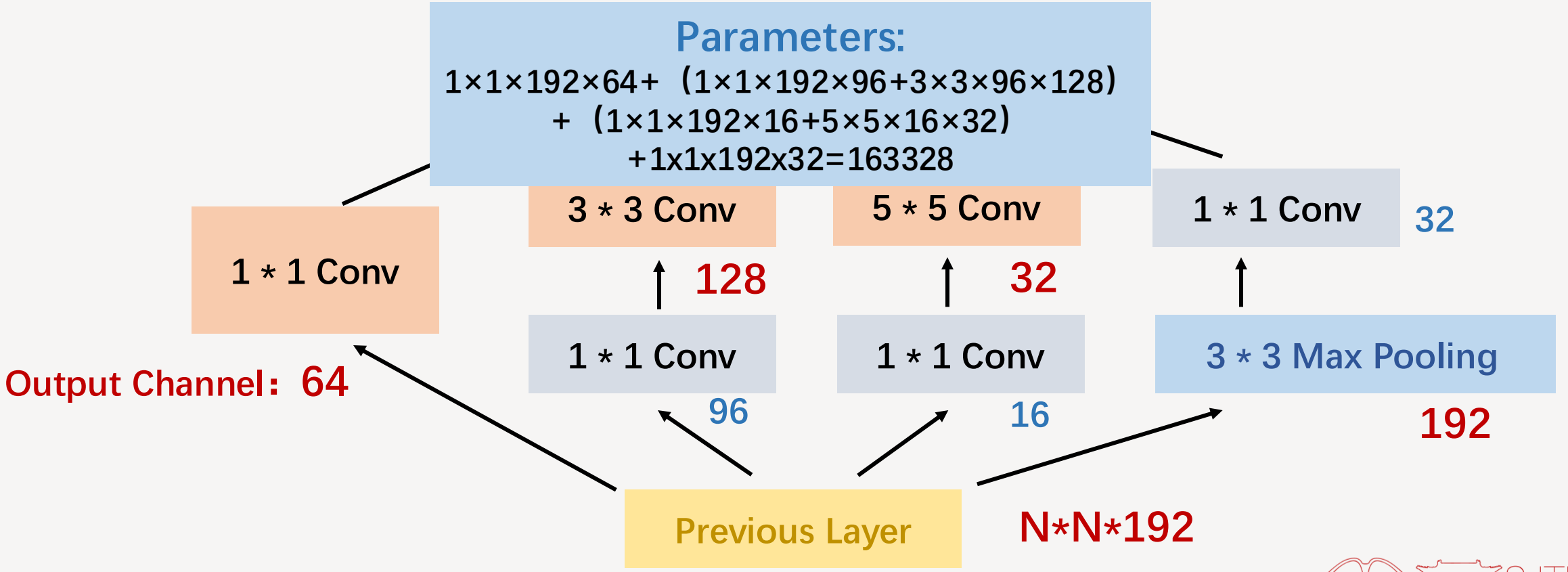


Inception V1

Inception V1

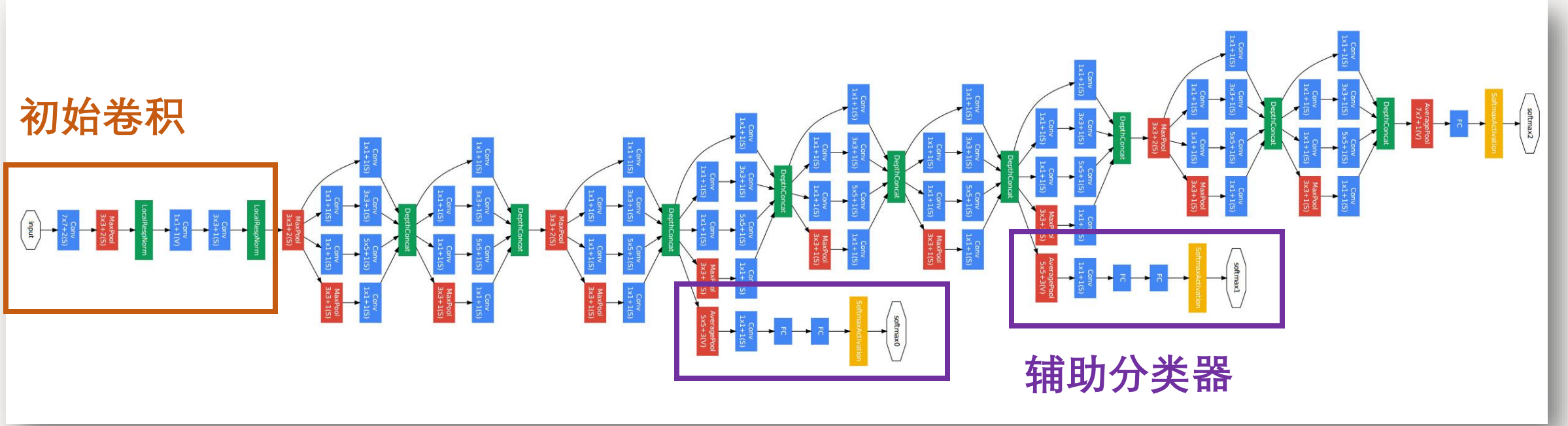
Bottleneck结构，大大减少了参数量

在3x3、5x5前、max pooling后分别加上1x1的卷积核，降低特征图的厚度



采用 Inception 的 GoogLeNet

使用了9个线性堆叠的Inception模块。它有22层（包括池化层的话是27层）。
在最后一个Inception模块使用全局平均池化（GAP）。

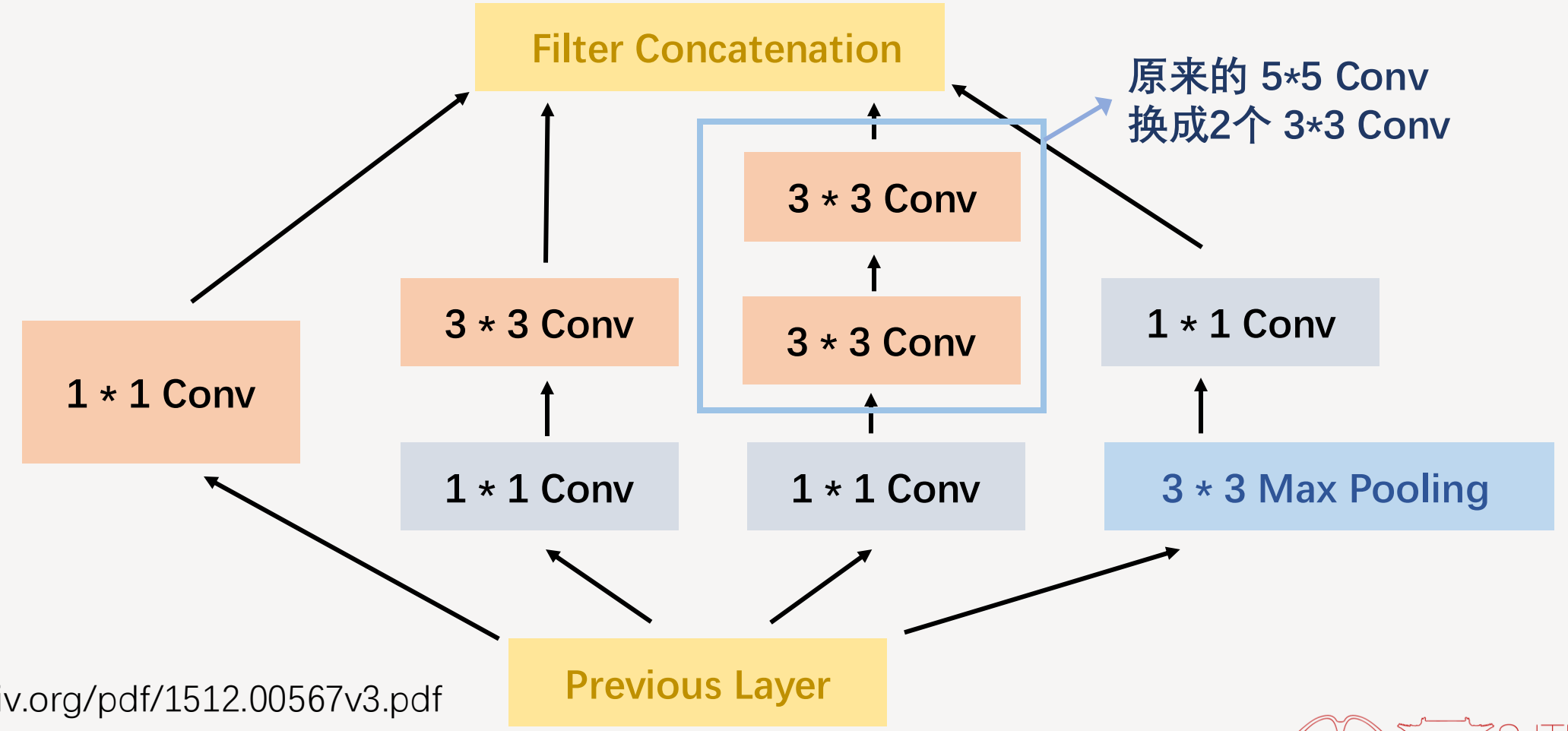


The total loss = real_loss + 0.3 * aux_loss_1 + 0.3 * aux_loss_2

InceptionV2 – 更少的参数

Inception V2

《Rethinking the Inception Architecture for Computer Vision》，提出了一系列能增加准确度和减少计算复杂度的修正方法。



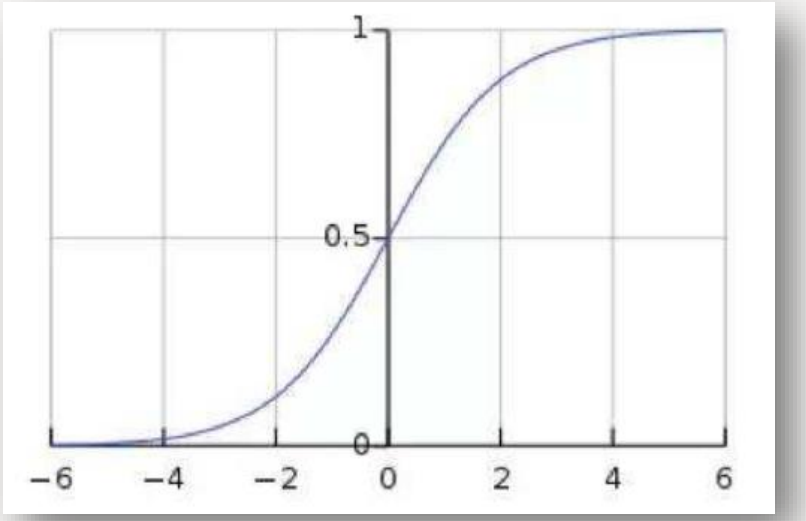
https://arxiv.org/pdf/1512.00567v3.pdf

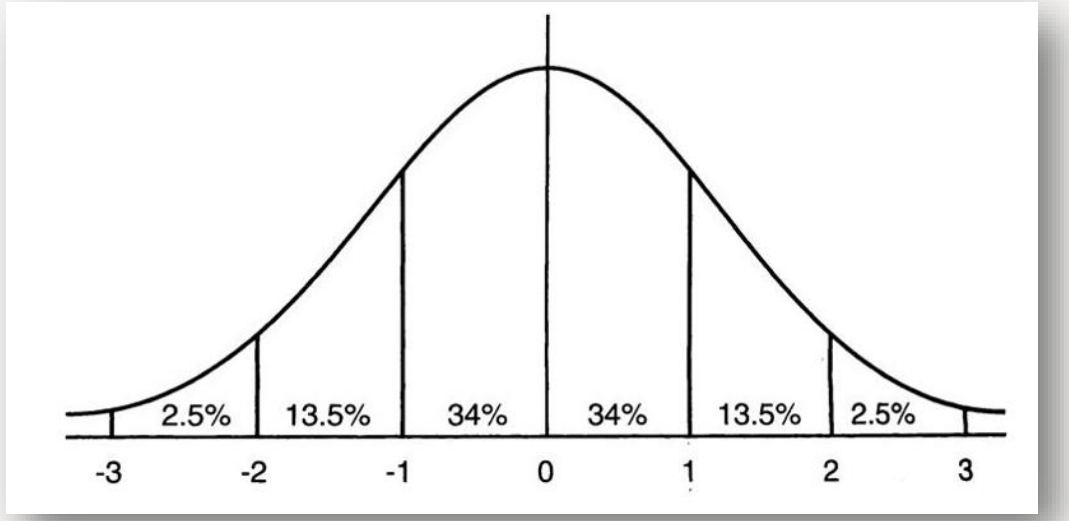
InceptionV2 - BN

Inception V2 提出了Batch Normalization（批标准化）

机器学习领域的重要假设：训练数据和测试数据满足独立同分布

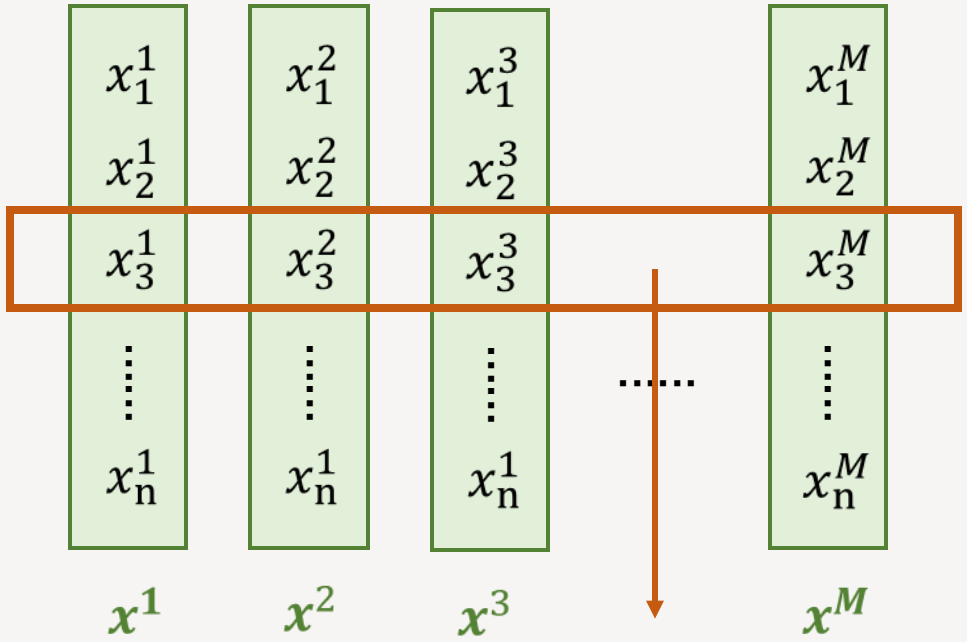
随着网络加深，在训练过程中，样本分布逐渐发生偏移或变动，训练收敛慢。
一般是整体分布逐渐往非线性函数的取值区间的上下限两端靠近。





InceptionV2 - BN

Batch Normalization 的做法



$$x_i^r = \frac{x_i^r - m_i}{\sigma_i}$$

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1\ldots m}\}$;
Parameters to be learned: γ, β

Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

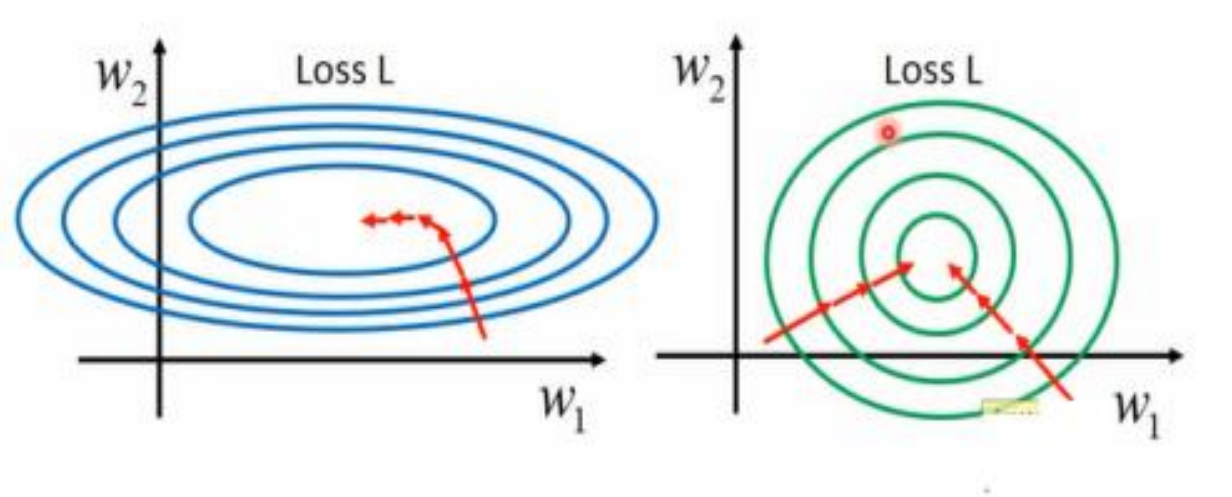
$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

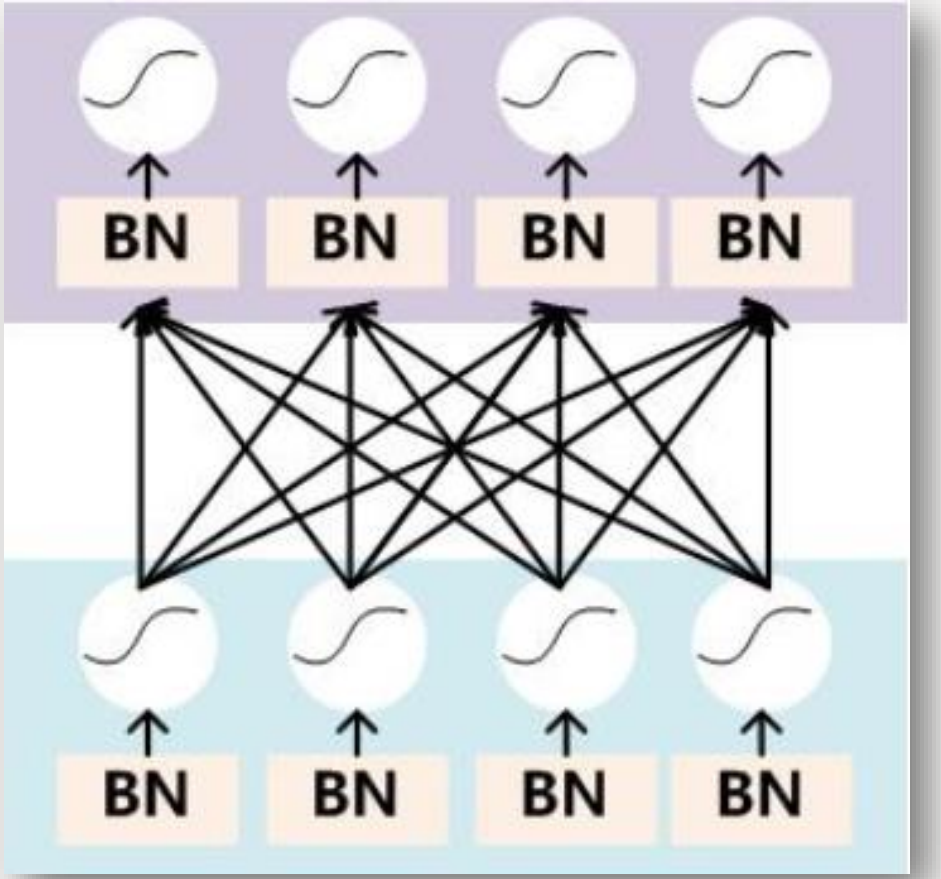
$$m_i = \frac{1}{M}\sum_{r=1}^{M} x_i^r \qquad \sigma_i^2 = \frac{1}{M}\sum_{r=1}^{M}(x_i^r - m_i)^2$$

InceptionV2 - BN

Inception V2

提出了Batch Normalization（批标准化）





BN的主要作用：

- 使每一层输入保持同分布
- 加速训练效率
- 降低对初始化的要求
- 轻微的正则化效果，一定程度上改善过拟合

References

Inception Family Reference

Going deeper with convolutions:
https://arxiv.org/pdf/1409.4842.pdf

Batch Normalization:
https://arxiv.org/pdf/1502.03167.pdf

Rethinking the Inception Architecture for Computer Vision: https://arxiv.org/pdf/1512.00567.pdf

Inception-v4, Inception-ResNet:
https://arxiv.org/pdf/1602.07261.pdf

目录 / CONTENTS

02

残差网络 ResNet

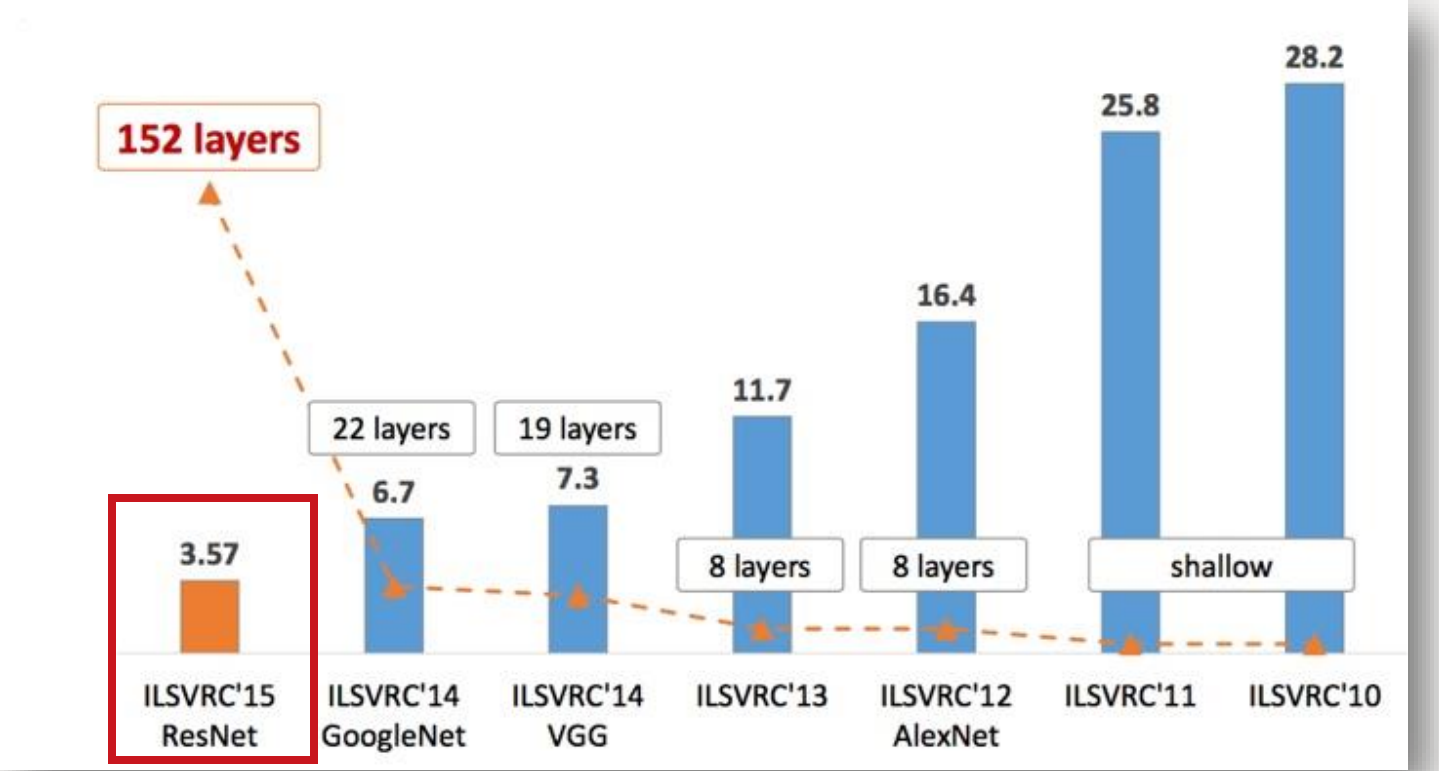
本节内容：

- **理解**残差网络的提出解决的关键问题
- **掌握**残差网络结构，分析网络创新点
- 完成**ResNet识别案例**

图像分类丨ILSVRC历届冠军网络
从AlexNet到SENet

ILSVRC（ImageNet Large-Scale Visual Recognition Challenge）

使用的数据集为ImageNet（李飞飞团队从2007年起，耗费大量人力，收集制作而成，CVPR-2009）的子集——约1.2 million的训练集（约1000类）



- ImageNet：一个超过15 million的图像数据集，约有22,000类。
- ILSVRC：2010-2017

ResNet

何恺明

2003广东省理科高考状元

本科：清华大学

博士：香港中文大学多媒体实验室

2024 MIT 计算机学院副教授

微软亚洲研究院（MSRA）

Facebook AI Research（FAIR）

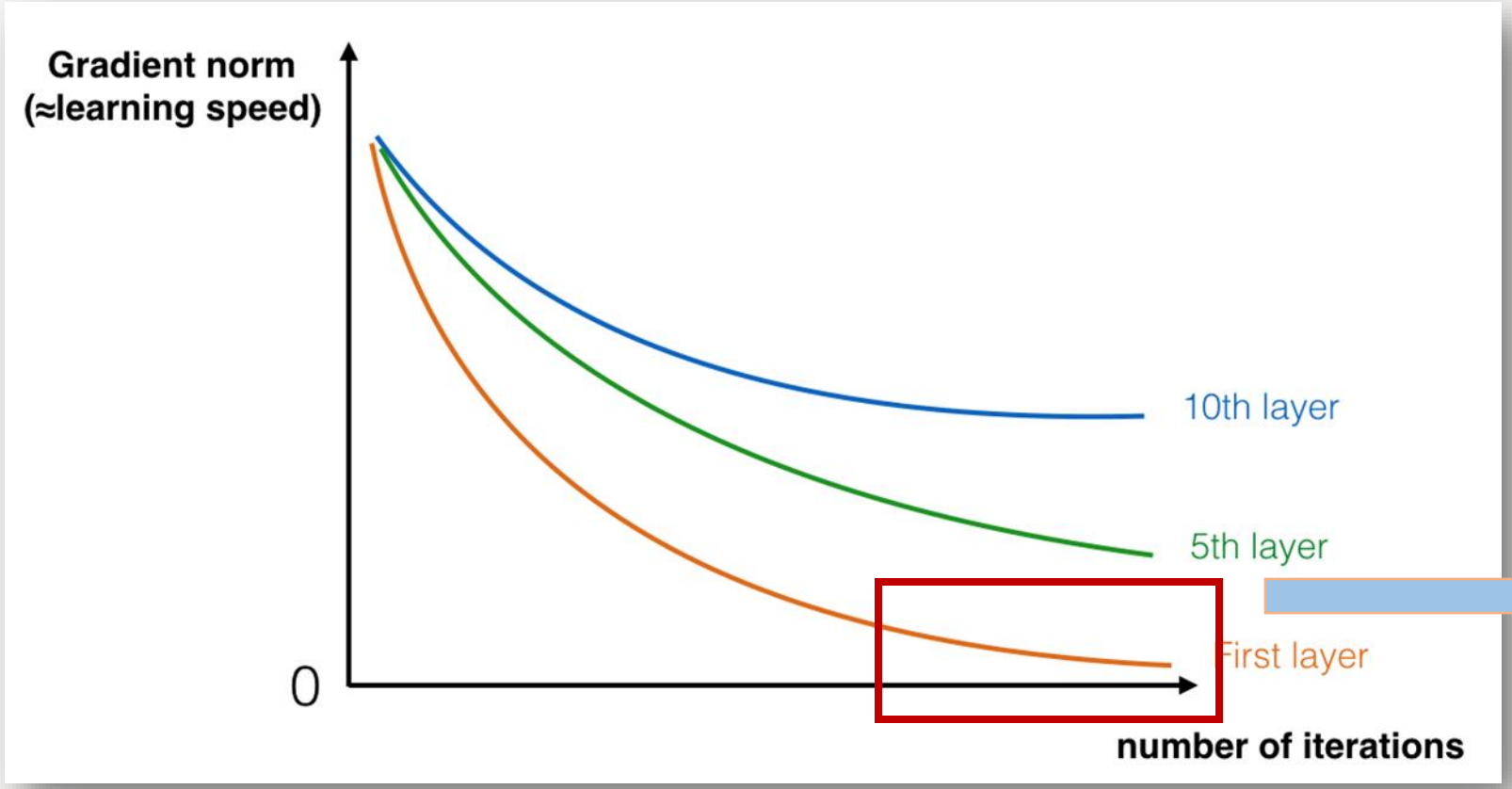
2022，荣登AI 2000全球最具影响力学者榜单

- CVPR 2009首位华人得主
- CVPR 2016和ICCV 2017（Marr Prize）最佳论文奖
- Identity Mappings in Deep Residual Networks（2016ECCV，ResNet的后续原理分析及改进）
- ResNet（一作）Faster-RCNN（二作）
- Focal Loss（三作），R-FCN（三作）
- Single Image去雾，SPP-Net(ECCV2014),
- Instance-Aware Semantic Segmentation via Multi-task Network Cascades（2016CVPR，2015MSCOCO语义分割冠军）,Mask-RCNN
- Aggregated Residual Transformations for Deep Neural Networks（2017CVPR，ResNeXt）

Why ResNet?

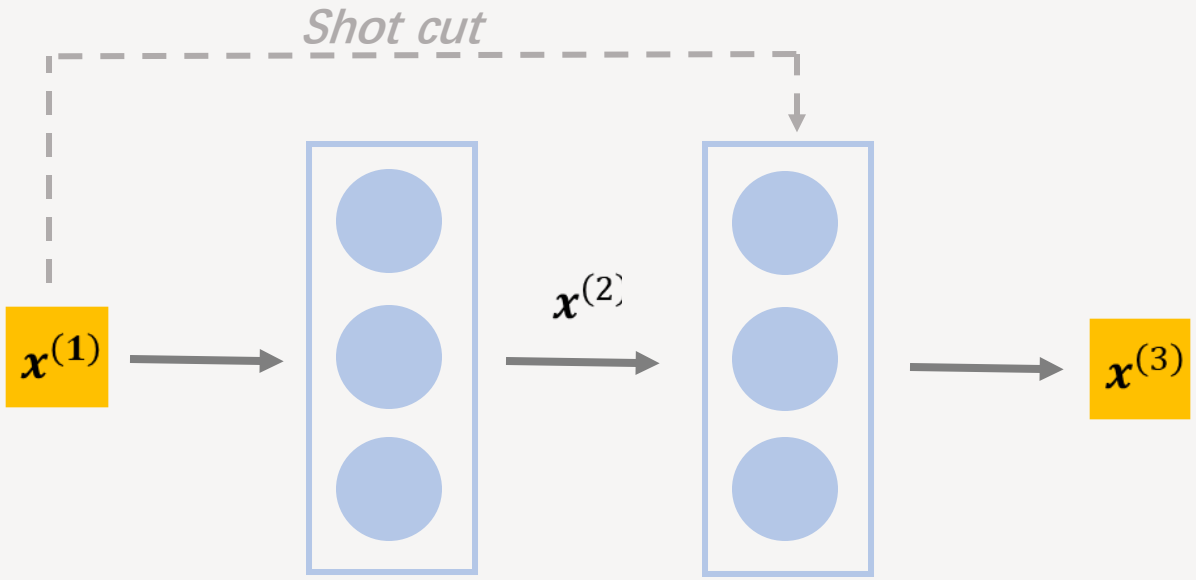
思考1

网络的加深可以提取更多的特征，达到更好的效果。
深度网络变得流行的同时，也带来了一些问题…



Gradient Vanishing

What’s the ResNet?



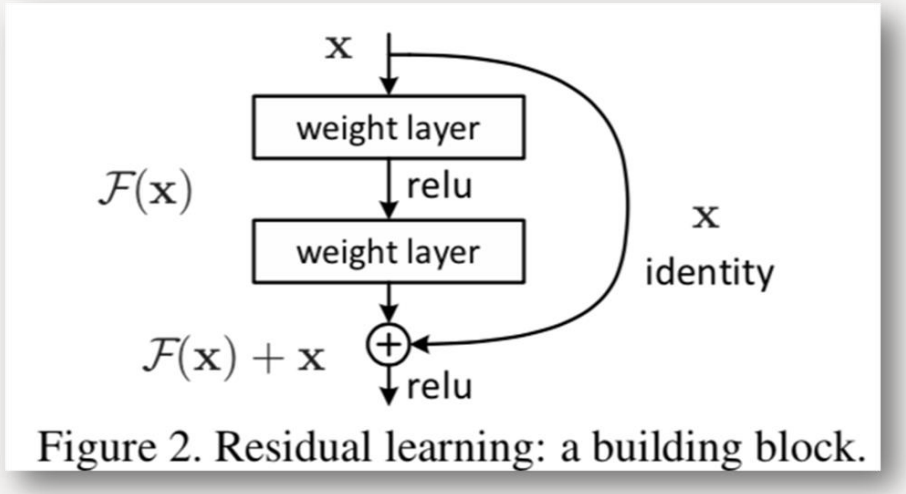
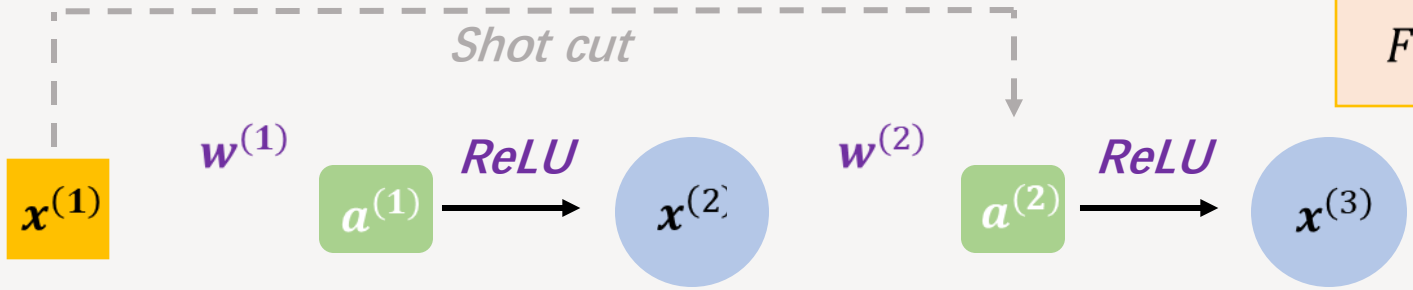


Figure 2. Residual learning: a building block.

$$F(x) = Linear(\sigma(Linear(x)))$$



$$a^{(1)} = w^{(1)}x^{(1)} + b^{(1)}$$

$$x^{(2)} = ReLU(a^{(1)})$$

$$a^{(2)} = w^{(2)}x^{(2)} + b^{(2)}$$

$$x^{(3)} = ReLU(a^{(2)})$$

$$x^{(3)} = ReLU(a^{(2)} + x^{(1)})$$

What’s the ResNet?

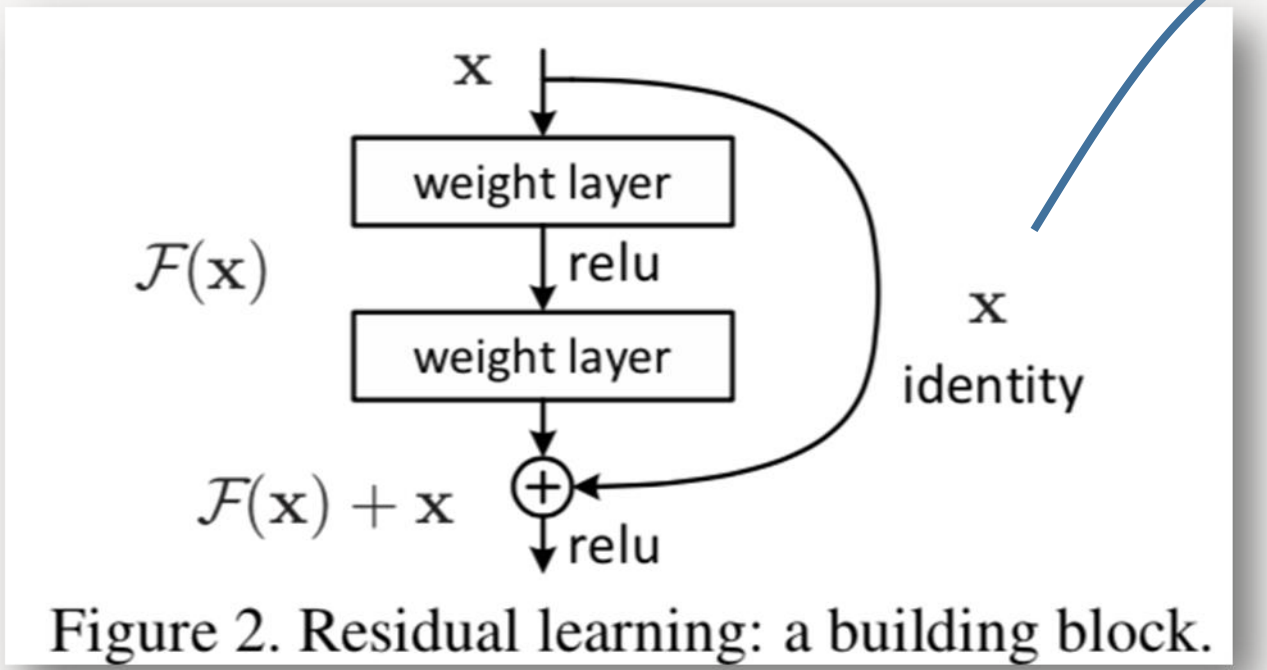


Figure 2. Residual learning: a building block.

Residual block

$$y_l = h(x_l) + F(x_l, W_l)$$

$$x_{l+1} = f(y_l)$$ ReLU

如果 $\boldsymbol{h(x)}$ 是恒等映射的话，那么：

$$x_{l+1} = x_l + F(x_l, W_l)$$

因此循环递归得到以下公式：

$$x_{l+2} = x_{l+1} + F(x_{l+1}, W_{l+1})$$

$$= x_l + F(x_l, w_l) + F(x_{l+1}, W_{l+1})$$

What’s the ResNet?

根据循环递归公式：

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i)$$

Residual block

对于L层的输出而言，可看作任何一个之前的x_l和中间残差块的叠加。
因此，对于整个网络来说，任何一层和该层之前的任何一层都可以看成残差模块，
这样保证了整个网络的前向传播畅通，改进后网络的反向传播公式如下：

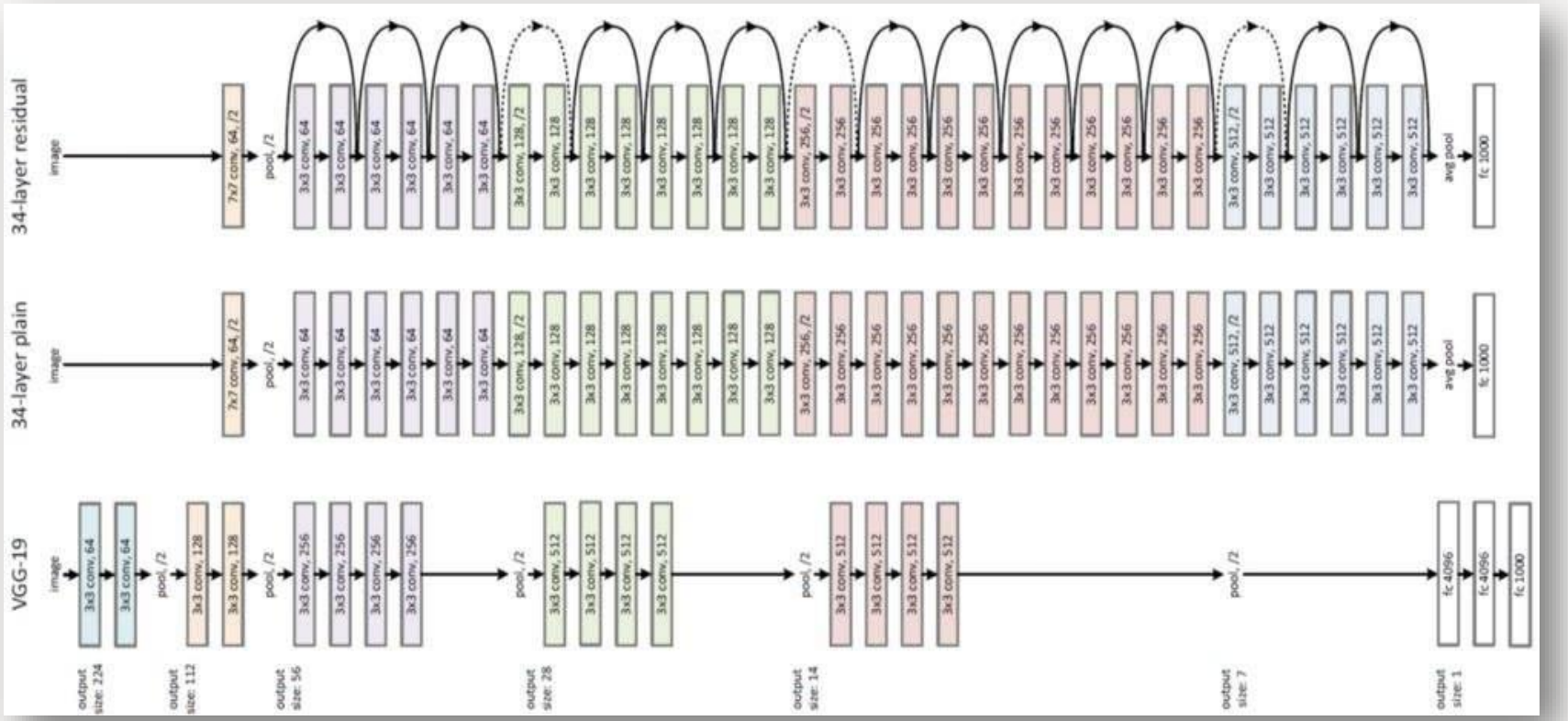
$$\frac{\partial l}{\partial x_l} = \frac{\partial l}{\partial x_L}\frac{\partial x_L}{\partial x_l} = \frac{\partial l}{\partial x_L}\left(1 + \frac{\partial}{\partial x_l}\sum_{i=l}^{L-1} F(x_i, W_i)\right)$$

链式法则的累乘变成了累加，
保证了梯度的有效传播。

Why ResNet

思考 2

当某一层已经达到了最佳状态，剩下层应该不做任何改变，自动学成恒等映射(identity mapping)，该网络的浅层解空间应是深度解空间的子集。



网络退化的问题

- 层数的增加会让误差变大，accuracy不降反升
- 冗余的层很难学习恒等映射

残差网络：深度网络至少和浅网络实现相同性能，让后面的层能更好的实现恒等映射。

How it works

x为浅层输出

$F(x)$为深浅之间的变换

$H(x)$为深层输出

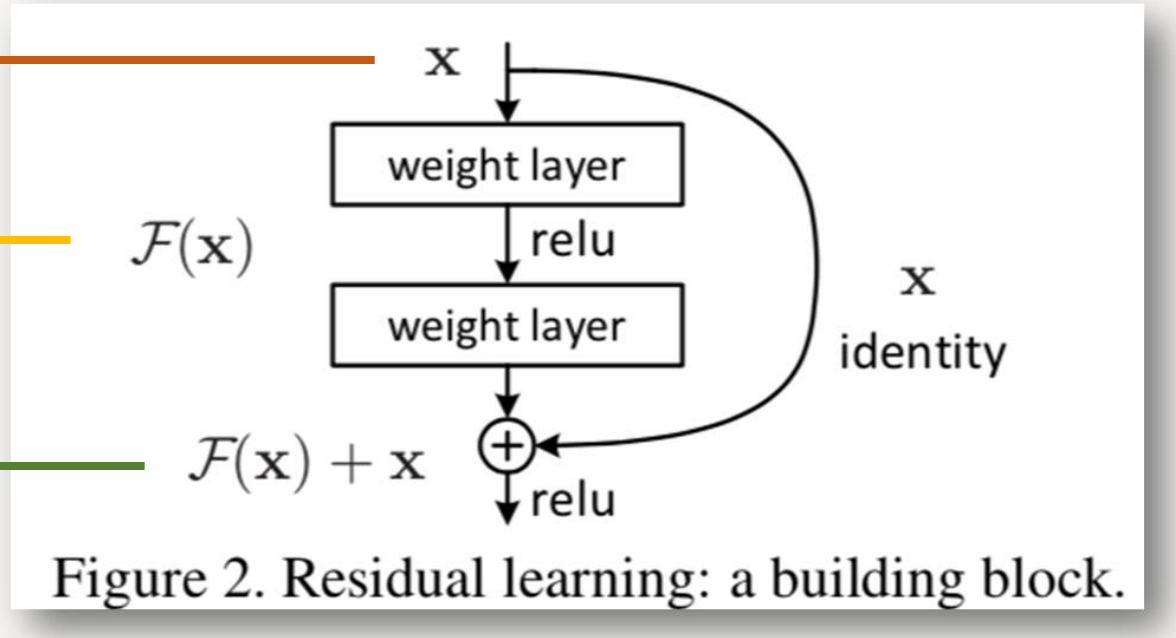


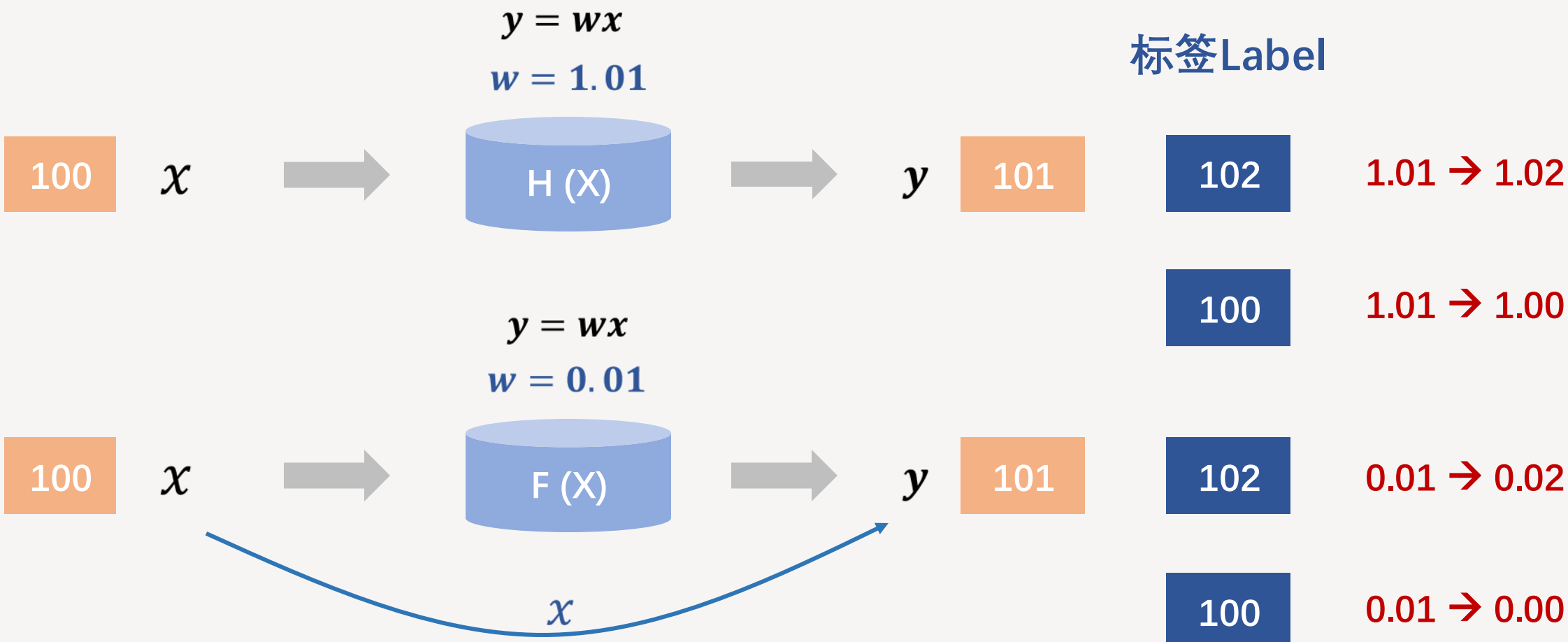
Figure 2. Residual learning: a building block.

当浅层的特征已经足够成熟，
任何对于它的改变都会让Loss变大，
F(x)会自动趋向于学成0，实现恒等映射。

ResNet模块将输出分成 F(x)+x 两部分：
原任务：学习x → H(x) 的恒等映射
现任务：学习x和H(x)之间的残差——F(x)=0

How it works

换个角度理解 Residual会减小模块中参数的值，让参数对反向传导的损失更敏感。



ResNet Structure

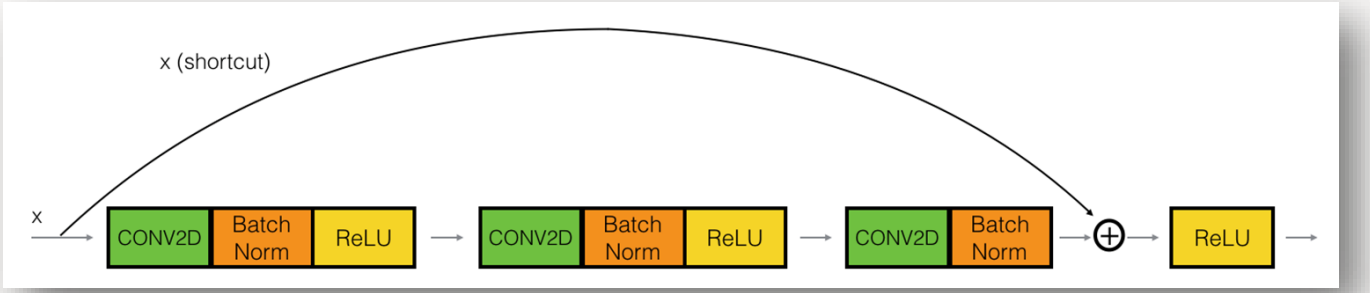
layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
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

ResNet设计非常巧妙：加深网络层数？担心梯度退化？不想增加复杂度？——都没问题！

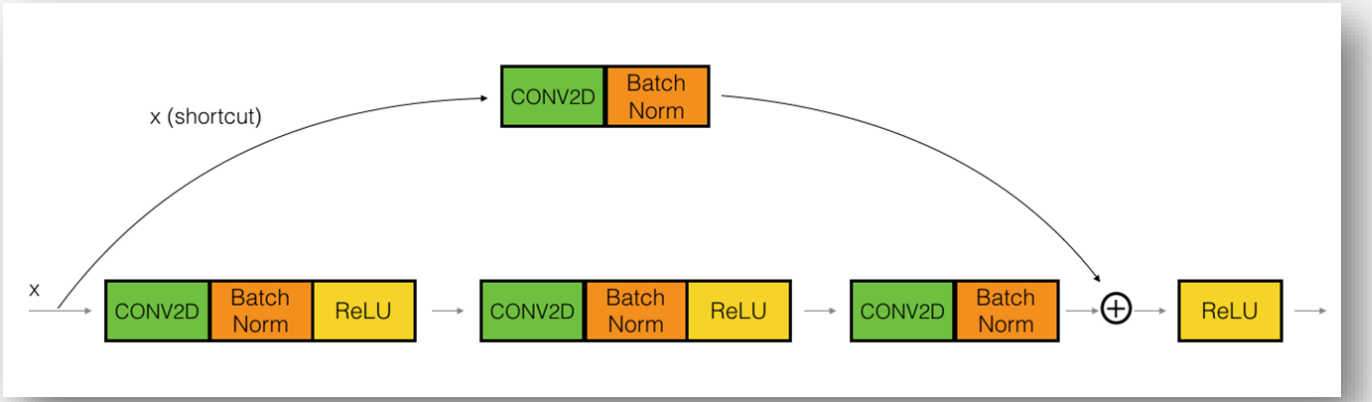
ResNet in CNN

卷积运算
中的
残差模块

The identity block



The convolutional block



多选题 5分

设置

ResNet残差网络的优点主要有：

A 帮助梯度传播，缓解梯度消失

B 虽然增加了参数，但简化了学习过程，提升了网络性能

C 使用跳层连接，缓解网络性能退化问题

提交

单选题 5分

设置

ResNet中的残差块到底指的是哪一部分?

A A框的部分

B B框的部分

C C框的部分

D 以上都不是

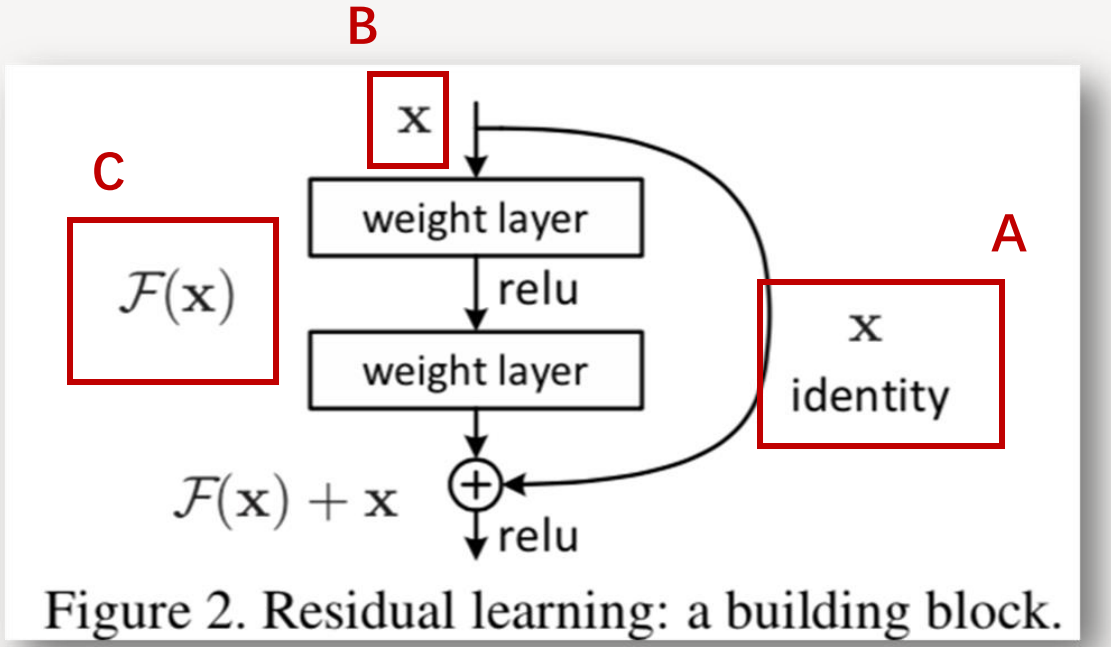


Figure 2. Residual learning: a building block.

提交

·Thanks·

学生创新中心：肖雄子彦