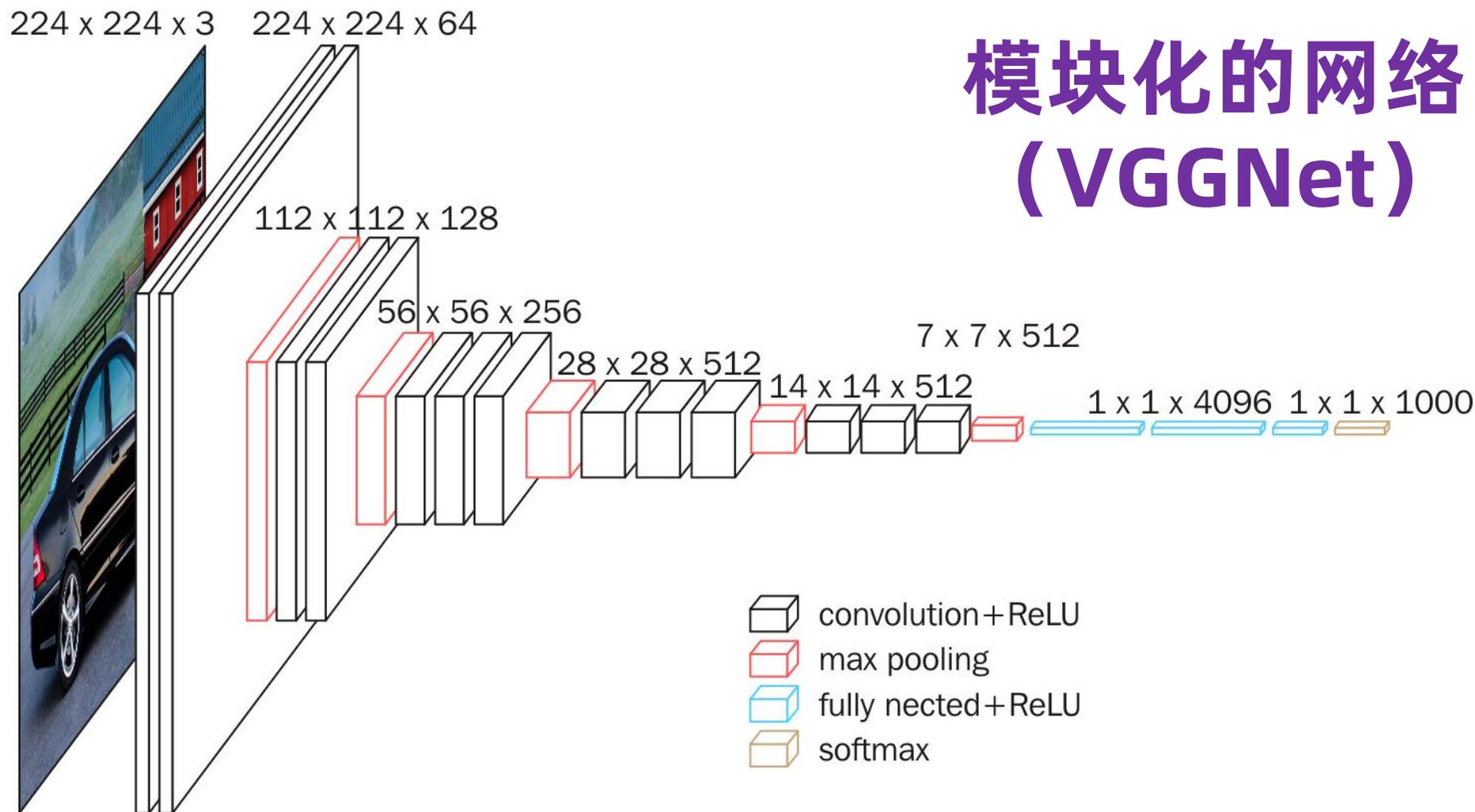


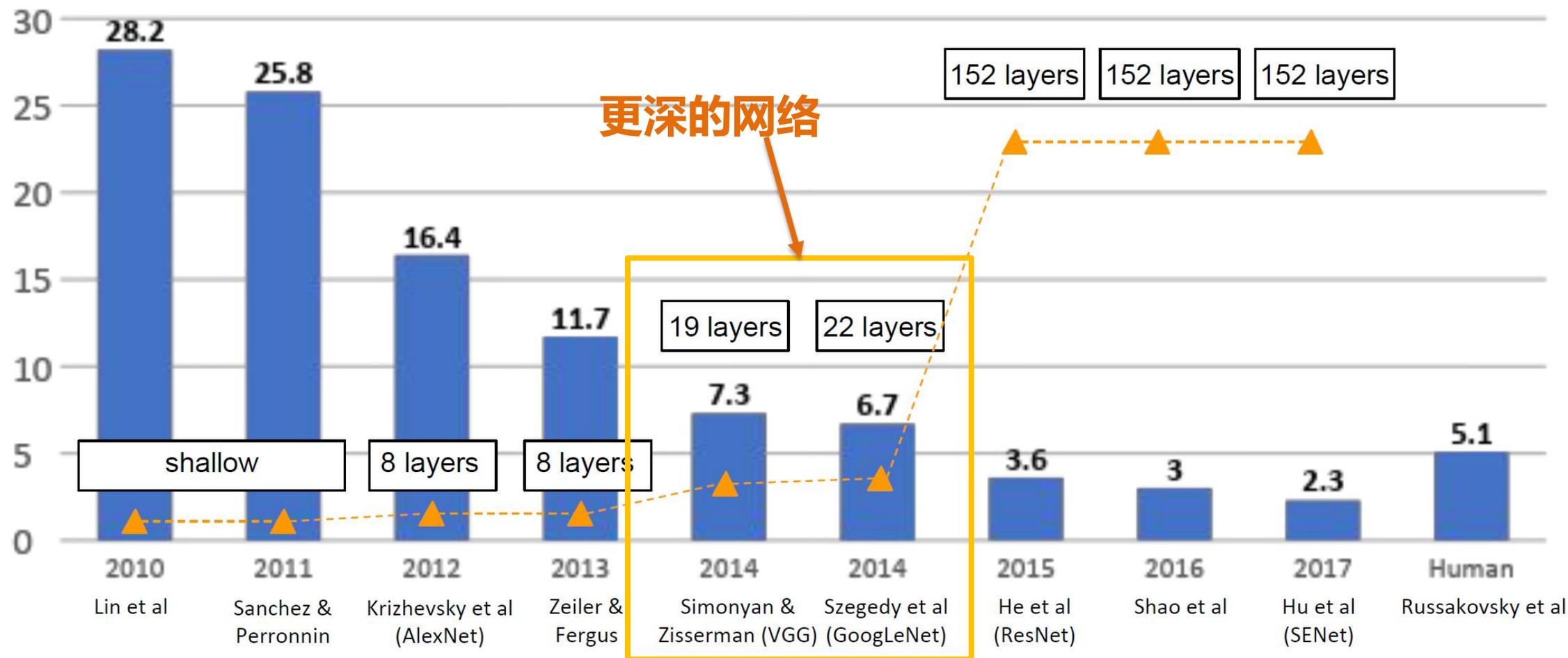
第07章 现代卷积神经网络

欧新宇



模块化的网络 (VGGNet)

ImageNet 大规模视觉识别挑战赛(ILSVRC)





VGG的特点

VGG的特点

简单就是王道 - 基于模块的设计

AlexNet被认为是更大、更深的LeNet，通过深度化实现了更好的分类精度。

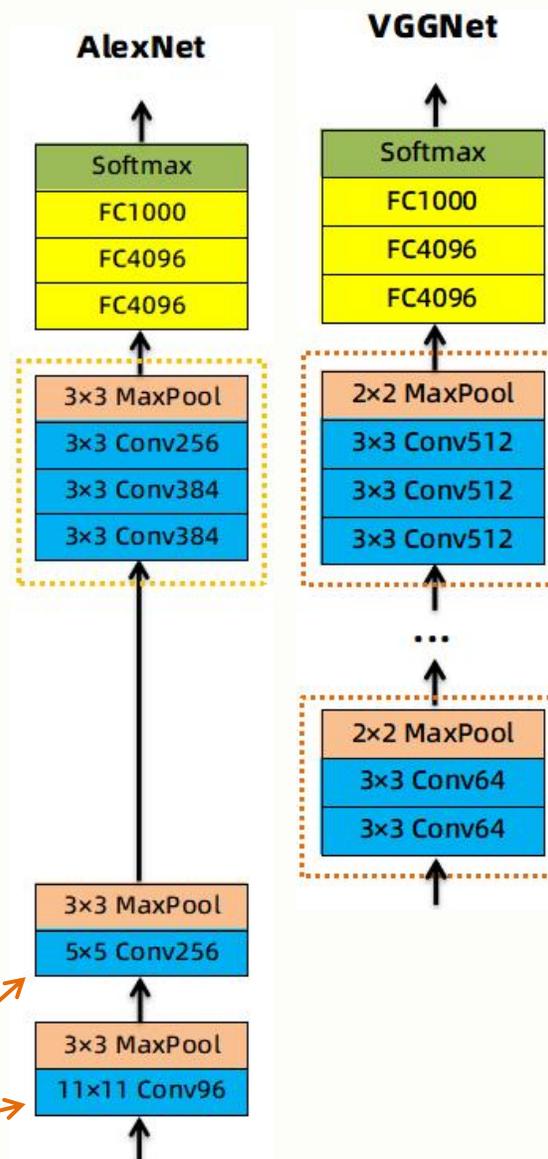
Q: 是否还能通过继续加深模型，实现精度的进一步提升？

A1: 堆叠更多的全连接层 → 太昂贵

A2: 堆叠更多的卷积层 → 超参数难以确定

A3: 精心的网络设计方法 → 设计难度较大 → GoogLeNet

A4: 模块化地进行卷积层组合 → 简单、使用 → VGG



不规则的结构

VGG的特点

小核心也能有大视野

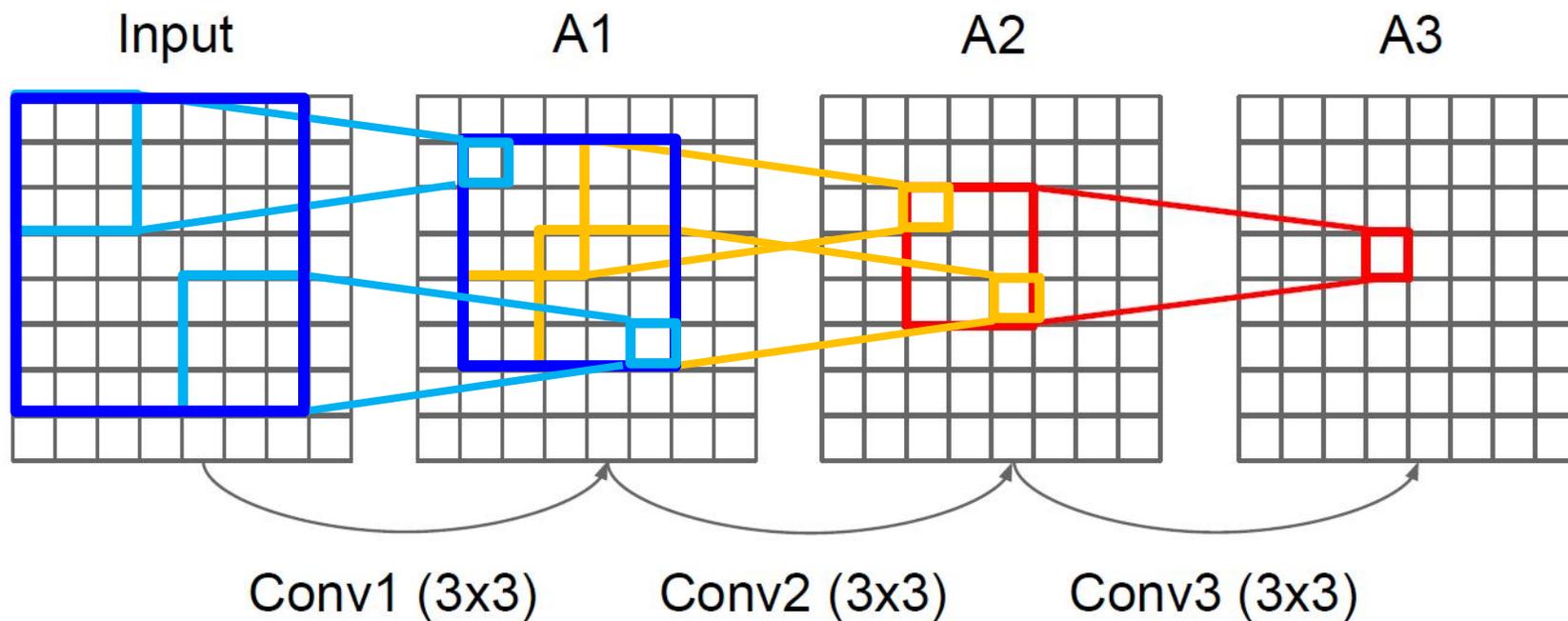
- 全部采用 3×3 的卷积核
 - 增加了更多的非线性，提高了网络的拟合/表达能力
 - 减少了参数的数量
- 采用堆叠卷积层的方法构建卷积层组，来保证较大的感受野 (receptive field)
- 全部使用 2×2 的最大池化核

Question: 3×3 的小卷积核是否能和原来的 7×7 的卷积核具有相同有效的感知域?

VGG的特点

小核心也能有大视野

Q: 3×3 的小卷积核是否能和 7×7 的卷积核具有相同有效的感知域?



- 两个 3×3 的卷积层有和 1 个 5×5 的卷积层相同的感受野
- 三个 3×3 的卷积层有和 1 个 7×7 的卷积核相同的感受野
- 四个 3×3 的卷积层有和 1 个 9×9 的卷积核相同的感受野

VGG16	VGG19
Softmax	Softmax
FC1000	FC1000
FC4096	FC4096
FC4096	FC4096
2x2 MaxPool	2x2 MaxPool
3x3 Conv512	3x3 Conv512
3x3 Conv512	3x3 Conv512
3x3 Conv512	3x3 Conv512
2x2 MaxPool	3x3 Conv512
3x3 Conv512	2x2 MaxPool
3x3 Conv512	3x3 Conv512
3x3 Conv512	3x3 Conv512
2x2 MaxPool	3x3 Conv512
3x3 Conv256	3x3 Conv512
3x3 Conv256	2x2 MaxPool
3x3 Conv256	3x3 Conv256
2x2 MaxPool	3x3 Conv256
3x3 Conv128	3x3 Conv256
3x3 Conv128	3x3 Conv256
2x2 MaxPool	2x2 MaxPool
3x3 Conv64	3x3 Conv128
3x3 Conv64	3x3 Conv128
Input	2x2 MaxPool
	3x3 Conv64
	3x3 Conv64
	Input

VGG的特点

更多的通道、更强的特征、更少的参数

堆叠小卷积核可以实现大卷积核相同的感受野，但是

堆叠小卷积核到底有什么好处呢？

- 更强的非线性：1 to 3 -> 更深的网络
- 更少的参数： $3*(3^2C)=27C$ vs. $7^2C=49C$ ，C通道数
- 更多的通道数：更多的特征
- 更小的卷积核：更大的特征图，更多的信息

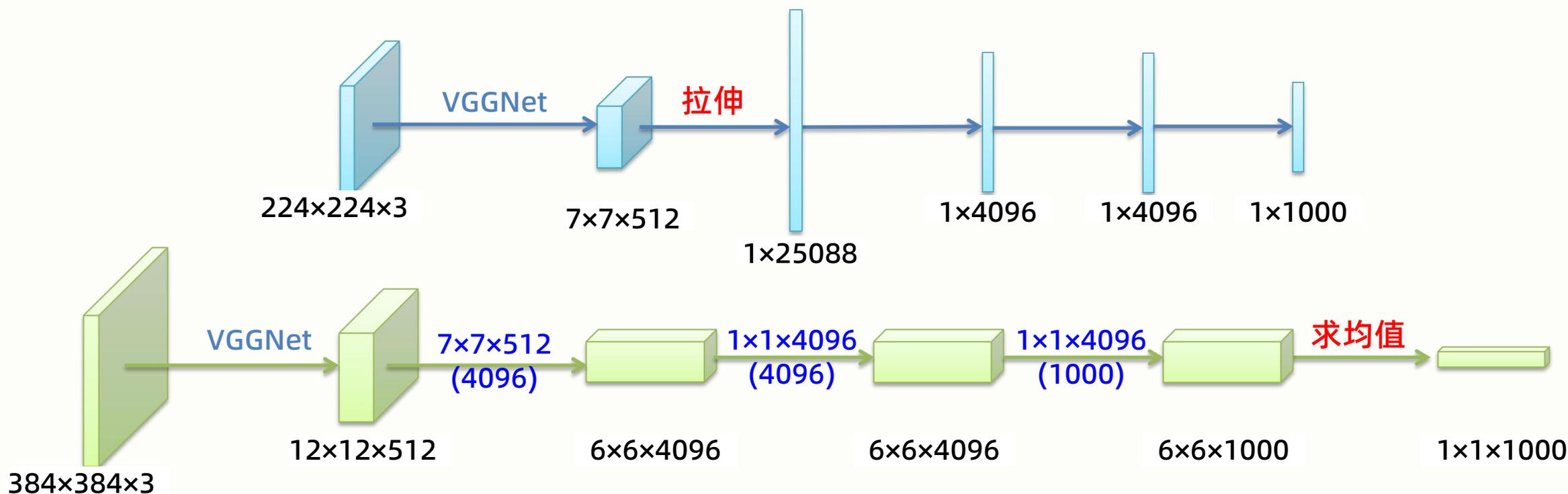
AlexNet	VGG16	VGG19
Softmax	Softmax	Softmax
FC1000	FC1000	FC1000
FC4096	FC4096	FC4096
FC4096	FC4096	FC4096
3×3 MaxPool	2×2 MaxPool	2×2 MaxPool
3×3 Conv256	3×3 Conv512	3×3 Conv512
3×3 Conv384	3×3 Conv512	3×3 Conv512
3×3 Conv384	3×3 Conv512	3×3 Conv512
3×3 MaxPool	2×2 MaxPool	3×3 Conv512
5×5 Conv256	3×3 Conv512	2×2 MaxPool
3×3 MaxPool	3×3 Conv512	3×3 Conv512
11×11 Conv256	3×3 Conv512	3×3 Conv512
Input	2×2 MaxPool	3×3 Conv512
	3×3 Conv256	3×3 Conv512
	3×3 Conv256	2×2 MaxPool
	3×3 Conv256	3×3 Conv256
	2×2 MaxPool	3×3 Conv256
	3×3 Conv128	3×3 Conv256
	3×3 Conv128	3×3 Conv256
	2×2 MaxPool	2×2 MaxPool
	3×3 Conv64	3×3 Conv128
	3×3 Conv64	3×3 Conv128
	Input	2×2 MaxPool
		3×3 Conv64
		3×3 Conv64
		Input

VGG的特点

支持任意尺度的测试

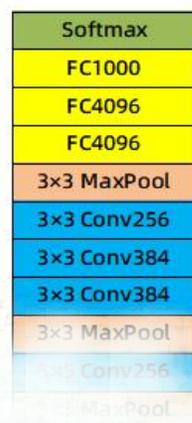
无论是Alexnet还是VGGNet都有一个问题，即无法实现任意尺度的训练/推理。

- 卷积层和全连接层的运算都是矩阵点乘，因此运算规则上没有发生变化
- 非标准尺度的输出是一个Scoremap，而不是一个值。需要对其求平均才能获得one-hot 标签向量

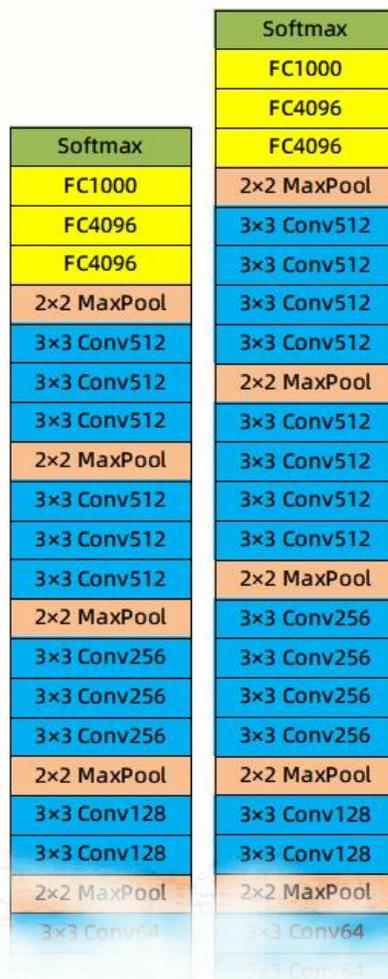


VGG的特点

- 8层的AlexNet
- 11×11 , 5×5 , 3个 3×3 的卷积核
- 16.4% top 5 错误率 in ILSVRC12



简单就是王道



与Alexnet一样，VGG同样采取输入层-卷积层-全连接层-输出层的结构。形式上包含5个卷积层组，3个全连接层和1个Softmax输出层，每个卷积层组后面紧跟一个Max-Pooling最大池化层，所有隐层的激活函数都是ReLU。

- 16~19层VGGNet
- 5组 3×3 的卷积核

16.8% top 5 错误率 in ILSVRC14



VGG架构的详细分析

VGG架构的详细分析

VGGNet的6种网络体系结构

论文中总共设计了6种不同的网络结构，都是由**5个**卷积层组，**3个**全连接层组成。

- 所有的卷积核都是**3×3卷积**，且**步长、填充都为1**；
- 每一组卷积核内的**通道数都是相同的**；
- 每个卷积核组后紧跟一个**maxpooling**层；
- 从A到E卷积层组内堆叠卷积数量逐渐增多，从11到19，其中D为著名的**VGG16**，E为著名的**VGG19**。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
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
maxpool					
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
maxpool					
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
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

VGG16 VGG19

Conv3_1
Conv3_2
Conv3_3
Conv3_4

VGG架构的详细分析

INPUT: [224x224x3] memory: $224*224*3=150K$ params: 0
 CONV3-64: [224x224x64] memory: $224*224*64=3.2M$ params: $(3*3*3)*64 = 1,728$
 CONV3-64: [224x224x64] memory: $224*224*64=3.2M$ params: $(3*3*64)*64 = 36,864$
 POOL2: [112x112x64] memory: $112*112*64=800K$ params: 0
 CONV3-128: [112x112x128] memory: $112*112*128=1.6M$ params: $(3*3*64)*128 = 73,728$
 CONV3-128: [112x112x128] memory: $112*112*128=1.6M$ params: $(3*3*128)*128 = 147,456$
 POOL2: [56x56x128] memory: $56*56*128=400K$ params: 0
 CONV3-256: [56x56x256] memory: $56*56*256=800K$ params: $(3*3*128)*256 = 294,912$
 CONV3-256: [56x56x256] memory: $56*56*256=800K$ params: $(3*3*256)*256 = 589,824$
 CONV3-256: [56x56x256] memory: $56*56*256=800K$ params: $(3*3*256)*256 = 589,824$
 POOL2: [28x28x256] memory: $28*28*256=200K$ params: 0
 CONV3-512: [28x28x512] memory: $28*28*512=400K$ params: $(3*3*256)*512 = 1,179,648$
 CONV3-512: [28x28x512] memory: $28*28*512=400K$ params: $(3*3*512)*512 = 2,359,296$
 CONV3-512: [28x28x512] memory: $28*28*512=400K$ params: $(3*3*512)*512 = 2,359,296$
 POOL2: [14x14x512] memory: $14*14*512=100K$ params: 0
 CONV3-512: [14x14x512] memory: $14*14*512=100K$ params: $(3*3*512)*512 = 2,359,296$
 CONV3-512: [14x14x512] memory: $14*14*512=100K$ params: $(3*3*512)*512 = 2,359,296$
 CONV3-512: [14x14x512] memory: $14*14*512=100K$ params: $(3*3*512)*512 = 2,359,296$
 POOL2: [7x7x512] memory: $7*7*512=25K$ params: 0
 FC: [1x1x4096] memory: 4096 params: $7*7*512*4096 = 102,760,448$
 FC: [1x1x4096] memory: 4096 params: $4096*4096 = 16,777,216$
 FC: [1x1x1000] memory: 1000 params: $4096*1000 = 4,096,000$

大多数内存
消耗在低层
的卷积层

大多数参数在
高层的FC层

总存储: $24M * 4 \text{ bytes} \approx 96MB$ / image (only forward! $\sim *2$ for bwd)

总参数: 138M parameters

VGG架构的详细分析

不同结构的性能对比

1. LRN层无性能增益 (A-LRN)

VGG作者通过网络A-LRN发现, AlexNet曾用到的LRN层并没有带来性能的提升, 因此在其它组的网络中均没再出现LRN层。

2. 多尺度训练可以降低错误率 (C、D、E)

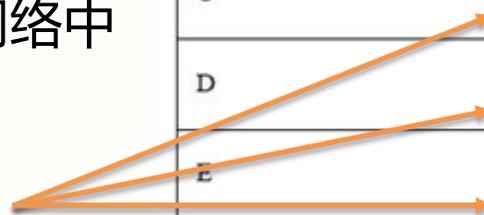
3. 随着深度增加, 分类性能逐渐提高 (A、B、C、D、E)

从11层的A到19层的E, 网络深度增加对top1和top5的错误率下降很明显。

4. 多个小卷积核比单个大卷积核性能好 (B)

作者用B和自己一个不在实验组里的较浅网络比较, 较浅网络用conv5x5来代替B的两个conv3x3, 结果显示多个小卷积核比单个大卷积核效果要好。

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Z)		
A	256	256	29.6	10.4
A-LRN	256	256	29.7	10.5
B	256	256	28.7	9.9
C	256	256	28.1	9.4
	384	384	28.1	9.3
D	[256;512]	384	27.3	8.8
	256	256	27.0	8.8
	384	384	26.8	8.7
E	[256;512]	384	25.6	8.1
	256	256	27.3	9.0
	384	384	26.9	8.7
	[256;512]	384	25.5	8.0



VGGNet架构的详细分析

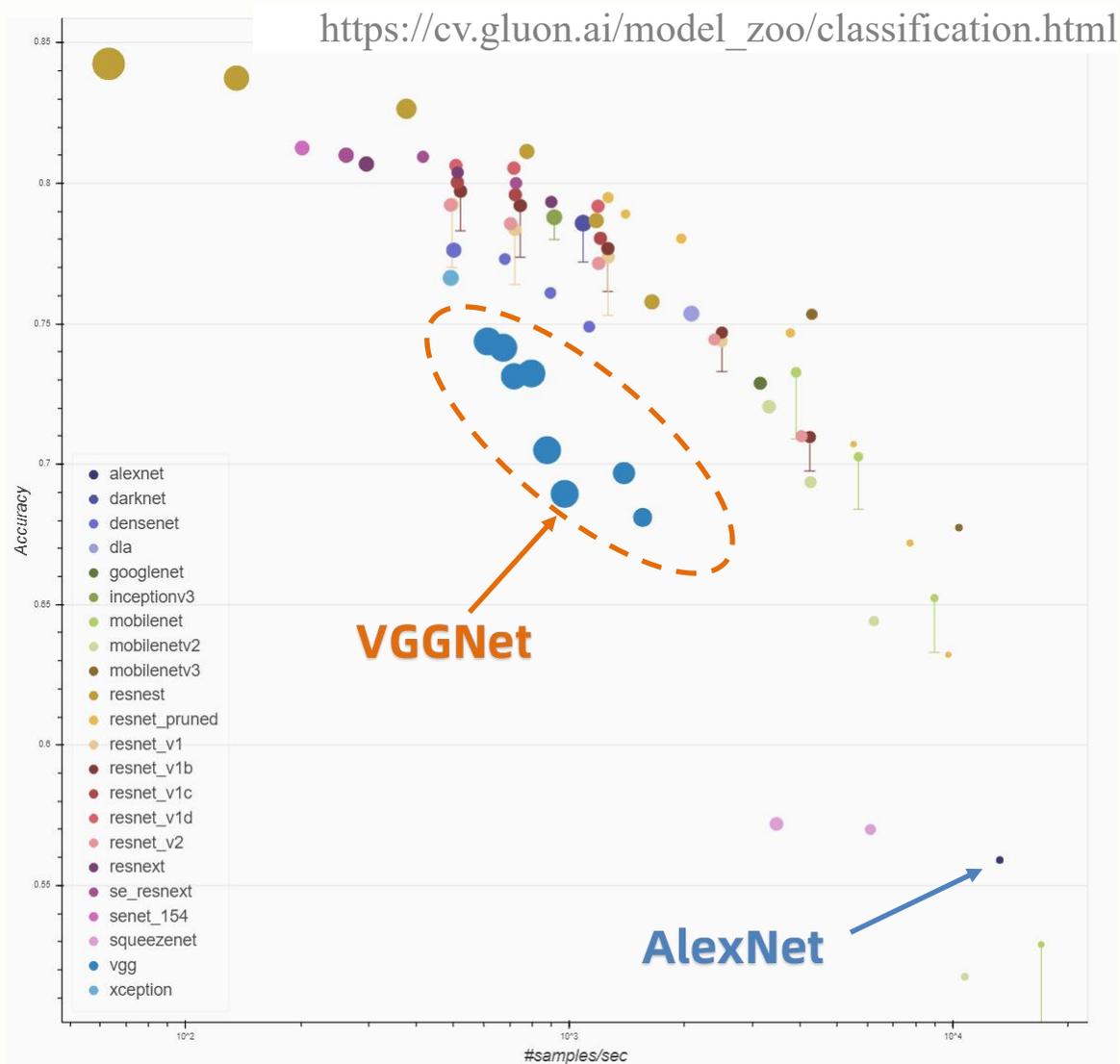
Visualization of Inference on GluonCV

● AlexNet

- 精度：0.559
- 内存：202Mb
- 速度：13270 张/秒

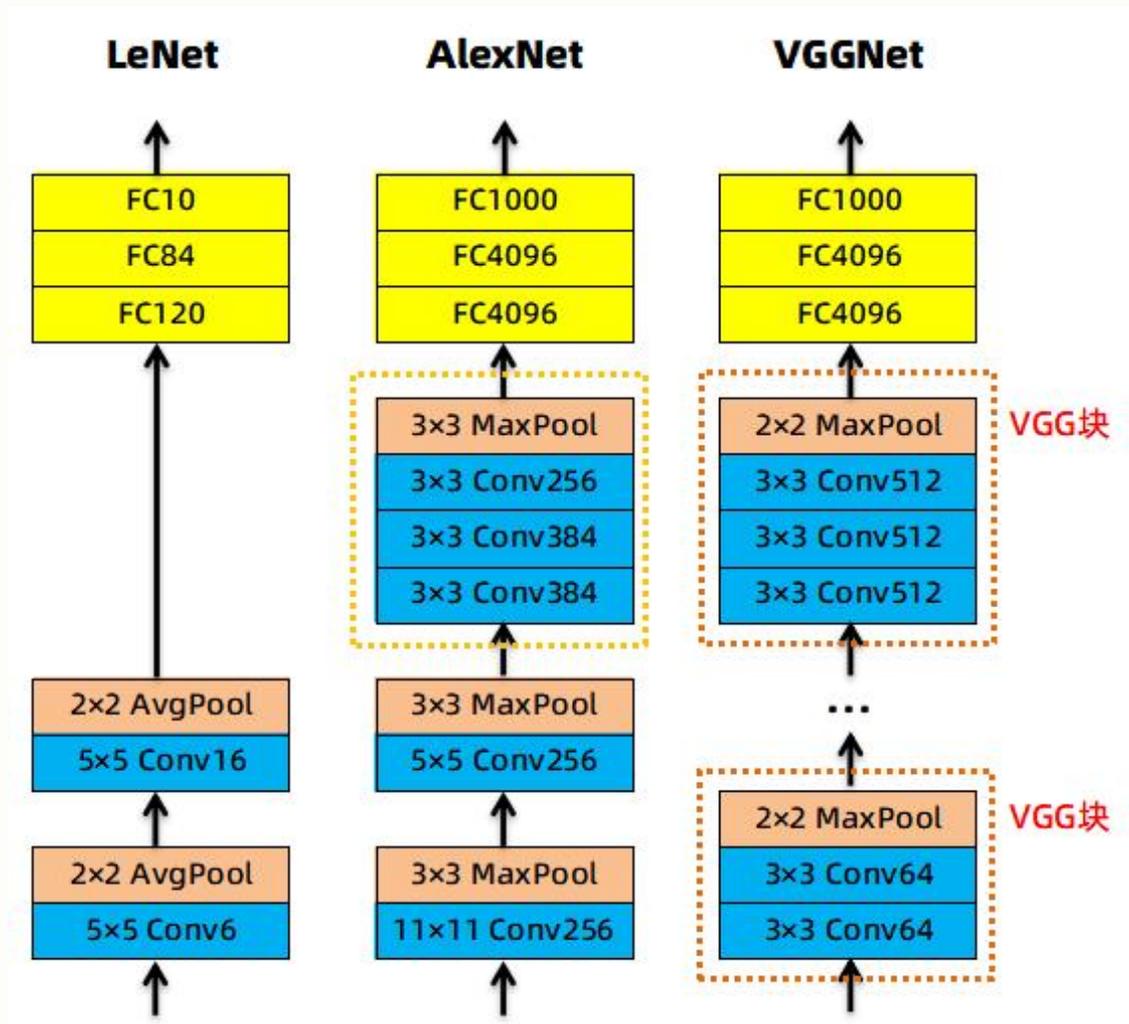
● VGG16

- 精度：0.732
- 内存：3422Mb
- 速度：797 张/秒

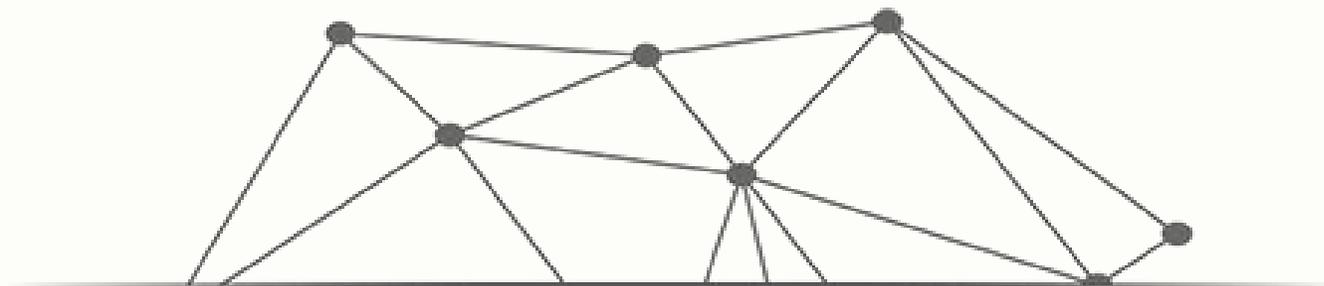


模块化的网络VGGNet

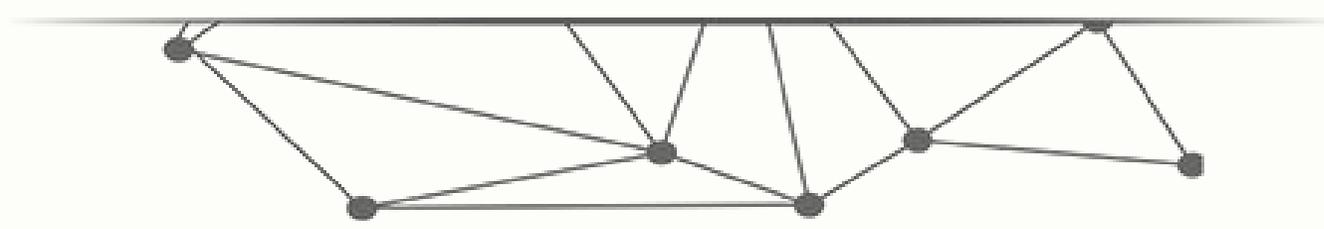
小结



- VGG使用模块化的设计思想，通过重复堆叠相同的卷积块来构建深度神经网络；
- 不同数量卷积块的组合可以得到不同复杂度的网络变种；
- 深度的增加能够有效提高性能；
- 通过堆叠，小卷积核能够具有和大卷积核相同的感受野，并增加非线性特性；
- 使用卷积替代全连接层，可适应不同尺度的样本；
- 性价比最佳模型：VGG16，从头到尾只有3x3卷积与2x2池化，简洁优美，性能优异。



课堂互动



读万卷书 行万里路 只为最好的修炼



QQ: 14777591 (宇宙骑士)

Email: ouxinyu@alumni.hust.edu.cn

Website: <http://ouxinyu.cn>

Tel: 18687840023

地址: 安宁校区 诚远楼201

南院 智能应用研究院A306-2