[Simonyan and Zisserman, 2014]

Small filters, Deeper networks

8 layers (AlexNet)
-> 16 - 19 layers (VGG16Net)

Only 3x3 CONV stride 1, pad 1 and 2x2 MAX POOL stride 2

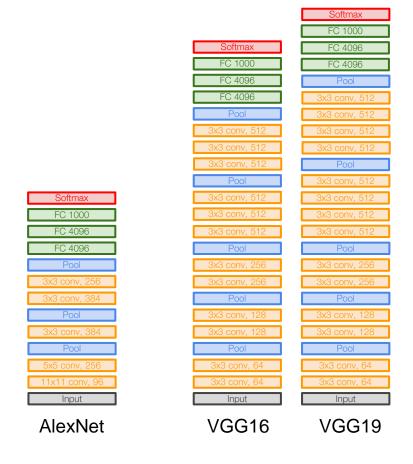
11.7% top 5 error in ILSVRC'13 (ZFNet)

-> 7.3% top 5 error in ILSVRC'14



[Simonyan and Zisserman, 2014]

Q: Why use smaller filters? (3x3 conv)



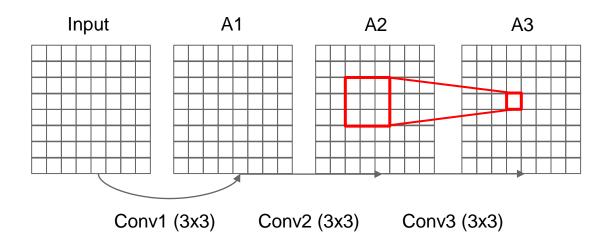
[Simonyan and Zisserman, 2014]

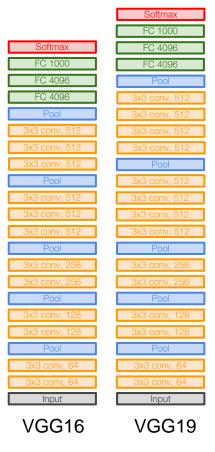
Q: Why use smaller filters? (3x3 conv)

Stack of three 3x3 conv (stride 1) layers has same **effective receptive field** as one 7x7 conv layer

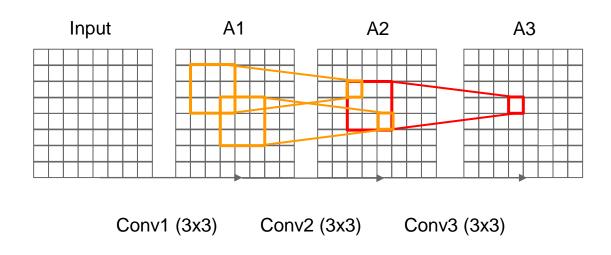


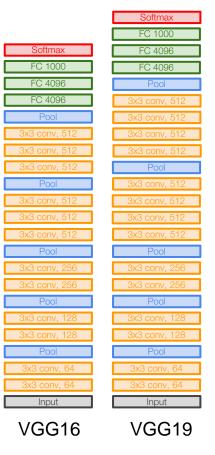
[Simonyan and Zisserman, 2014]



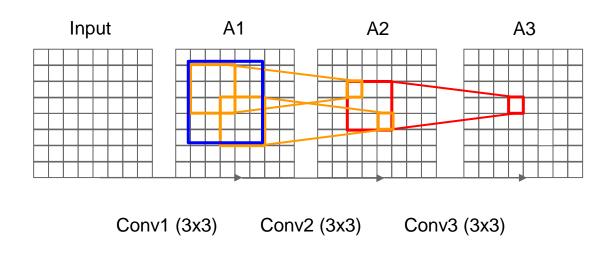


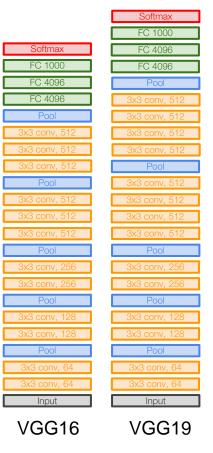
[Simonyan and Zisserman, 2014]



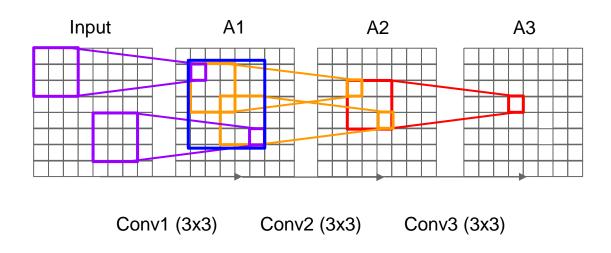


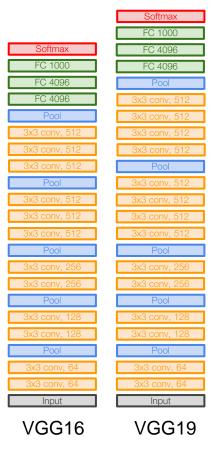
[Simonyan and Zisserman, 2014]



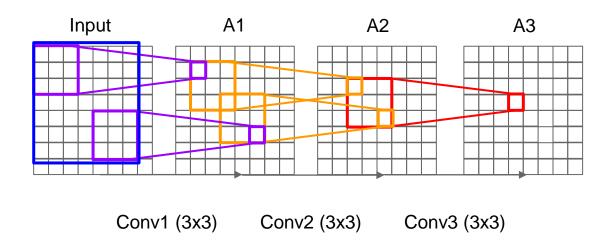


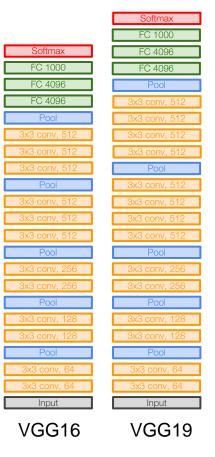
[Simonyan and Zisserman, 2014]





[Simonyan and Zisserman, 2014]





[Simonyan and Zisserman, 2014]

Q: Why use smaller filters? (3x3 conv)

Stack of three 3x3 conv (stride 1) layers has same **effective receptive field** as one 7x7 conv layer

[7x7]



FC 1000

[Simonyan and Zisserman, 2014]

Q: Why use smaller filters? (3x3 conv)

Stack of three 3x3 conv (stride 1) layers has same **effective receptive field** as one 7x7 conv layer

But deeper, more non-linearities

And fewer parameters: 3 * (3²C²) vs. 7²C² for C channels per layer



INPUT: [224x224x3] memory: 224*224*3=150K params: 0 (not counting biases)	
CONV3-64: [224x224x64] memory: 224*224*64=3.2M params: (3*3*3)*64 = 1,728	Softmax
CONV3-64: [224x224x64] memory: 224*224*64=3.2M params: (3*3*64)*64 = 36,864	FC 1000
POOL2: [112x112x64] memory: 112*112*64=800K params: 0	FC 4096
	FC 4096
CONV3-128: [112x112x128] memory: 112*112*128=1.6M params: (3*3*64)*128 = 73,728	Pool 3x3 conv, 512
CONV3-128: [112x112x128] memory: 112*112*128=1.6M params: (3*3*128)*128 = 147,456	3x3 conv, 512
POOL2: [56x56x128] memory: 56*56*128=400K params: 0	3x3 conv, 512
CONV3-256: [56x56x256] memory: 56*56*256=800K params: (3*3*128)*256 = 294,912	Pool
CONV3-256: [56x56x256] memory: 56*56*256=800K params: (3*3*256)*256 = 589,824	3x3 conv, 512
CONV3-256: [56x56x256] memory: 56*56*256=800K params: (3*3*256)*256 = 589,824	3x3 conv, 512
POOL2: [28x28x256] memory: 28*28*256=200K params: 0	3x3 conv, 512
CONV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*256)*512 = 1,179,648	Pool
CONV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*512)*512 = 2,359,296	3x3 conv, 256
CONV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*512)*512 = 2,359,296	3x3 conv, 256
POOL2: [14x14x512] memory: 14*14*512=100K params: 0	Pool
CONV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512 = 2,359,296	3x3 conv, 128
CONV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512 = 2,359,296	3x3 conv, 128 Pool
CONV3-512: [14x14x512] memory: $14*14*512=100$ K params: $(3*3*512)*512=2,359,296$	3x3 conv, 64
POOL2: [7x7x512] memory: 7*7*512=25K params: 0	3x3 conv, 64
FC: [1x1x4096] memory: 4096 params: 7*7*512*4096 = 102,760,448	Input
FC: [1x1x4096] memory: 4096 params: 4096*4096 = 16,777,216	
	VGG16
FC: [1x1x1000] memory: 1000 params: 4096*1000 = 4,096,000	
Fei-Fei Li, Ehsan Adeli, Zane Durante Lecture 6 - 52	April 17, 2024
	7 (prin 11 ; 202 i

(not counting biases)

```
CONV3-64: [224x224x64] memory: 224*224*64=3.2M params: (3*3*3)*64 = 1,728
                                                                                            FC 1000
CONV3-64: [224x224x64] memory: 224*224*64=3.2M params: (3*3*64)*64 = 36,864
                                                                                            FC 4096
POOL2: [112x112x64] memory: 112*112*64=800K params: 0
                                                                                            FC 4096
CONV3-128: [112x112x128] memory: 112*112*128=1.6M params: (3*3*64)*128 = 73,728
                                                                                            Pool
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
                                                                                            Pool
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
                                                                                          3x3 conv, 512
CONV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*256)*512 = 1,179,648
                                                                                            Pool
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
                                                                                            Pool
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
                                                                                            Pool
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
                                                                                             Input
FC: [1x1x4096] memory: 4096 params: 4096*4096 = 16,777,216
                                                                                          VGG16
FC: [1x1x1000] memory: 1000 params: 4096*1000 = 4,096,000
TOTAL memory: 24M * 4 bytes ~= 96MB / image (for a forward pass)
TOTAL params: 138M parameters
Fei-Fei Li, Ehsan Adeli, Zane Durante
                                                            Lecture 6 - 53
                                                                                            April 17, 2024
```

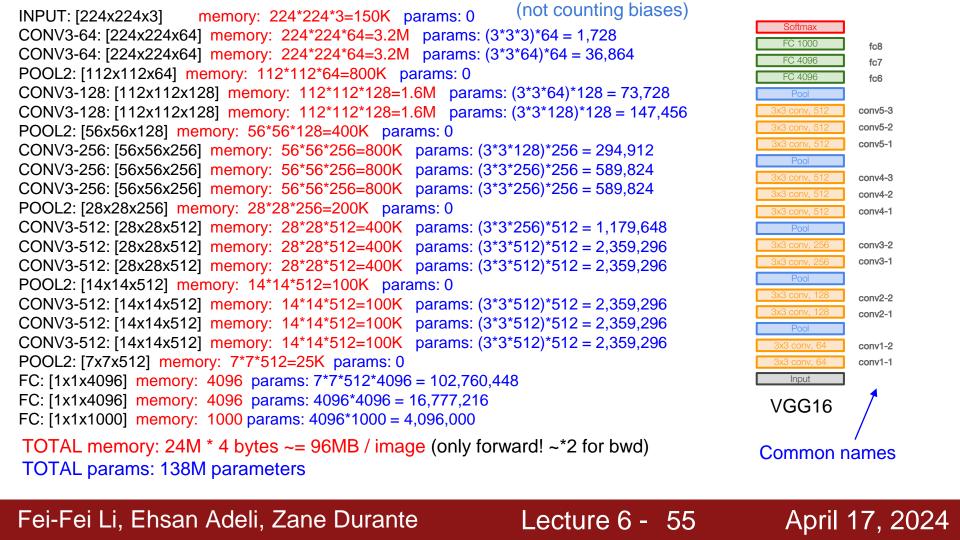
INPUT: [224x224x3] memory: 224*224*3=150K params: 0

(not counting biases)

```
CONV3-64: [224x224x64] memory: 224*224*64=3.2M params: (3*3*3)*64 = 1,728
                                                                                        Note:
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
                                                                                        Most memory is in
CONV3-128: [112x112x128] memory: 112*112*128=1.6M params: (3*3*64)*128 = 73,728
                                                                                        early CONV
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
                                                                                        Most params are
CONV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512=2,359,296
                                                                                        in late FC
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
TOTAL memory: 24M * 4 bytes ~= 96MB / image (only forward! ~*2 for bwd)
TOTAL params: 138M parameters
Fei-Fei Li, Ehsan Adeli, Zane Durante
                                                            Lecture 6 - 54
                                                                                           April 17, 2024
```

INPUT: [224x224x3] memory: 224*224*3=150K params: 0

(not counting biases)



[Simonyan and Zisserman, 2014]

Details:

- ILSVRC'14 2nd in classification, 1st in localization
- Similar training procedure as Krizhevsky 2012
- No Local Response Normalisation (LRN)
- Use VGG16 or VGG19 (VGG19 only slightly better, more memory)
- Use ensembles for best results
- FC7 features generalize well to other tasks



ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners

