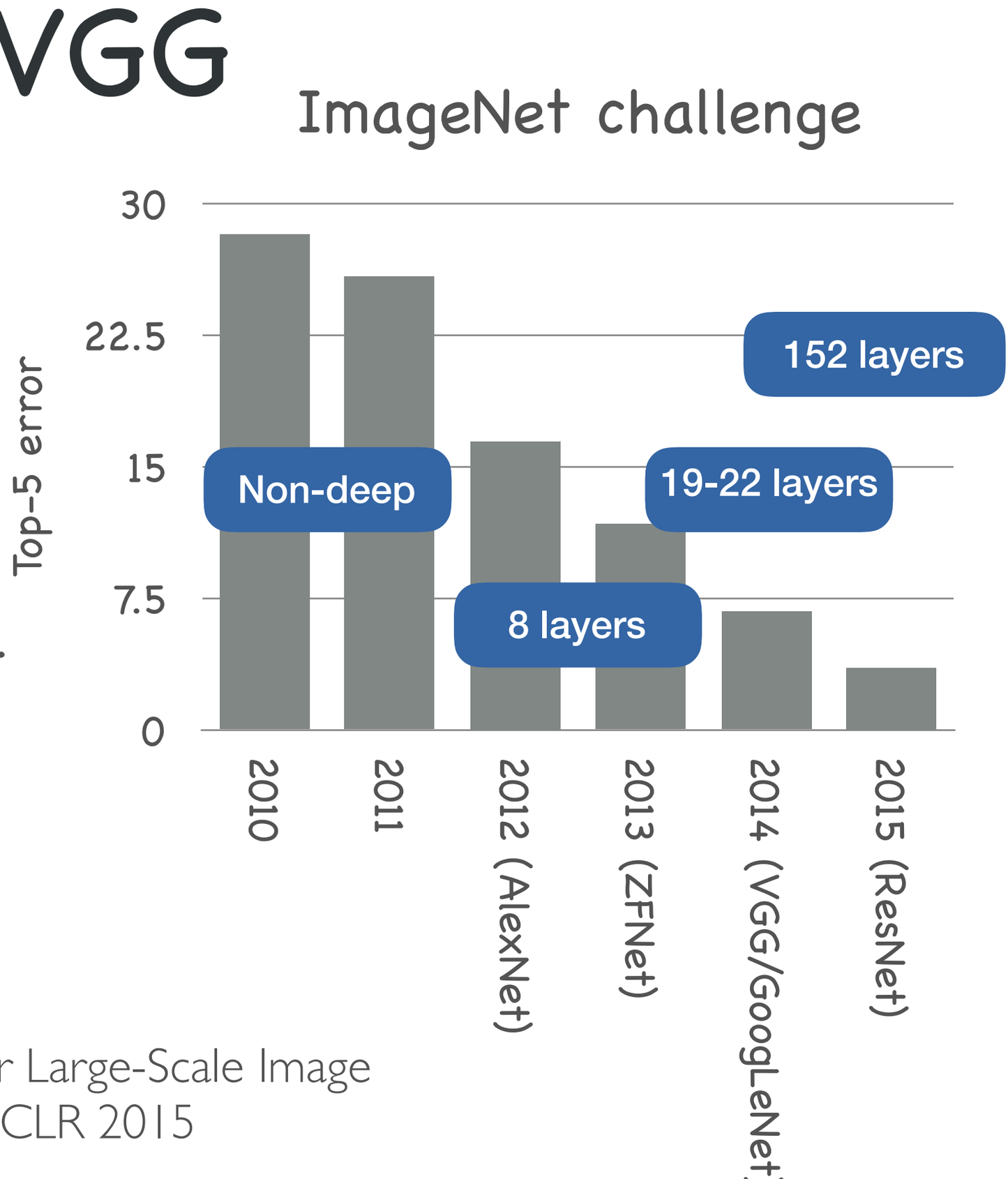
Case Study: VGG

VGG

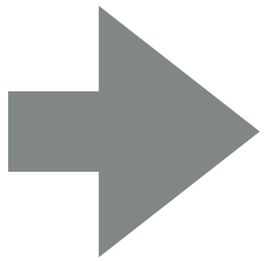
- Deeper AlexNet/ZFNet



Very Deep Convolutional Networks for Large-Scale Image Recognition, Simonyan and Zisserman, ICLR 2015

ZFNet to VGG

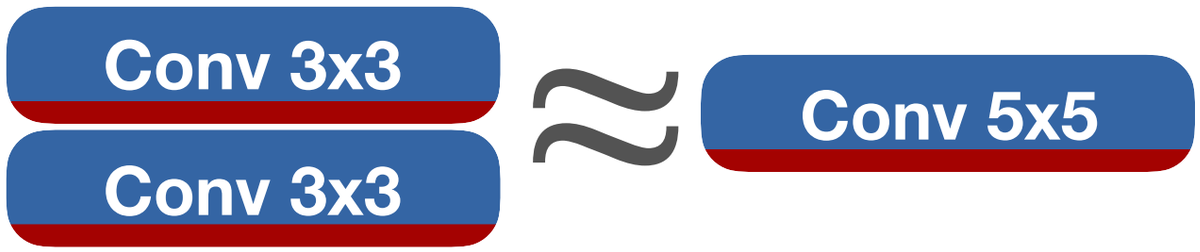
Conv 7x7	stride=2, C=96
LRN	
Max Pool 3x3	stride=2
Conv 5x5	stride=2, C=256
LRN	
Max Pool 3x3	pad=0, stride=2
Conv 3x3	C=512
Conv 3x3	C=1024
Conv 3x3	C=512
Max Pool 3x3	pad=0, stride=2
Linear (4096)	
Linear (4096)	
Linear (1000)	



Conv 3x3	C=64
Conv 3x3	C=64
Max Pool 2x2	stride=2
Conv 3x3	C=128
Conv 3x3	C=128
Max Pool 2x2	stride=2
Conv 3x3	C=256
Conv 3x3	C=256
Conv 3x3	C=256
Max Pool 2x2	stride=2
Conv 3x3	C=512
Conv 3x3	C=512
Conv 3x3	C=512
Max Pool 2x2	stride=2
Conv 3x3	C=512
Conv 3x3	C=512
Conv 3x3	C=512
Max Pool 2x2	stride=2
Linear (4096)	
Linear (4096)	
Linear (1000)	

Insights in VGG

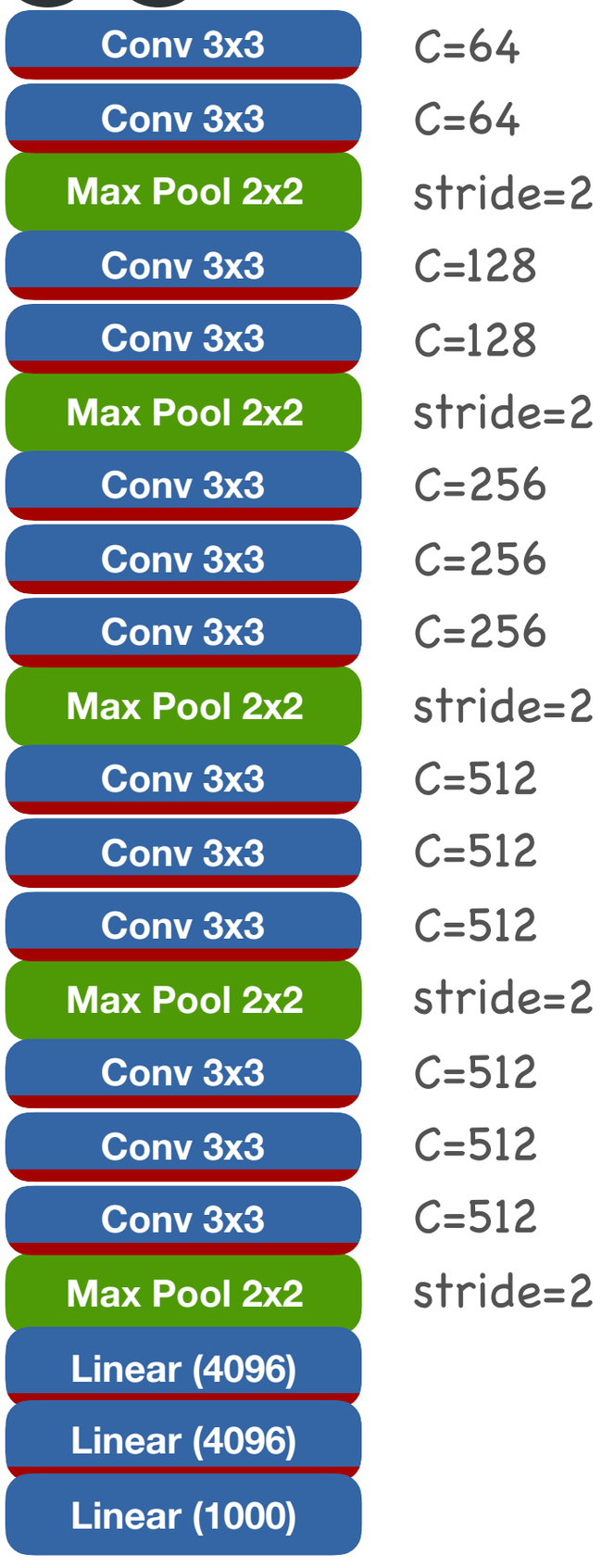
- Why use smaller filters?



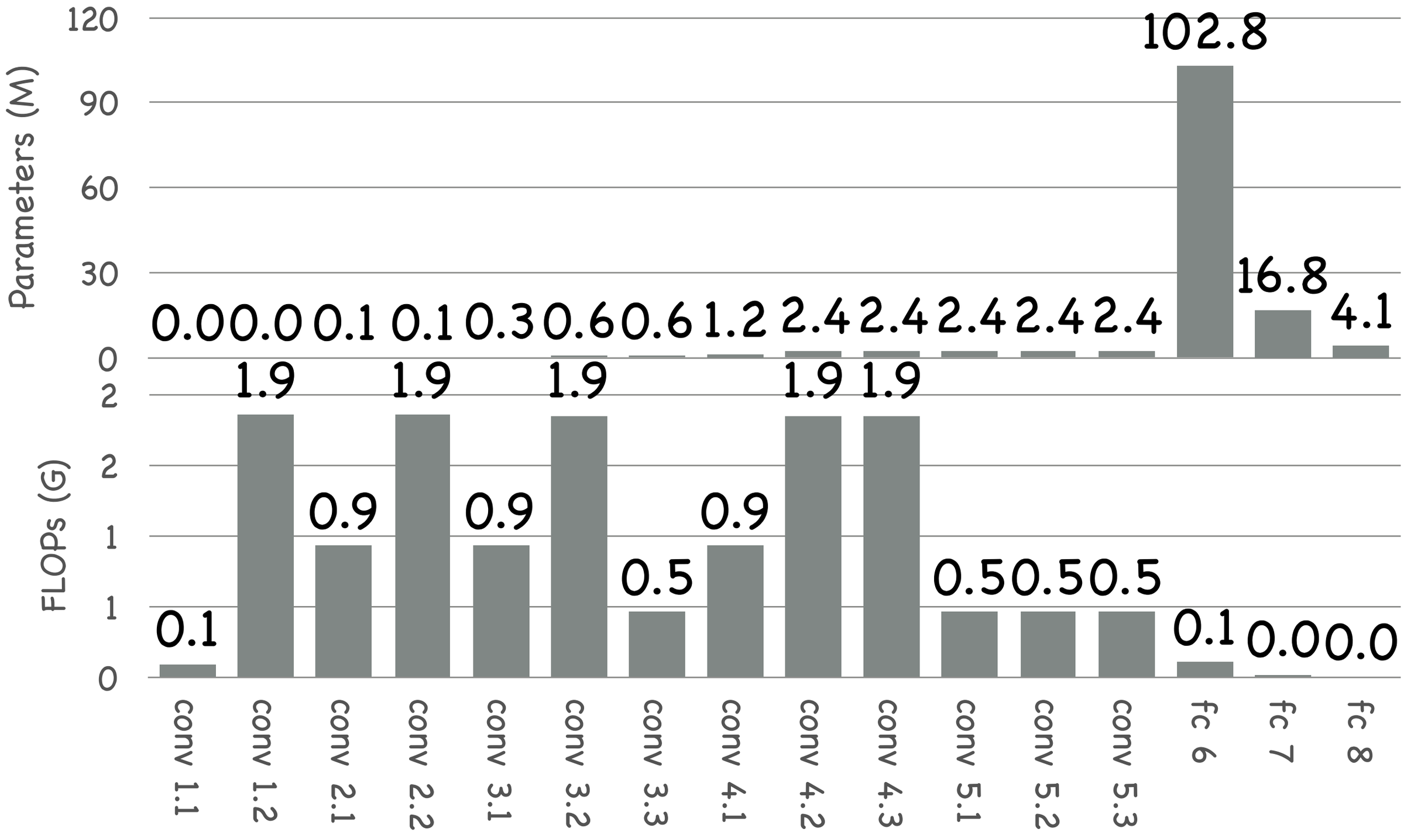
  - Factorization

Training VGG

- Vanishing gradients

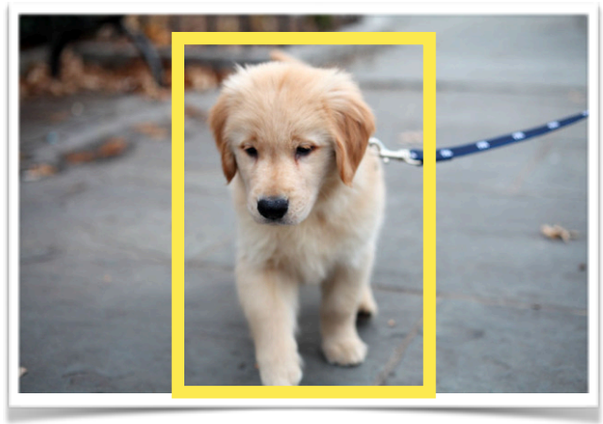


Parameters and computation



VGG

- Generalizes well to other tasks
  - Detection
  - Segmentation
  - Style-transfer (graphics)



Fully Convolutional Networks for Semantic Segmentation, Shelhamer et al., CVPR, 2015
Image style transfer using convolutional neural networks, Gatys et al., CVPR, 2016
Rich feature hierarchies for accurate object detection and semantic segmentation, Girshick et al., CVPR, 2014