DTU Course 02456 Deep learning
2 Convolutional neural networks
2017 Updates

Ole Winther

Dept for Applied Mathematics and Computer Science
Technical University of Denmark (DTU)

September 7, 2017

Objectives of CNN 2017 lecture

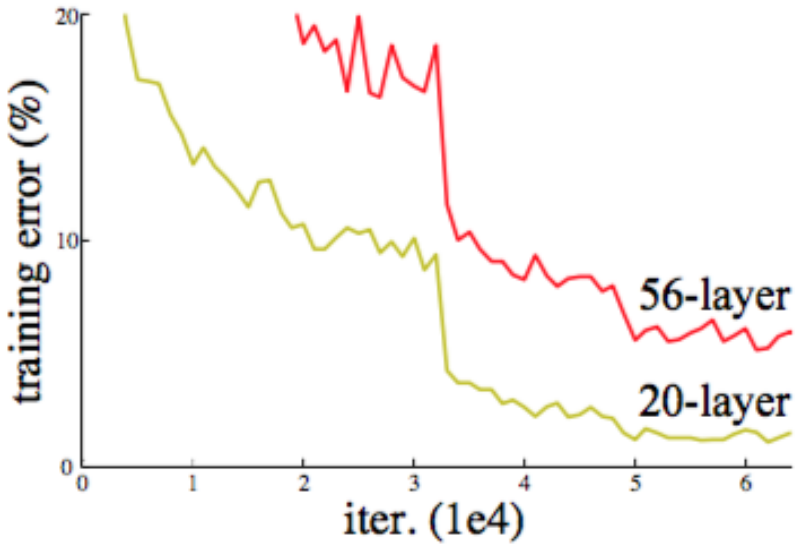
- New CNN architectures introduced in 2016-2017.
- Image segmentation - important application for autonomous systems and medical imaging
- New activation functions

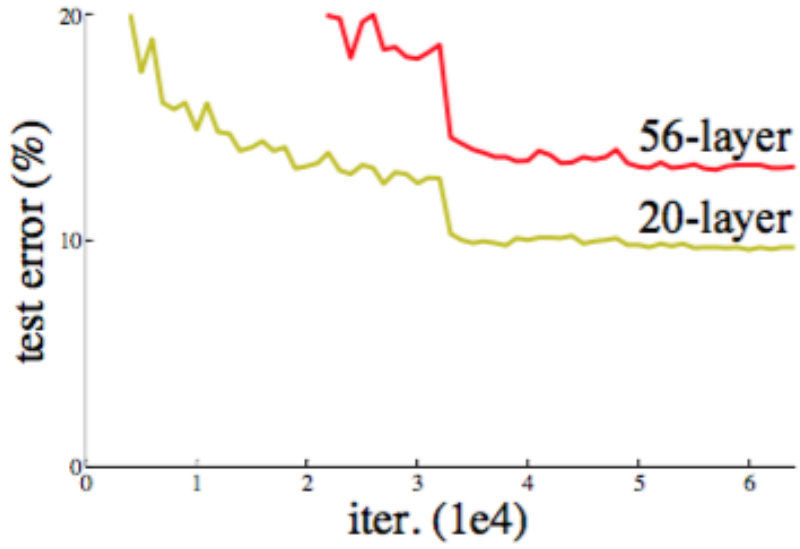
Part 1:

Res-, Dense- and WaveNets

Deep residual nets

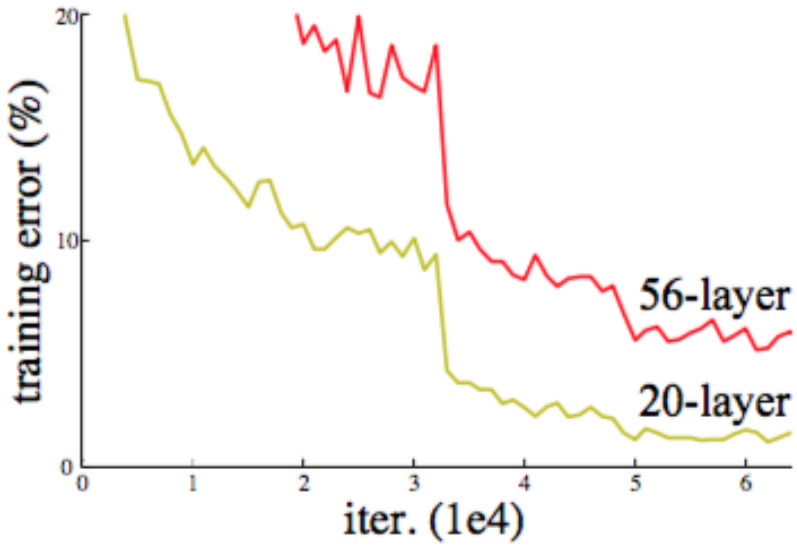
- Observation - it is difficult to fit really deep nets:

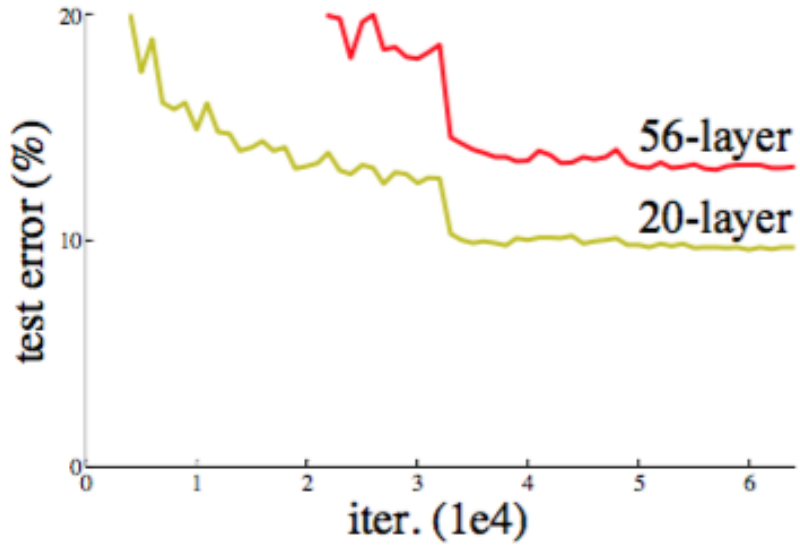




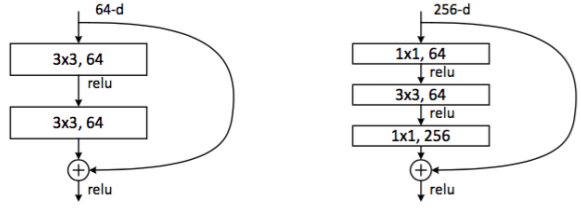
Deep residual nets

- Observation - it is difficult to fit really deep nets:



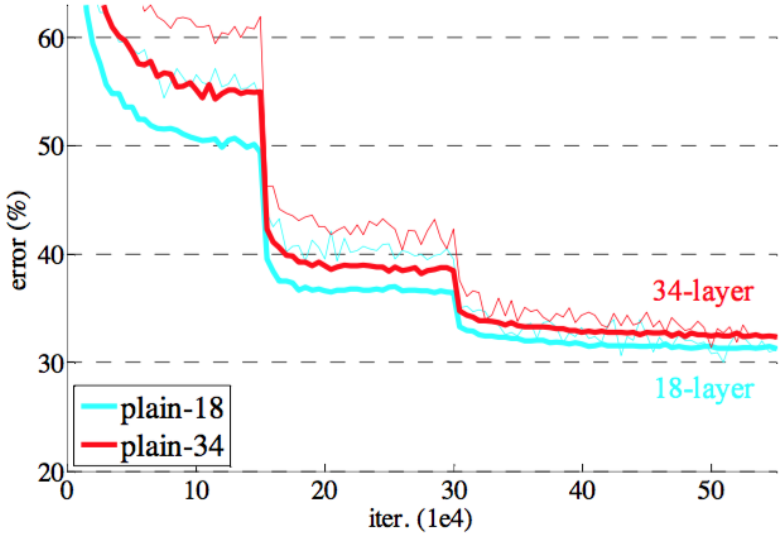


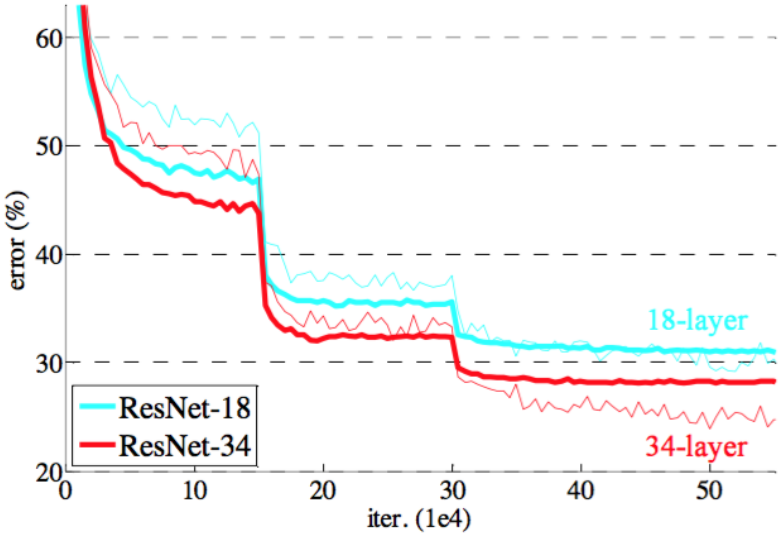
- Solution - bypass layers: $\mathbf{h}_l = \mathcal{F}(\mathbf{h}_{l-1}) + \mathbf{h}_{l-1}$



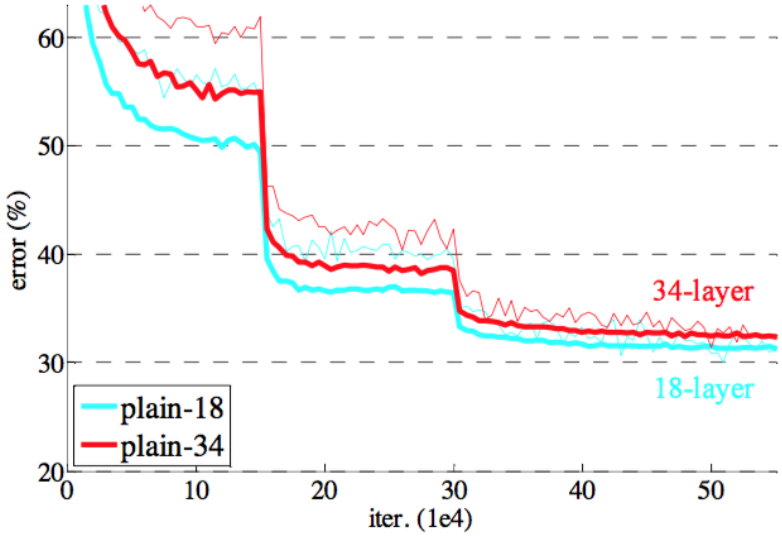
- Initially $\mathcal{F}(\mathbf{h}_{l-1}) \approx 0$ and net $\approx$ linear

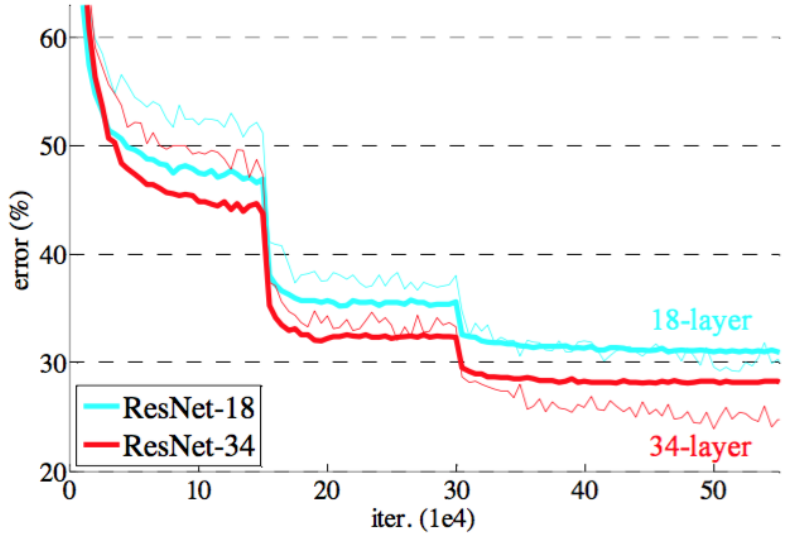
Deep residual nets results - ImageNet results





Deep residual nets results - ImageNet results

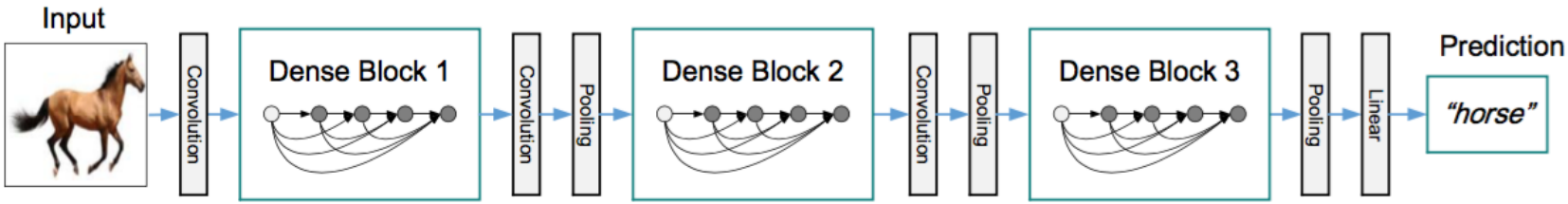




method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43†
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

- 10 ensemble models: ResNets 3.57% top 5 error.
- GoogleNet (2014) 6.66%

DenseNets



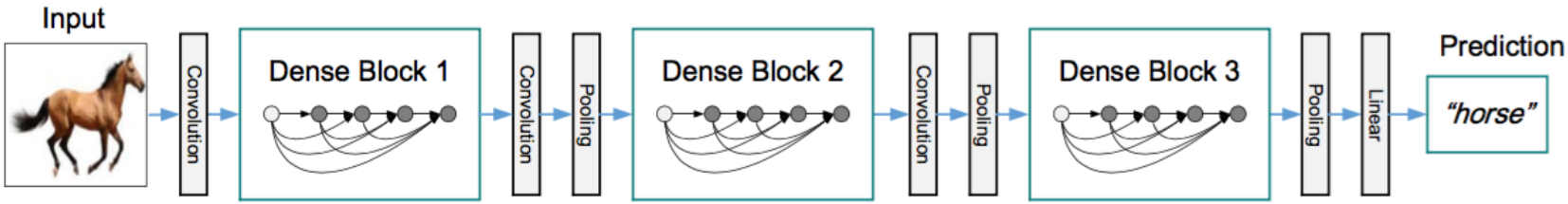
- Reuse features extracted in previous layers

$$\mathbf{h}_l = \mathcal{F}([\mathbf{h}_0, \ldots, \mathbf{h}_{l-1}])$$

- Compare this with ResNets

$$\mathbf{h}_l = \mathcal{F}(\mathbf{h}_{l-1}) + \mathbf{h}_{l-1}$$

DenseNets



- Reuse features extracted in previous layers

$$\mathbf{h}_l = \mathcal{F}([\mathbf{h}_0, \ldots, \mathbf{h}_{l-1}])$$

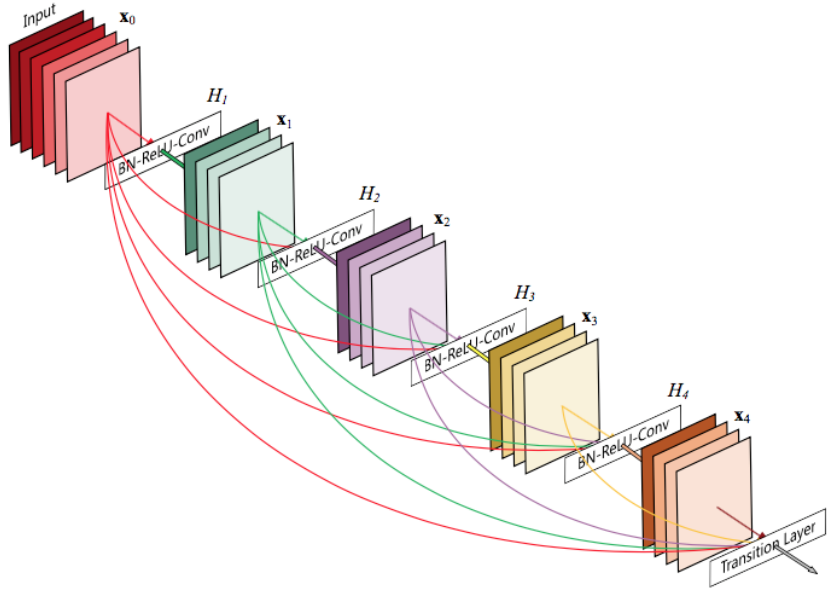
- Compare this with ResNets

$$\mathbf{h}_l = \mathcal{F}(\mathbf{h}_{l-1}) + \mathbf{h}_{l-1}$$

- Avoids parameter explosion by: (details next slide)
  - Having few filters in each layer and
  - use of transition layers
- State-of-the-art on CIFAR and SVHN
- (not tested on ImageNet)
- with fewer parameters.

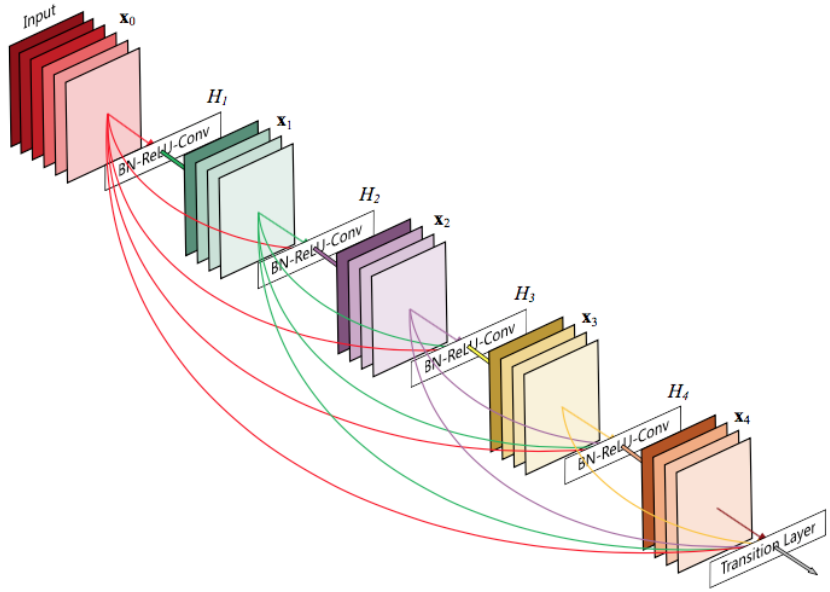
DenseNets details

- L layers in a dense block - $L = 5$ in figure
- Growth factor (feature maps in a layer) k - $k = 4$ in figure
- Feature maps are connected to all subsequent layers in subsequent layers: $l \times k$ channels in layer l.
- Total $\frac{L(L+1)}{2}k$ vs LK in standard.



DenseNets details

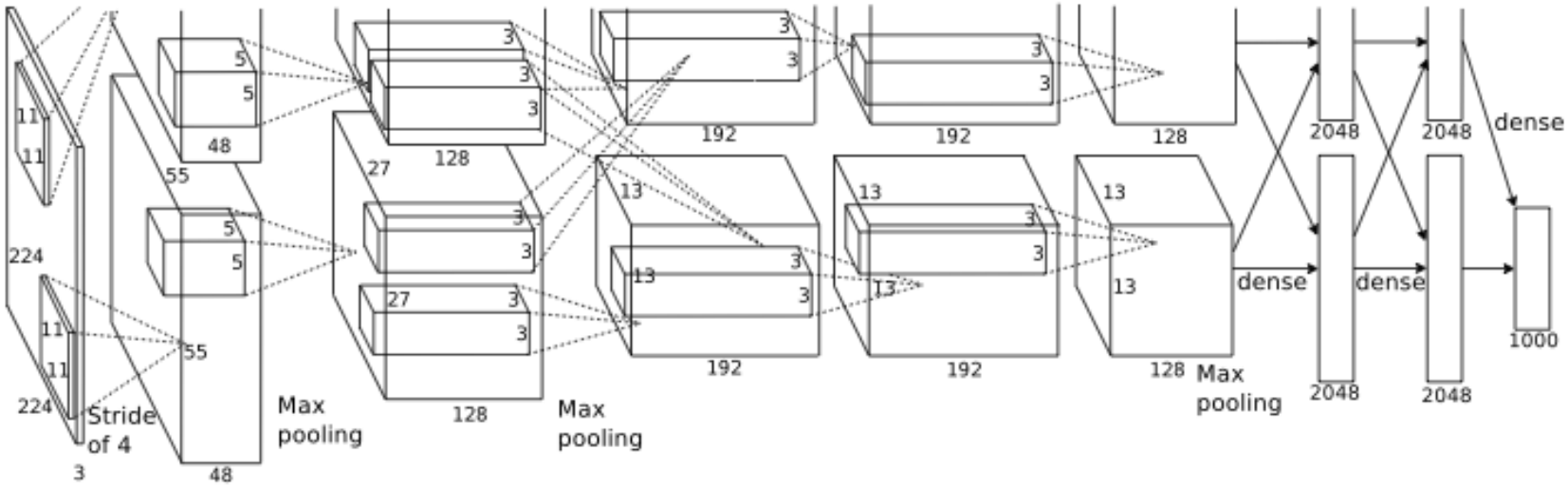
- L layers in a dense block - $L = 5$ in figure
- Growth factor (feature maps in a layer) k - $k = 4$ in figure
- Feature maps are connected to all subsequent layers in subsequent layers: $l \times k$ channels in layer l.
- Total $\frac{L(L+1)}{2}k$ vs LK in standard.



- Keep number of parameters by small k and
- use transition layers (with 1×1 convolutions and pooling)
- between dense blocks.

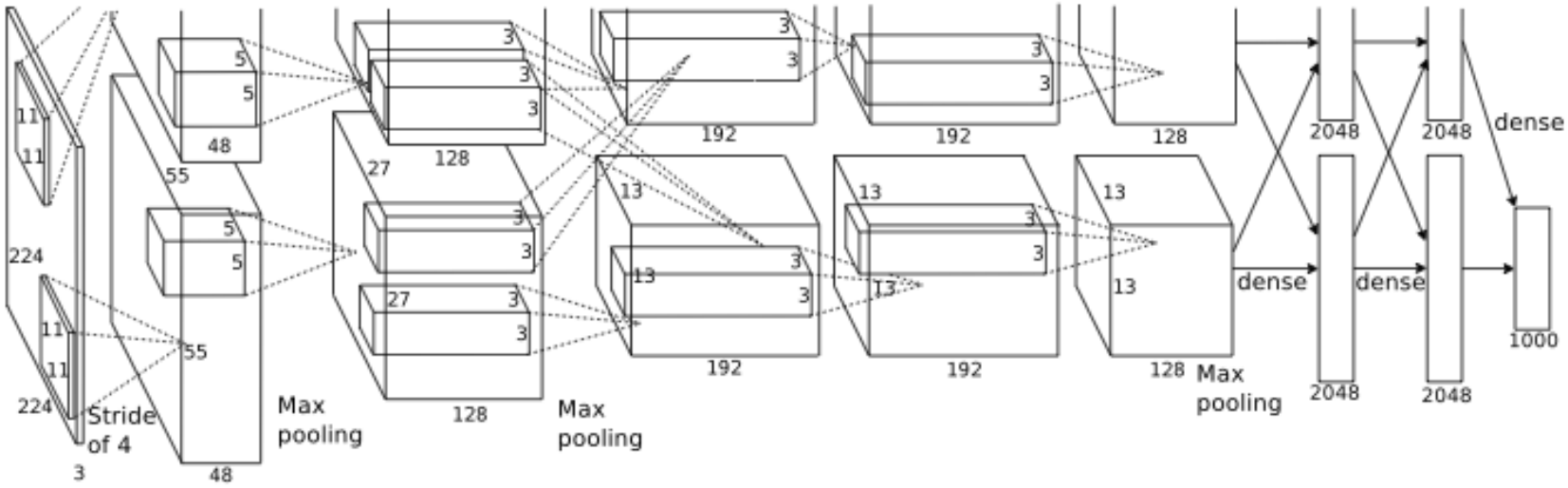
We need bigger brains

- AlexNet (2012): 16.4% error, 8 layers, 1.4 Gflop

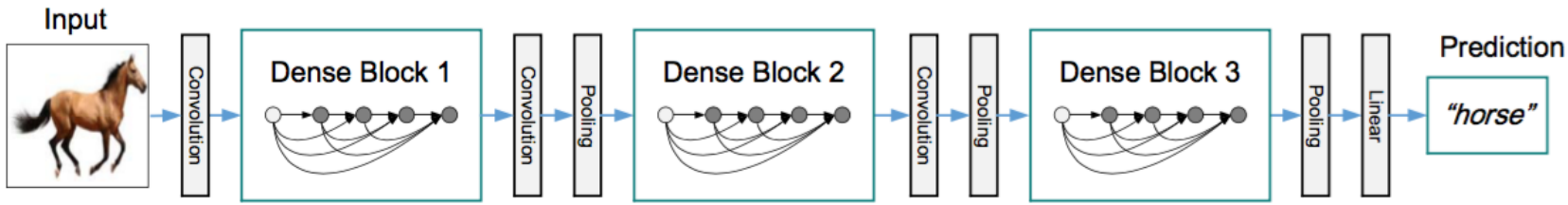


We need bigger brains

- AlexNet (2012): 16.4% error, 8 layers, 1.4 Gflop

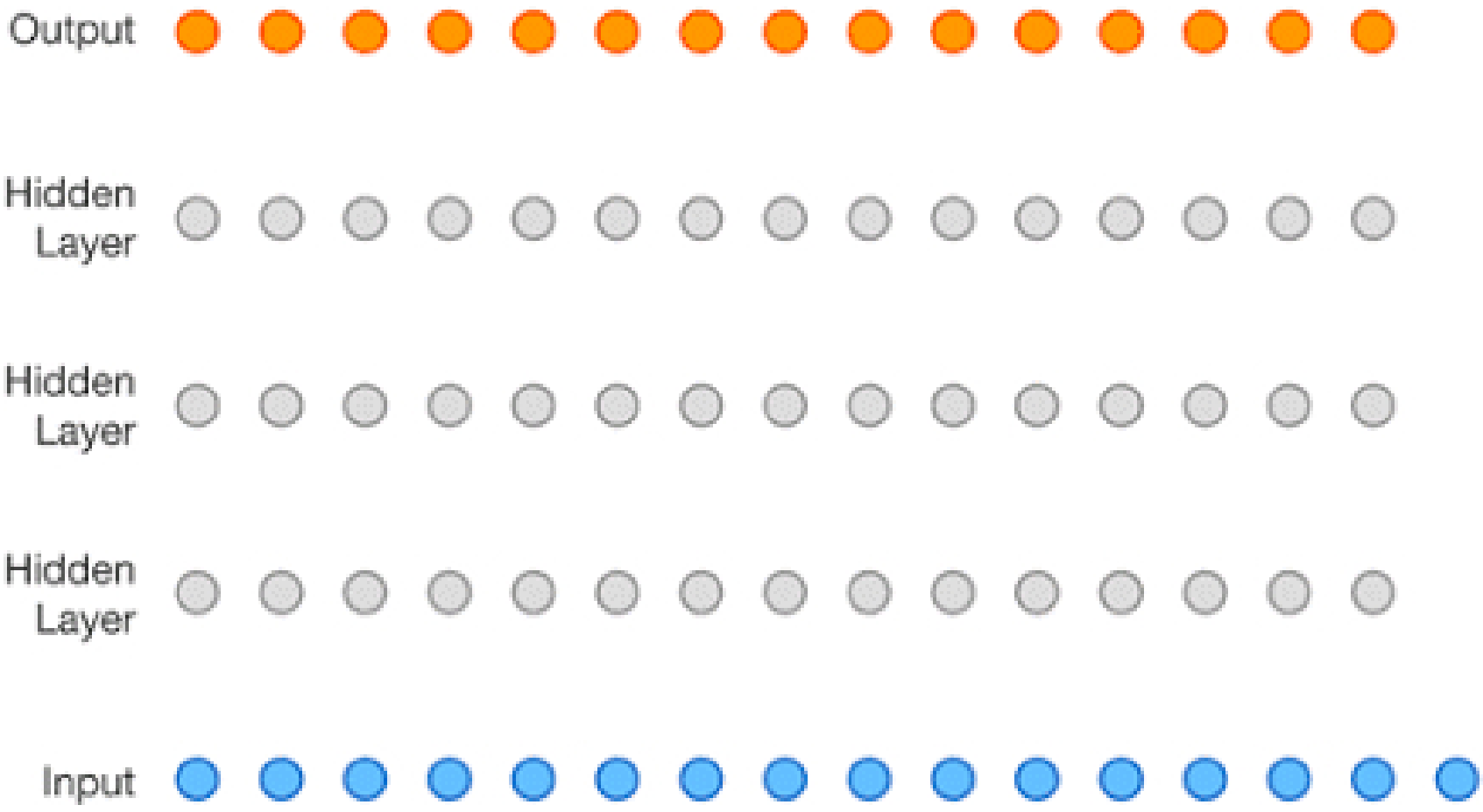


- ResNet (2016): 3.5% error, 152 layers, 22.6 Gflop.



- (This is a so-called DenseNet and not a ResNet.)
- Source: Source Jen-Hsun Huang, CEO NVIDIA, GTC Europe, 2016

WaveNet



DeepMind blogpost and

`https://arxiv.org/pdf/1609.03499.pdf`

Part 2:
New activation functions

Activation functions

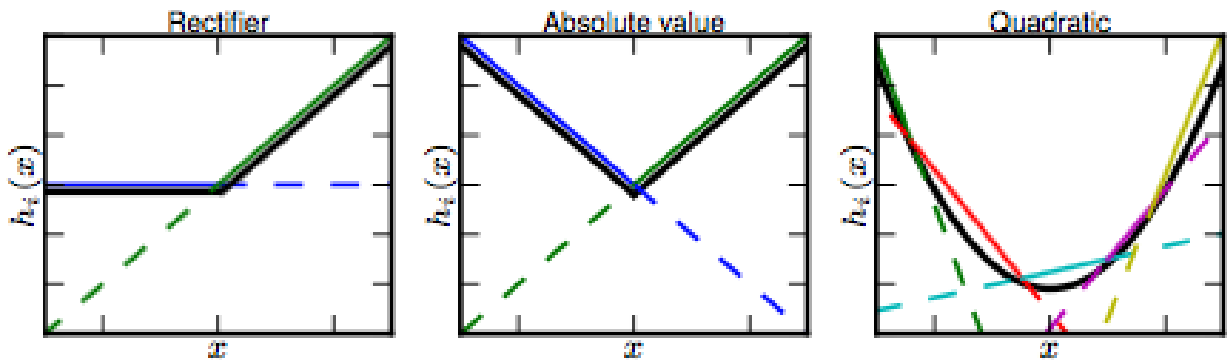
- Gated linear units (GLU) works well with conv nets

$$\mathrm{GLU}(x) = (Wx + b) \otimes \sigma(Vx + c)$$

- MaxOut for unit i. Each unit has K inputs

$$\mathrm{MaxOut}_i(x) = \max_{k \in 1,\ldots,K} z_{ik}$$

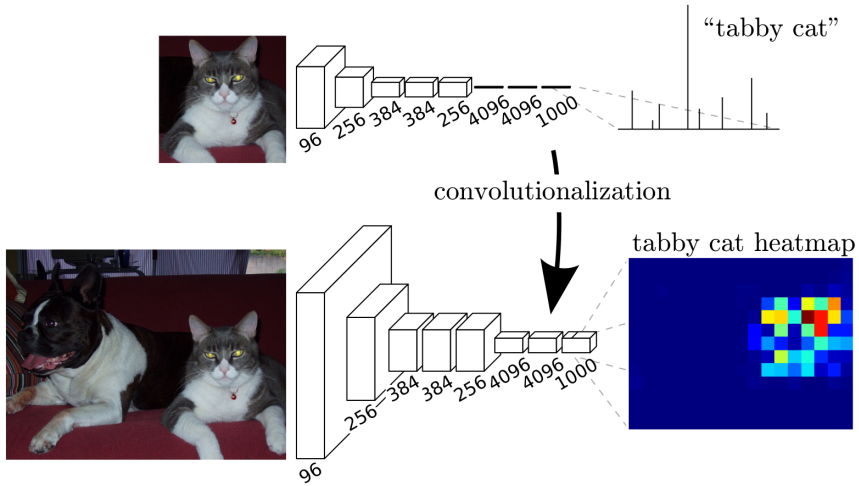
- Works well with dropout

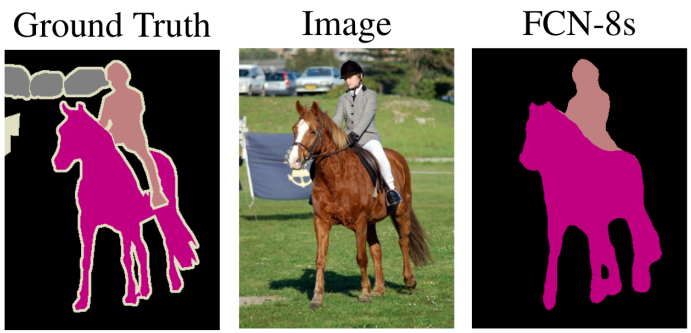


- $\mathrm{LeakyRelu}(z) = \max(az, z)$ has
- slope $a \in [0, 1]$ for $z < 0$
- Other in ReLu family: SoftPLus, ELU, SELU, etc

Part 3:
Image Segmentation

Semantic segmentation (Long et al., 2015)

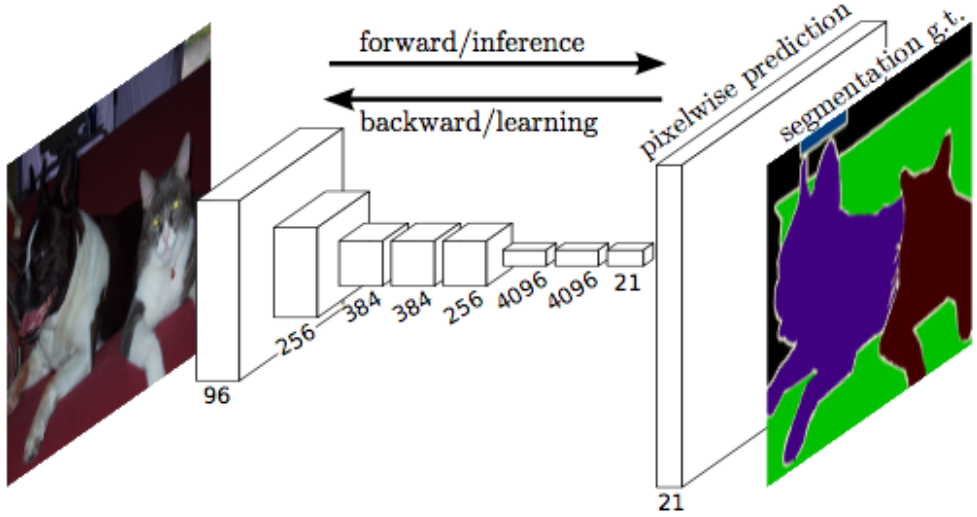




- Sliding a convolutional network to classify each location.
- Lots of shared computation.

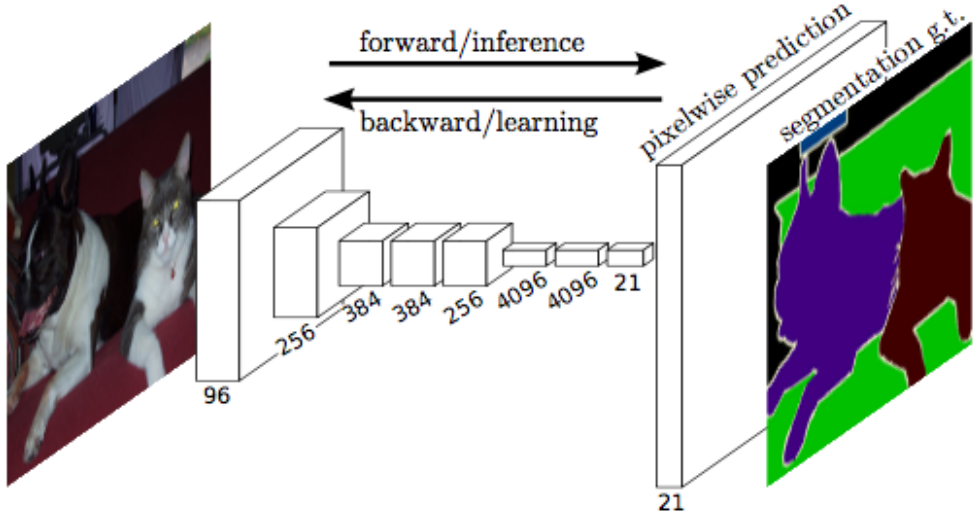
End-to-end image segmentation

- Classify each pixel

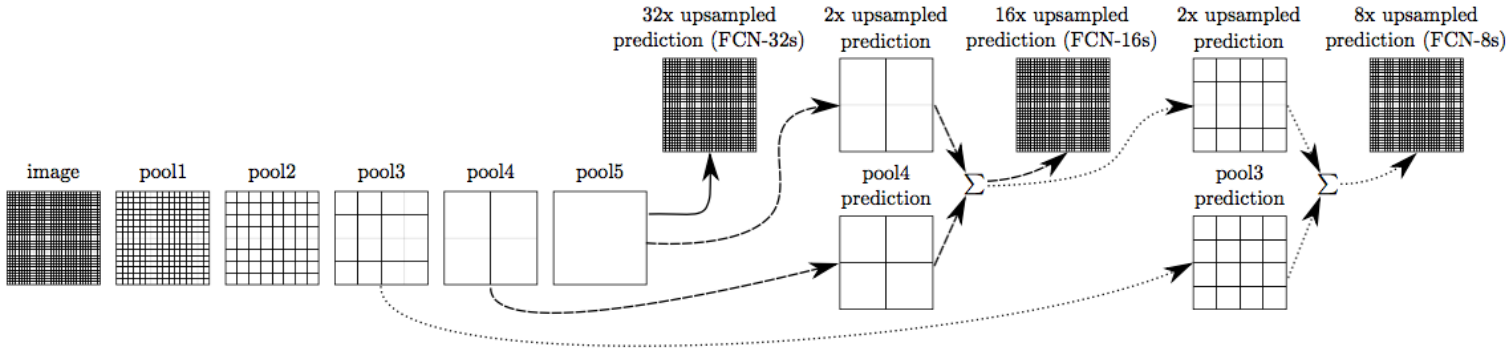


End-to-end image segmentation

- Classify each pixel

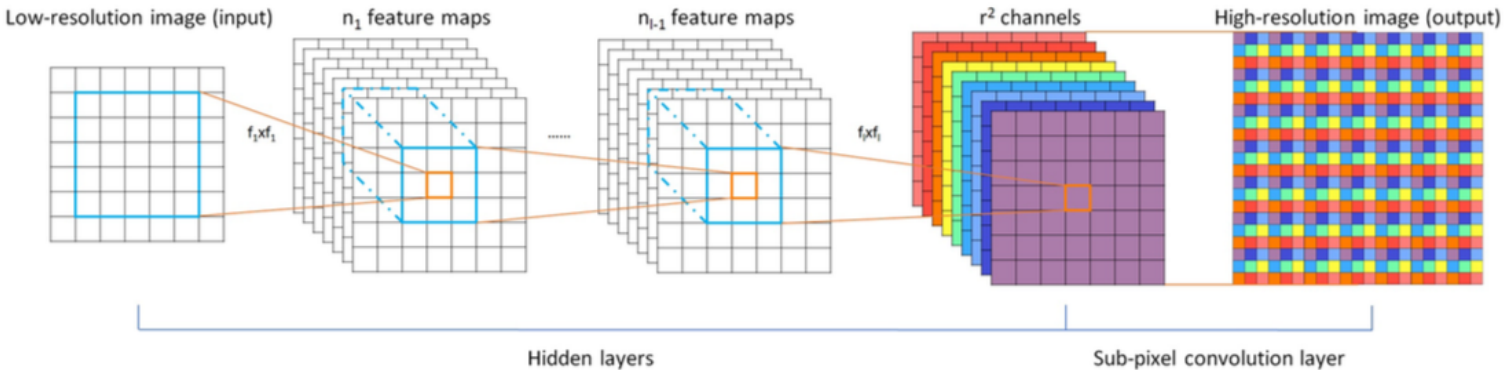


- Use convolutions, pooling and upsampling

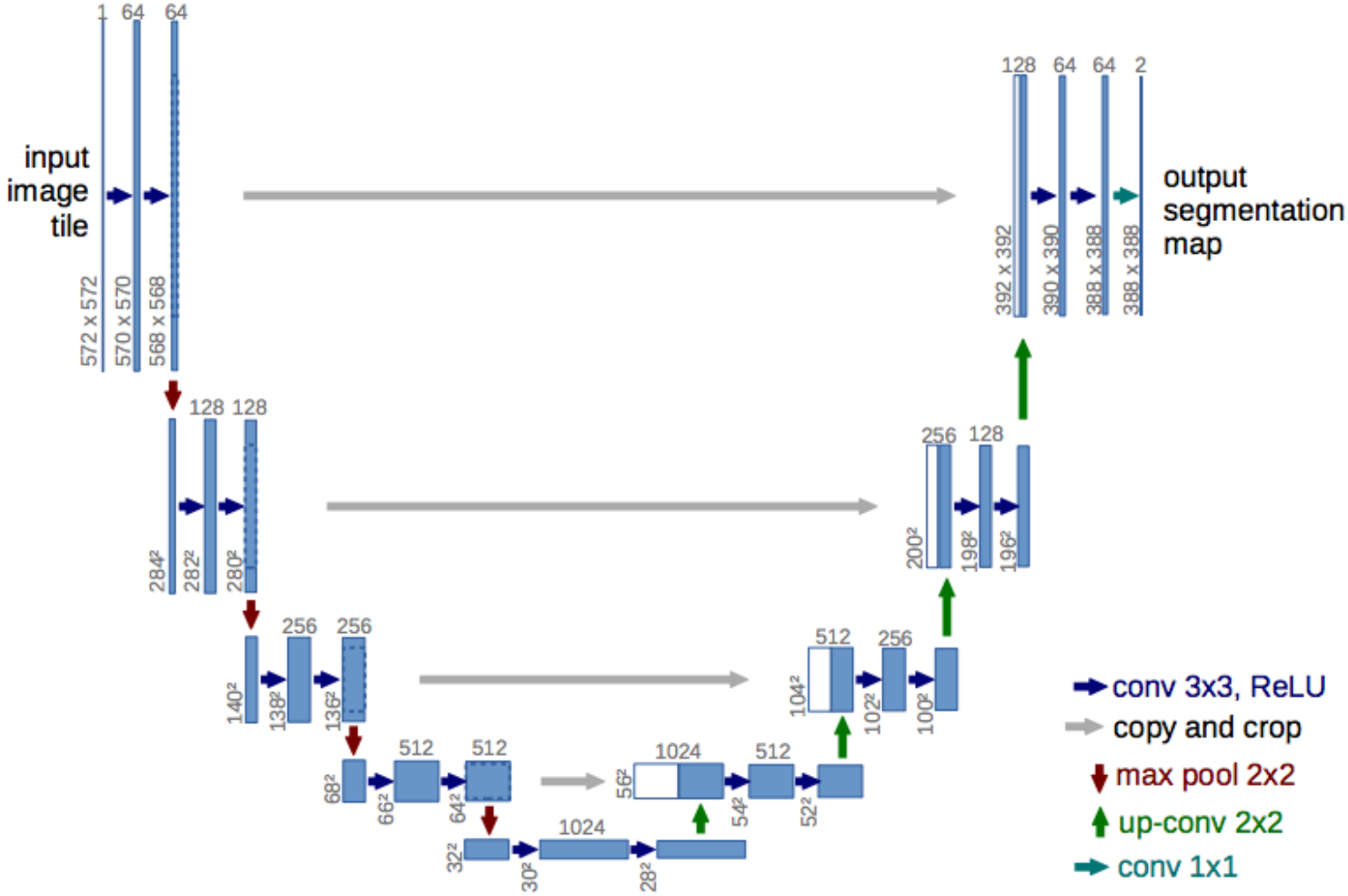


Sub-pixel deconvolutions

- Convert $H \times H \times C \rightarrow 2H \times 2H \times C/4$:



U-Net


input image tile
572 x 572
570 x 570
568 x 568
1
64
64
128
128
284²
282²
280²
256
256
140²
138²
136²
512
512
68²
66²
64²
1024
32²
30²
28²
1024
512
56²
54²
512
256
104²
102²
100²
256
128
200²
198²
196²
128
64
64
2
392 x 392
390 x 390
388 x 388
388 x 388
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

The Cityscapes dataset

Cityscapes leaderboard

- Performance metric

$$\text{Intersection over Union} = IoU = \frac{TP}{TP + FP + FN}$$

name	fine	coarse	16-bit	depth	video	sub	IoU class	iIoU class	IoU category	iIoU category	Runtime [s]	code
motovis	yes	yes	no	no	no	no	81.3	57.7	91.5	80.7	n/a	no
PSPNet	yes	yes	no	no	no	no	81.2	59.6	91.2	79.2	n/a	yes
NetWarp	yes	yes	no	no	yes	no	80.5	59.5	91.0	79.8	n/a	no
ResNet-38	yes	yes	no	no	no	no	80.6	57.8	91.0	79.1	n/a	yes
tek-Ifly	yes	no	no	no	no	no	81.1	60.1	90.9	79.6	n/a	no
ResNet-38	yes	no	no	no	no	no	78.4	59.1	90.9	81.1	n/a	yes
TuSimple_Coarse	yes	yes	no	no	no	no	80.1	56.9	90.7	77.8	n/a	yes
SAC-multiple	yes	no	no	no	no	no	78.1	55.2	90.6	78.3	n/a	no
SegModel	yes	yes	no	no	no	no	79.2	56.4	90.4	77.0	n/a	no
TuSimple	yes	no	no	no	no	no	77.6	53.6	90.1	75.2	n/a	no
Global-Local-Refinement	yes	no	no	no	no	no	77.3	53.4	90.0	76.8	n/a	no

References

- He et al, Deep Residual Learning for Image Recognition, https://arxiv.org/pdf/1512.03385v1.pdf
- Huang et al, Densely connected convolutional networks, http://arxiv.org/pdf/1608.06993.pdf
- Oord et al, WaveNet a generative model for raw audio, https://arxiv.org/pdf/1609.03499.pdf
- Dauphin et al, Language Modeling with Gated Convolutional Networks, https://arxiv.org/pdf/1612.08083.pdf
- Goodfellow et al, MaxOut networks, https://arxiv.org/pdf/1302.4389.pdf
- Long et al, Fully convolutional networks for semantic segmentation, https://arxiv.org/abs/1411.4038
- Shi et al, Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network https://arxiv.org/abs/1609.05158
- Ronneberger et al, U-Net: Convolutional Networks for Biomedical Image Segmentation, https://arxiv.org/abs/1505.04597

Thanks!
Ole Winther