Deep Learning II: Advanced Neural Network Architectures

8DM50 Machine Learning in Medical Imaging and Biology

Jelmer Wolterink

28-09-2020

UNIVERSITY
OF TWENTE.

Deep Learning II

Me

-Assistant professor @ University of Twente

-Deep learning for medical image analysis - cardiovascular image analysis (CT, MR)

Today

1. **Advanced neural network architectures**

2. Interpretability and generative adversarial networks

3. Practical assignment in Keras

In this lecture

Convolutional neural networks

- Recap
- Advanced architectures

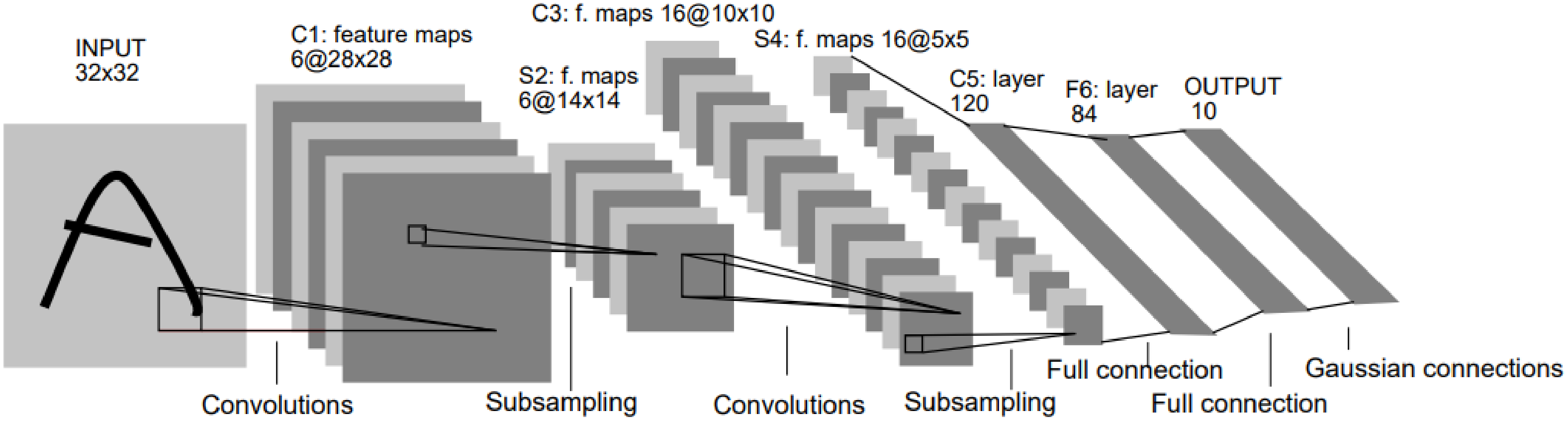
Neural networks for sequential data

- Recurrent neural networks
- Long short term memory (LSTM) units

CNNs for pixelwise prediction

- Patch-based segmentation
- Encoder-decoder architectures

Recap: Convolutional neural networks



[Demo]

A standard convolut[illegible] network consists of

- **Convolutional** laye[illegible]nsform input into feature maps
- **Subsampling** oper[illegible] reduce size of feature maps, e.g. max pooling
- **Fully-connected** la[illegible]erform classification (multi-layered perceptron)

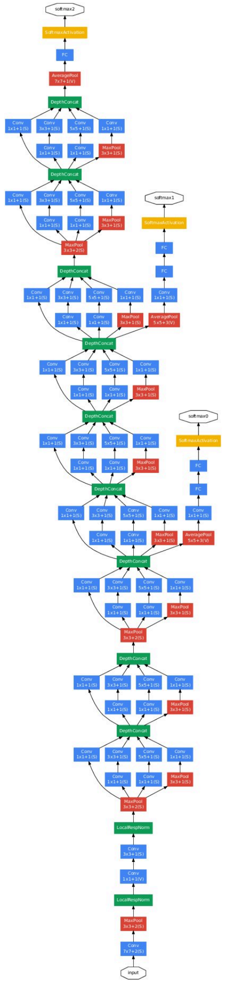
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.

UNIVERSITY
OF TWENTE.

What happened between 1998 and now?

Compute

Data



Algorithms

UNIVERSITY
OF TWENTE.

Data: ImageNet challenge

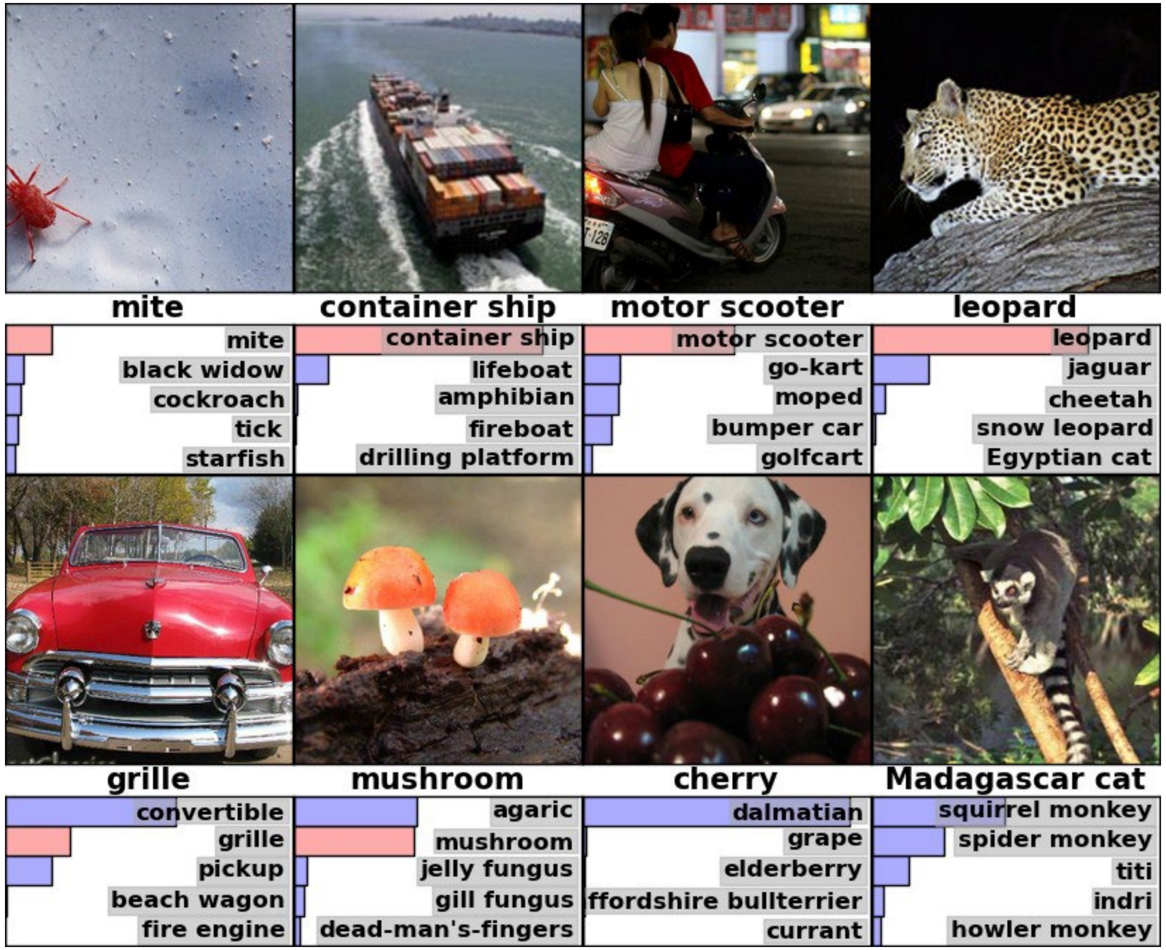
Benchmark for image classification/object detection

Data

- > 1,200,000 RGB images
- Images show one of **1000** classes

Task

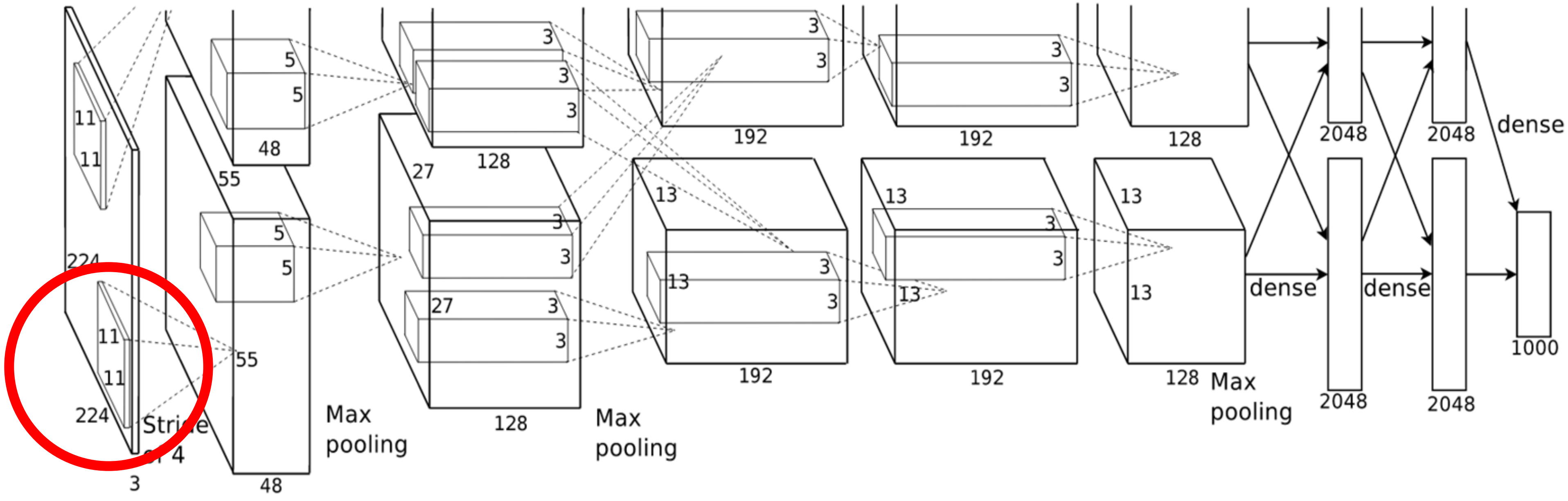
- Detect label of image
- Top-1\top-5 accuracy



UNIVERSITY OF TWENTE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

AlexNet



- Substantially outperformed ‘conventional’ methods in 2012
- Convolutional + subsampling + fully connected layers
- Trained in parallel on two GPUs
- Training time
  - **2012**: 5 to 6 days (2 x GTX 580 3GB GPU)
  - **2017**: 24 minutes (supercomputer 32,000 cores)
- Large 11 x 11 convolution kernels

UNIVERSITY OF TWENTE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

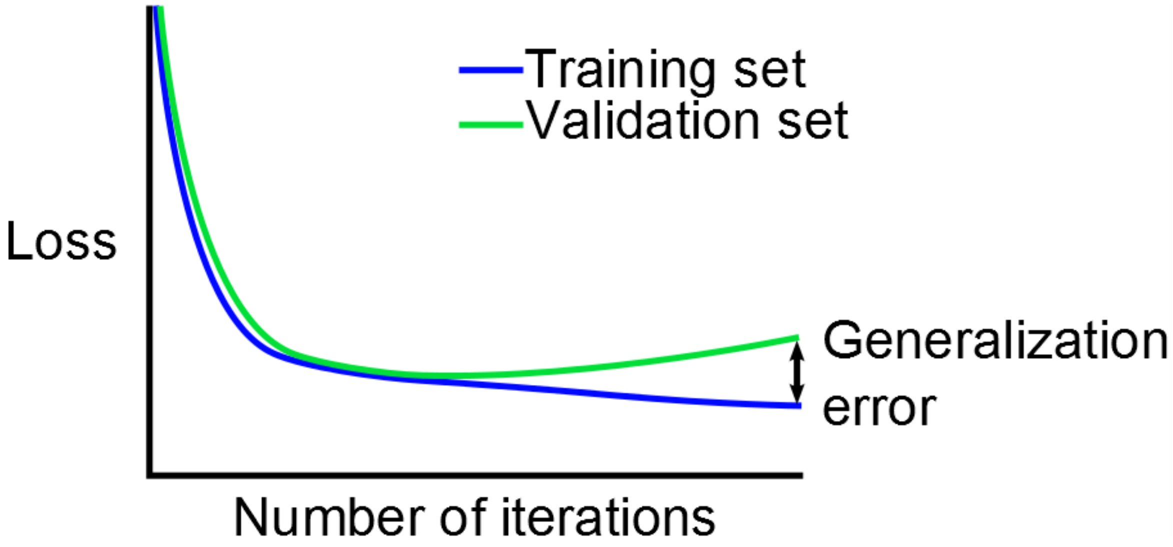
Overfitting

Reasons

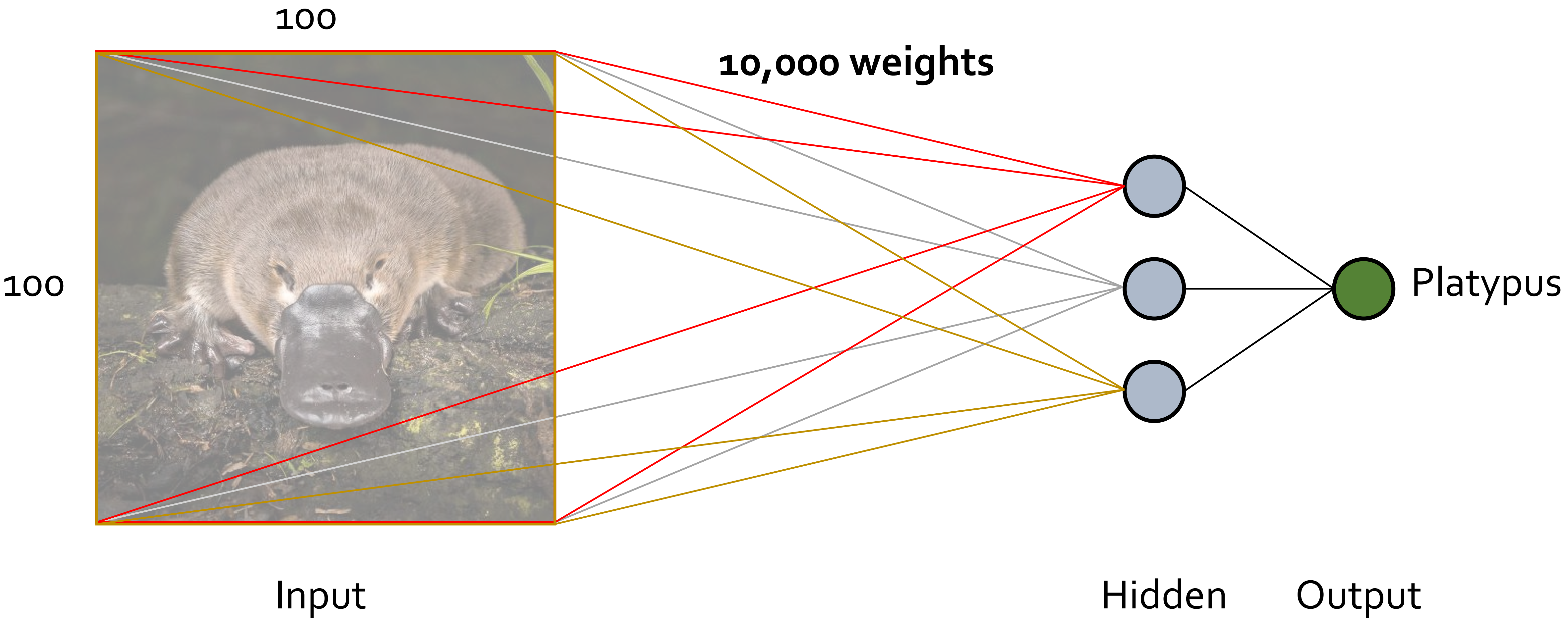
- Too many parameters
- Not enough data

Solution

- Reduce number of parameters

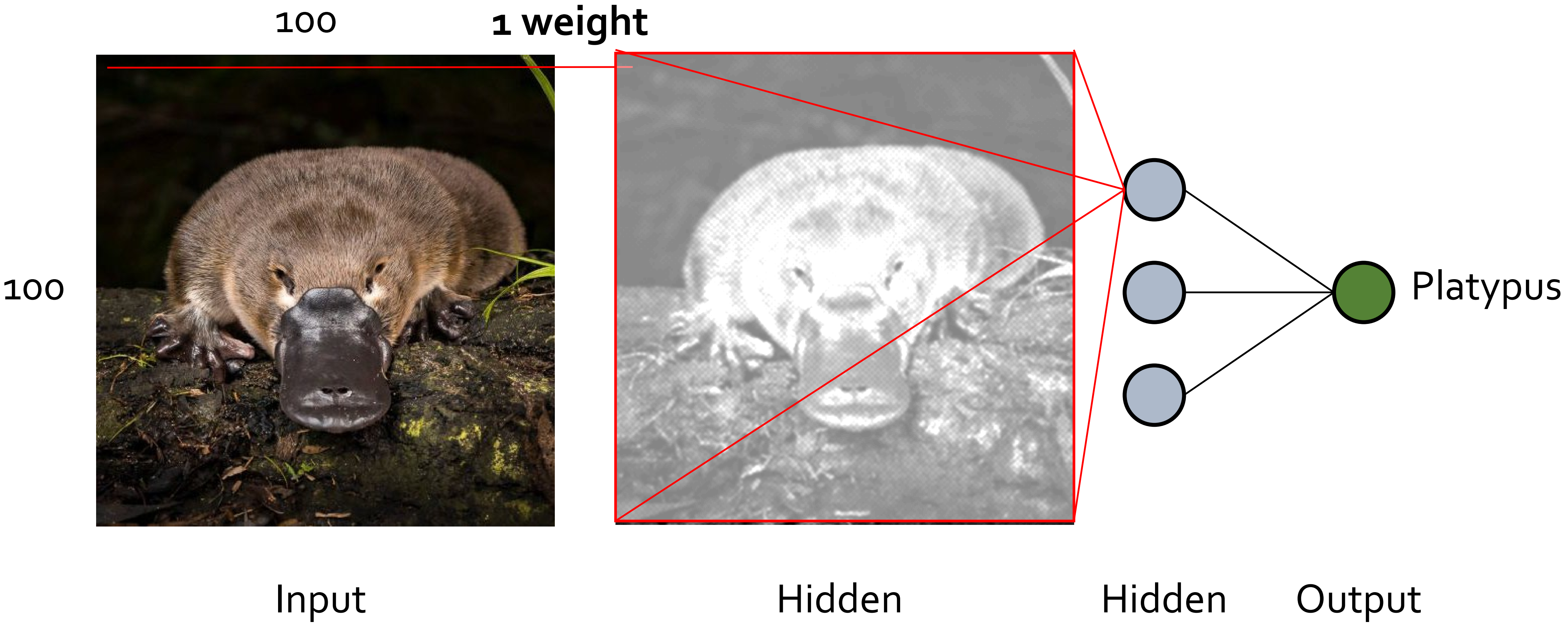


Kernel size



https://www.abc.net.au/news/2018-11-10/platypus-vulnerable-to-extinction-researchers-say/10477902

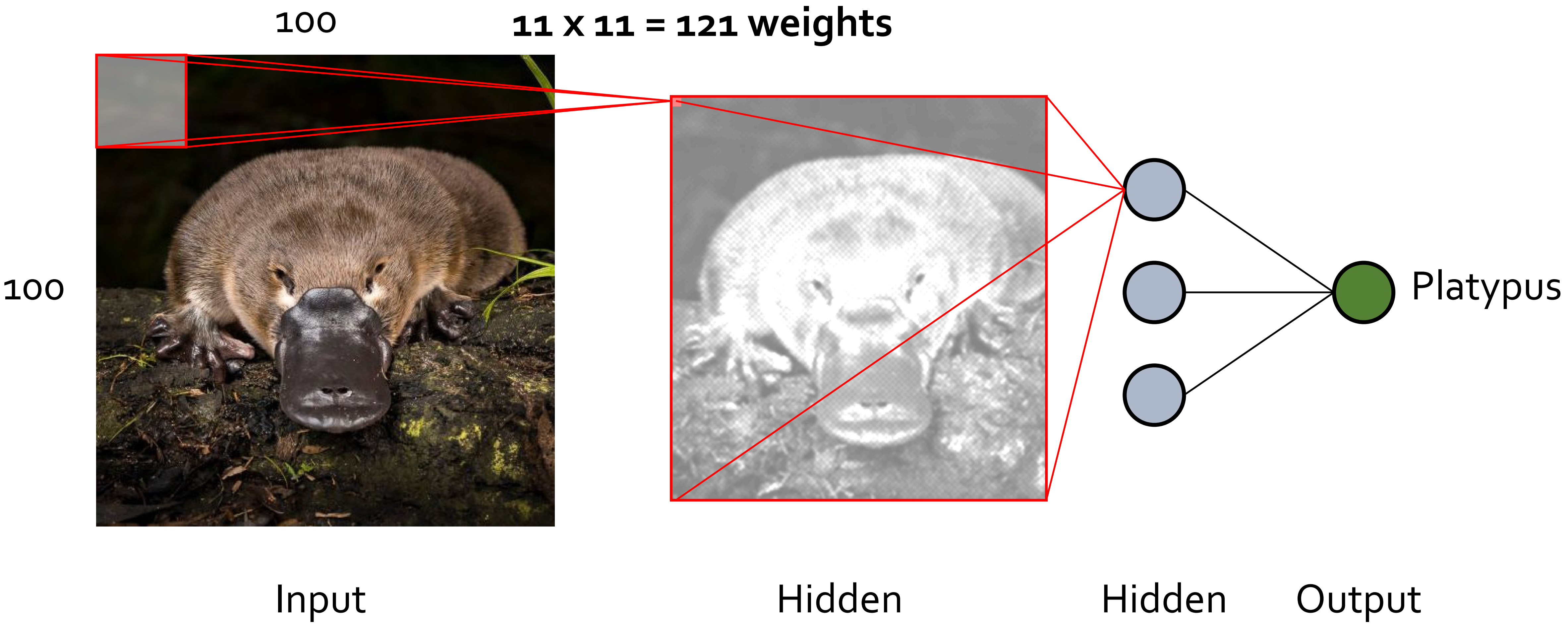
Kernel size



https://www.abc.net.au/news/2018-11-10/platypus-vulnerable-to-extinction-researchers-say/10477902

UNIVERSITY OF TWENTE.

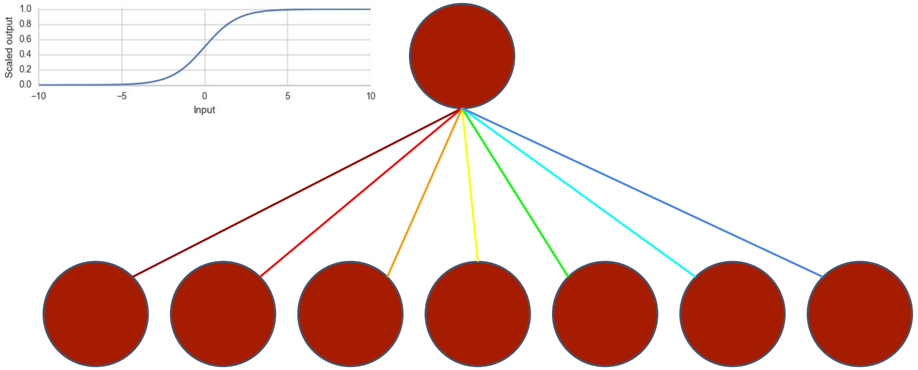
Kernel size

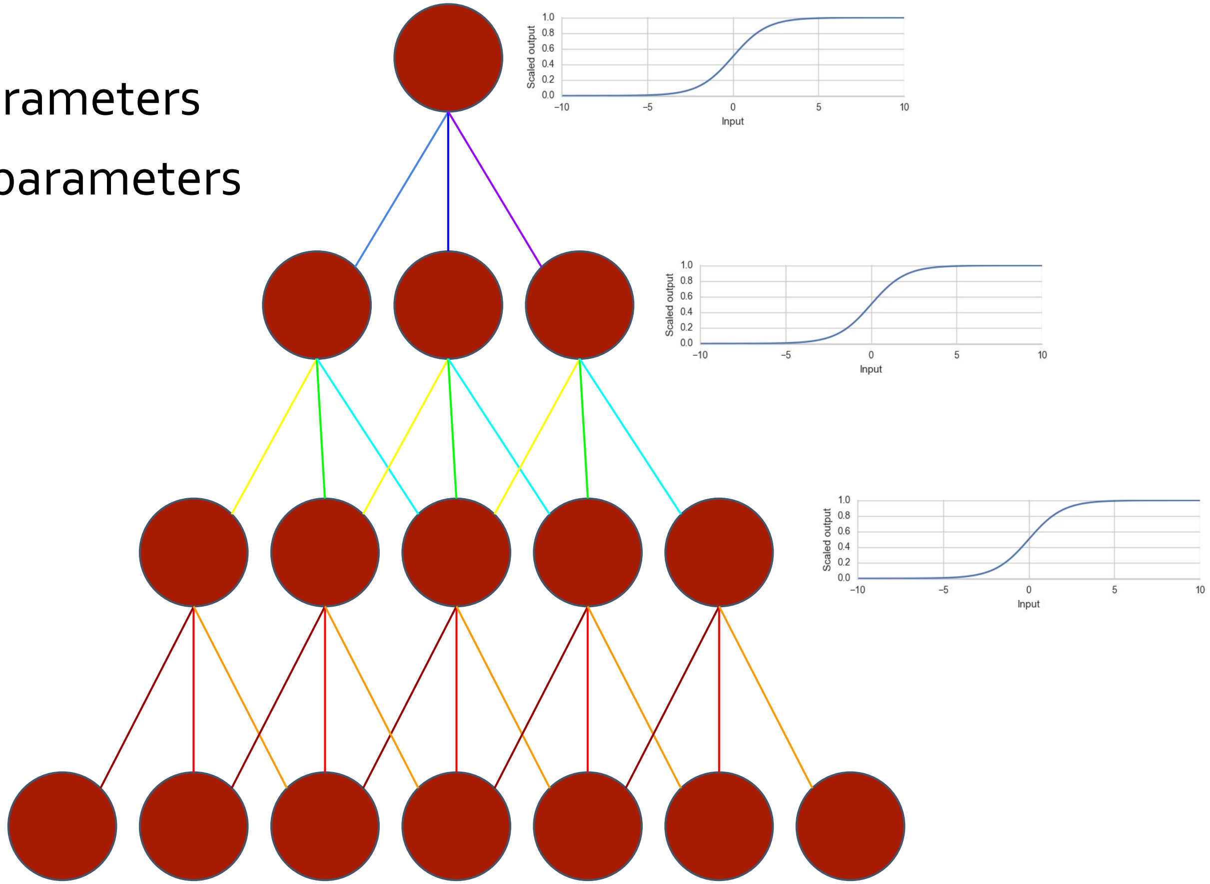


https://www.abc.net.au/news/2018-11-10/platypus-vulnerable-to-extinction-researchers-say/10477902

Using smaller kernels

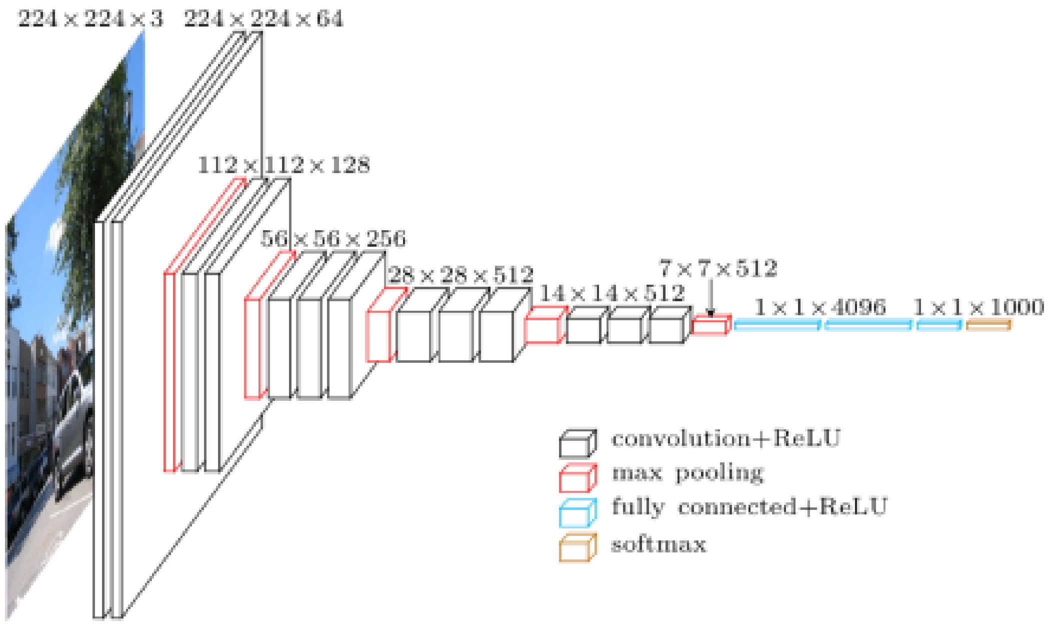
- Large kernels have many parameters: 7 x 7 = 49 parameters
- Smaller kernels reduce parameters: 3 x (3 x 3) = 27 parameters
- More nonlinearities means more abstraction





Simonyan et al. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

VGG-Net



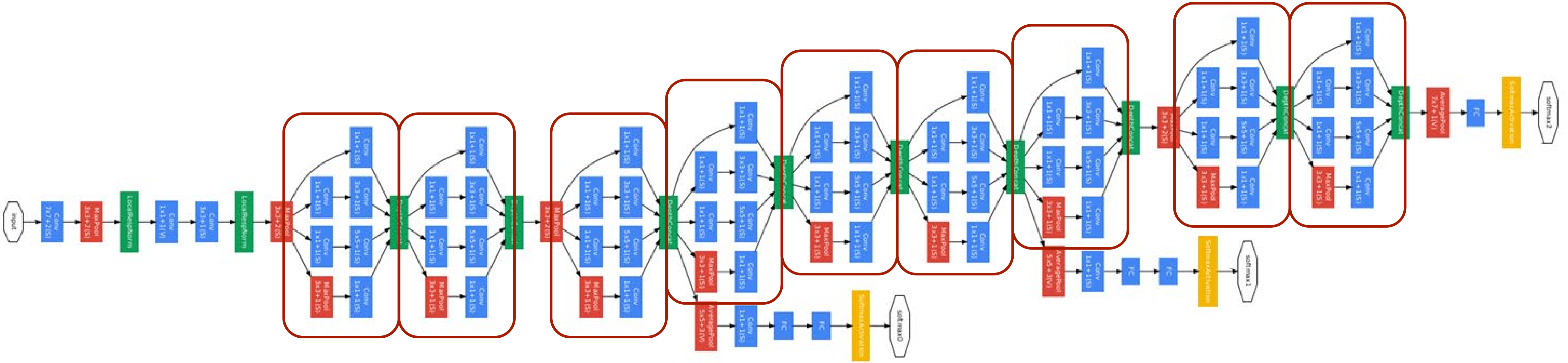
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	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 **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 **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 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Simonyan et al. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

UNIVERSITY OF TWENTE.

GoogLeNet (Inception v1)

- 22 layer-network
- Very deep compared to LeNet/AlexNet
- SOTA on ImageNet (when published)



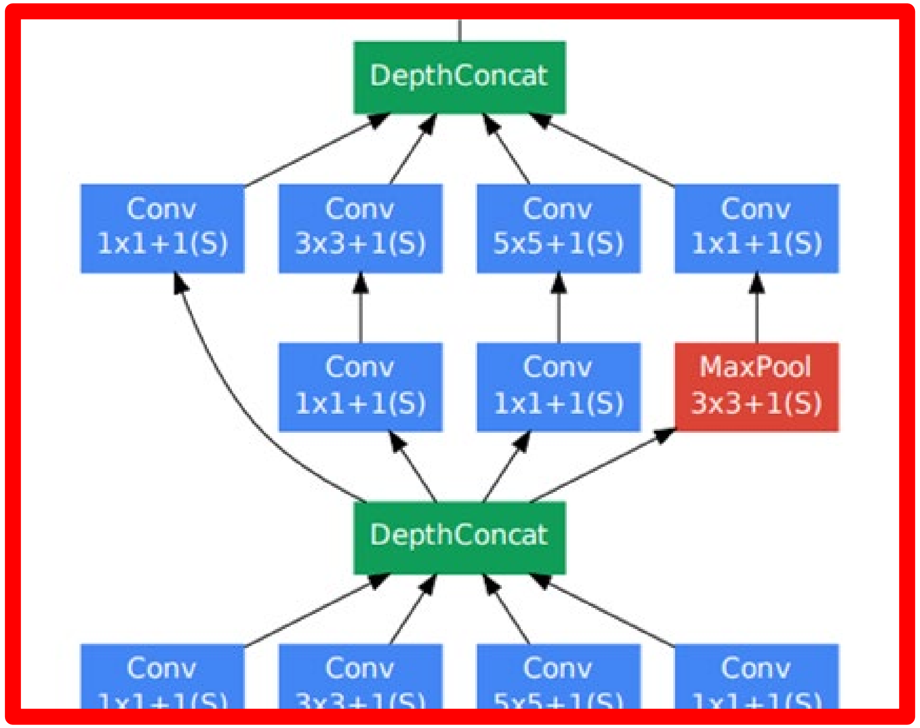
Szegedy et al. "Going deeper with convolutions." CVPR 2015.

Inception module

- Combine parallel multi-scale convolutions
- Let the model pick best filter size

1x1 3x3 5x5

- Bottleneck layers: 1 x 1 convolutions
  - Aggregate feature maps
  - Prevent explosion in number of parameters



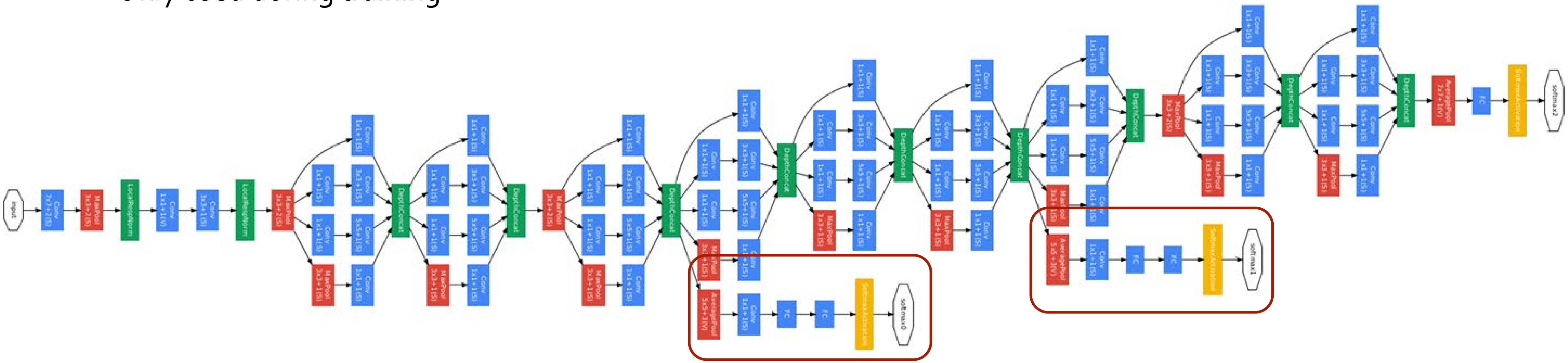
Szegedy et al. "Going deeper with convolutions." CVPR 2015.

UNIVERSITY OF TWENTE.

Auxiliary classifiers

Auxiliary classifiers provide extra supervision

- Vanishing gradients
- Enforce useful features at intermediate layers
- Only used during training



Szegedy et al. "Going deeper with convolutions." CVPR 2015.

Residual connections

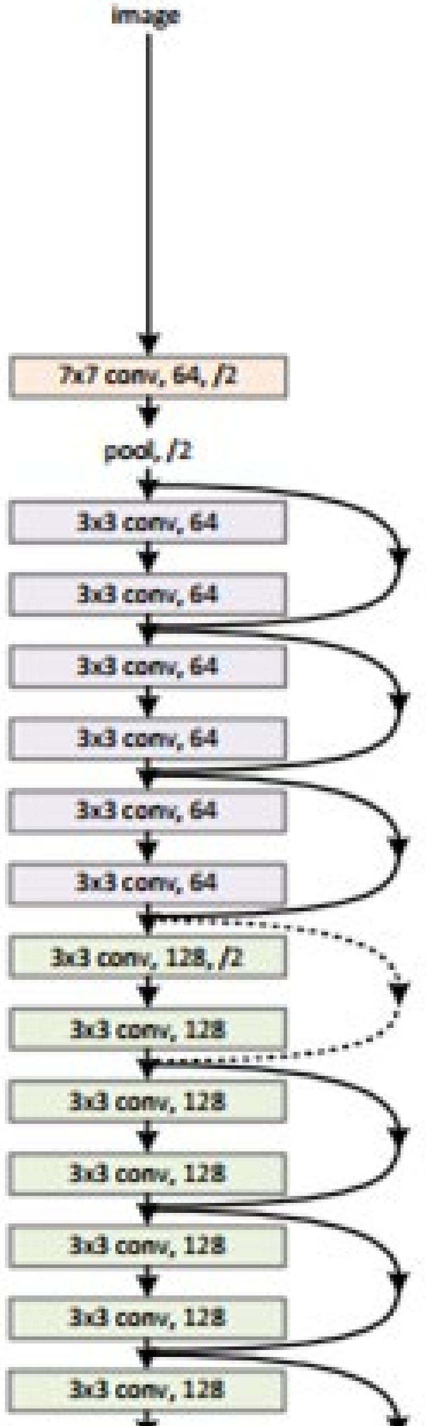
Very deep networks

- allow learning of better representations
- are difficult to optimize due to vanishing gradients

Residual connections can skip layers $H(x) = x + F(x)$

A deep network is at least as strong as it's shallower variant

- If adding layers doesn't help, just use the skip connection



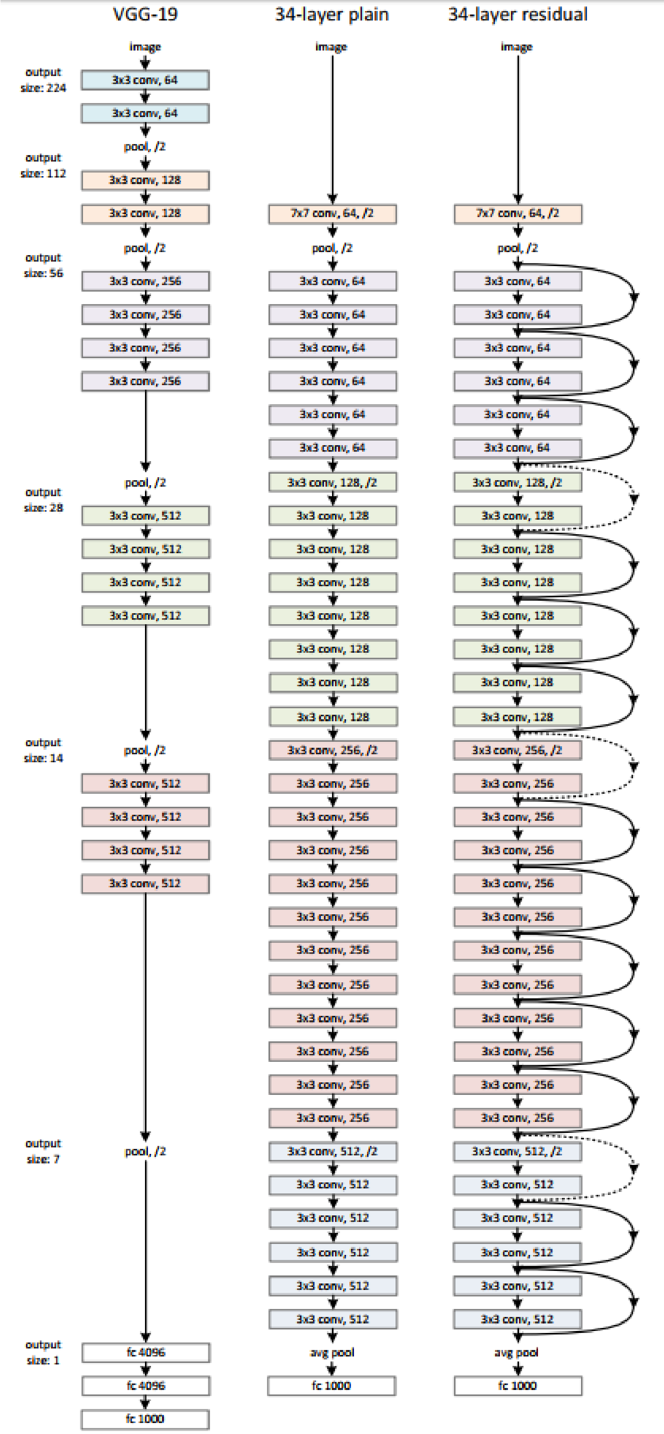
He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In CVPR (pp. 770-778).

Residual network (ResNet)

- Organize layers in blocks
- Use bottleneck layers
- Residual connections barely add computational complexity

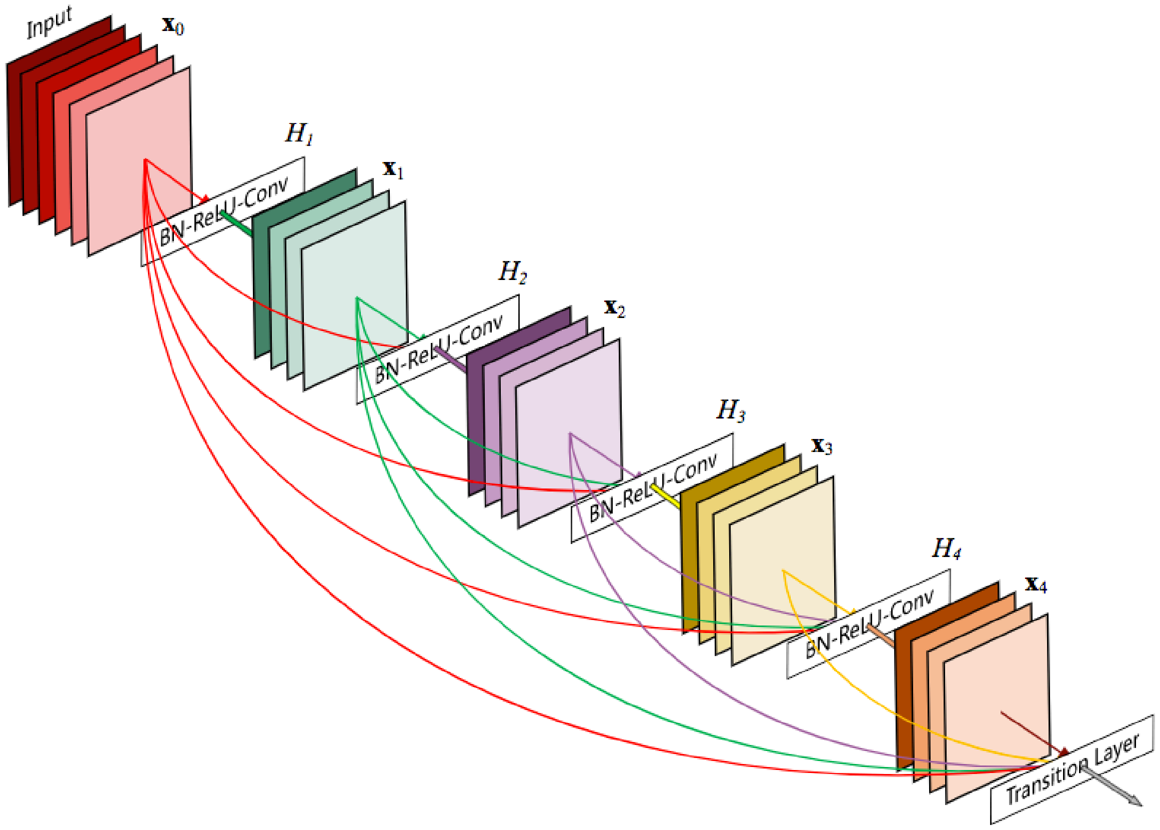
- SOTA on ImageNet (when published)

- Inspired
  - Wide residual nets (50-layer wide ResNet > 152-layer ResNet)
  - DenseNets: get identity mapping from all previous layers



He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In CVPR (pp. 770-778).

Dense networks

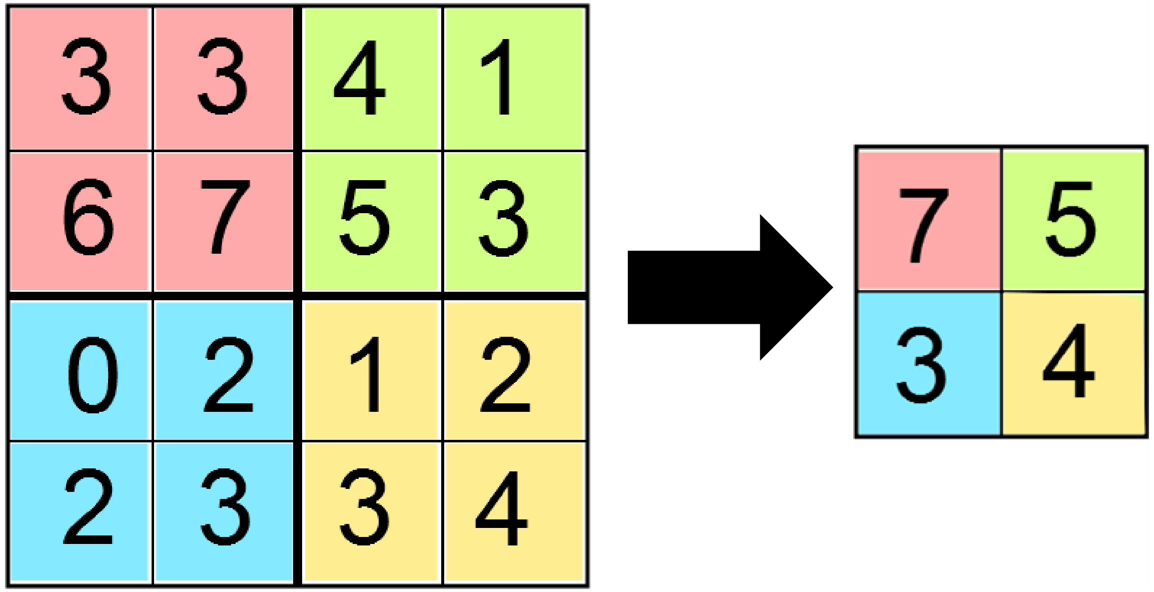


UNIVERSITY OF TWENTE.

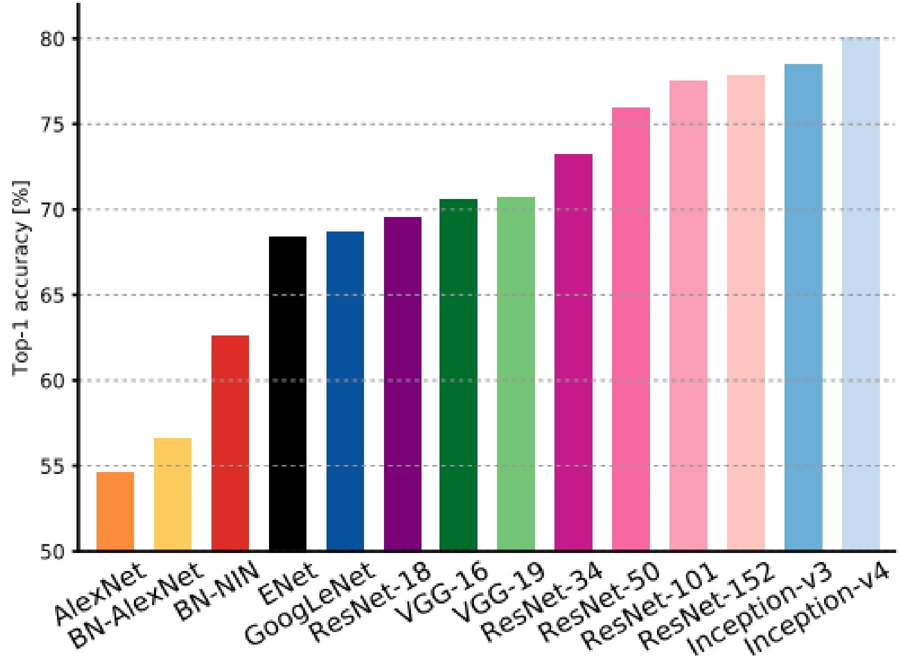
Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

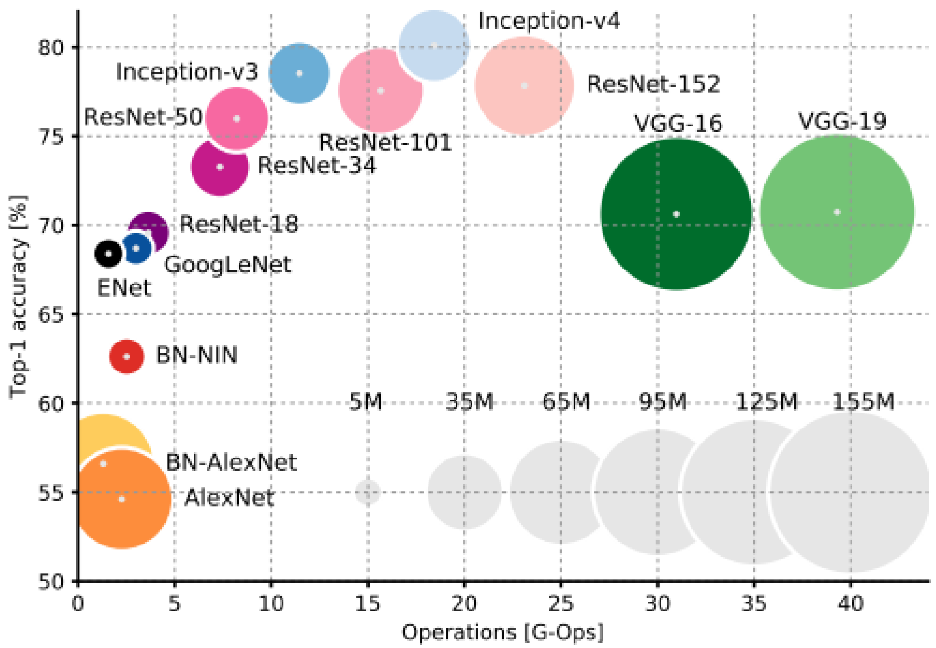
Downsampling

- We often want to go from a large image to a single prediction
- Use downsampling operations like pooling
- Pooling is not trainable



Complexity vs. accuracy



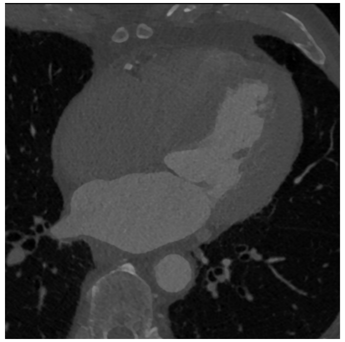


Canziani et al.. "An analysis of deep neural network models for practical applications." arXiv preprint arXiv:1605.07678 (2016).

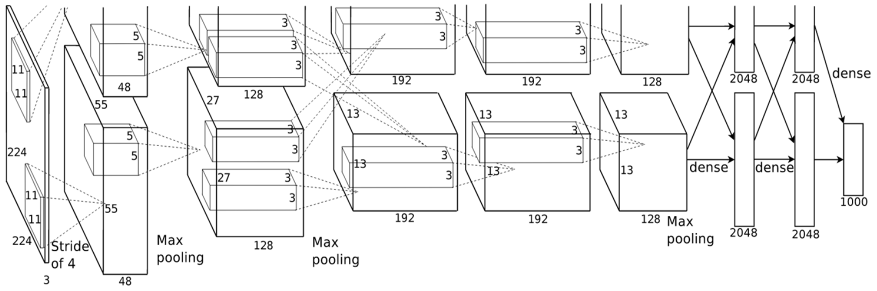
UNIVERSITY
OF TWENTE.

Example: Organ localization in CT

2D image



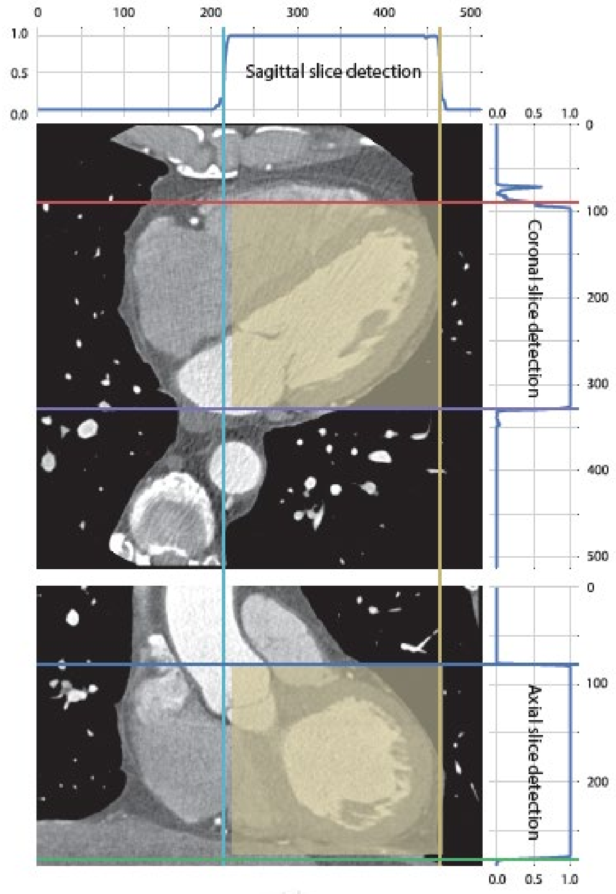
AlexNet

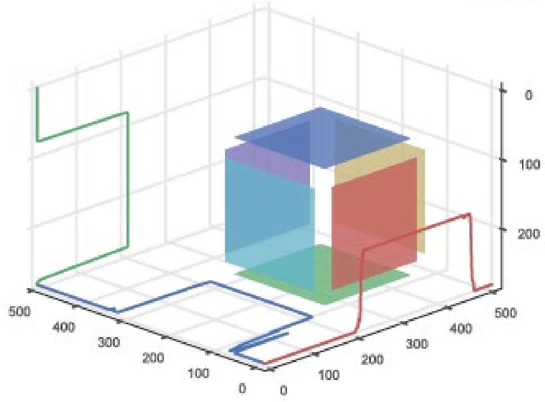


Ventricle?

- Yes
- No

For each image slice

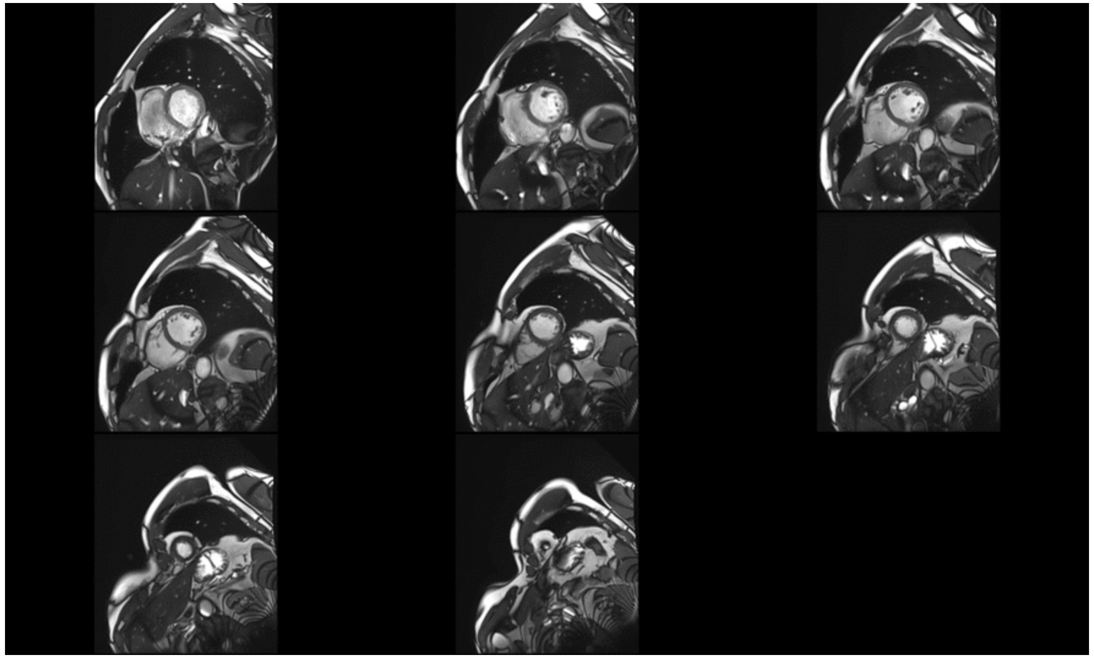




de Vos, B. D., Wolterink, J. M., de Jong, P. A., Leiner, T., Viergever, M. A., & Isgum, I. (2017). ConvNet-based localization of anatomical structures in 3-D medical images. IEEE TMI, 36(7), 1470-1481.

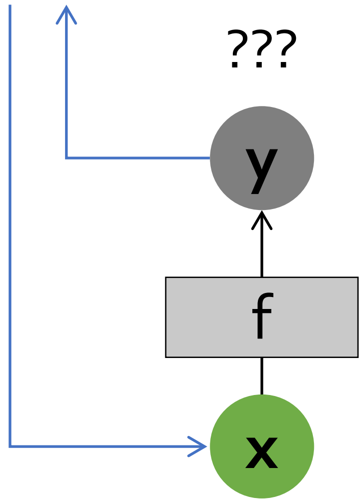
Sequences

- A lot of data is sequential
- E.g. videos, audio, text, ECG, medical images, …
- Can we use this in our neural network?



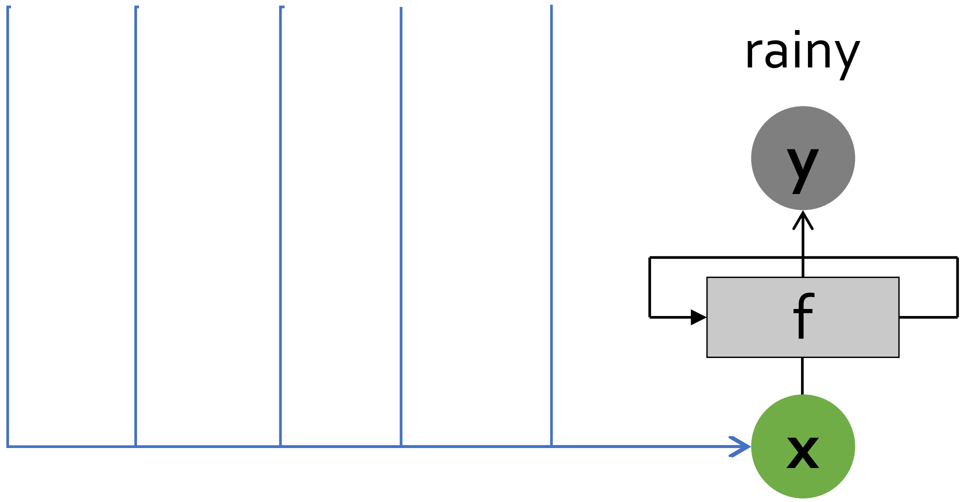
Recurrent neural networks (RNNs)

- It would be good to use information that came before
- A feedforward neural network has no ‘memory’
- Consider training a neural network to predict the next word
    - “It’s September, the weather is …”



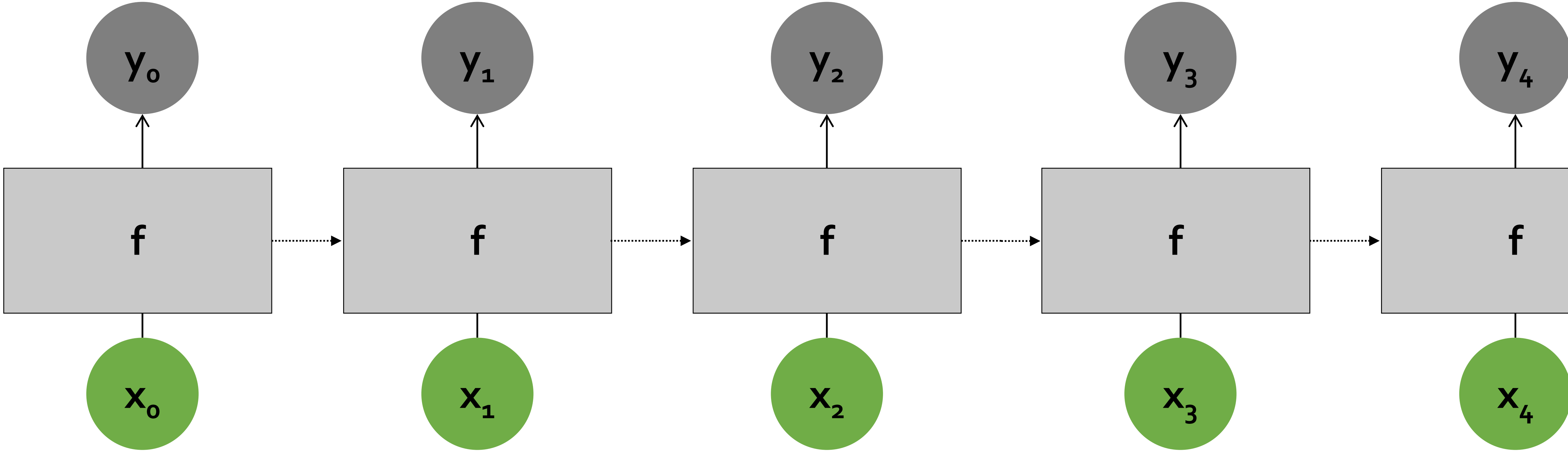
Recurrent neural networks (RNNs)

- It would be good to use information that came before
- A feedforward neural network has no ‘memory’
- Consider training a neural network to predict the next word
    - “It’s September, the weather is …”



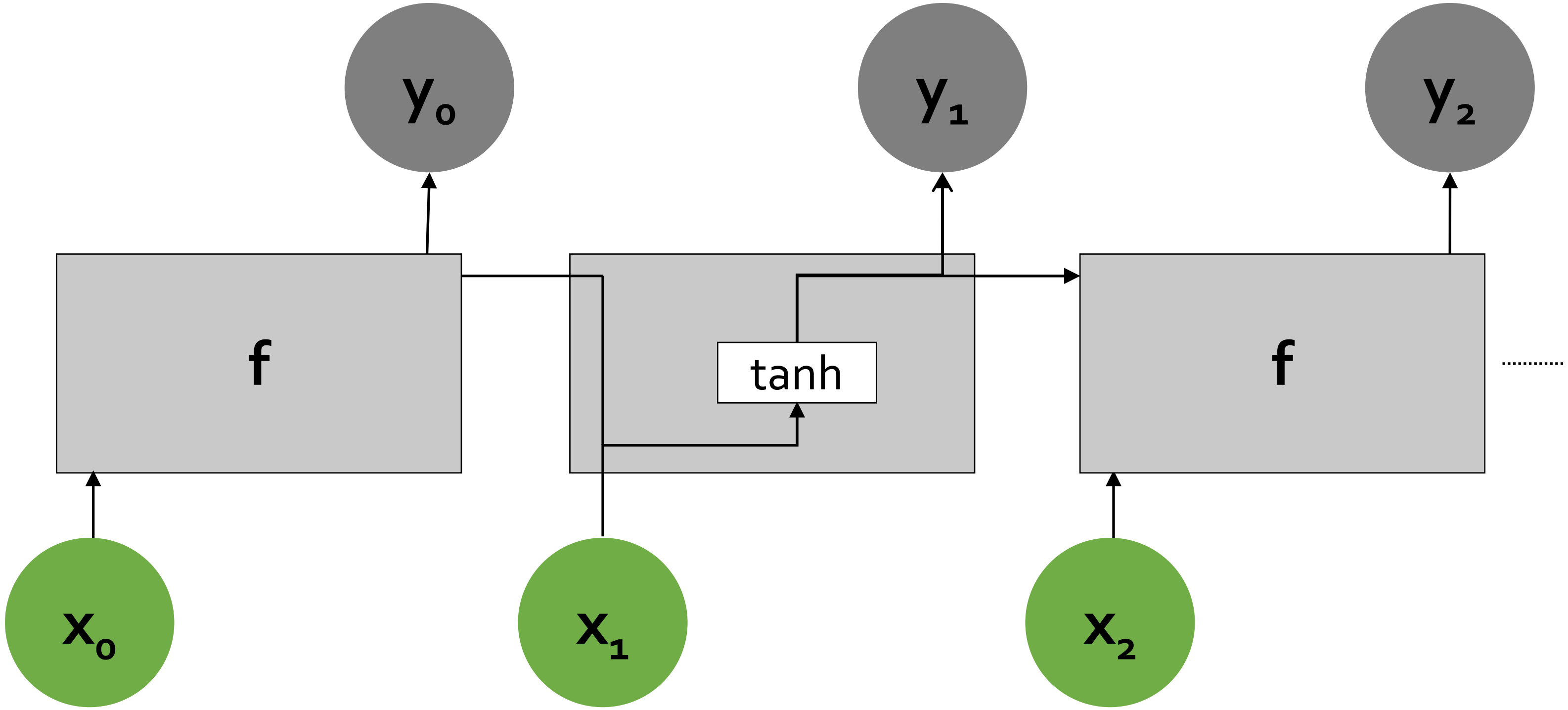
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Unrolling a recurrent neural network



Unrolling a recurrent neural network

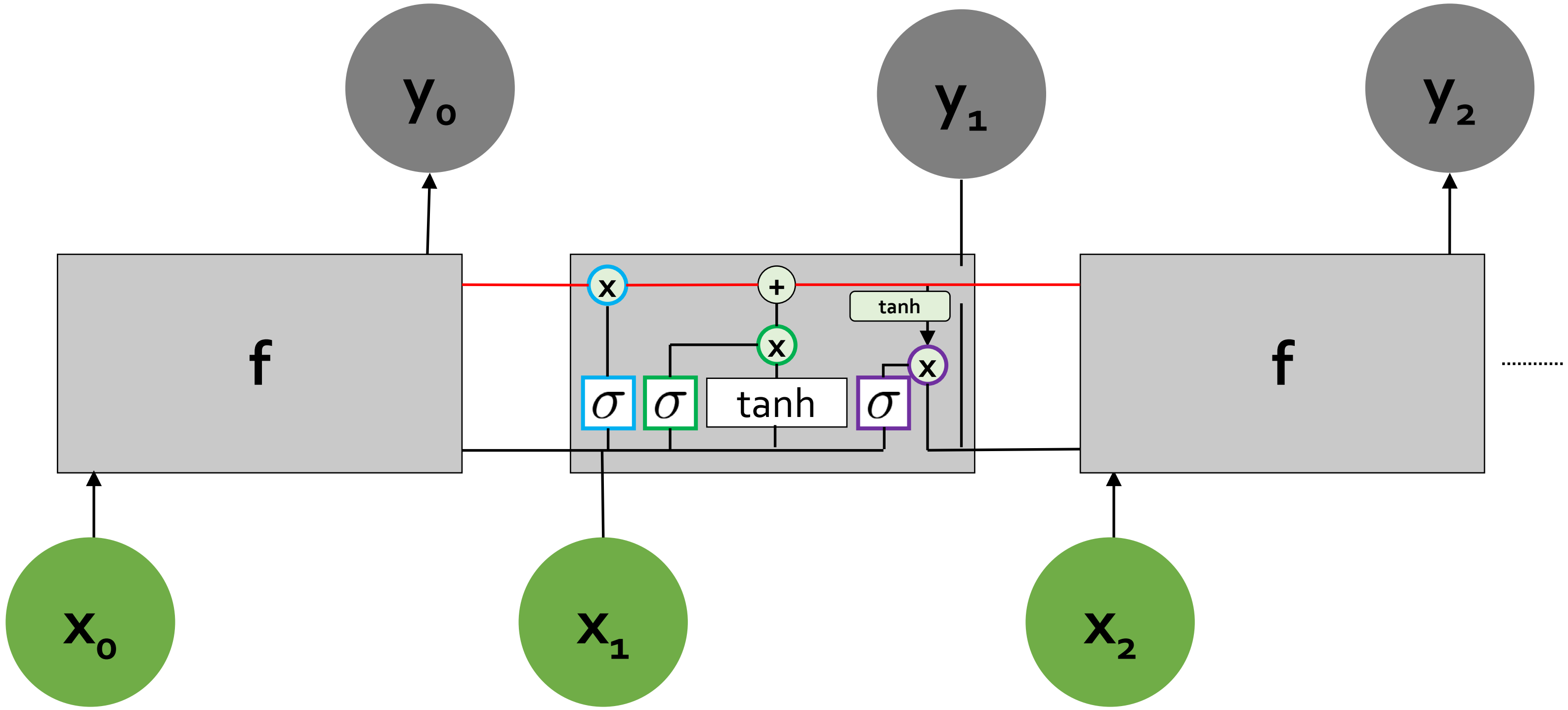
- Traditional RNNs have poor memory
- Previous outputs will get overwritten



http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Long short-term memory (LSTM)

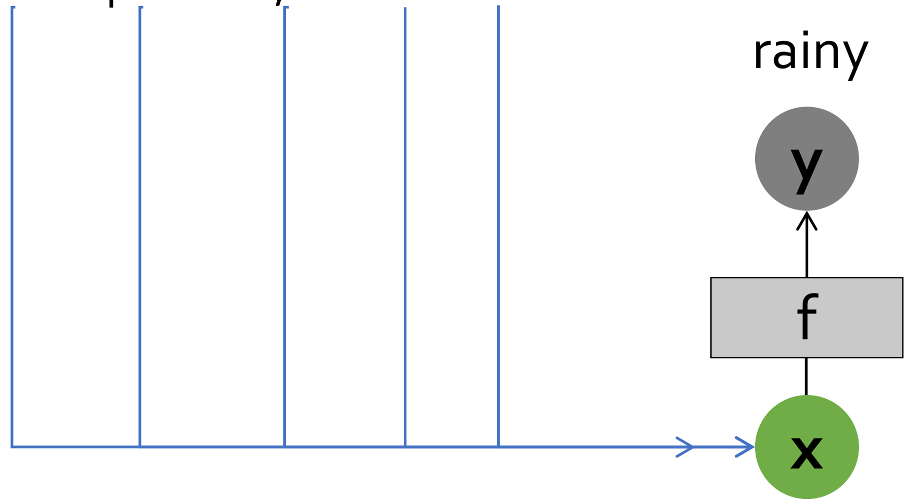
1. Cell state
2. Gates
- Forget gate
- Input gate
- Output gate



Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Recurrent vs. feedforward

- Recurrent networks are intuitively appealing, but
  - feedforward networks are faster (parallel), simpler and they often very competitive
  - “It’s September, the weather is …”



- One way to deal with large contexts in feedforward networks
  - dilated convolutions

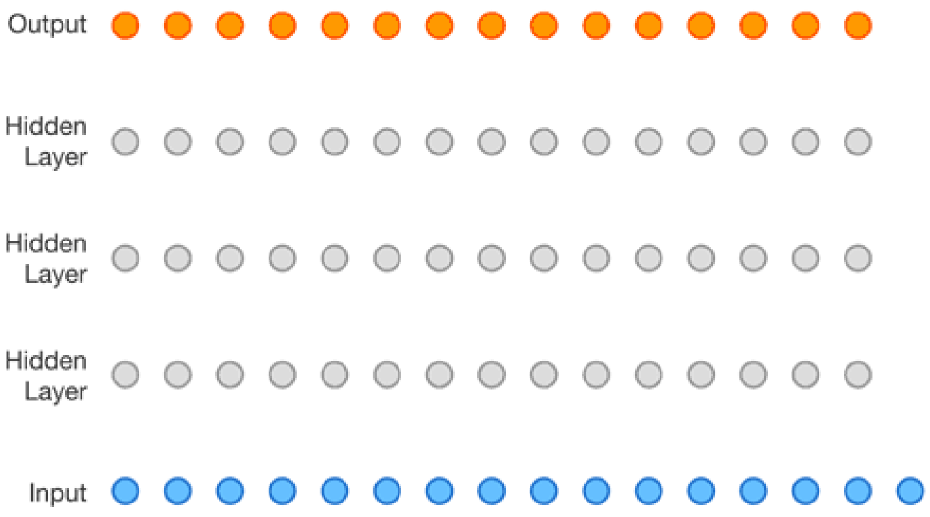
UNIVERSITY OF TWENTE.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.

Dilated convolutions

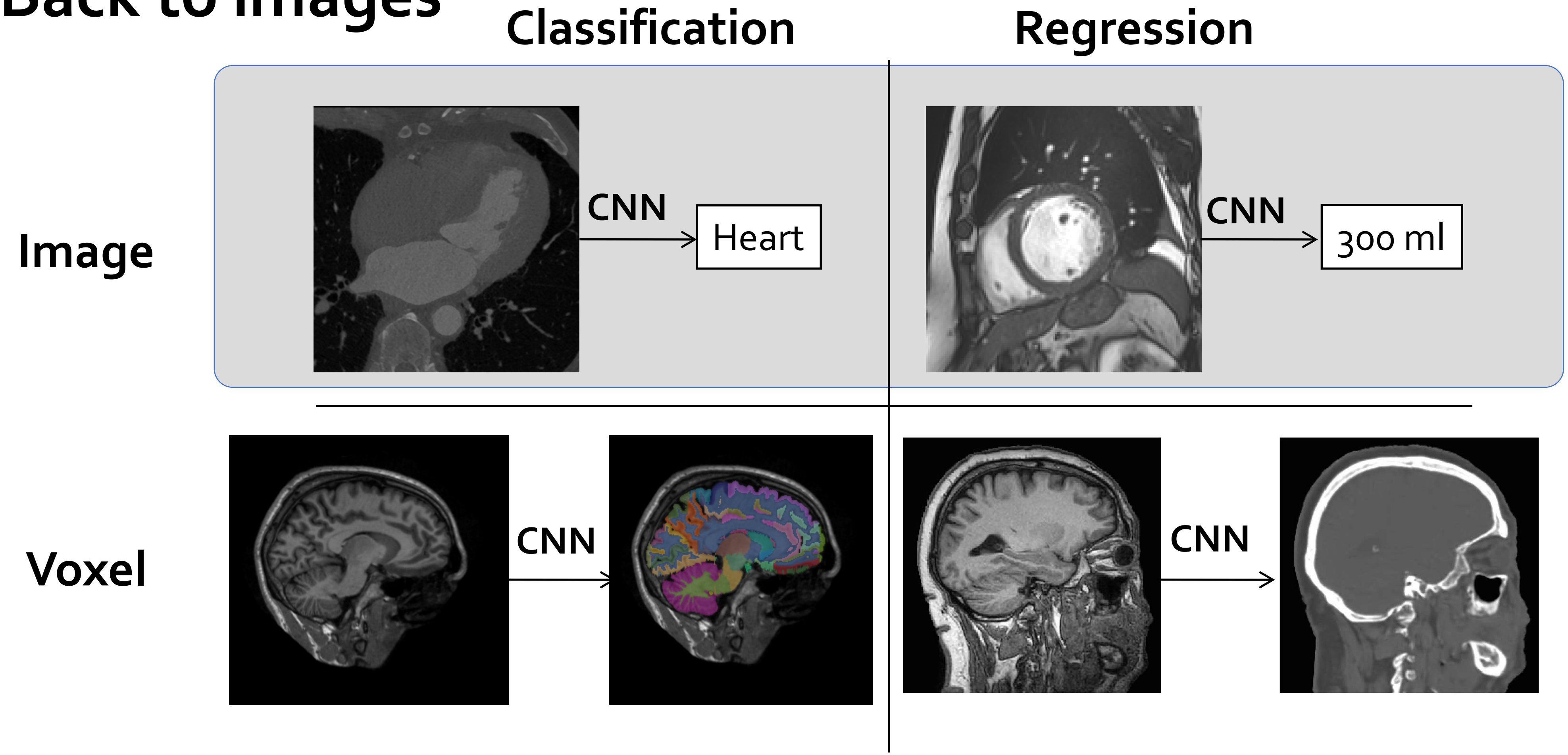
In each layer, add more spacing between elements

- increase receptive field
- prevent explosion in number of parameters



https://deepmind.com/blog/wavenet-generative-model-raw-audio/
Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.

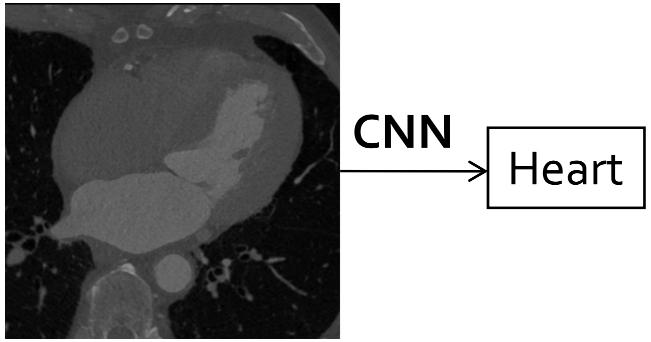
Back to images



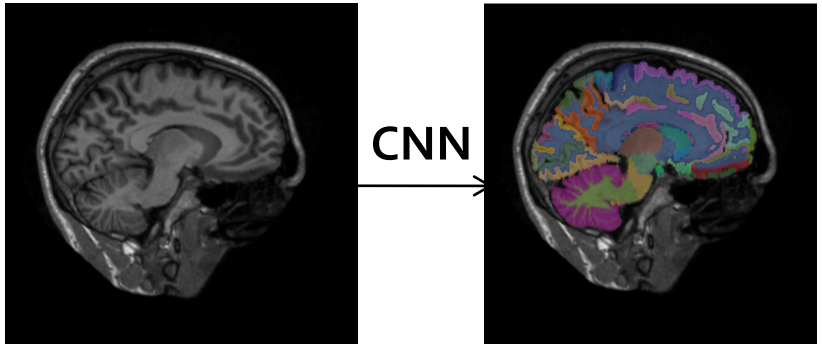
de Vos, Bob D., et al. "ConvNet-based localization of anatomical structures in 3-D medical images." IEEE Trans Med Imaging 36.7 (2017): 1470-1481.
Luo, Gongning, et al. "Multi-views fusion CNN for left ventricular volumes estimation on cardiac MR images." IEEE Transactions on Biomedical Engineering 65.9 (2018): 1924-1934.
Moeskops, Pim, et al. "Automatic segmentation of MR brain images with a convolutional neural network." IEEE transactions on medical imaging 35.5 (2016): 1252-1261.
Wolterink, Jelmer M., et al. "Deep MR to CT synthesis using unpaired data." International Workshop on Simulation and Synthesis in Medical Imaging. Springer, Cham, 2017.

CNNs for pixelwise prediction

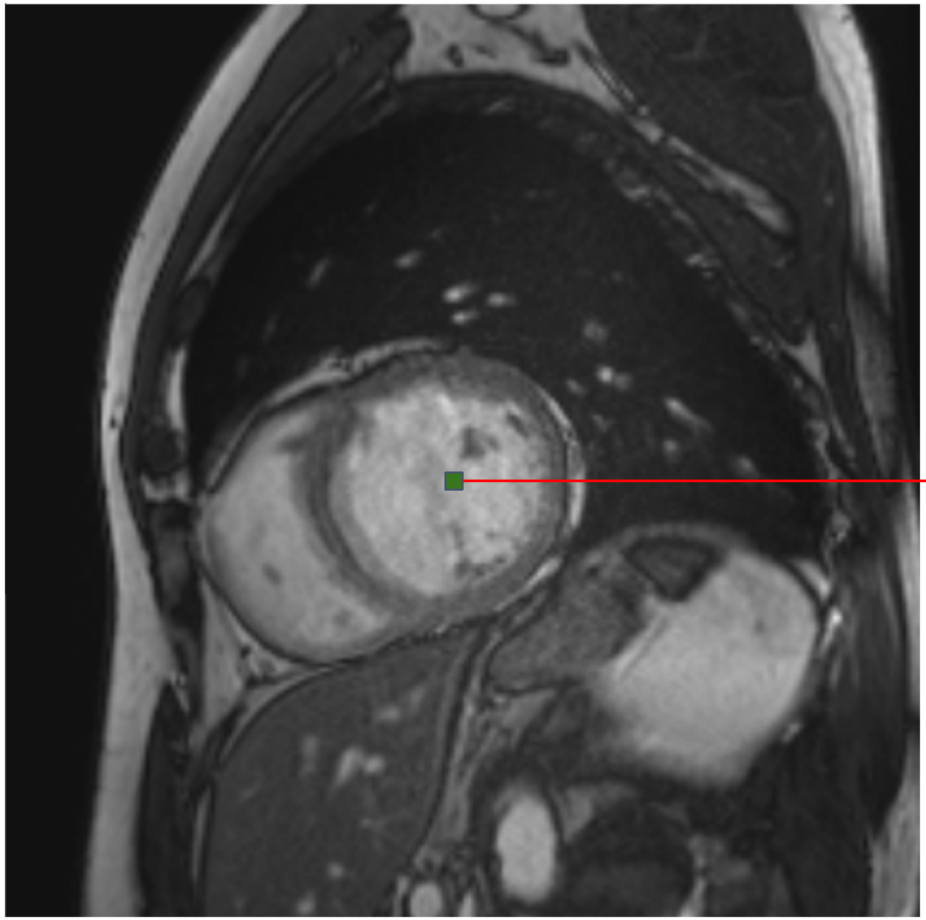
LeNet, AlexNet, VGG-Net, GoogLeNet all predict one value per **image**



Often, we want to predict one value per **pixel/voxel**



Sliding window

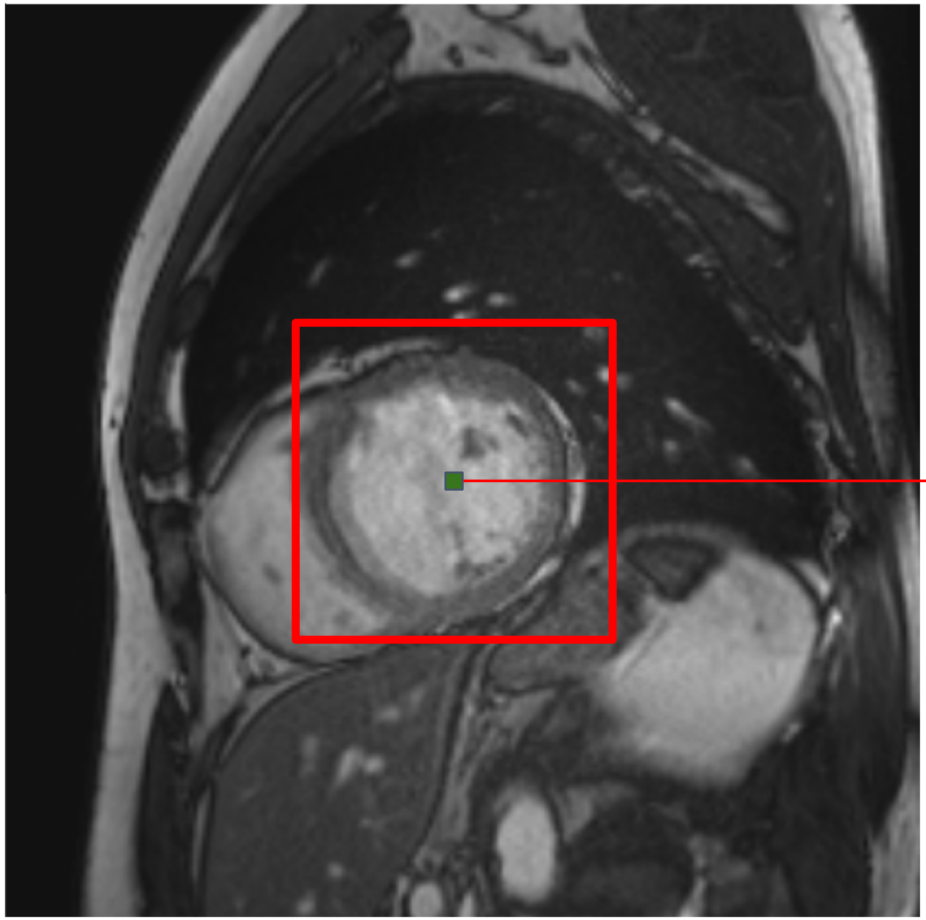


- **Image** = patch centered at voxel
- **Label** = class of center voxel

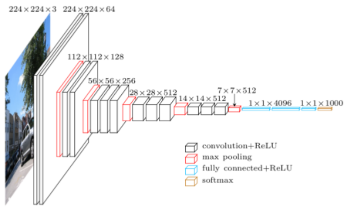
Background
Left ventricle
Myocardium
Right ventricle

UNIVERSITY OF TWENTE.

Sliding window



- **Image** = patch centered at voxel
- **Label** = class of center voxel



Background

Left ventricle

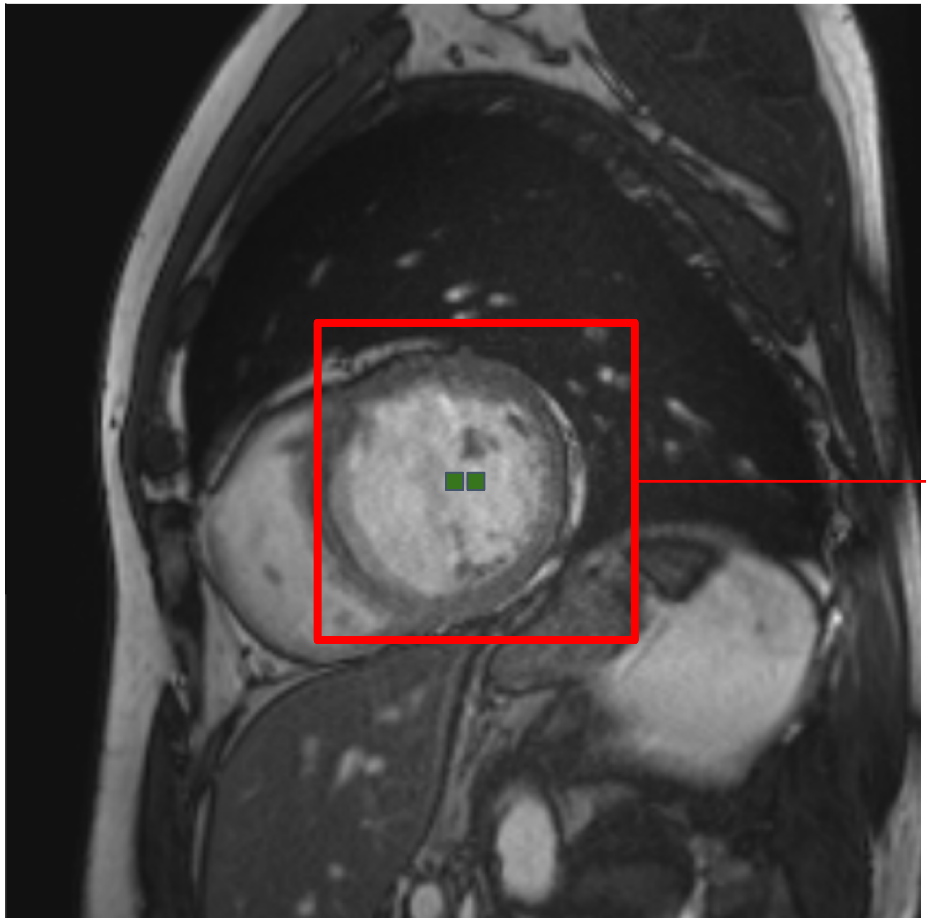
Myocardium

Right ventricle

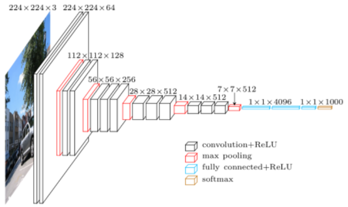
UNIVERSITY
OF TWENTE.

Sliding window

Combination of thousands of image classification tasks



- **Image** = patch centered at voxel
- **Label** = class of center voxel



Background

Left ventricle

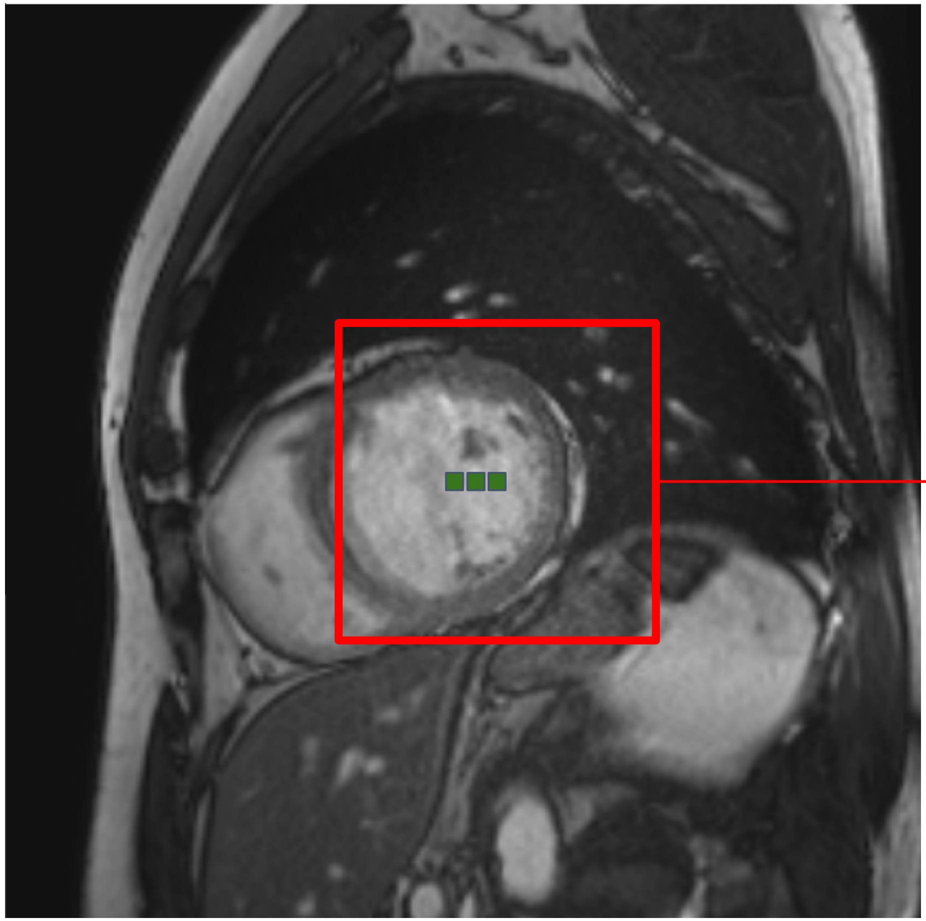
Myocardium

Right ventricle

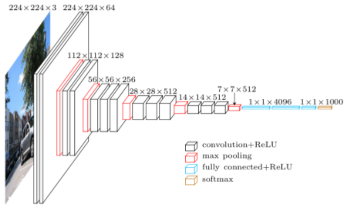
UNIVERSITY
OF TWENTE.

Sliding window

Combination of thousands of image classification tasks

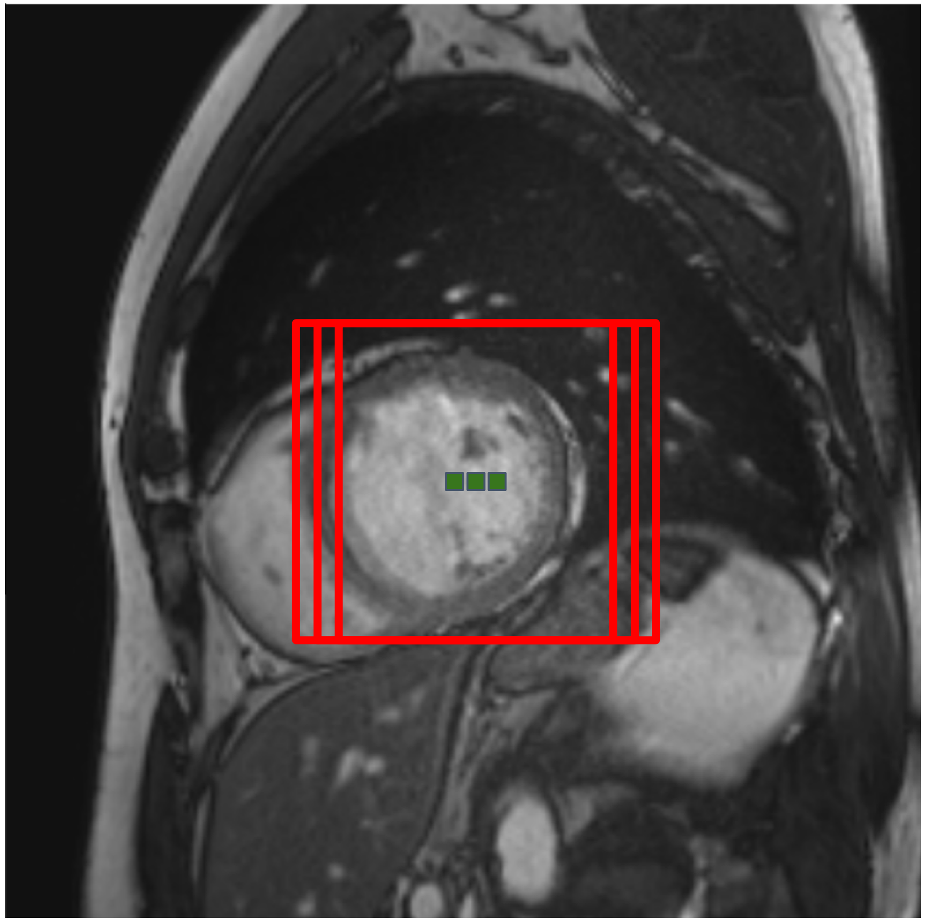


- **Image** = patch centered at voxel
- **Label** = class of center voxel



- Background
- Left ventricle
- Myocardium
- Right ventricle

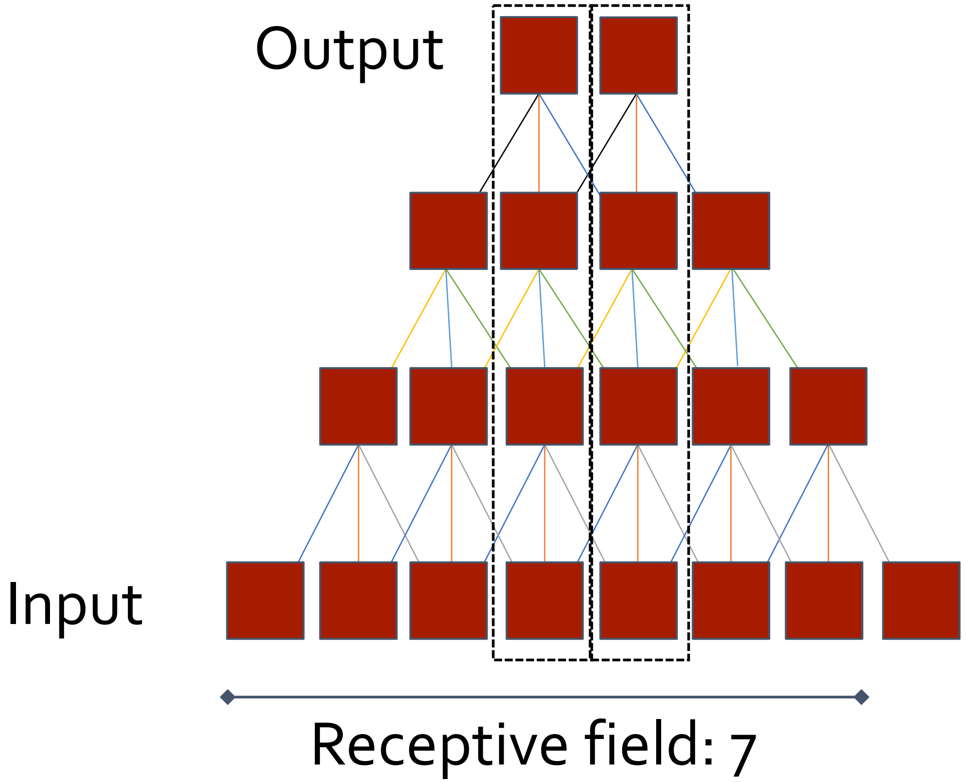
Sliding window



Sliding window approaches are inefficient

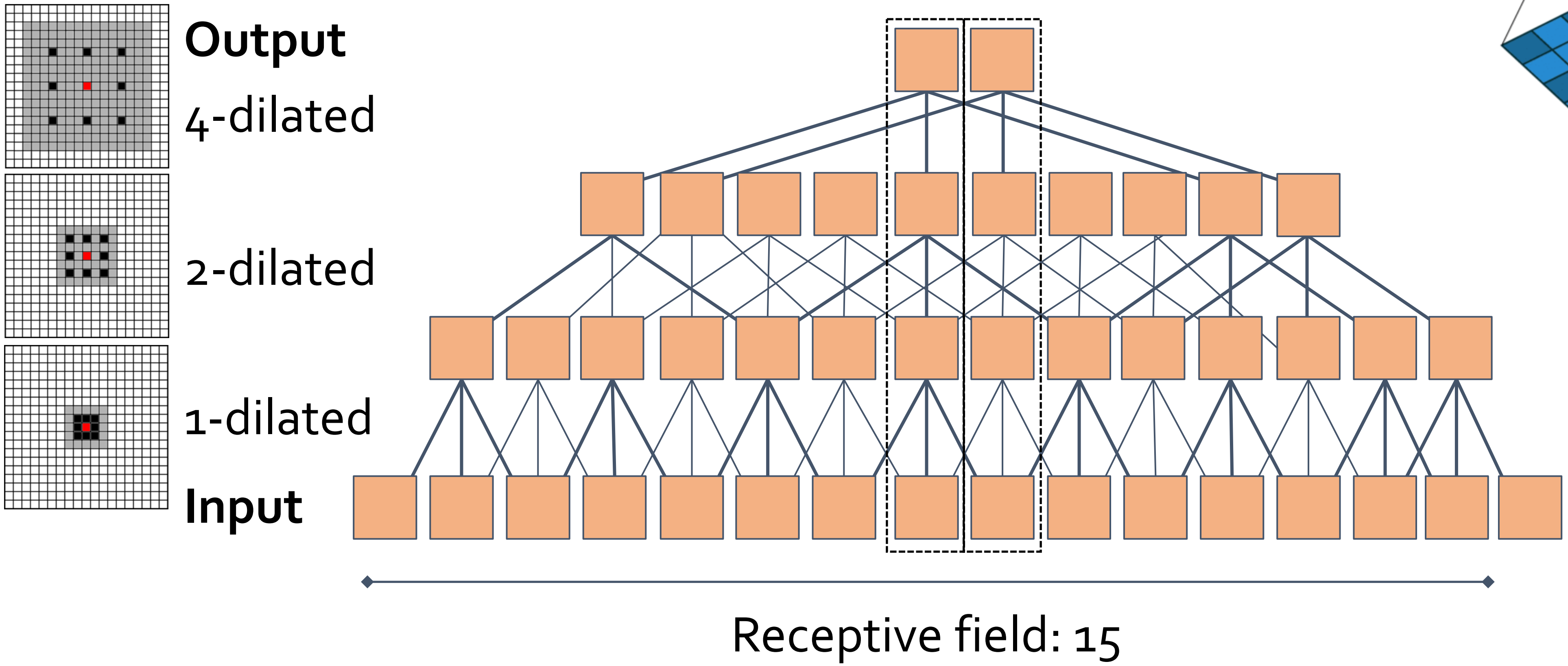
- Each patch is processed separately
- Lots of redundant operations
- We would like to re-use/share operations

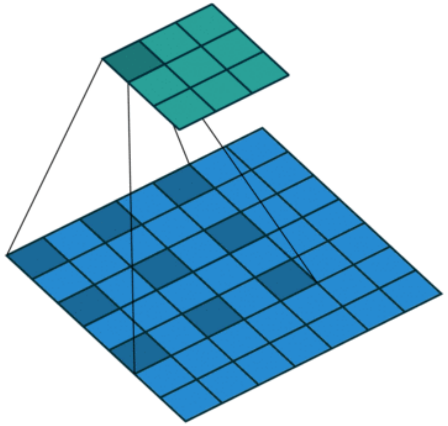
All-convolutional network



UNIVERSITY OF TWENTE.

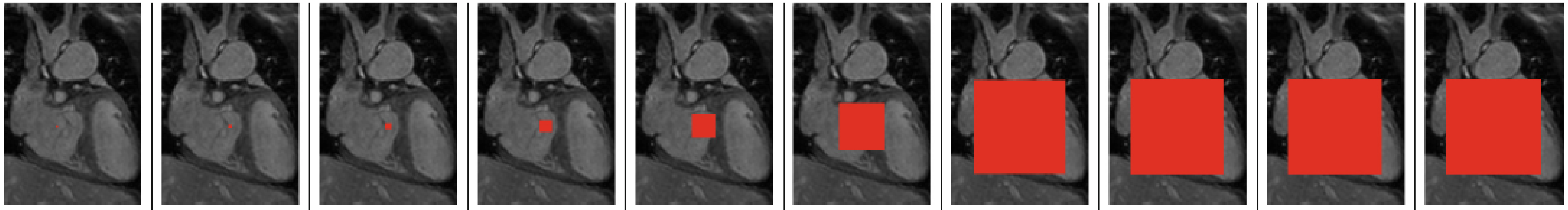
Dilated convolutions





Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

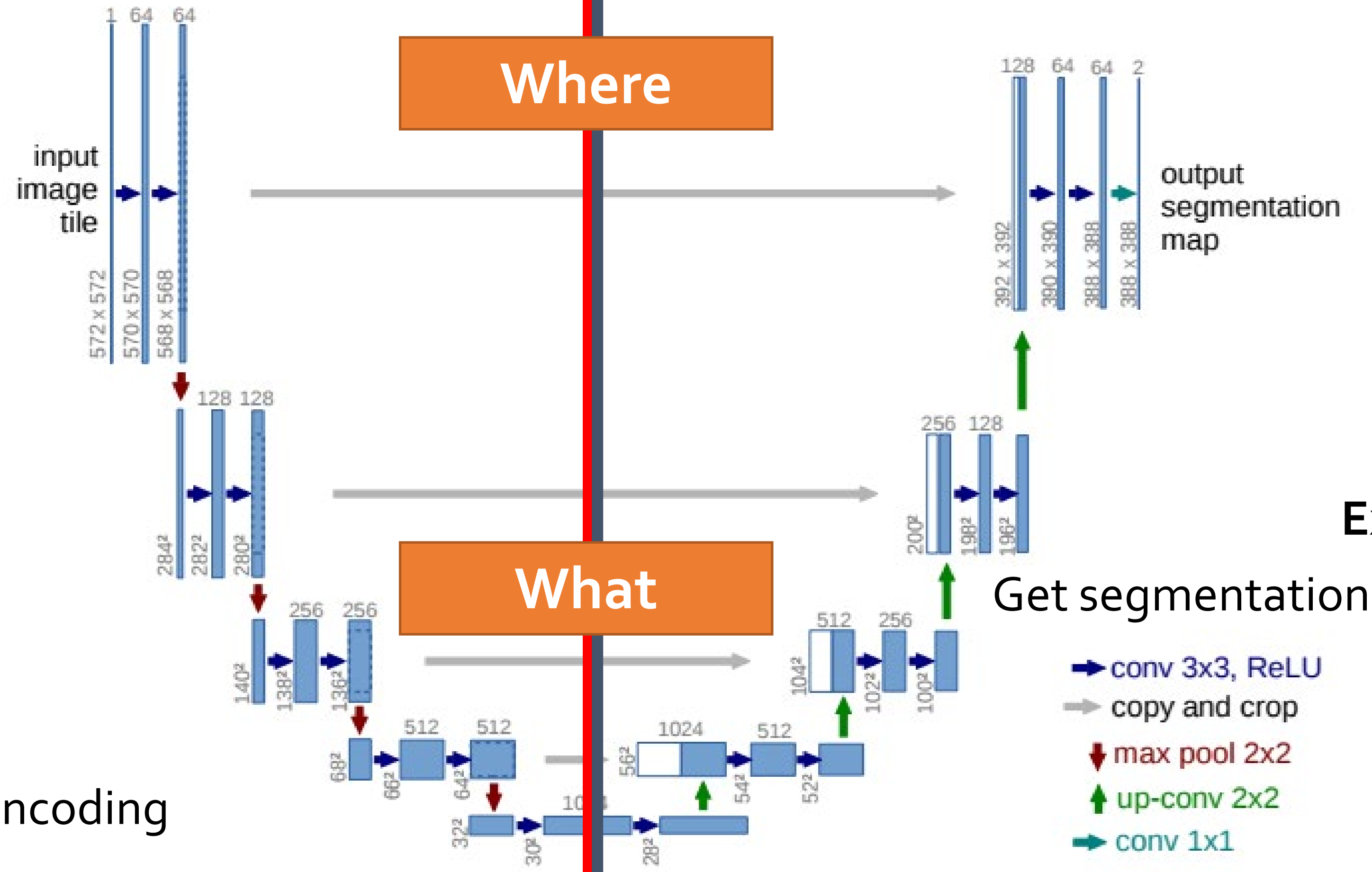
Example



Layer	1	2	3	4	5	6	7	8	9	10
Convolution	3×3	3×3	3×3	3×3	3×3	3×3	3×3	3×3	1×1	1×1
Dilation	1	1	2	4	8	16	32	1	1	1
Field	3×3	5×5	9×9	17×17	33×33	65×65	129×129	131×131	131×131	131×131
Channels	32	32	32	32	32	32	32	32	192	3
Parameters	320	9248	9248	9248	9248	9248	9248	9344	6912	579

Wolterink, J. M., Leiner, T., Viergever, M. A., & Išgum, I. (2016). Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. In RAMBO (pp. 95-102). Springer, Cham.

Encoder-decoder architecture: U-Net



Contracting path

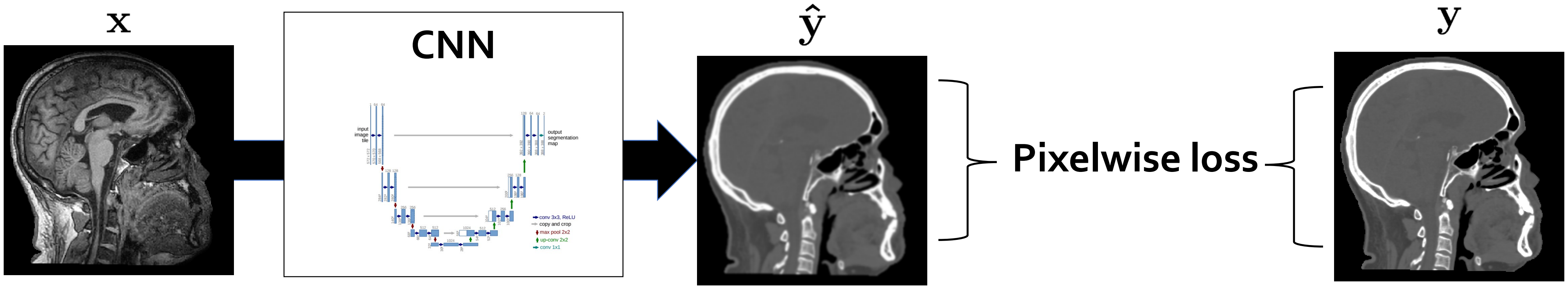
Get multi-resolution encoding

Expanding path

Get segmentation from encoding

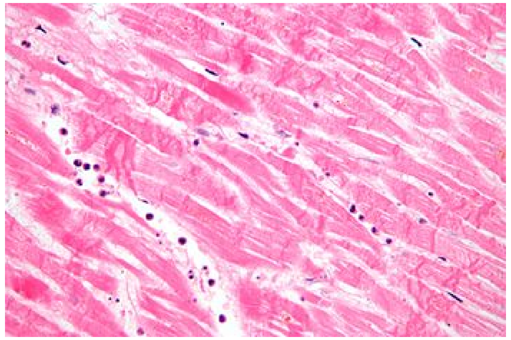
Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In MICCAI (pp. 234-241). Springer, Cham.

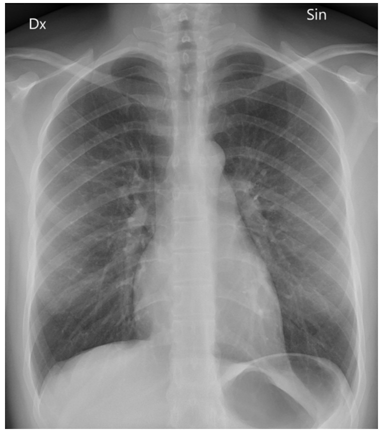
Training



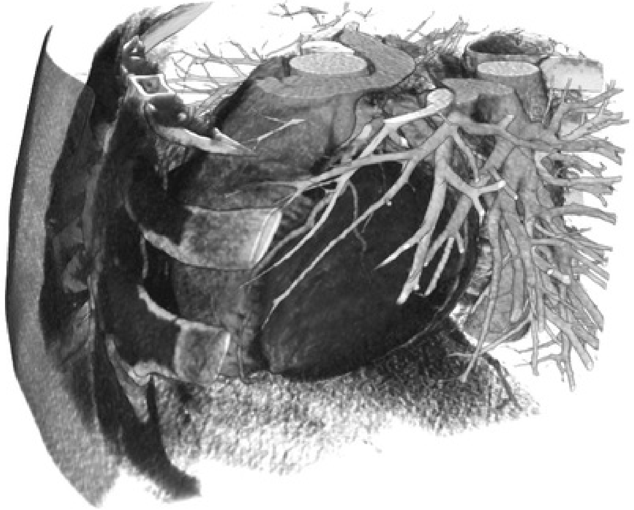
2D or 3D images

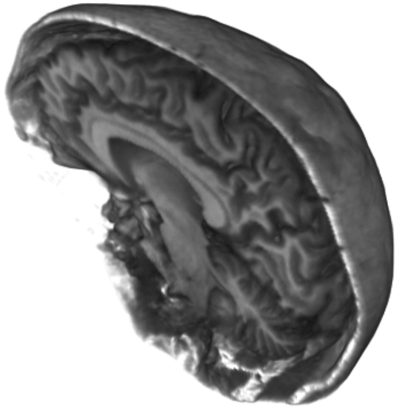
2D data





3D data





https://en.wikipedia.org/wiki/Histopathology
https://en.wikipedia.org/wiki/Fundus_photography
https://en.wikipedia.org/wiki/Chest_radiograph

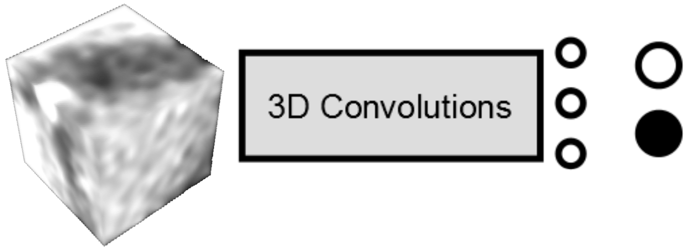
3D networks

Many medical images are 3D instead of 2D

- MR images
- CT images

Can we just use 3D layers instead of 2D layers? Sure!

- 3D convolution layers in Keras, TensorFlow, PyTorch, etc.
- 3D network architectures (e.g. U-Net, V-Net)



But

- Is your data really 3D (think about acquisition)? Isotropy?
- Increase in memory consumption + operations + parameters

Summary

Advanced architectures

- AlexNet, GoogleNet, VGG-Net, ResNet
- Deeper, larger, better + some tricks

Recurrent neural networks

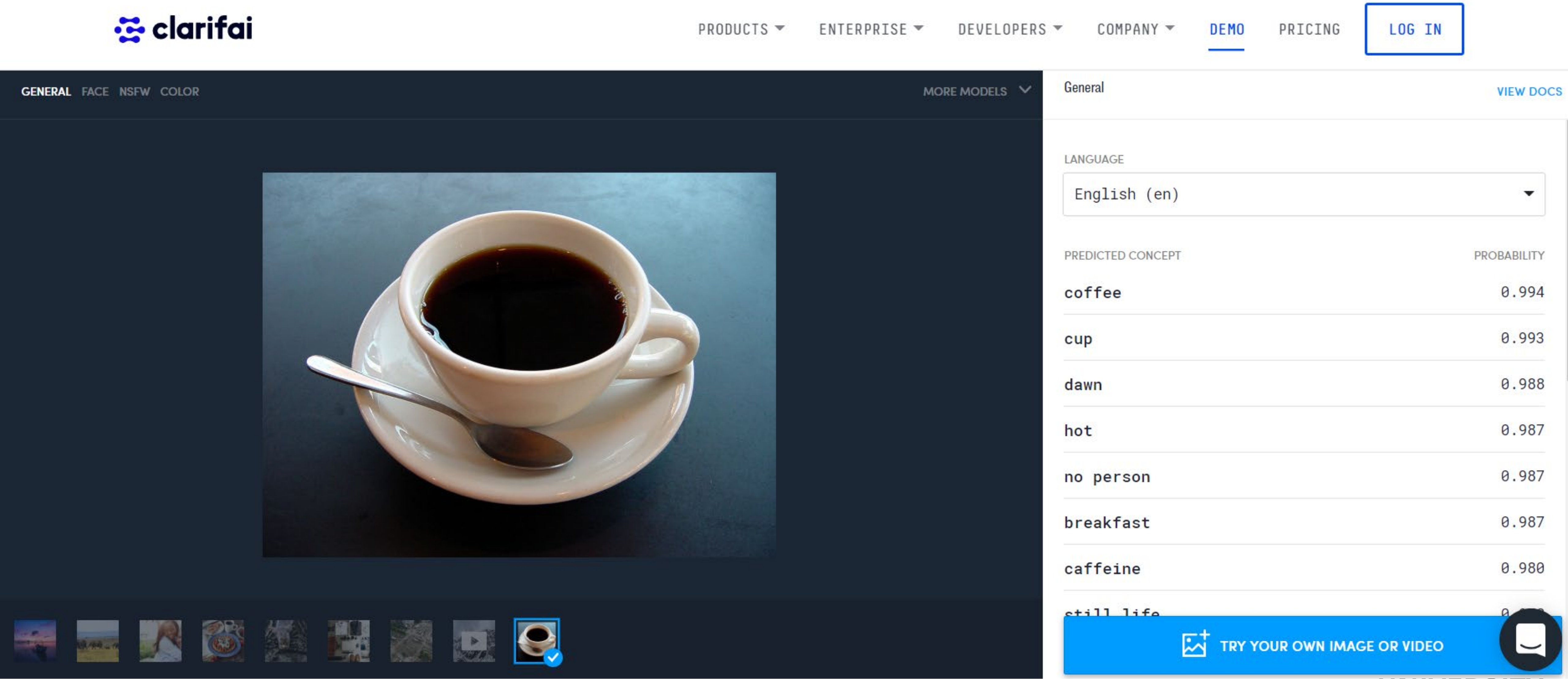
- RNNs + LSTMs

Per image prediction != per voxel prediction

- All-convolutional networks
- Encoder-decoder architectures

2D/3D data

- Most 2D neural networks extend to 3D


clarifai
PRODUCTS
ENTERPRISE
DEVELOPERS
COMPANY
DEMO
PRICING
LOG IN
GENERAL FACE NSFW COLOR
MORE MODELS
General
VIEW DOCS
LANGUAGE
English (en)
PREDICTED CONCEPT
PROBABILITY
coffee 0.994
cup 0.993
dawn 0.988
hot 0.987
no person 0.987
breakfast 0.987
caffeine 0.980
TRY YOUR OWN IMAGE OR VIDEO