LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

# Deep Learning

## Chapter 6: CNNs - Famous Architectures

**Mina Rezaei**

Department of Statistics – LMU Munich

Winter Semester 2020

# LECTURE OUTLINE

# FAMOUS ARCHITECTURES: LENET

- Pioneering work on CNNs by Yann Lecun in 1998[Le Cun et. al, 1998].
- Applied on the MNIST dataset for automated handwritten digit recognition.
- Consists of convolutional, "subsampling" and dense layers.
- Complexity and depth of the net was mainly restricted by limited computational power back in the days.

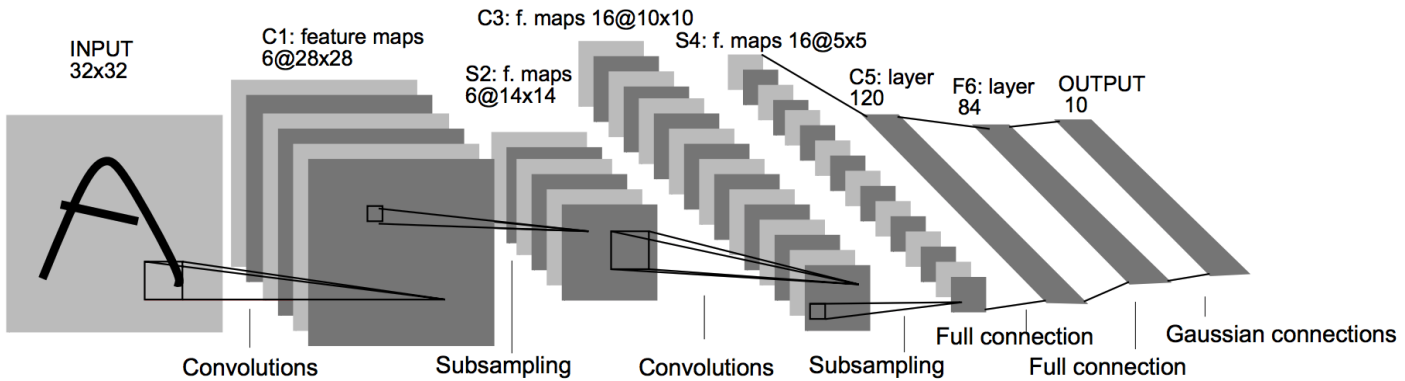



**Figure:** LeNet architecture: two conv layers with subsampling, followed by dense layers and a 'Gaussian connections' layer.

# FAMOUS ARCHITECTURES: LENET

- A neuron in a subsampling layer looks at a $2 \times 2$ region of a feature map, sums the four values, multiplies it by a trainable coefficient, adds a trainable bias and then applies a sigmoid activation.
- A stride of 2 ensures that the size of the feature map reduces by about a half.
- The 'Gaussian connections' layer has a neuron for each possible class.
- The output of each neuron in this layer is the (squared) Euclidean distance between the activations from the previous layer and the weights of the neuron.

# FAMOUS ARCHITECTURES: ALEXNET

- Introduced by [Krizhevsky et al., 2012], won the ImageNet challenge in 2012 and is basically a deeper version of the LeNet.
- Trained in parallel on two GPUs, using two streams of convolutions which are partly interconnected.
- Contains ReLU activations and makes use of data set augmentation strategies.

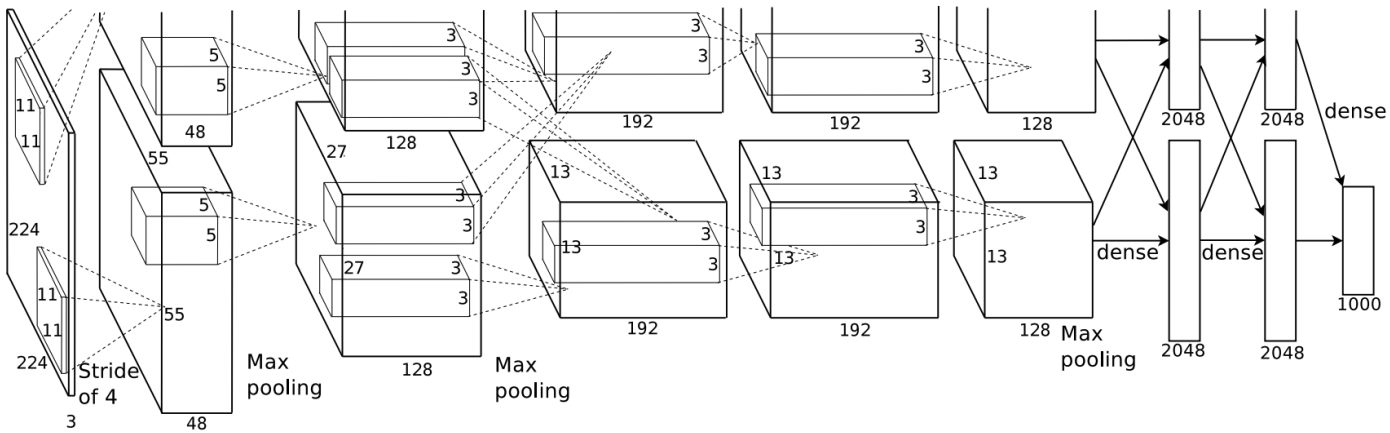
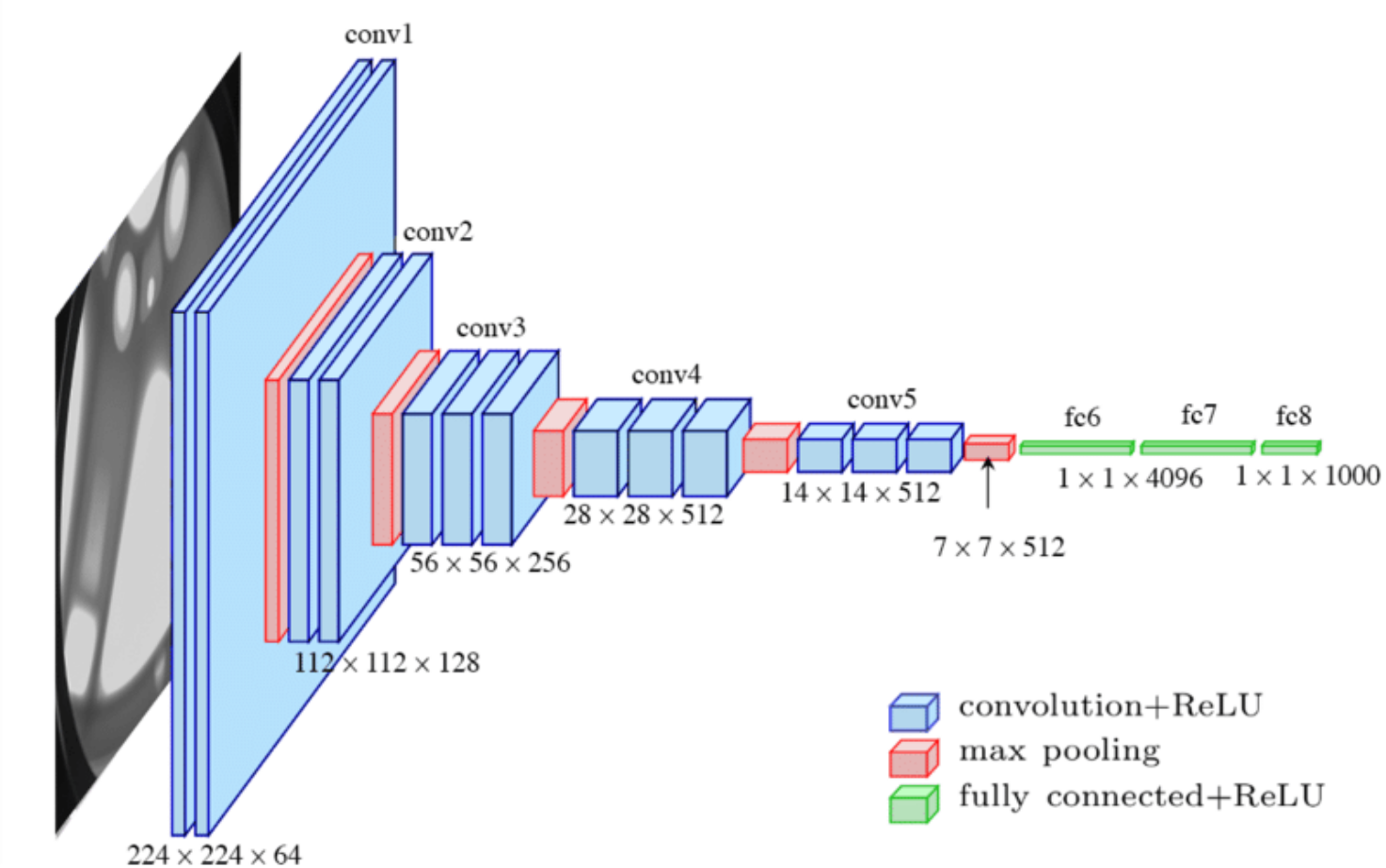



**Figure:** AlexNet architecture.

# FAMOUS ARCHITECTURES: VGG

- Architecture introduced by [Simoyan et. al, 2014] as “Very Deep Convolutional Network”.
- A deeper variant of the AlexNet.
- Mainly uses many convolutional layers with a small kernel size 3x3.
- Performed very well in the ImageNet Challenge 2014.
- Exists in a small version (VGG16) with a total of 16 layers and a larger version (VGG19) with 19 layers.

# FAMOUS ARCHITECTURES: VGG




credit : Max Ferguson

**Figure:** VGG Net 16 with 13 conv layers.

# FAMOUS ARCHITECTURES: VGG

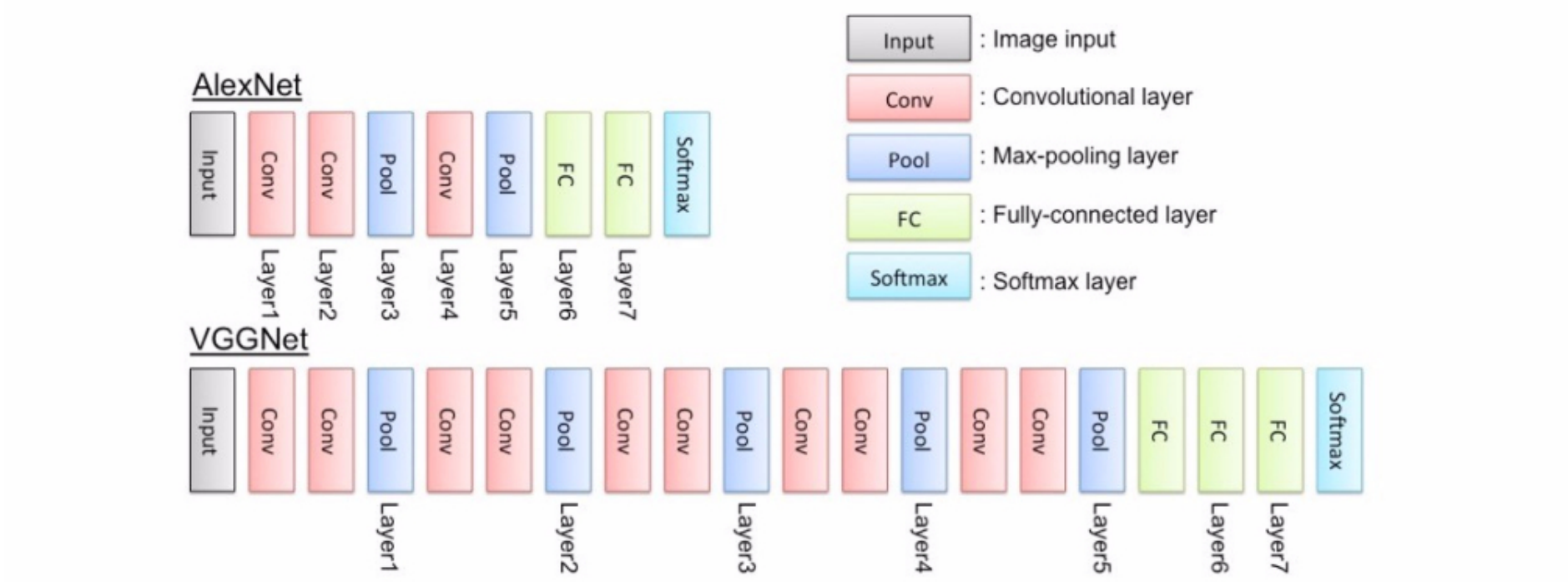



**Figure:** VGG Net with 10 conv layers in comparison with AlexNet.

# FAMOUS ARCHITECTURES: GOOGLENET

- The net architecture that first made use of inception modules [Szegedy et. al , 2014].
- Also referred to as Inception Net.
- Also includes Batch Normalization to improve training.
- Uses auxiliary losses: branches throughout the net that consist of softmax layers that make predictions using early stage layers.
- Those losses are jointly optimized and the output from the final head is used for deployment and testing of the architecture.
- Inception modules allow the net to “choose” between different kernels.

# FAMOUS ARCHITECTURES: GOOGLENET

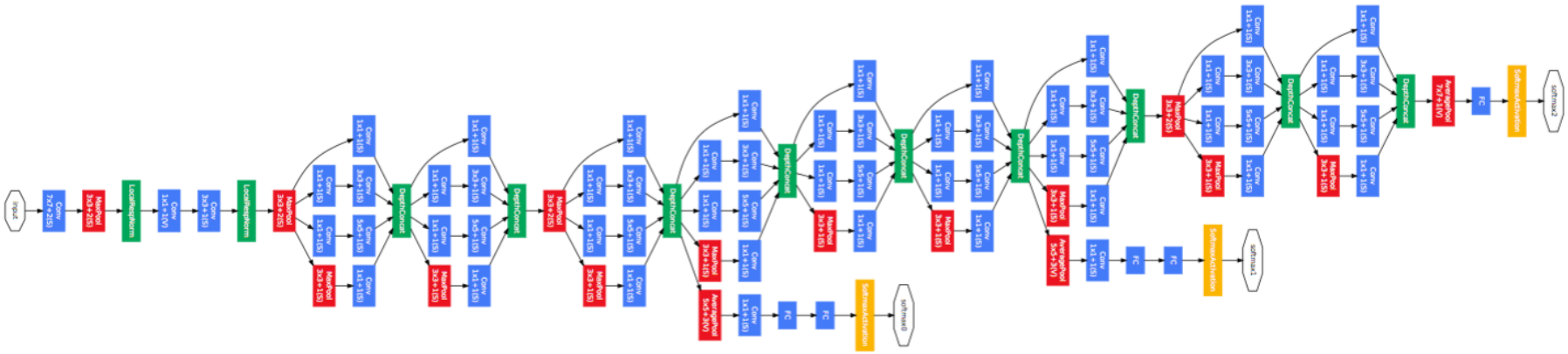


**Figure:** GoogleNet architecture. Yellow rectangles mark the auxiliary losses which are combined during training.

# FAMOUS ARCHITECTURES: GOOGLENET

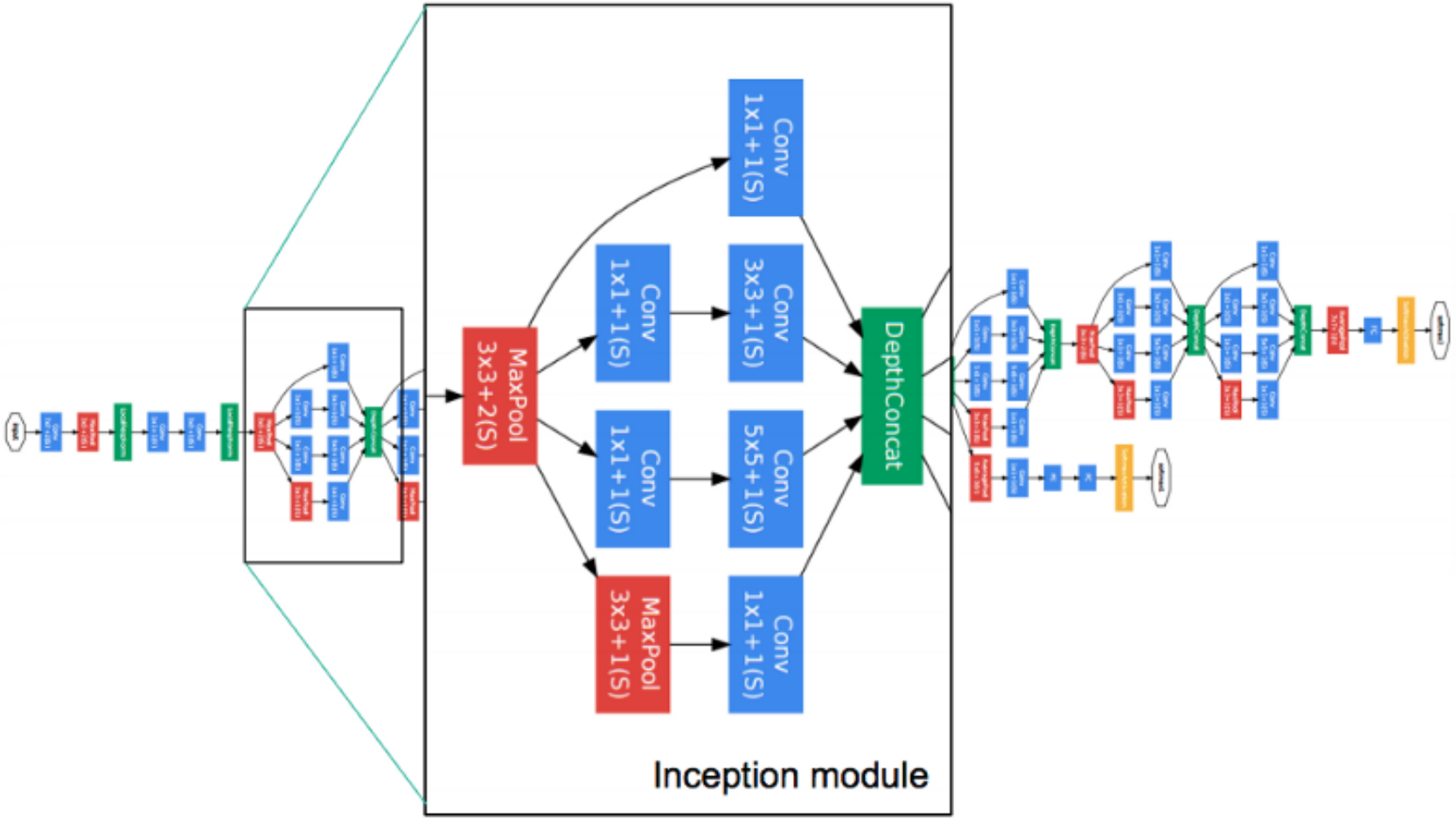



**Figure:** Zoom view in one of the inception modules upon which the GoogLeNet architecture is build.

# FAMOUS ARCHITECTURES: RESNET

- Net architecture that makes use of skip connections / residual learning.
- This allowed [He et. al , 2015] to create a very deep net architectures of up to 152 layers.
- Batch normalization and global average pooling is used.

# FAMOUS ARCHITECTURES: RESNET

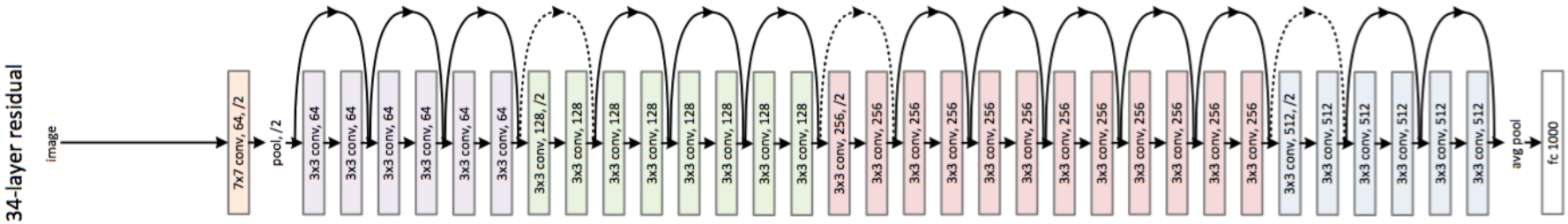



**Figure:** A deep ResNet architecture with a total of 34 layers.

# FAMOUS ARCHITECTURES: RESNET

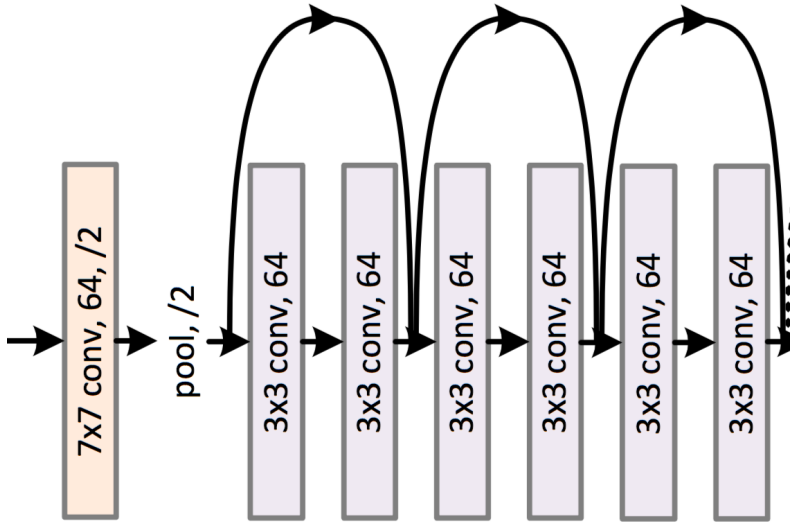



**Figure:** Zoom view in the ResNet architecture. The arrows mark residual connections which allow the net to “skip” layers which do not contribute to an improvement in the prediction performance.

# FAMOUS ARCHITECTURES: U-NET

- U-Net [Ronneberger et al., 2015] is a fully convolutional net that makes use of upsampling (via transposed convolutions, for example) as well as skip connections.
- Input images are getting convolved and down-sampled in the first half of the architecture.
- Then, they are getting upsampled and convolved again in the second half to get back to the input dimension.
- Skip connections throughout the net combine feature maps from earlier layers with those from later layers by concatenating both sets of maps along the depth/channel axis.
- Only convolutional and no dense layers are used.

# FAMOUS ARCHITECTURES: U-NET

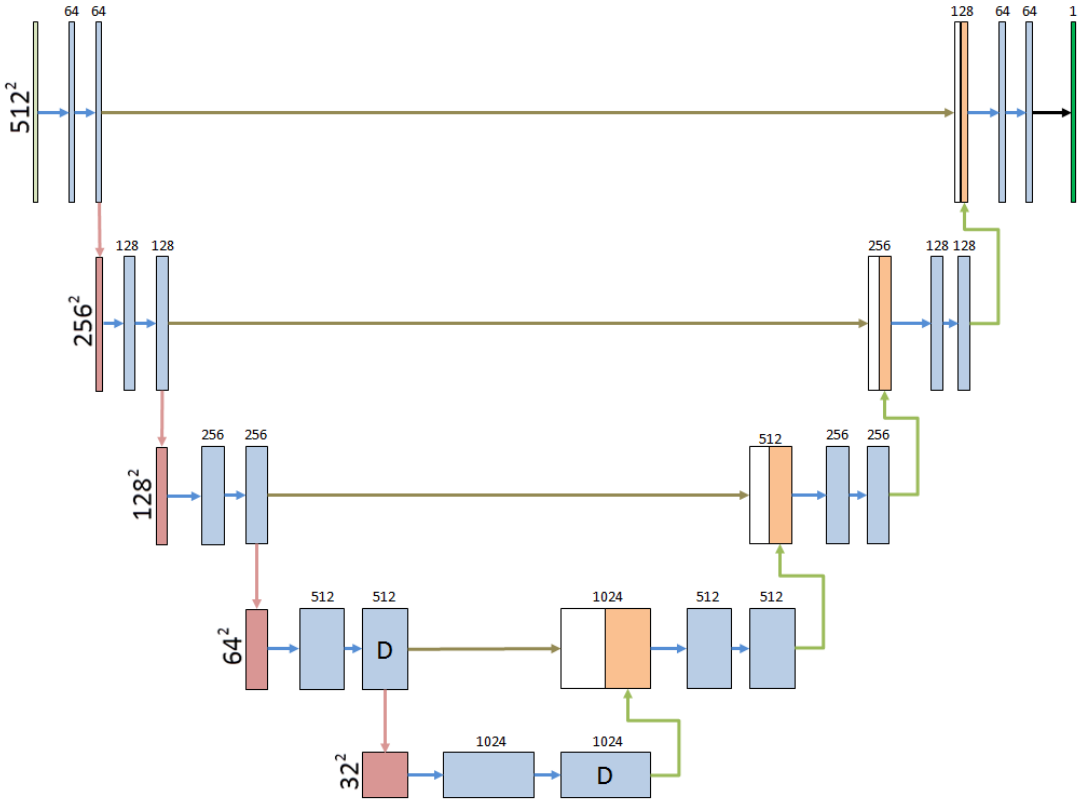



**Figure:** Illustration of the architecture. Blue arrows are convolutions, red arrows max-pooling operations, green arrows upsampling steps and the brown arrows merge layers with skip connections. The height and width of the feature blocks are shown on the vertical and the depth on the horizontal. D are dropout layers.

# FAMOUS ARCHITECTURES: U-NET - EXAMPLE

- Example problem setting: train a neural net to pixelwise segment roads in satellite imagery.
- Answer the question: **Where is the road map?**

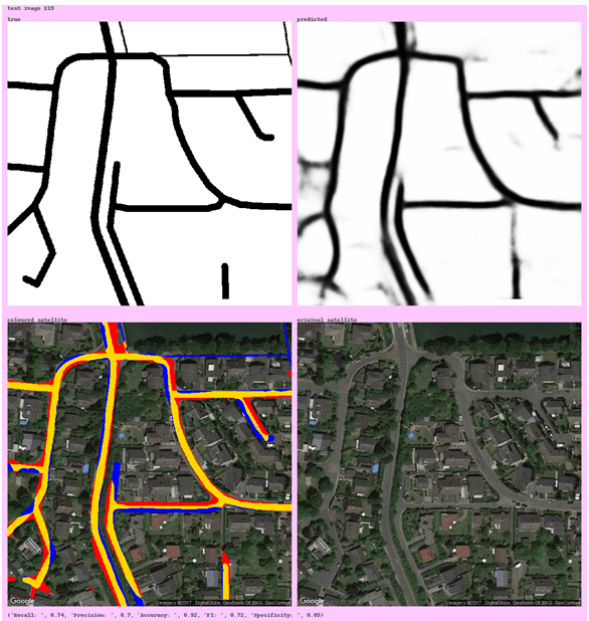
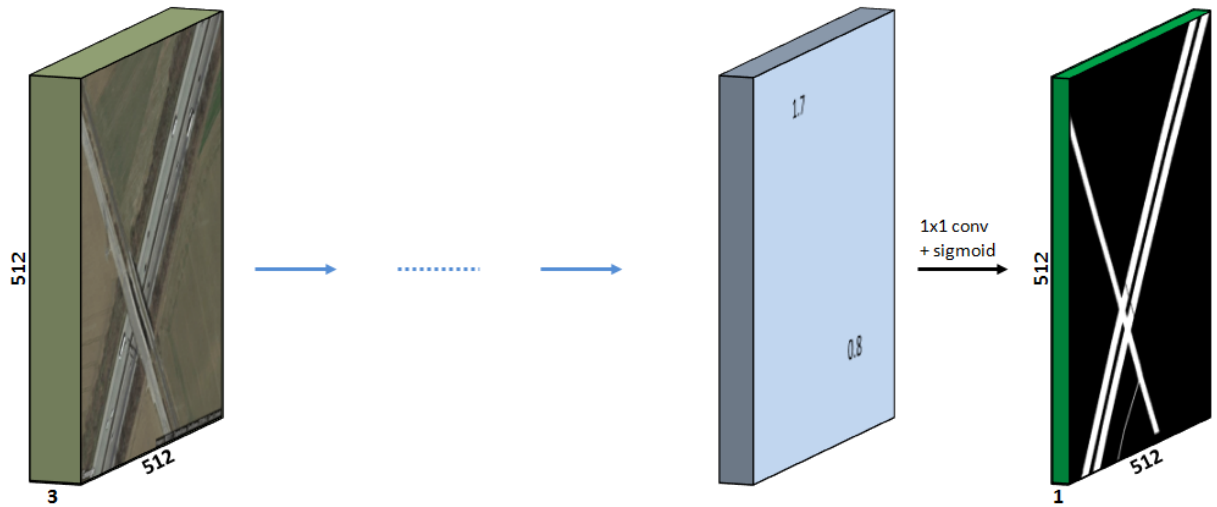


**Figure:** Model prediction on a test satellite image. Yellow are correctly identified pixels, blue false negatives and red false positives.

# FAMOUS ARCHITECTURES: U-NET - EXAMPLE

- The net takes an RGB image [512, 512, 3] and outputs a binary (road / no road) probability mask [512, 512, 1] for each pixel.
- The model is trained via a binary cross entropy loss which was combined over each pixel.




$$Output_{i,j} = \frac{1}{1+exp(-1.7 \cdot \mathbf{w})} \in (0,1)$$

**Figure:** Scheme for the input/ output of the net architecture.

# FAMOUS ARCHITECTURES: SUMMARY

- Main points from the previous architectures:
  - LeNet was one of the first successful applications of CNNs.
  - AlexNet is a deeper version of LeNet.
  - VGG is a deeper version of AlexNet.
  - ResNet makes heavy use of skip connections.
  - GoogLeNet uses inception modules.
  - Batch normalization is often used in modern architectures.
  - Global average pooling is also a prominent module in modern architectures.
- There exists a great variety of different architectures which perform very well on different tasks.
- Almost all of the above described architectures exist in extended versions such as the Inception V1 - V4, deeper and shallower ResNets, ResNets for time series ... it is up to you to discover them all.

# REFERENCES


Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton (2012)

ImageNet Classification with Deep Convolutional Neural Networks

https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke and Andrew Rabinovich (2014)

Going deeper with convolutions

https://arxiv.org/abs/1409.4842

Kaiming He, Zhang, Xiangyu, Ren, Shaoqing, and Jian Sun (2015)

Deep Residual Learning for Image Recognition

https://arxiv.org/abs/1512.03385

LeCunn Yann, Bottou, Leon, Bengio, Yoshua and Patrick Haffner (1998)

Global training of document processing systems using graph transformer networks

http://vision.stanford.edu/cs598_spring07/papers/Lecun98.pdf

# REFERENCES


Simoyan Karen and Andrew Zisserman (2014)

Very Deep Convolutional Networks for Large-Scale Image Recognition

https://arxiv.org/abs/1409.1556

Olaf Ronneberger, Philipp Fischer, Thomas Brox (2015)

U-Net: Convolutional Networks for Biomedical Image Segmentation

http://arxiv.org/abs/1505.04597