EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Image Analysis II

AlexNet, ZFNet, VGGNet, GoogLeNet, ResNet, …

Radovan Fusek

VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE | DEPARTMENT OF COMPUTER SCIENCE

CovNet Architectures

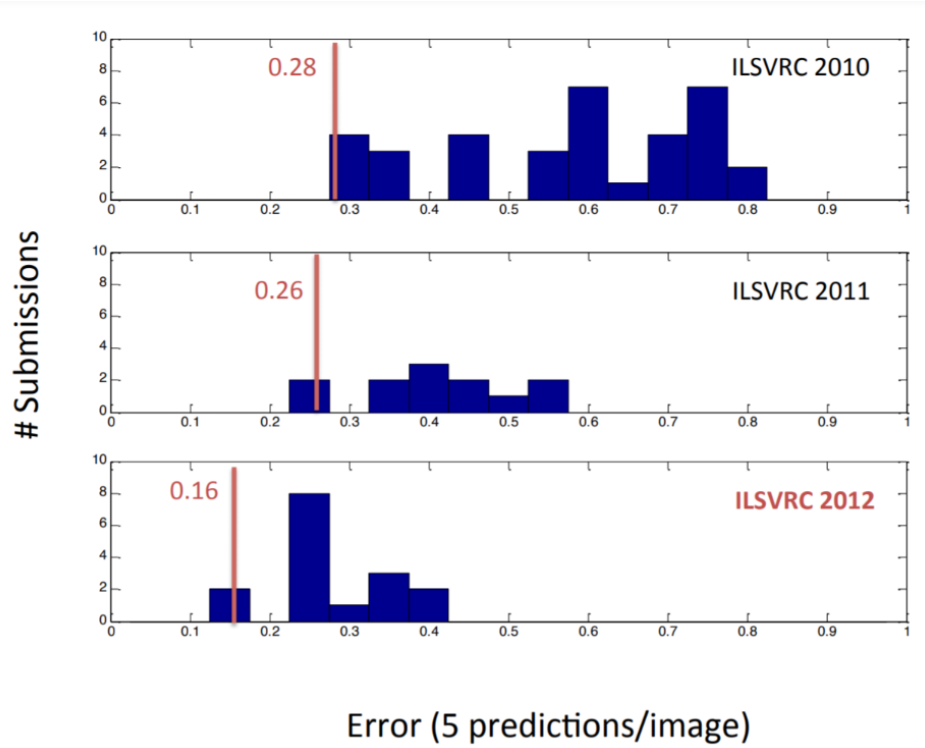
- LeNet (1990s)
- AlexNet (2012)
- ZF Net (2013)
- VGGNet (2014)
- GoogLeNet (2014)
- ResNets (2015)
- DenseNet (2017)

https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

CovNet Architectures

2012

- **ImageNet Database and competition**
- **http://image-net.org/challenges/LSVRC/2012/**
- **http://www.image-net.org/**
- **http://image-net.org/challenges/LSVRC/2012/ilsvrc2012.pdf**



https://en.wikipedia.org/wiki/ImageNet

AlexNet - 2012

- similar architecture as LeNet but deeper (5 convolutional, 3 fully-connected)
- 1000 different classes (e.g. cats, dogs etc.)

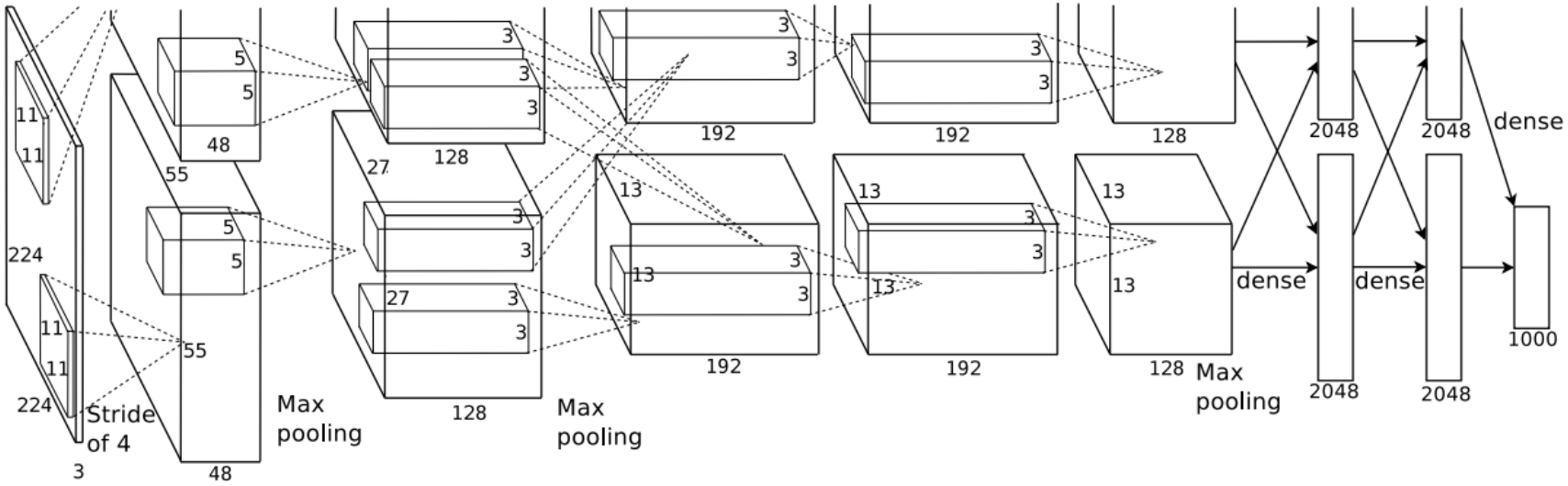
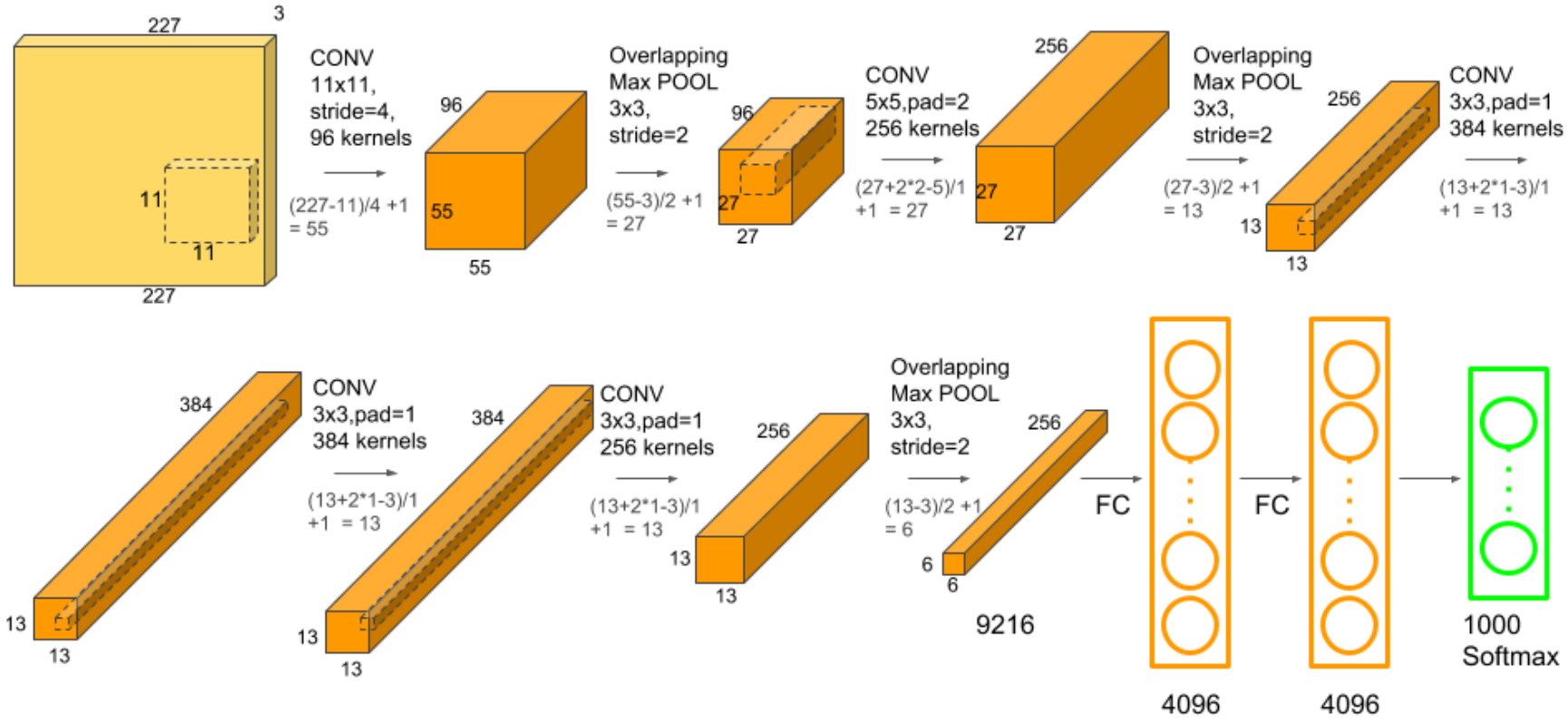


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

AlexNet - 2012

- 5-6 days to train on two GTX 580 GPUs (90 epochs)
- Image augmentation (flip, color changes, ..)
- Dropout

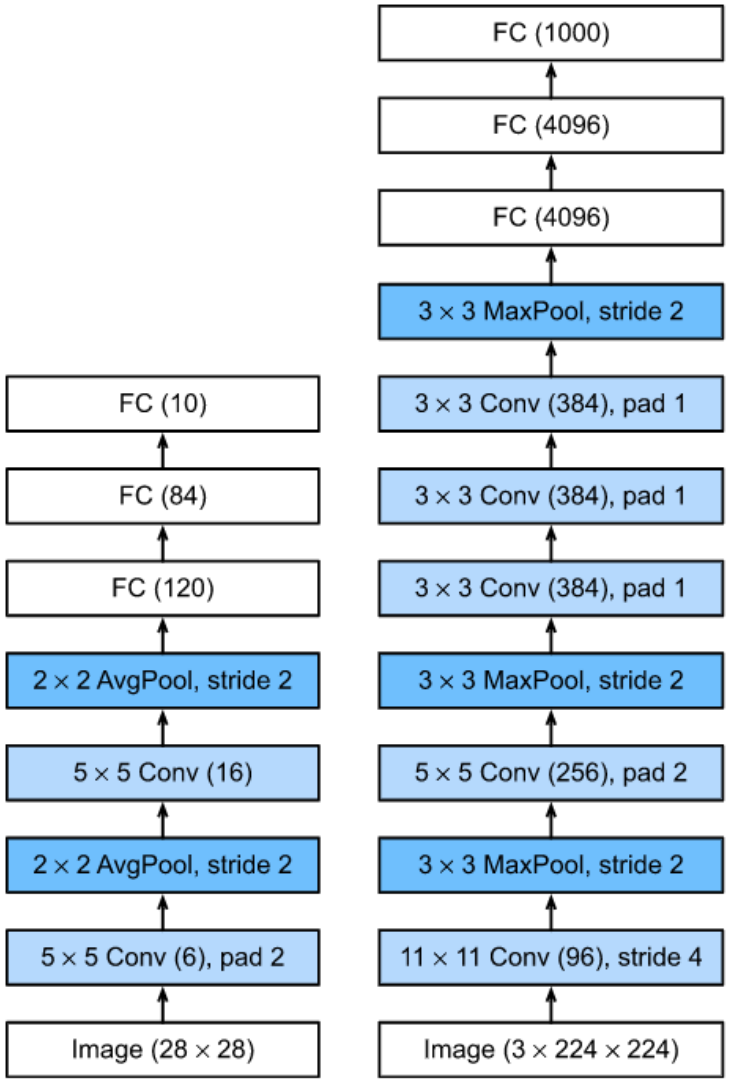


https://www.learnopencv.com/understanding-alexnet/
https://medium.com/@RaghavPrabhu/cnn-architectures-lenet-alexnet-vgg-googlenet-and-resnet-7c81c017b848

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012



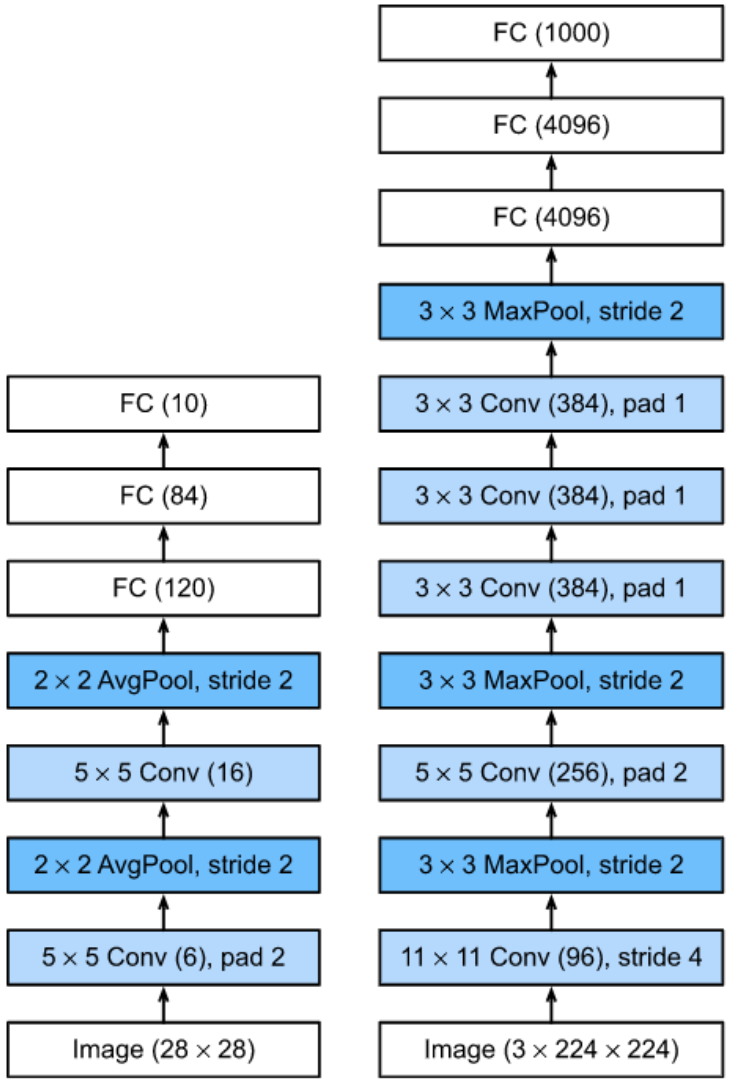
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```



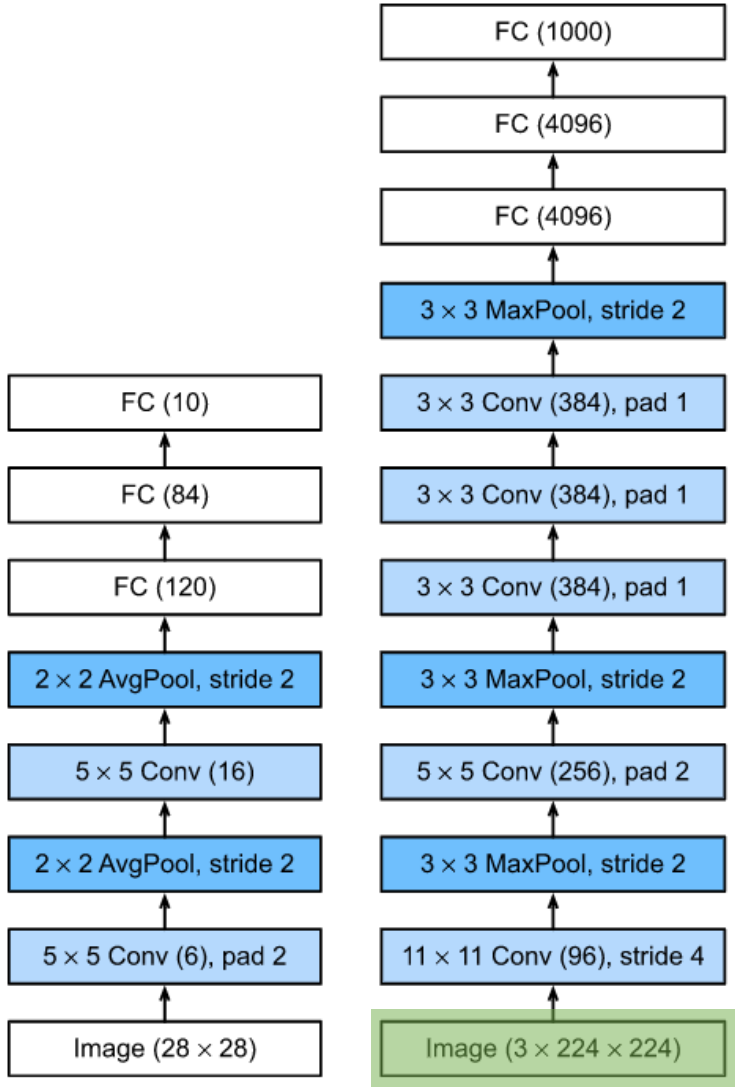
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```




```
input<matrix<unsigned char>>
```

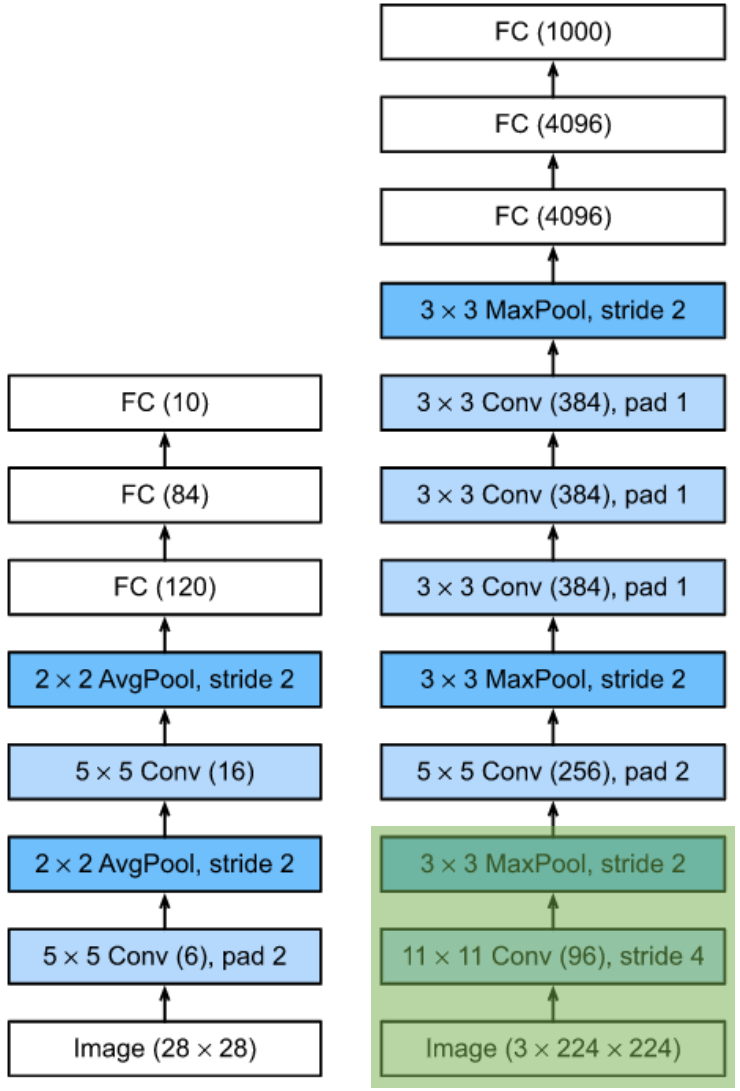
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```




```
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

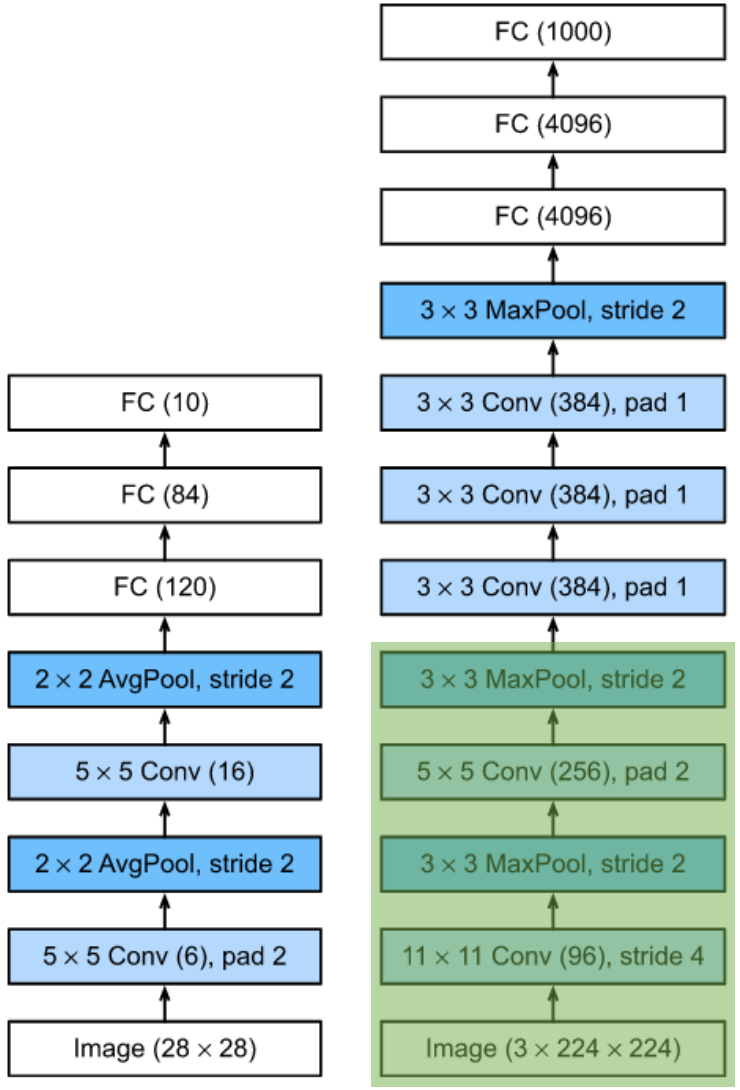
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```




```
max_pool<3,3,2,2,relu<con<256,5,5,1,1,
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

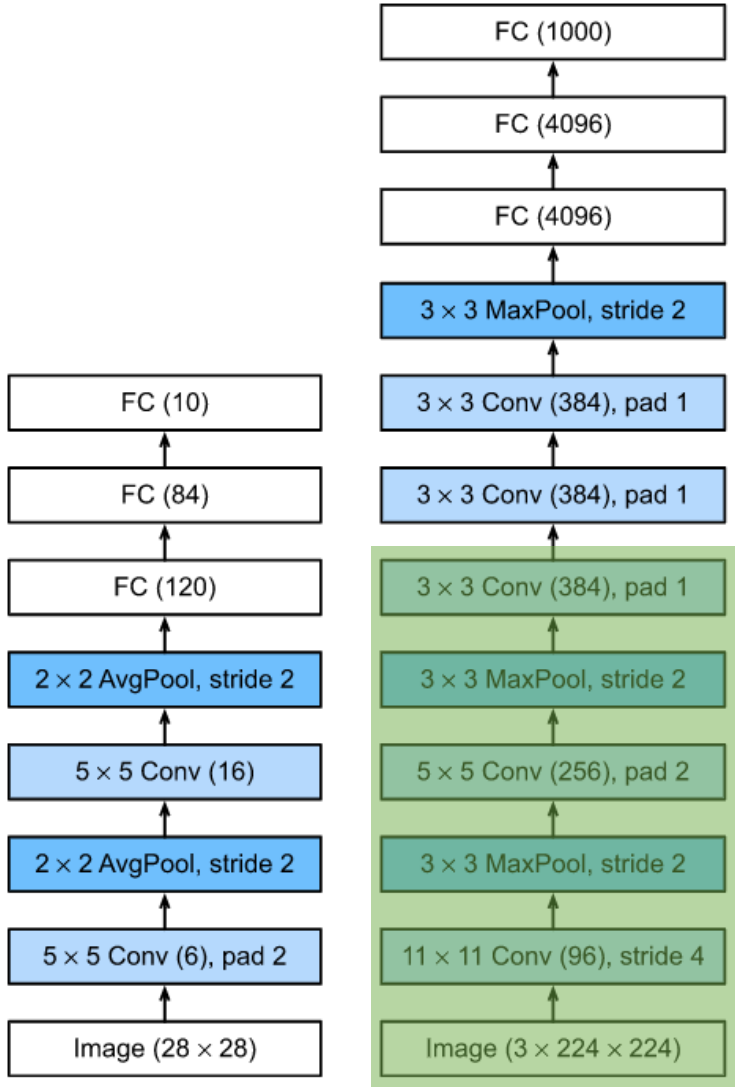
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```




```
relu<con<384,3,3,1,1,
max_pool<3,3,2,2,relu<con<256,5,5,1,1,
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

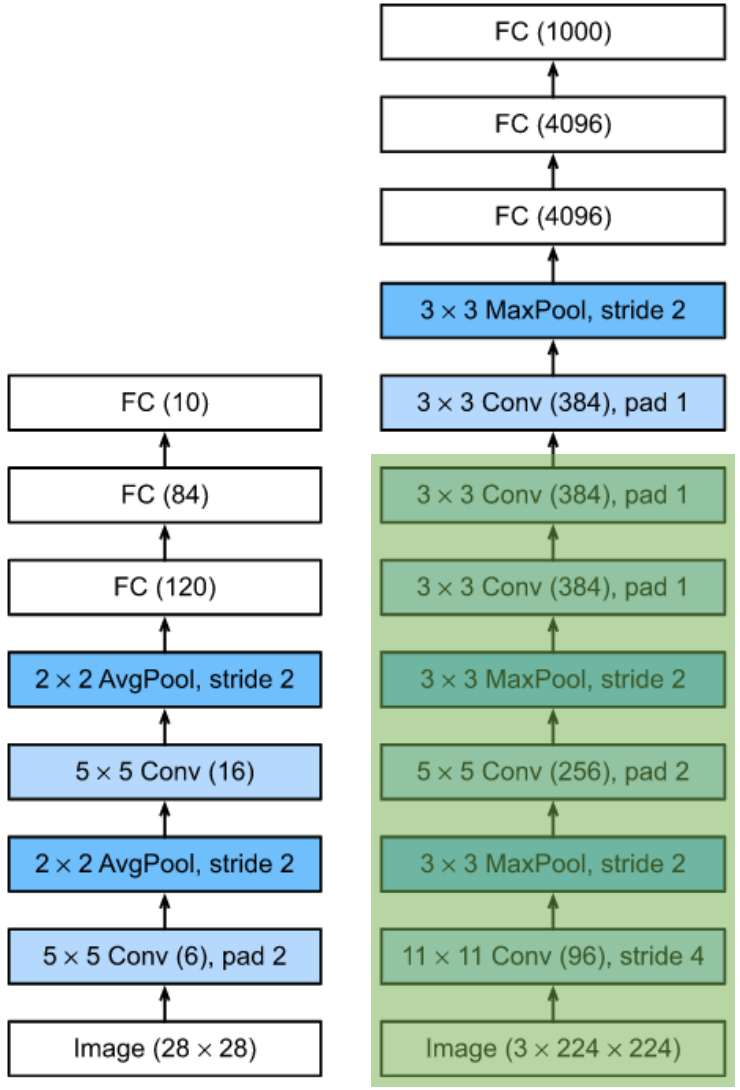
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```




```
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
max_pool<3,3,2,2,relu<con<256,5,5,1,1,
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

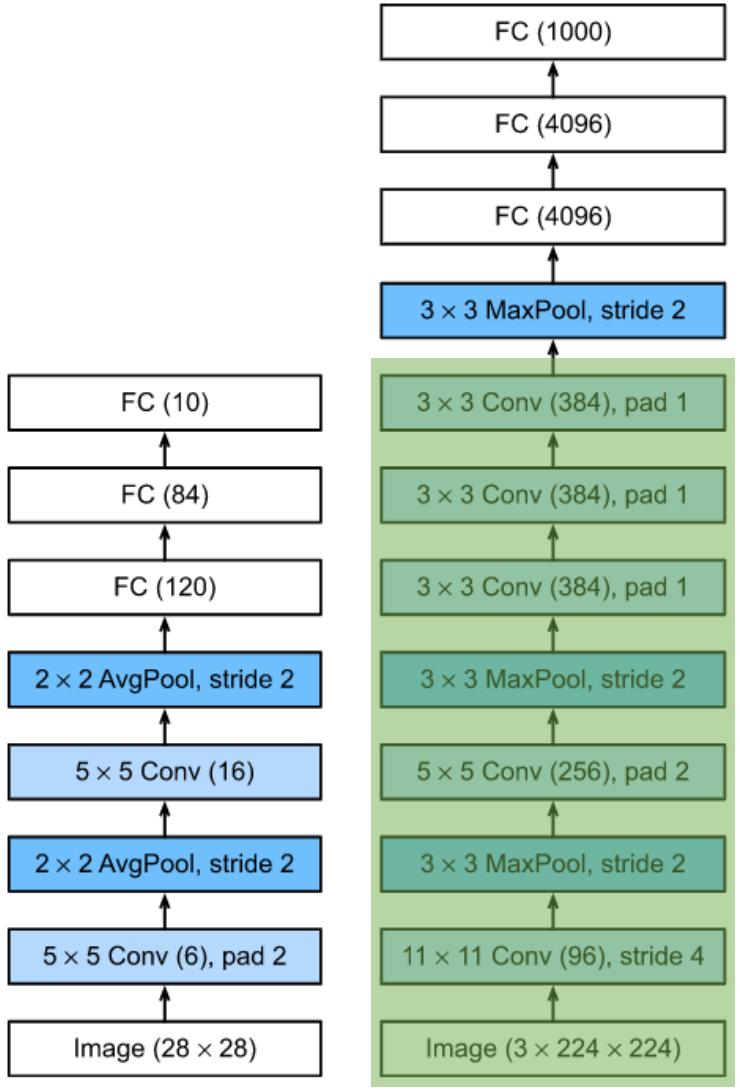
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```




```
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
max_pool<3,3,2,2,relu<con<256,5,5,1,1,
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

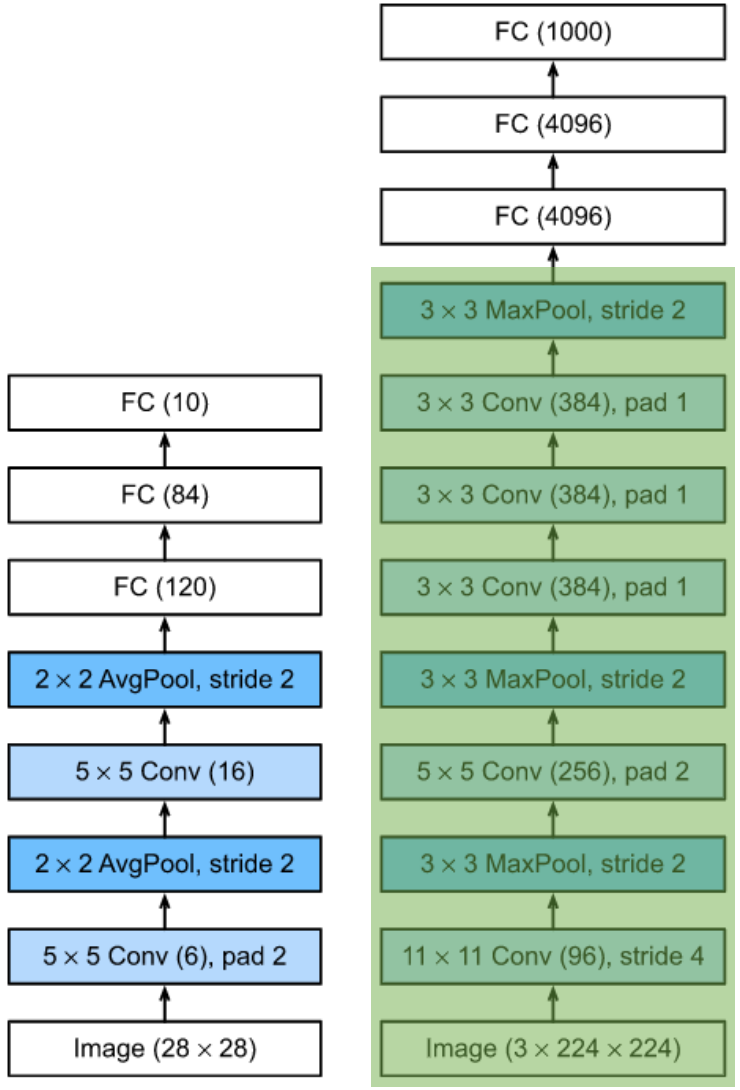
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```




```
max_pool<3,3,2,2,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
max_pool<3,3,2,2,relu<con<256,5,5,1,1,
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

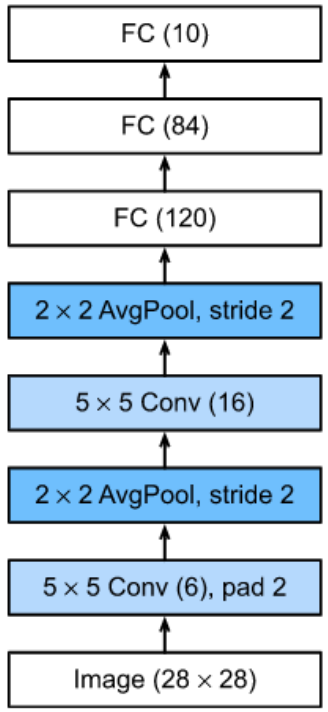
http://d2l.ai/chapter_convolutional-modern/alexnet.html

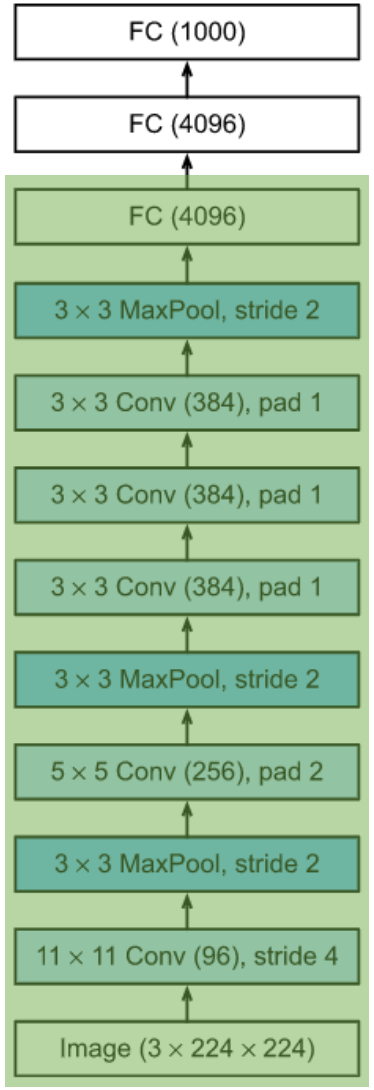
Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```






```
dropout<relu<fc<4096,
max_pool<3,3,2,2,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
max_pool<3,3,2,2,relu<con<256,5,5,1,1,
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

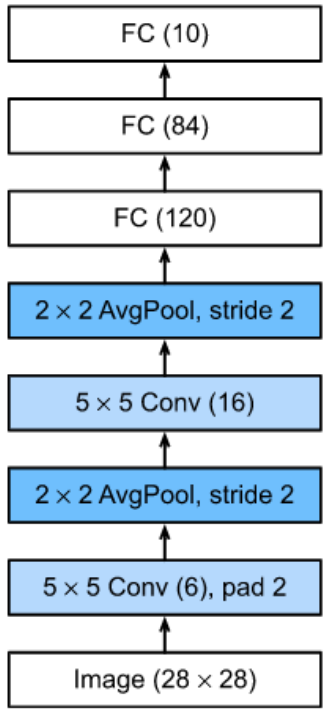
http://d2l.ai/chapter_convolutional-modern/alexnet.html

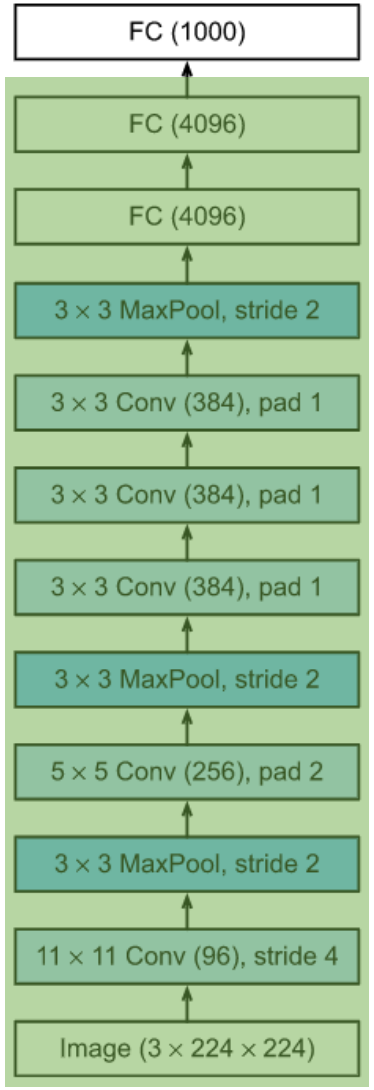
Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```






```
dropout<relu<fc<4096,
dropout<relu<fc<4096,
max_pool<3,3,2,2,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
max_pool<3,3,2,2,relu<con<256,5,5,1,1,
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

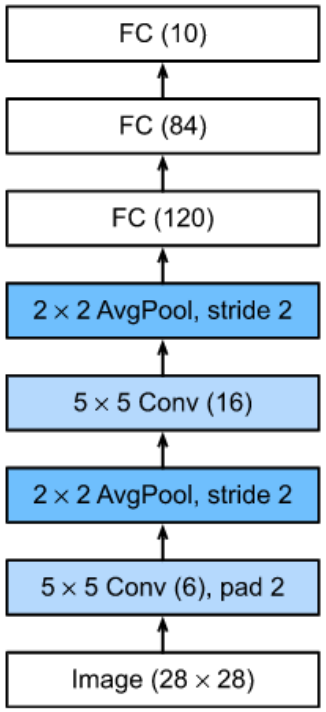
http://d2l.ai/chapter_convolutional-modern/alexnet.html

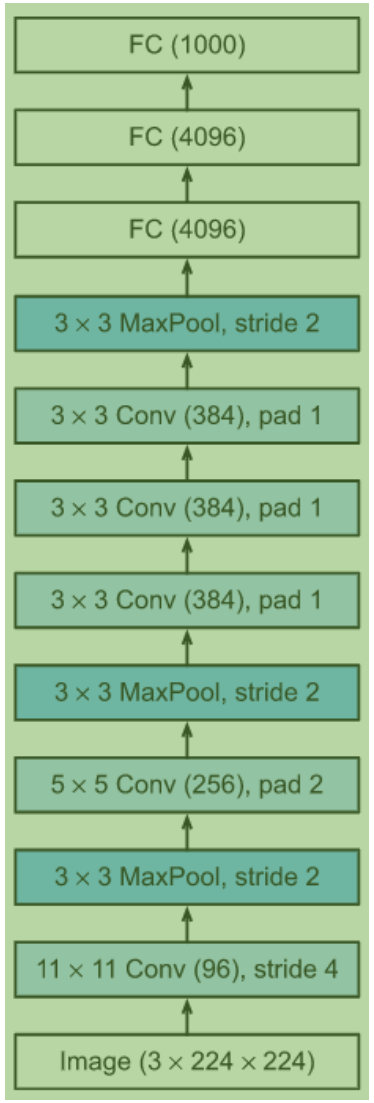
Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```






```
fc<1000,
dropout<relu<fc<4096,
dropout<relu<fc<4096,
max_pool<3,3,2,2,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
max_pool<3,3,2,2,relu<con<256,5,5,1,1,
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

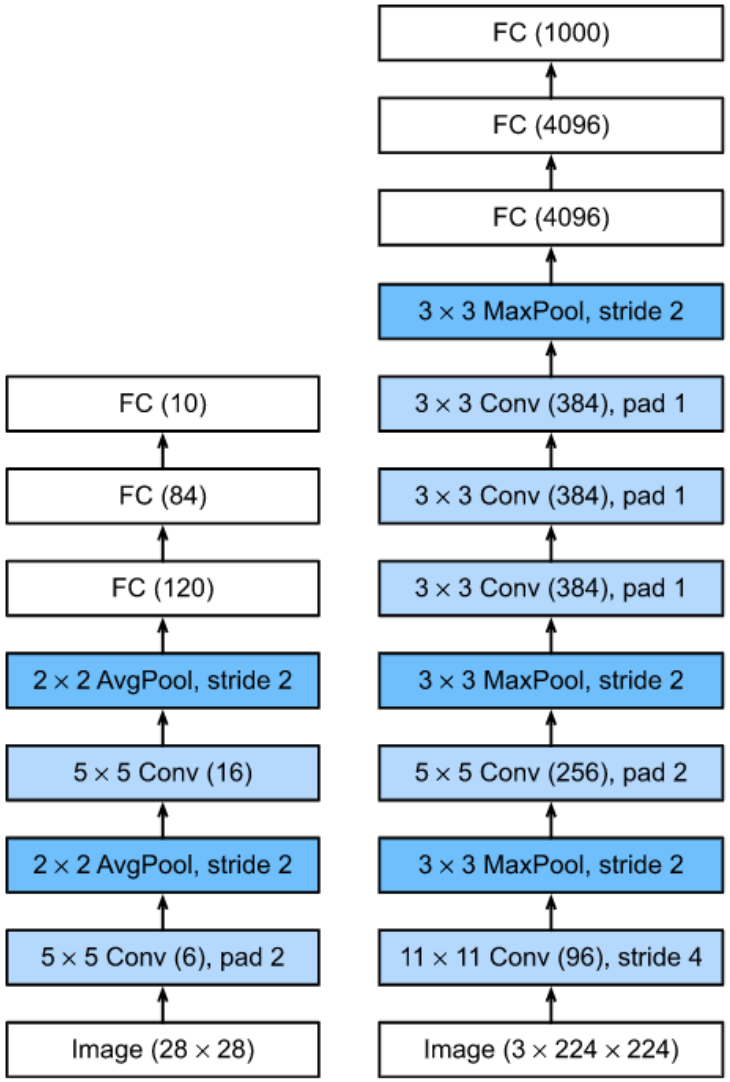
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (Dlib/C++)

```
fc<10,
relu<fc<84,
relu<fc<120,
max_pool<2,2,2,2,relu<con<16,5,5,1,1,
max_pool<2,2,2,2,relu<con<6,5,5,1,1,
input<matrix<unsigned char>>
```



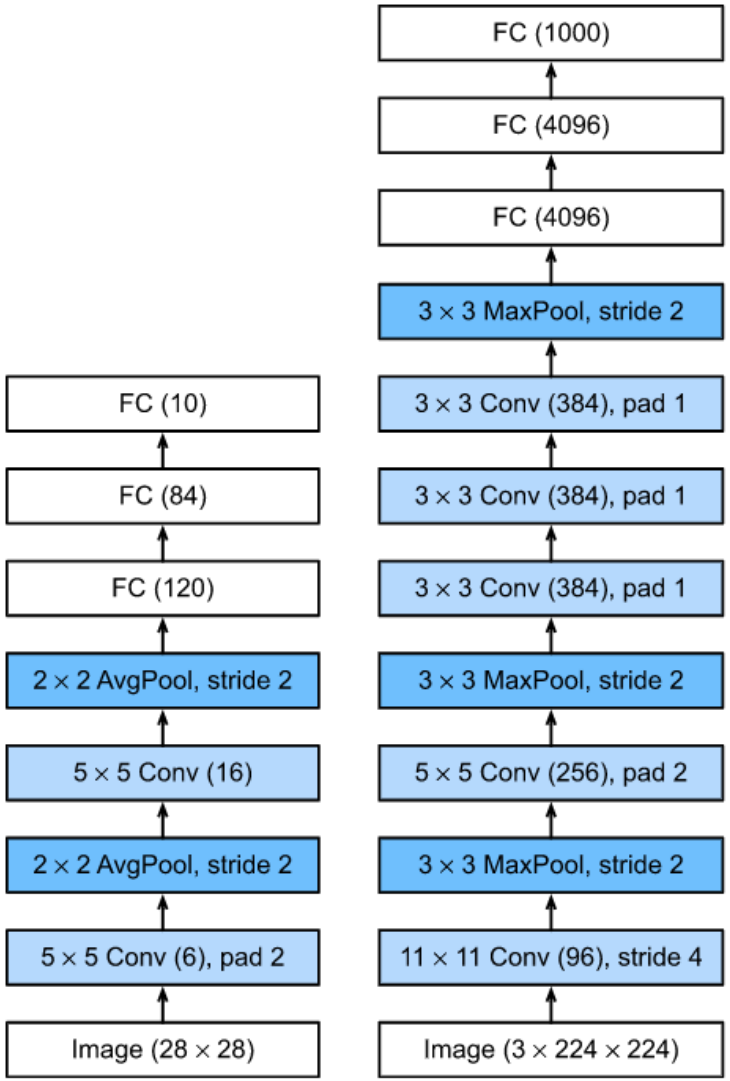

```
fc<1000,
dropout<relu<fc<4096,
dropout<relu<fc<4096,
max_pool<3,3,2,2,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
relu<con<384,3,3,1,1,
max_pool<3,3,2,2,relu<con<256,5,5,1,1,
max_pool<3,3,2,2,relu<con<96,11,11,4,4,
input<matrix<unsigned char>>
```

http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (PyTorch)



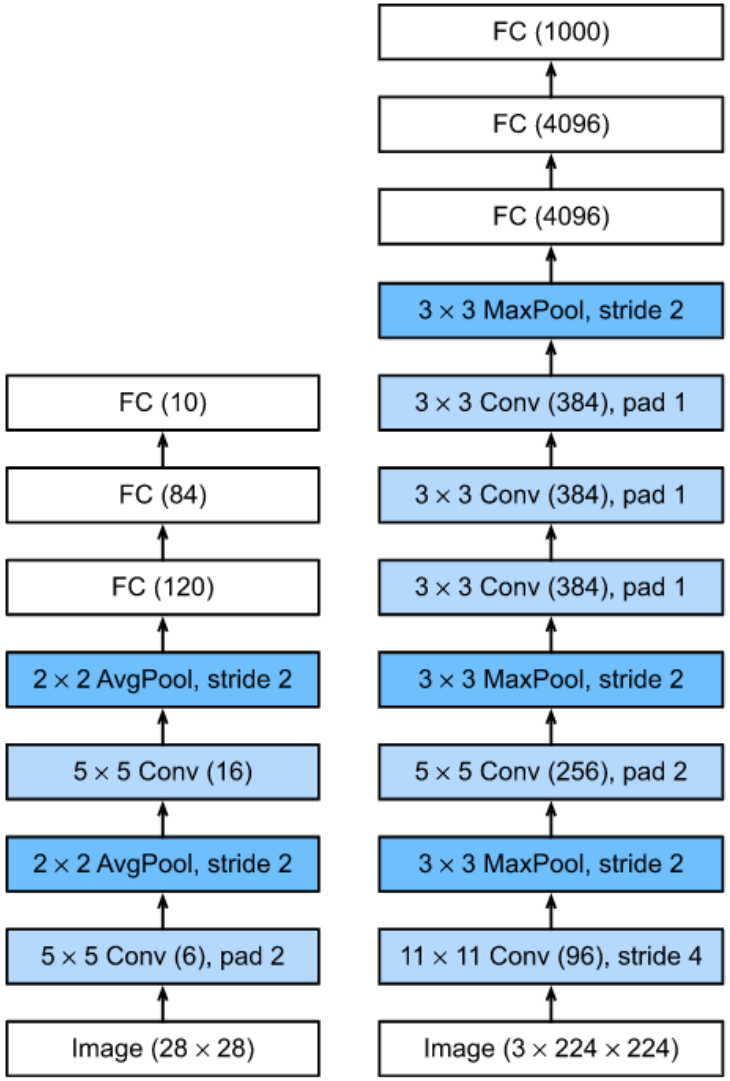
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (PyTorch)

```
net = nn.Sequential(
    nn.Conv2d(3, 6, kernel_size=5, padding=0, stride=1),
    nn.ReLU(),
    nn.AvgPool2d(kernel_size=2, stride=2),
    nn.Conv2d(6, 16, kernel_size=5, padding=0, stride=1),
    nn.ReLU(),
    nn.AvgPool2d(kernel_size=2, stride=2),
    nn.Flatten(),
    nn.LazyLinear(120),
    nn.ReLU(),
    nn.LazyLinear(84),
    nn.ReLU(),
    nn.LazyLinear(2))
```

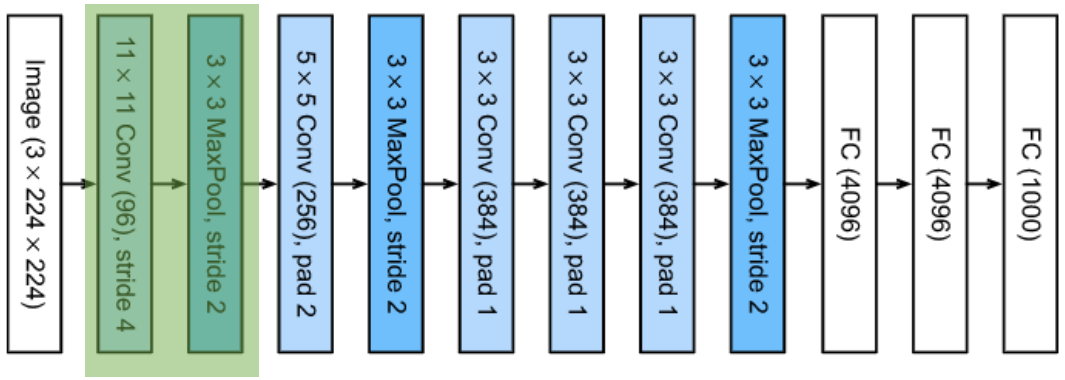


http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (PyTorch)



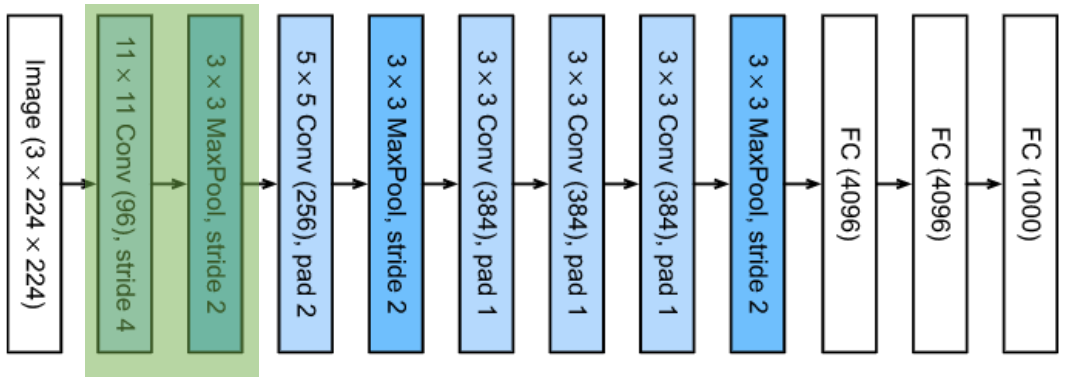

```
net = nn.Sequential(
    
)
```

http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (PyTorch)



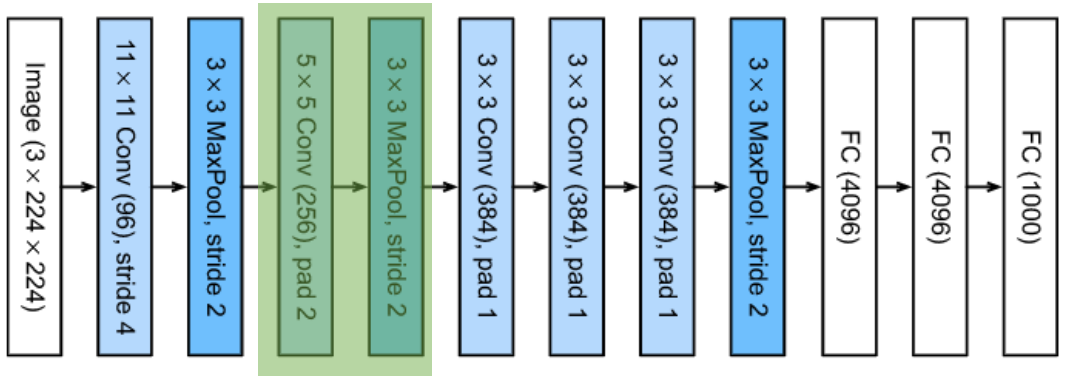

```
net = nn.Sequential(
    nn.Conv2d(3, 96, kernel_size=11, stride=4, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    )
```

http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (PyTorch)



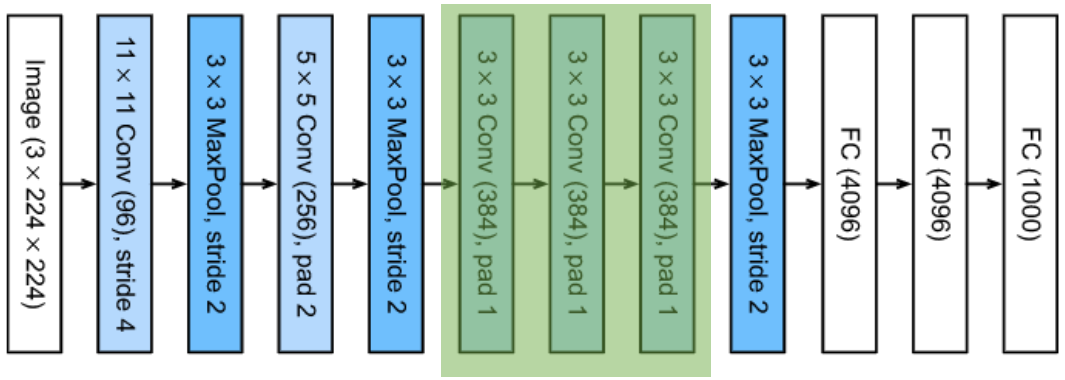

```
net = nn.Sequential(
    nn.Conv2d(3, 96, kernel_size=11, stride=4, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    nn.Conv2d(96, 256, kernel_size=5, padding=2),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    )
```

http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (PyTorch)



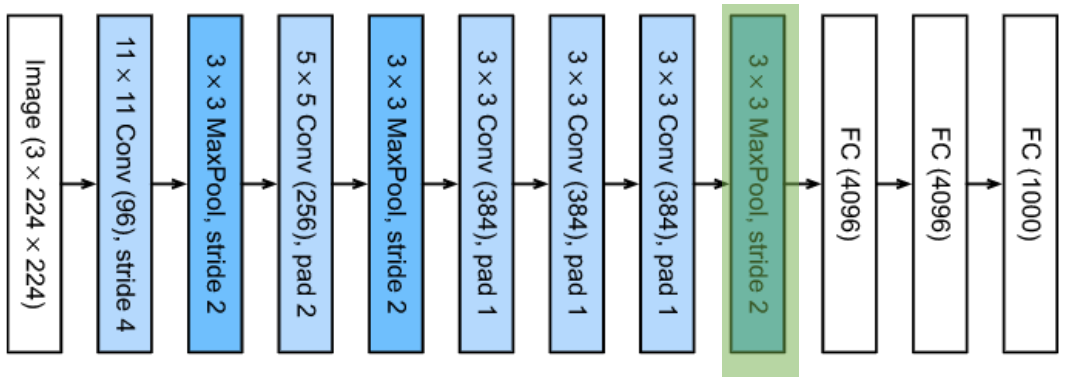

```
net = nn.Sequential(
    nn.Conv2d(3, 96, kernel_size=11, stride=4, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    nn.Conv2d(96, 256, kernel_size=5, padding=2),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    nn.Conv2d(256, 384, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.Conv2d(384, 384, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.Conv2d(384, 256, kernel_size=3, padding=1),
    nn.ReLU(),
    )
```

http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (PyTorch)



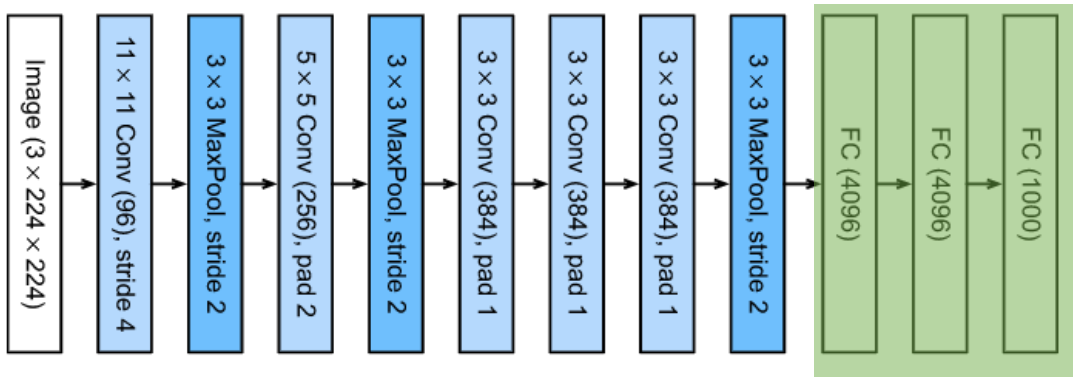

```
net = nn.Sequential(
    nn.Conv2d(3, 96, kernel_size=11, stride=4, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    nn.Conv2d(96, 256, kernel_size=5, padding=2),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    nn.Conv2d(256, 384, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.Conv2d(384, 384, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.Conv2d(384, 256, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    nn.Flatten(),
    
    )
```

http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet – 2012 (PyTorch)




```
net = nn.Sequential(
    nn.Conv2d(3, 96, kernel_size=11, stride=4, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    nn.Conv2d(96, 256, kernel_size=5, padding=2),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    nn.Conv2d(256, 384, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.Conv2d(384, 384, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.Conv2d(384, 256, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=3, stride=2),
    nn.Flatten(),
    nn.Linear(6400, 4096),
    nn.ReLU(),
    nn.Dropout(p=0.5),
    nn.Linear(4096, 4096),
    nn.ReLU(),
    nn.Dropout(p=0.5),
    nn.Linear(4096, 2)
    )
```

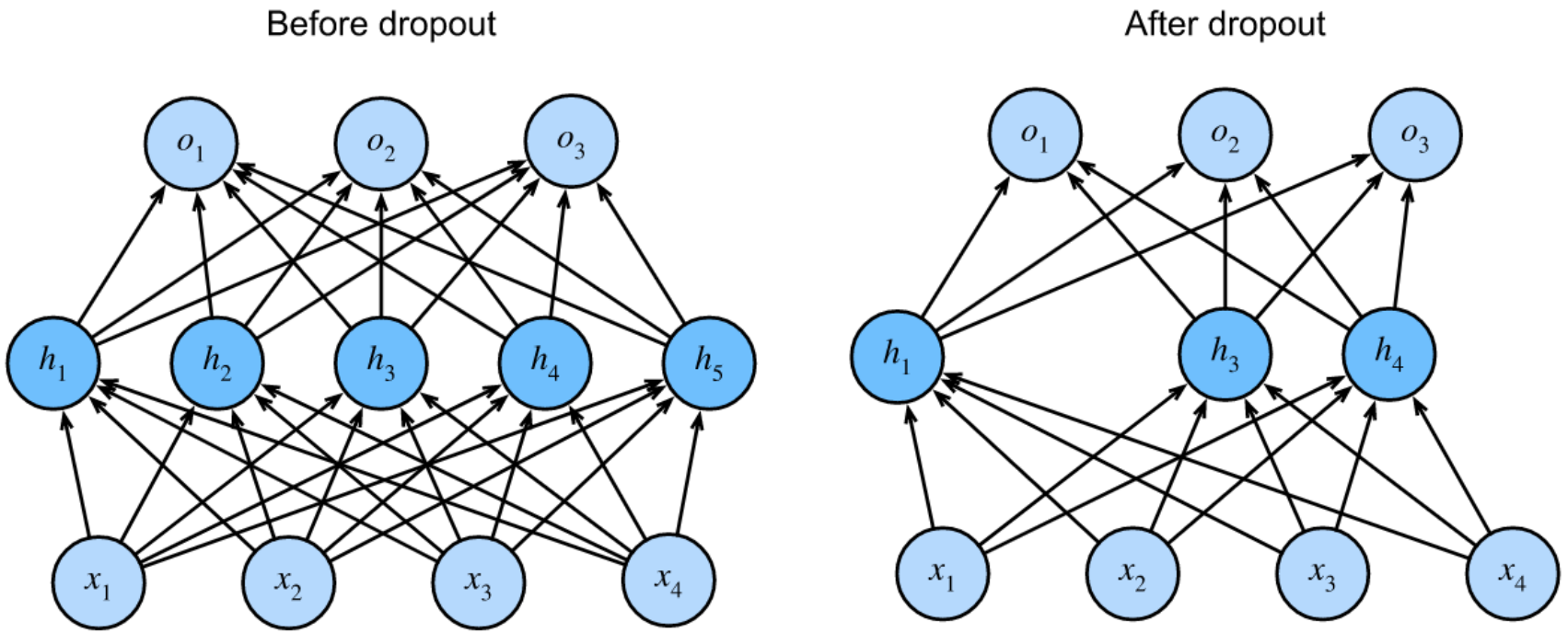
http://d2l.ai/chapter_convolutional-modern/alexnet.html

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

AlexNet - 2012

DROPOUT

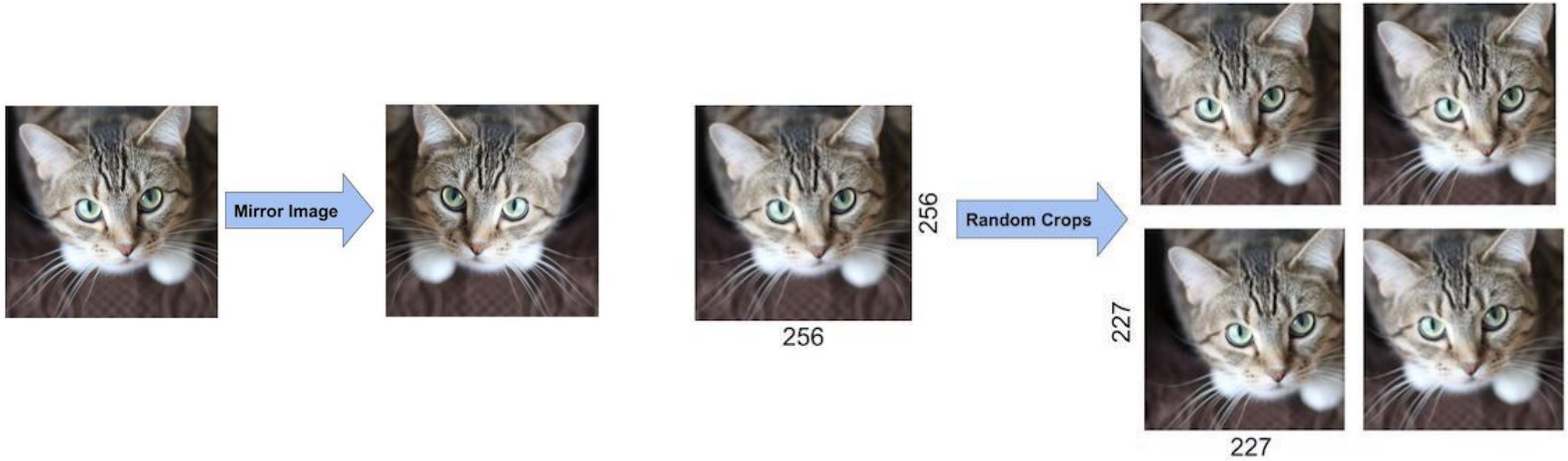
- in some situations, the model is larger than we need – the model can be modified manually or using dropout technique
- ***drop out*** some neurons during training to avoid overfitting
- randomly **dropping out neurons**
  - neuron is dropped from the network with a probability of 0.5



Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov: Improving neural networks by preventing co-adaptation of feature detectors. CoRR abs/1207.0580 (2012)
https://www.learnopencv.com/understanding-alexnet/
http://d2l.ai/chapter_multilayer-perceptrons/dropout.html#sec-dropout
Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

AlexNet - 2012

Data Augmentation



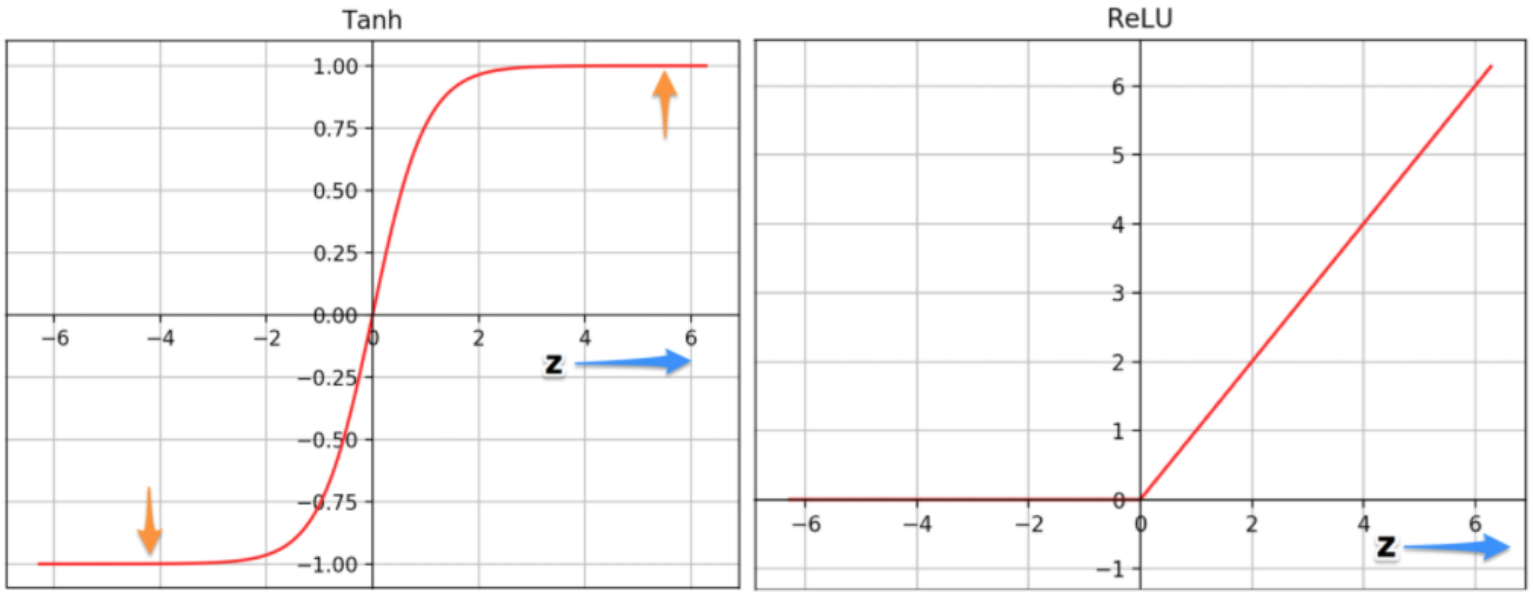
https://www.learnopencv.com/understanding-alexnet/
http://d2l.ai/chapter_multilayer-perceptrons/dropout.html#sec-dropout

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

AlexNet - 2012

ReLU (Rectified Linear Unit)



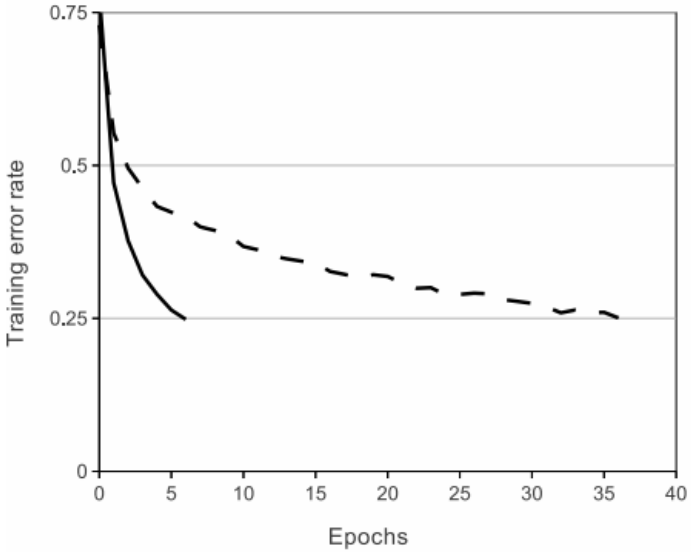


Figure 1: A four-layer convolutional neural network with ReLUs **(solid line)** reaches a 25% training error rate on CIFAR-10 six times faster than an equivalent network with tanh neurons **(dashed line)**. The learning rates for each network were chosen independently to make training as fast as possible. No regularization of any kind was employed. The magnitude of the effect demonstrated here varies with network architecture, but networks with ReLUs consistently learn several times faster than equivalents with saturating neurons.

https://www.learnopencv.com/understanding-alexnet/
http://d2l.ai/chapter_multilayer-perceptrons/dropout.html#sec-dropout

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

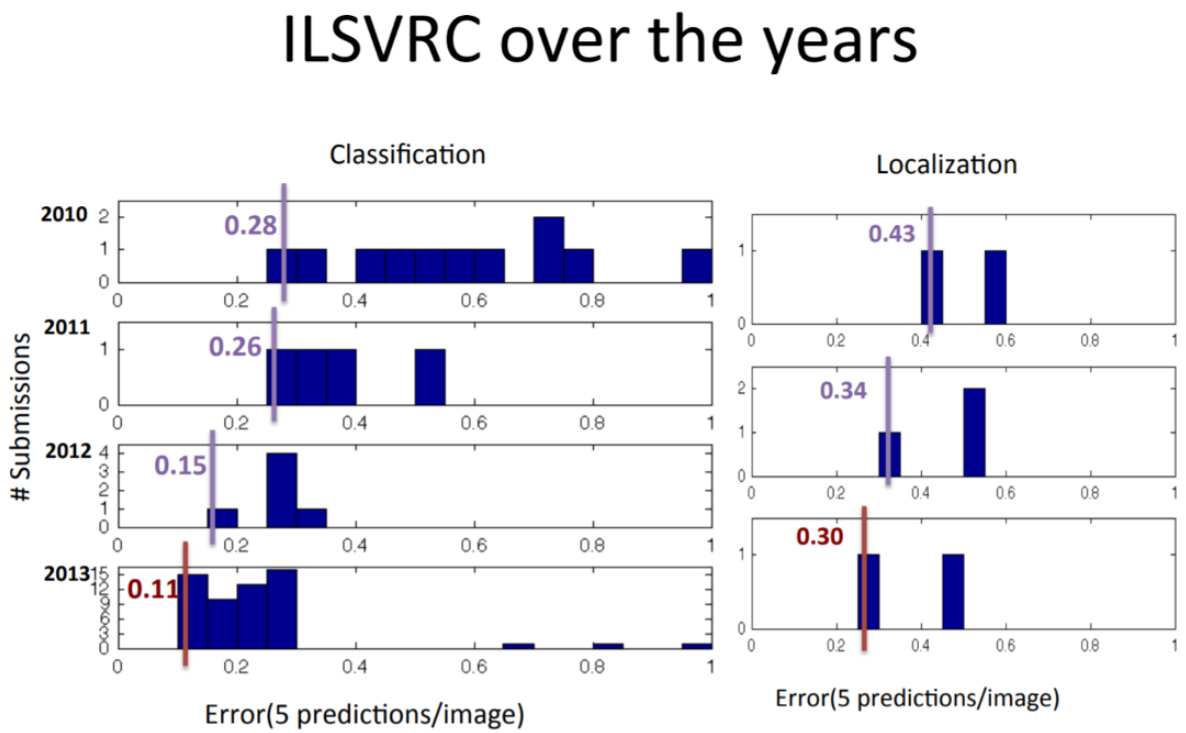
CovNet Architectures

- LeNet (1990s)
- AlexNet (2012)
- ZF Net (2013)
- VGGNet (2014)
- GoogLeNet (2014)
- ResNets (2015)
- DenseNet (2017)

ZFNet

2013

- http://image-net.org/challenges/LSVRC/2013/
- http://www.image-net.org/challenges/LSVRC/2013/slides/ILSVRC2013_12_7_13_clsloc.pdf



https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

Matthew D. Zeiler, Rob Fergus: Visualizing and Understanding Convolutional Networks

ZFNet

- Similar architecture to AlexNet
- 11x11 vs. 7x7 filters in the first layer
- Number of filters is increasing
- Visualization of features - Deconvolutional Network

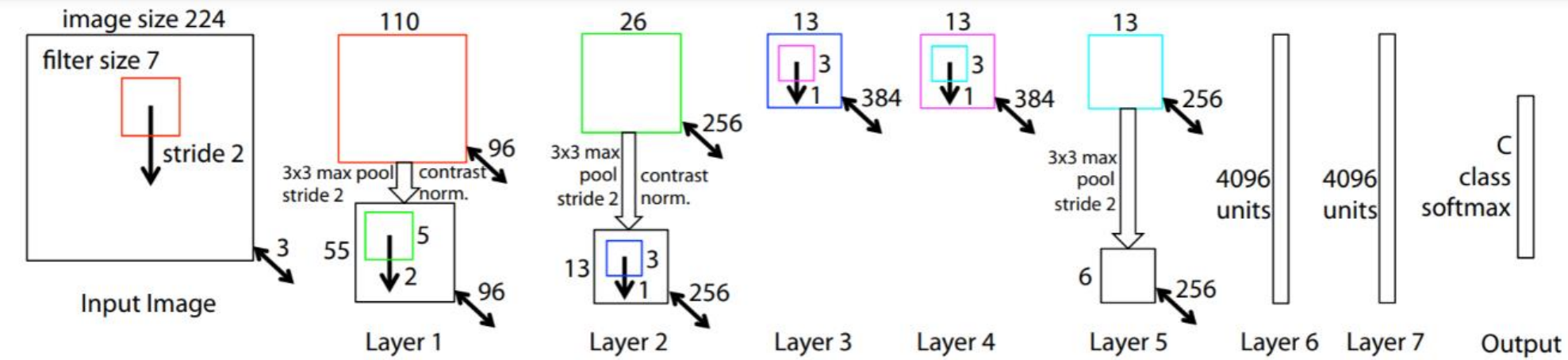


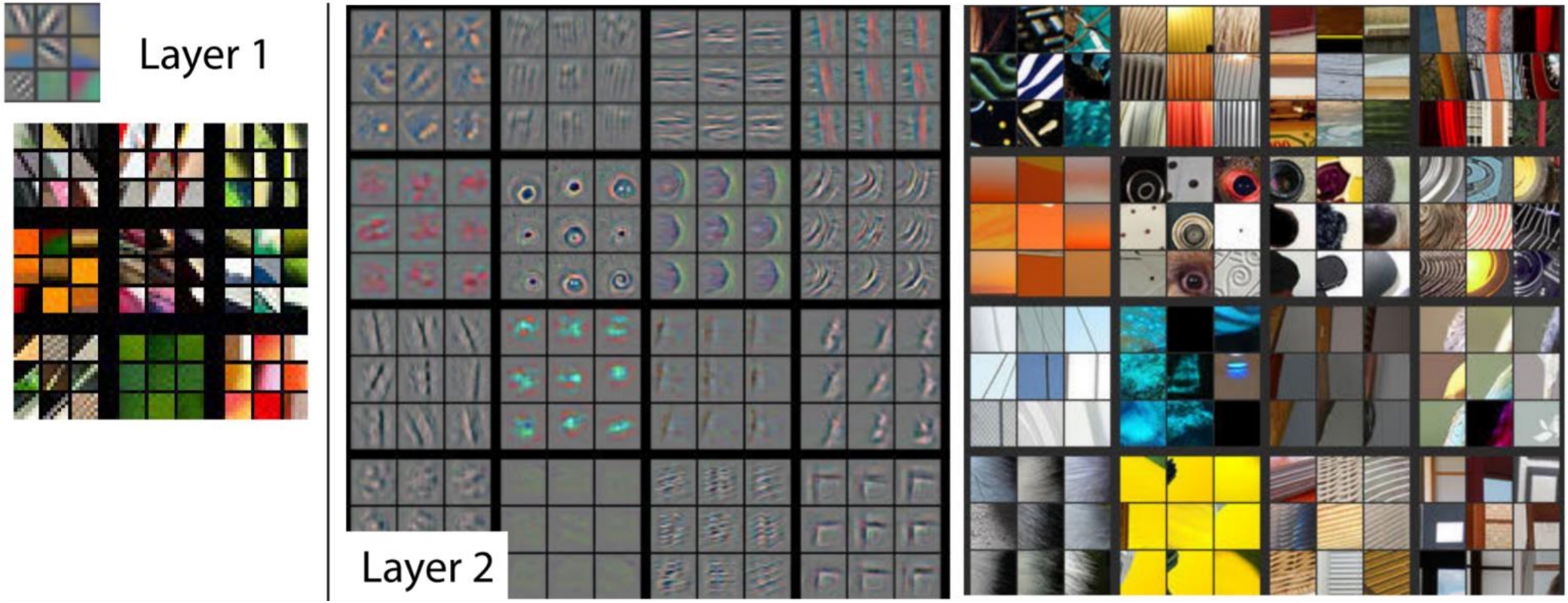
Figure 3. Architecture of our 8 layer convnet model. A 224 by 224 crop of an image (with 3 color planes) is presented as the input. This is convolved with 96 different 1st layer filters (red), each of size 7 by 7, using a stride of 2 in both x and y. The resulting feature maps are then: (i) passed through a rectified linear function (not shown), (ii) pooled (max within 3x3 regions, using stride 2) and (iii) contrast normalized across feature maps to give 96 different 55 by 55 element feature maps. Similar operations are repeated in layers 2,3,4,5. The last two layers are fully connected, taking features from the top convolutional layer as input in vector form ($6 \cdot 6 \cdot 256 = 9216$ dimensions). The final layer is a C-way softmax function, C being the number of classes. All filters and feature maps are square in shape.

https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

Matthew D. Zeiler, Rob Fergus: Visualizing and Understanding Convolutional Networks

ZFNet

First Layers - low level feature detector (edge detector, circular features)

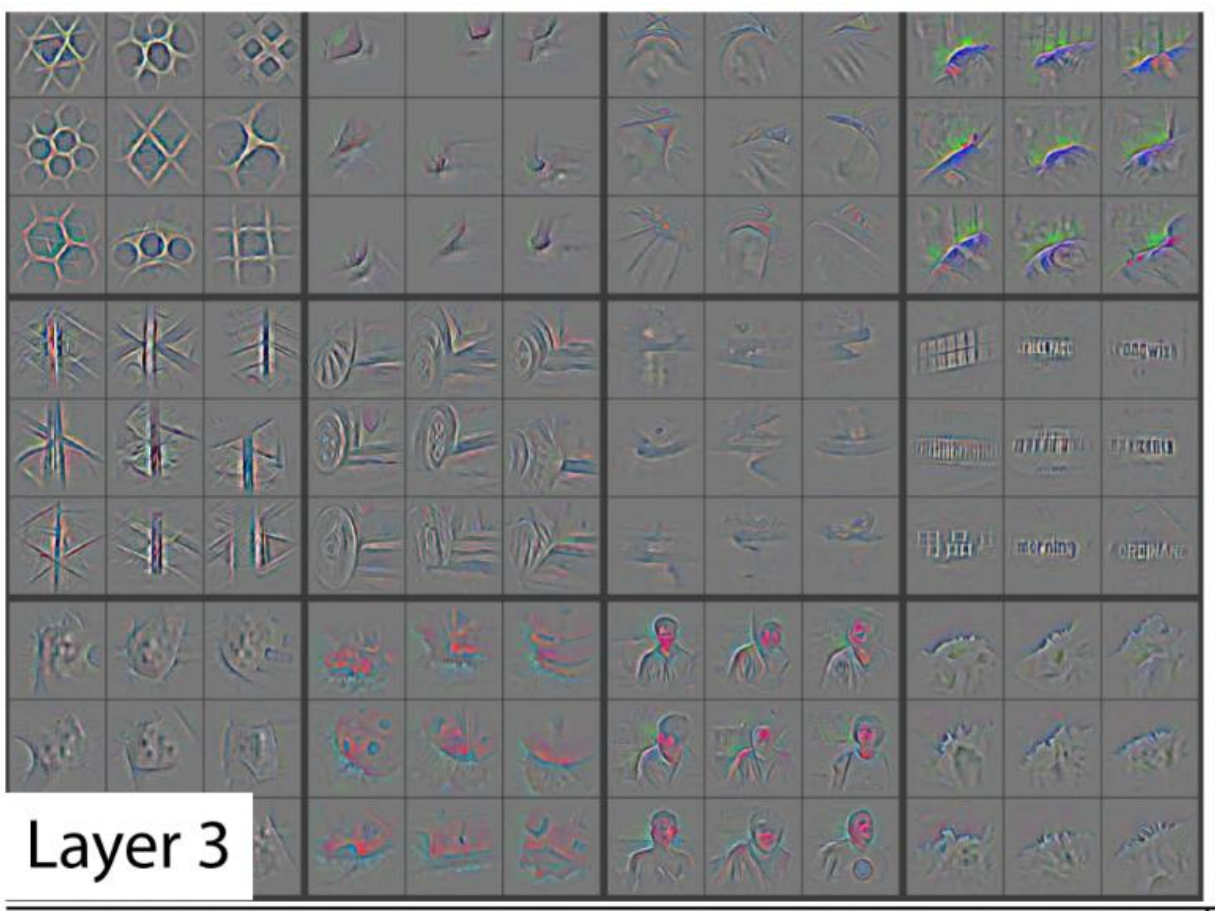


https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

Matthew D. Zeiler, Rob Fergus: Visualizing and Understanding Convolutional Networks

ZFNet

Deep Layers - higher level features such as dogs faces or flowers

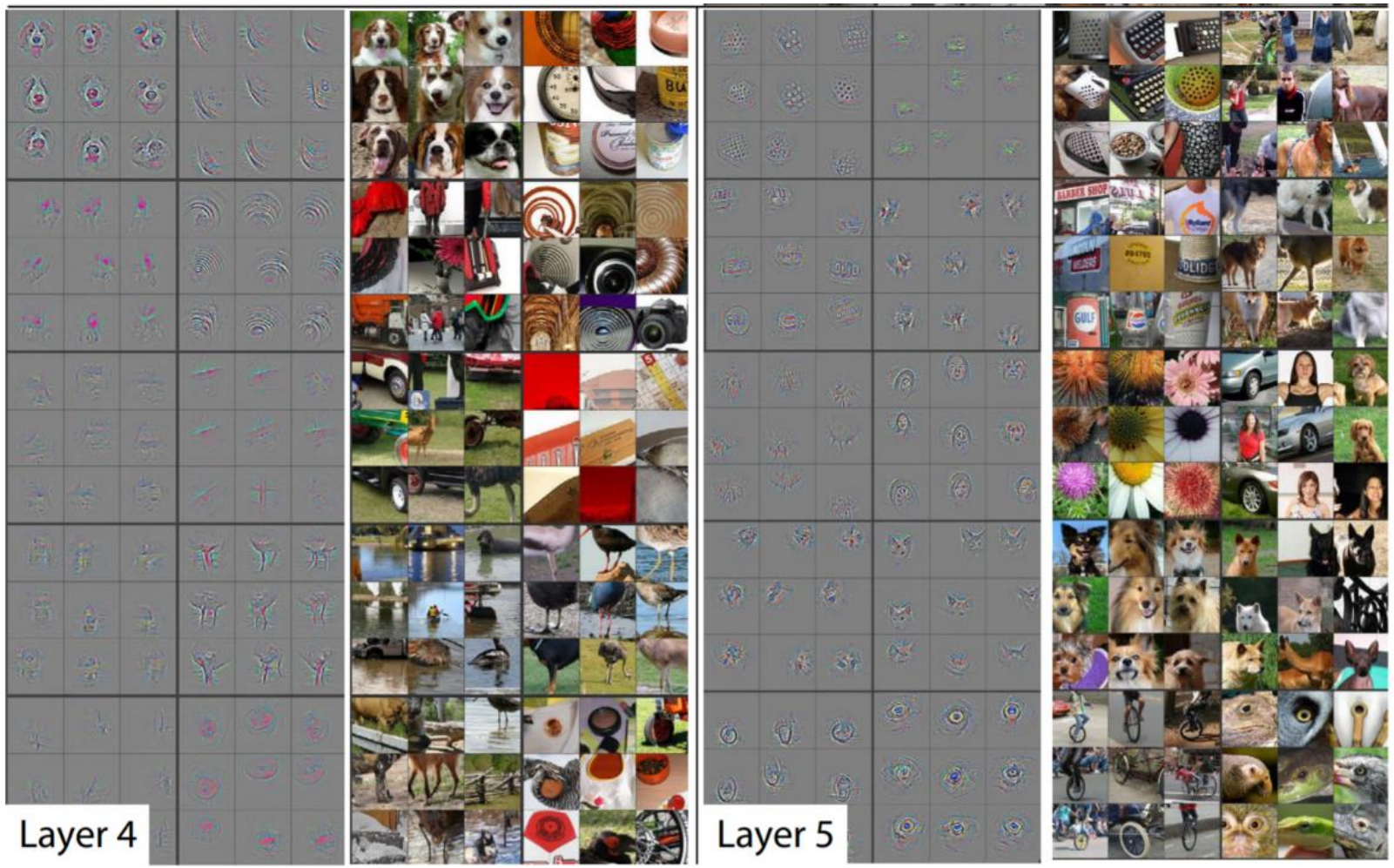


https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

Matthew D. Zeiler, Rob Fergus: Visualizing and Understanding Convolutional Networks

ZFNet

Deep Layers - higher level features such as dogs faces or flowers

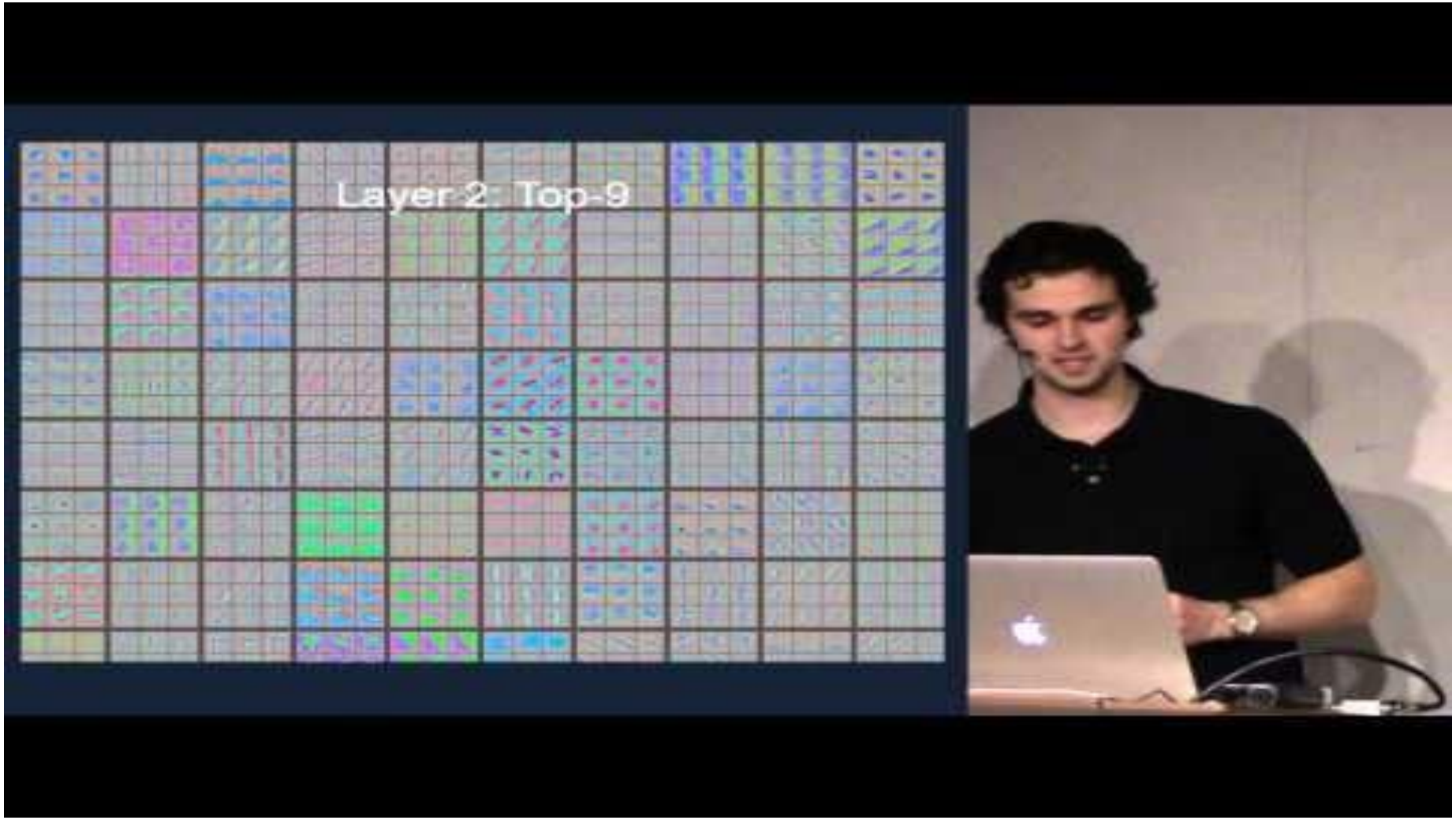


https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

Matthew D. Zeiler, Rob Fergus: Visualizing and Understanding Convolutional Networks

ZFNet

https://www.youtube.com/watch?reload=9&v=ghEmQSxT6tw



https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

Matthew D. Zeiler, Rob Fergus: Visualizing and Understanding Convolutional Networks

CovNet Architectures

- LeNet (1990s)
- AlexNet (2012)
- ZF Net (2013)
- VGGNet (2014)
- GoogLeNet (2014)
- ResNets (2015)
- DenseNet (2017)

VGG - 2014

2014

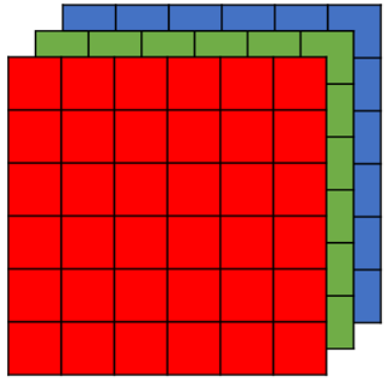
- ImageNet Challenge http://www.image-net.org/challenges/LSVRC/2014/
- http://www.image-net.org/challenges/LSVRC/2014/results
- **Idea:**
  - stack the convolutional layers with increasing filter sizes
  - 3 x 3 filter size stride of 1
  - MaxPooling 2 x 2 stride of 2
  - **VGG Block**
  - Dropout, MaxPooling, ReLu
  - On a system equipped with four NVIDIA Titan Black GPUs, training a single net took 2–3 weeks depending on the architecture

http://d2l.ai/chapter_convolutional-modern/vgg.html https://arxiv.org/pdf/1409.1556v6.pdf

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

VGG - 2014

CONV = 3 x 3 filter, s=1, same MAX-POOL = 2 x 2, s=2



$224 \times 224 \times 3$

[CONV 64] × 2 → $224 \times 224 \times 64$ → POOL → $112 \times 112 \times 64$

[CONV 128] × 2 → $112 \times 112 \times 128$ → POOL → $56 \times 56 \times 128$

[CONV 256] × 3 → $56 \times 56 \times 256$ → POOL → $28 \times 28 \times 256$

[CONV 512] × 3 → $28 \times 28 \times 512$ → POOL → $14 \times 14 \times 512$

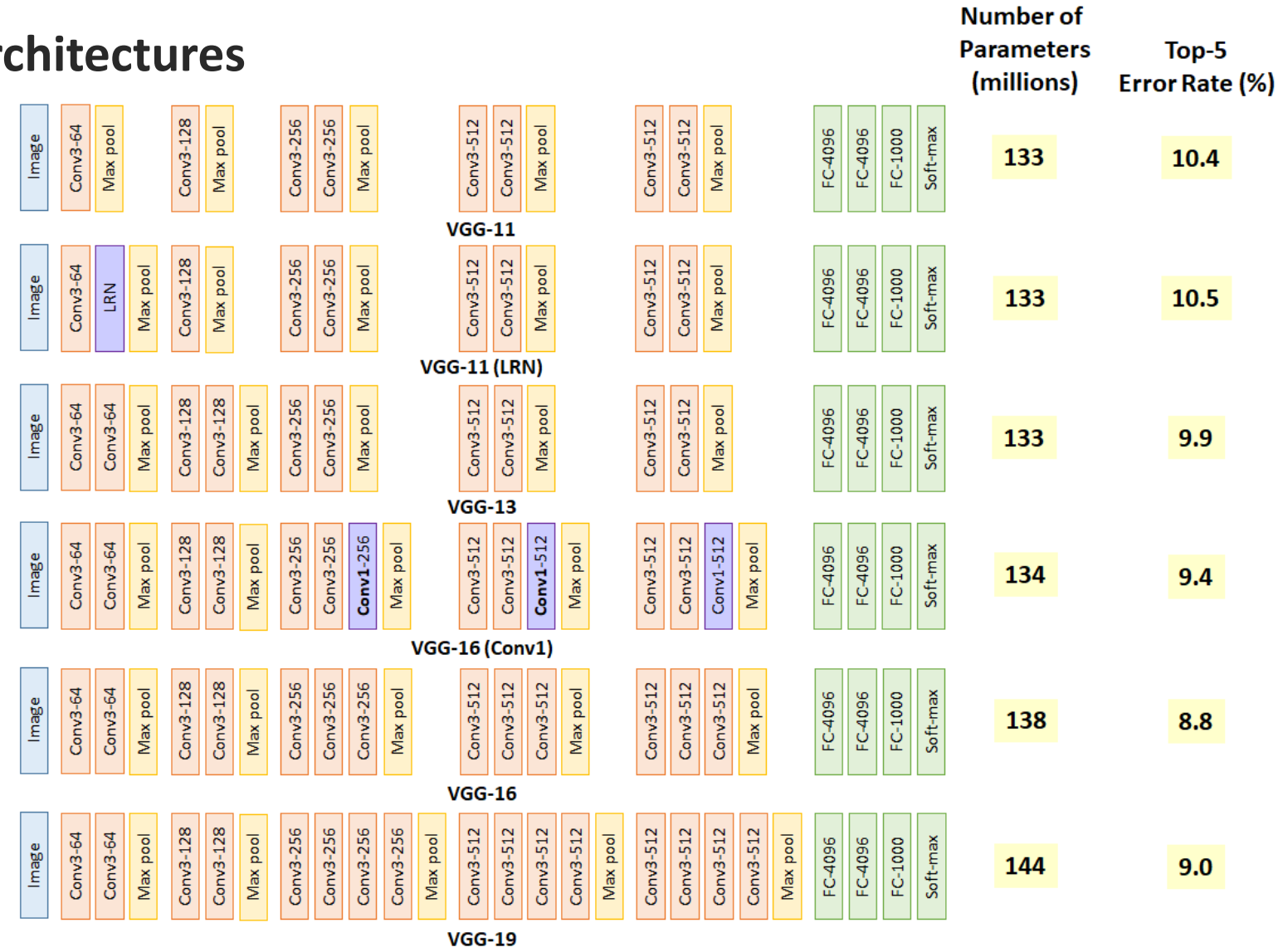
[CONV 512] × 3 → $14 \times 14 \times 512$ → POOL → $7 \times 7 \times 512$ → FC 4096 → FC 4096 → Softmax 1000

http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

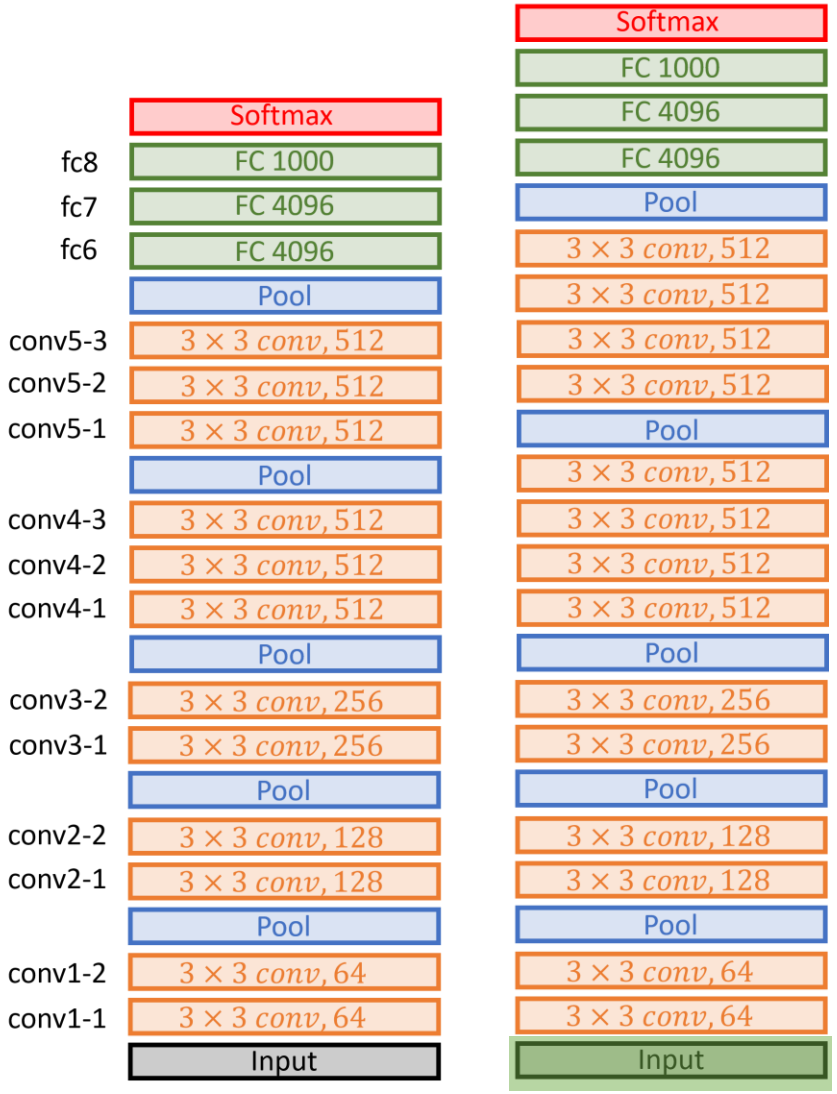
VGG - 2014

6 VGGNet Architectures



https://towardsdatascience.com/cnn-architectures-a-deep-dive-a99441d18049
http://d2l.ai/chapter_convolutional-modern/vgg.html
Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

VGG - 2014




```
input<matrix<unsigned char>>
```

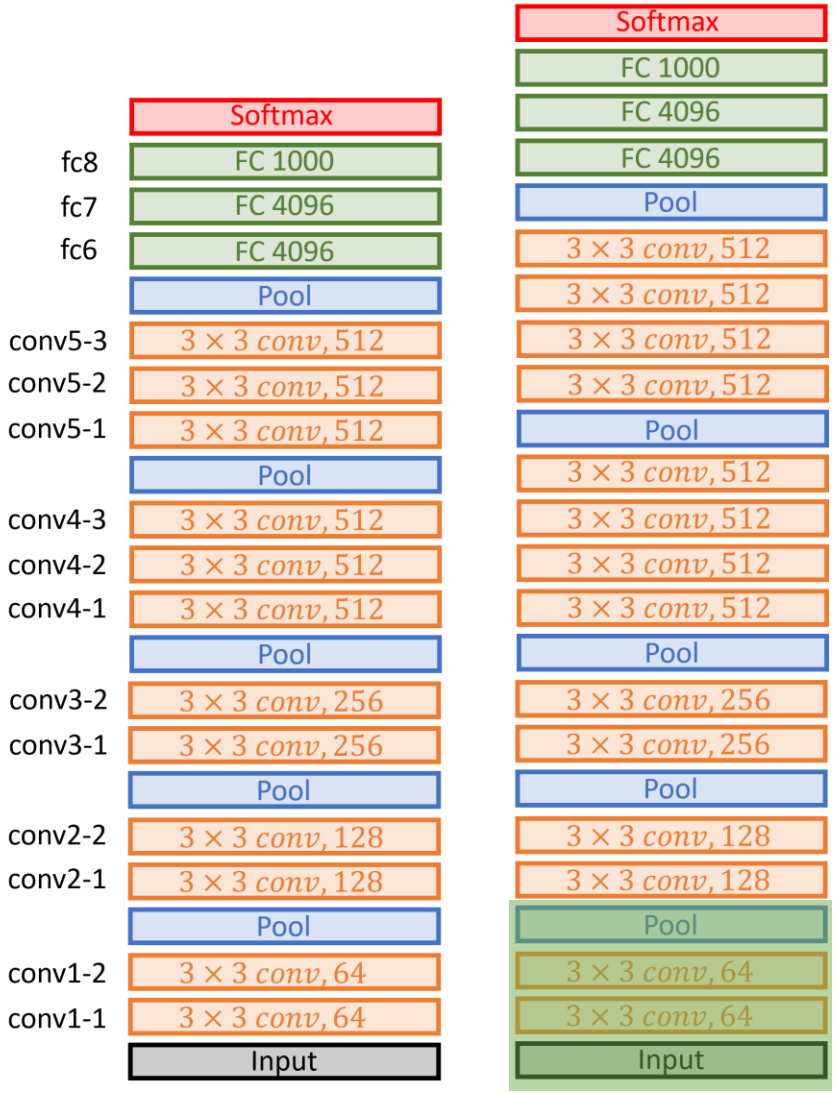
http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

http://d2l.ai/chapter_convolutional-modern/vgg.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

VGG - 2014




```
max_pool<2,2,2,2,
relu<con<64,3,3,1,1,
relu<con<64,3,3,1,1,
input<matrix<unsigned char>>
```

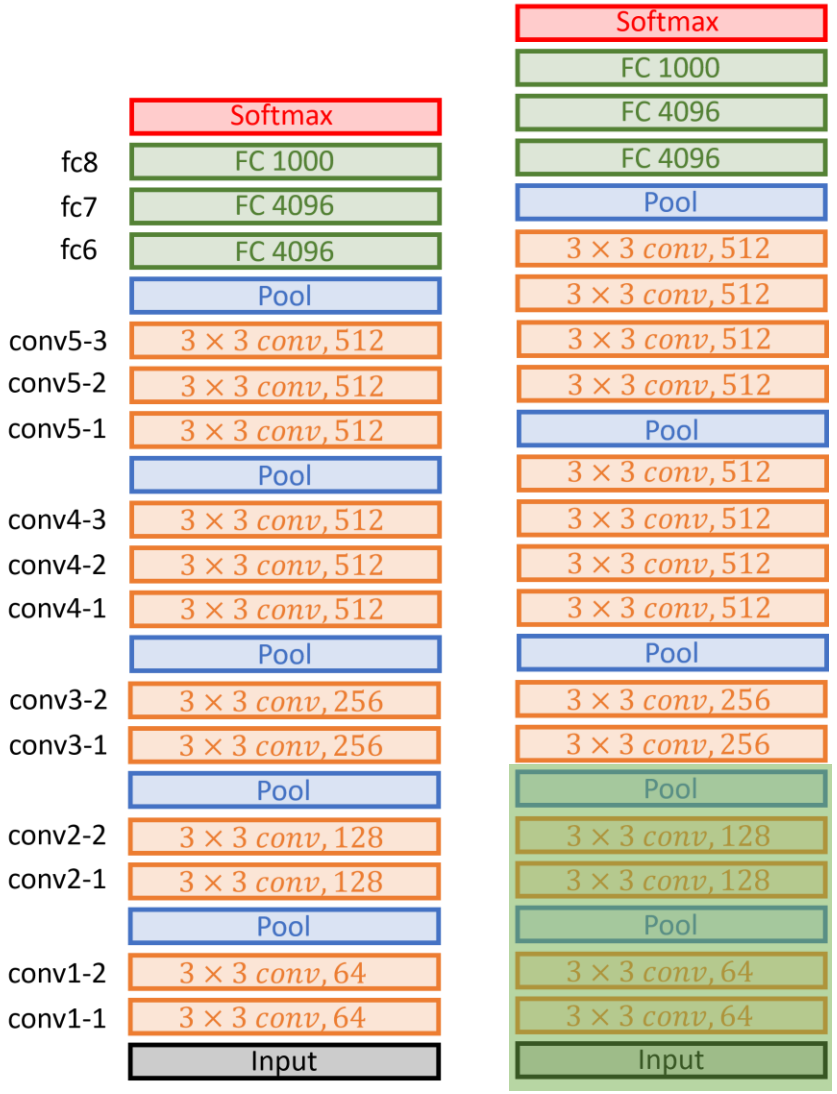
http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

http://d2l.ai/chapter_convolutional-modern/vgg.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

VGG - 2014



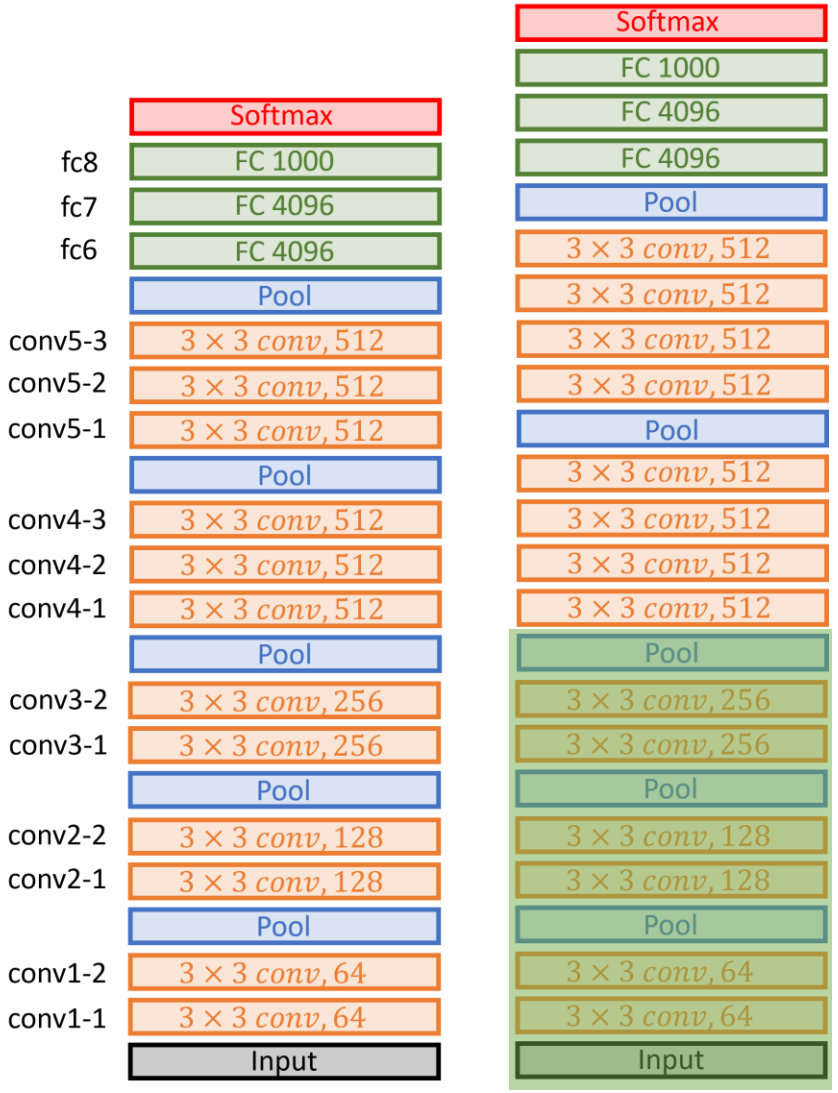

```
max_pool<2,2,2,2,
relu<con<128,3,3,1,1,
relu<con<128,3,3,1,1,
max_pool<2,2,2,2,
relu<con<64,3,3,1,1,
relu<con<64,3,3,1,1,
input<matrix<unsigned char>>
```

http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

http://d2l.ai/chapter_convolutional-modern/vgg.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

VGG - 2014




```
max_pool<2,2,2,2,
relu<con<256,3,3,1,1,
relu<con<256,3,3,1,1,
max_pool<2,2,2,2,
relu<con<128,3,3,1,1,
relu<con<128,3,3,1,1,
max_pool<2,2,2,2,
relu<con<64,3,3,1,1,
relu<con<64,3,3,1,1,
input<matrix<unsigned char>>
```

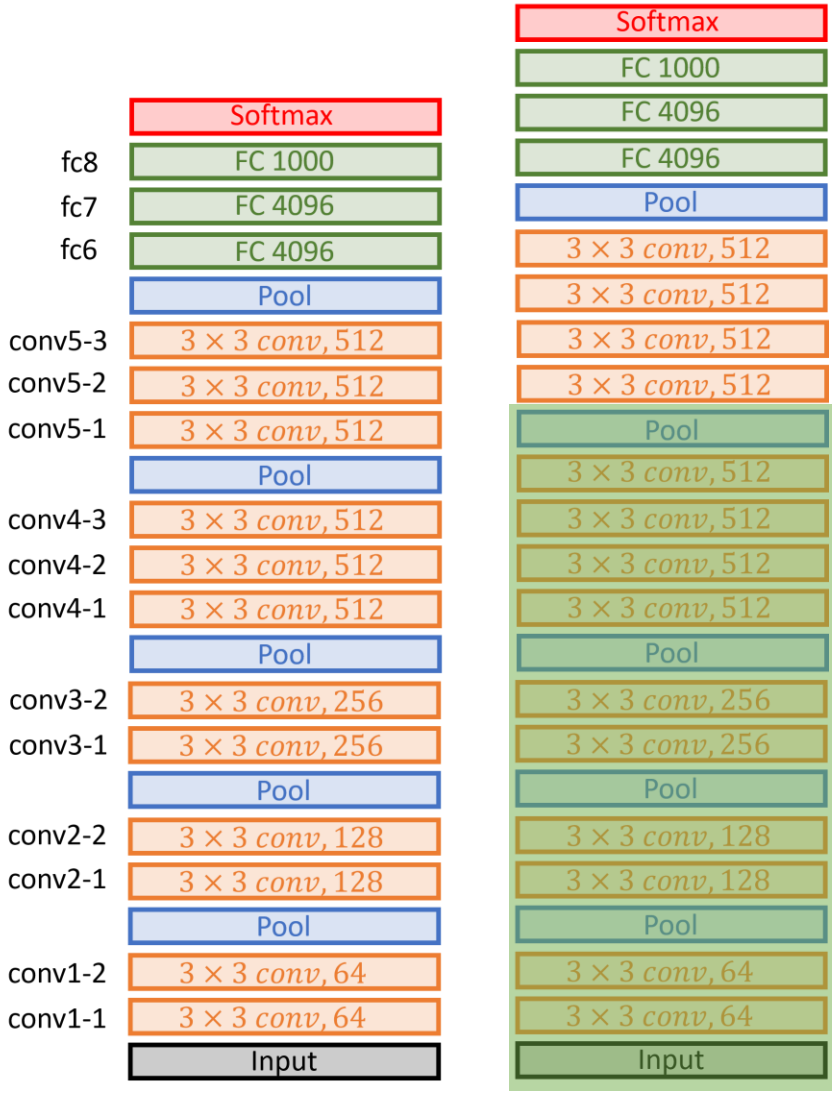
http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

http://d2l.ai/chapter_convolutional-modern/vgg.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

VGG - 2014



VGG16 **VGG19**

```
max_pool<2,2,2,2,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
max_pool<2,2,2,2,
relu<con<256,3,3,1,1,
relu<con<256,3,3,1,1,
max_pool<2,2,2,2,
relu<con<128,3,3,1,1,
relu<con<128,3,3,1,1,
max_pool<2,2,2,2,
relu<con<64,3,3,1,1,
relu<con<64,3,3,1,1,
input<matrix<unsigned char>>
```

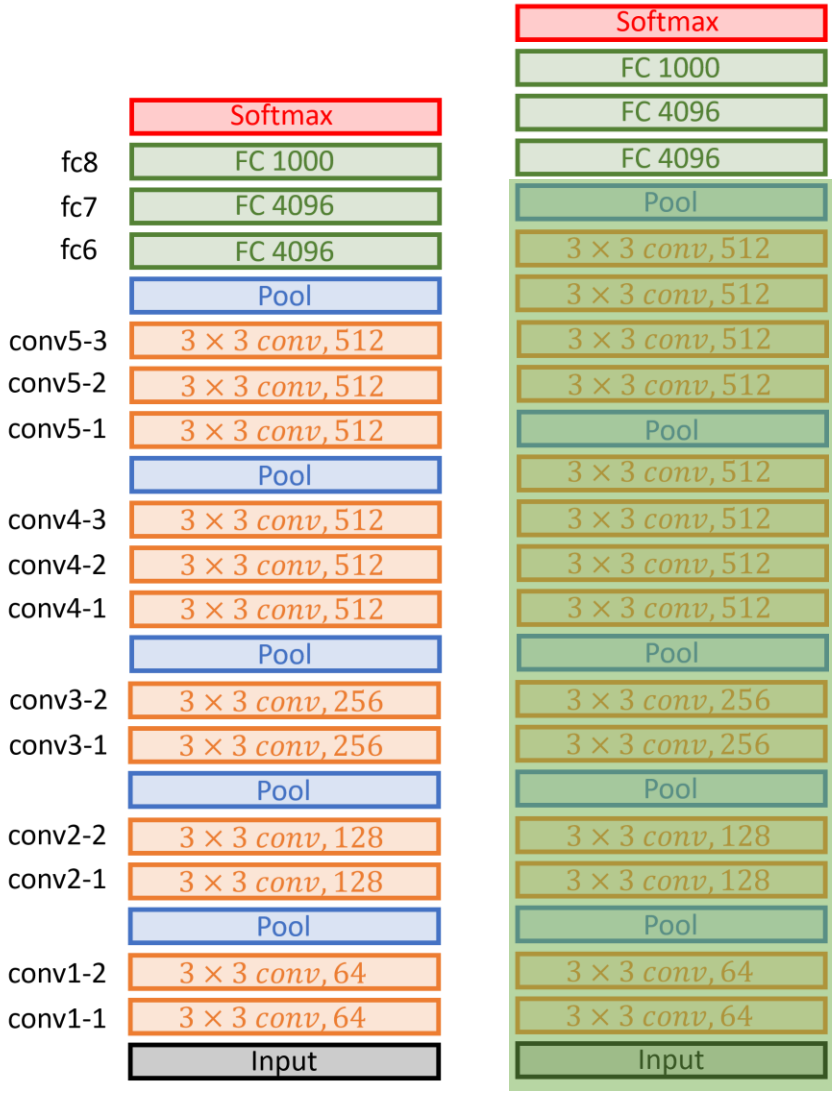
http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

http://d2l.ai/chapter_convolutional-modern/vgg.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

VGG - 2014




```
max_pool<2,2,2,2,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
max_pool<2,2,2,2,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
max_pool<2,2,2,2,
relu<con<256,3,3,1,1,
relu<con<256,3,3,1,1,
max_pool<2,2,2,2,
relu<con<128,3,3,1,1,
relu<con<128,3,3,1,1,
max_pool<2,2,2,2,
relu<con<64,3,3,1,1,
relu<con<64,3,3,1,1,
input<matrix<unsigned char>>
```

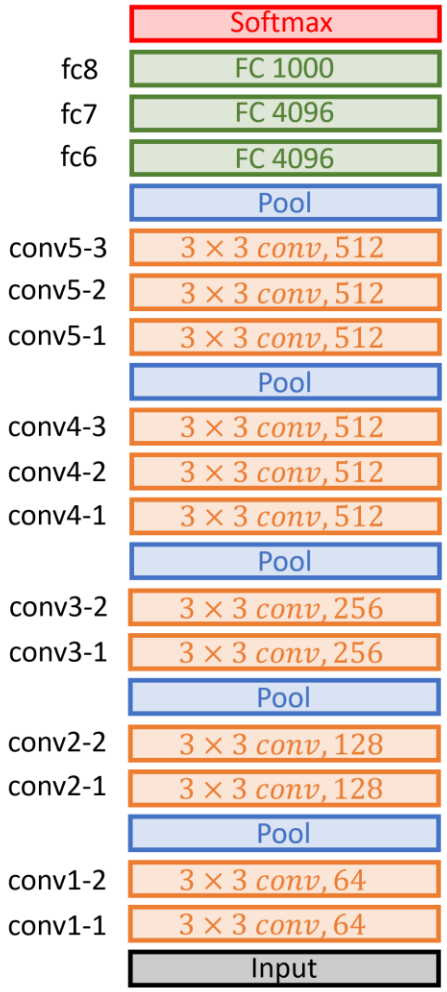
http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

http://d2l.ai/chapter_convolutional-modern/vgg.html

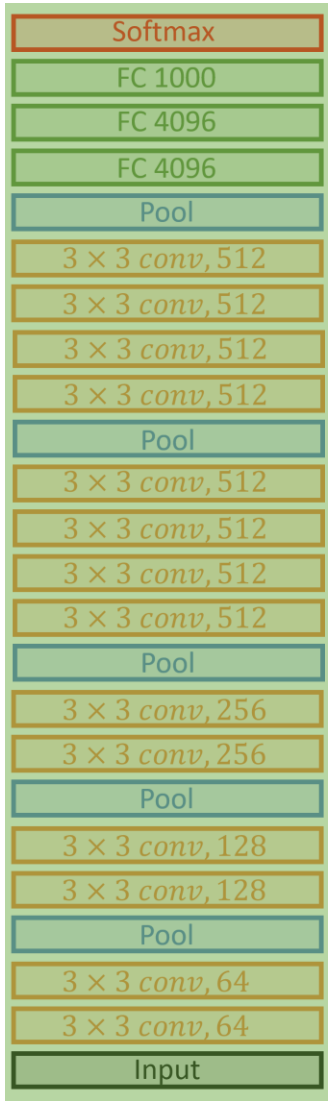
Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

VGG - 2014



VGG16



VGG19

```
fc<1000,
dropout<relu<fc<4096,
dropout<relu<fc<4096,
max_pool<2,2,2,2,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
max_pool<2,2,2,2,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
relu<con<512,3,3,1,1,
max_pool<2,2,2,2,
relu<con<256,3,3,1,1,
relu<con<256,3,3,1,1,
max_pool<2,2,2,2,
relu<con<128,3,3,1,1,
relu<con<128,3,3,1,1,
max_pool<2,2,2,2,
relu<con<64,3,3,1,1,
relu<con<64,3,3,1,1,
input<matrix<unsigned char>>
```

http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

http://d2l.ai/chapter_convolutional-modern/vgg.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

VGG - 2014

```
using net_type_vgg = loss_multiclass_log<
                                fc<2,
                                dropout<relu<fc<4096,
                                dropout<relu<fc<4096,
                                max_pool<2,2,2,2,
                                relu<con<128,3,3,1,1,
                                relu<con<128,3,3,1,1,
                                max_pool<2,2,2,2,
                                relu<con<64,3,3,1,1,
                                relu<con<64,3,3,1,1,
                                max_pool<2,2,2,2,
                                relu<con<32,3,3,1,1,
                                relu<con<32,3,3,1,1,
                                max_pool<2,2,2,2,
                                relu<con<16,3,3,1,1,
                                relu<con<16,3,3,1,1,
                                max_pool<2,2,2,2,
                                relu<con<8,3,3,1,1,
                                relu<con<8,3,3,1,1,
                                input<matrix<unsigned char>>
                                >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>;
```

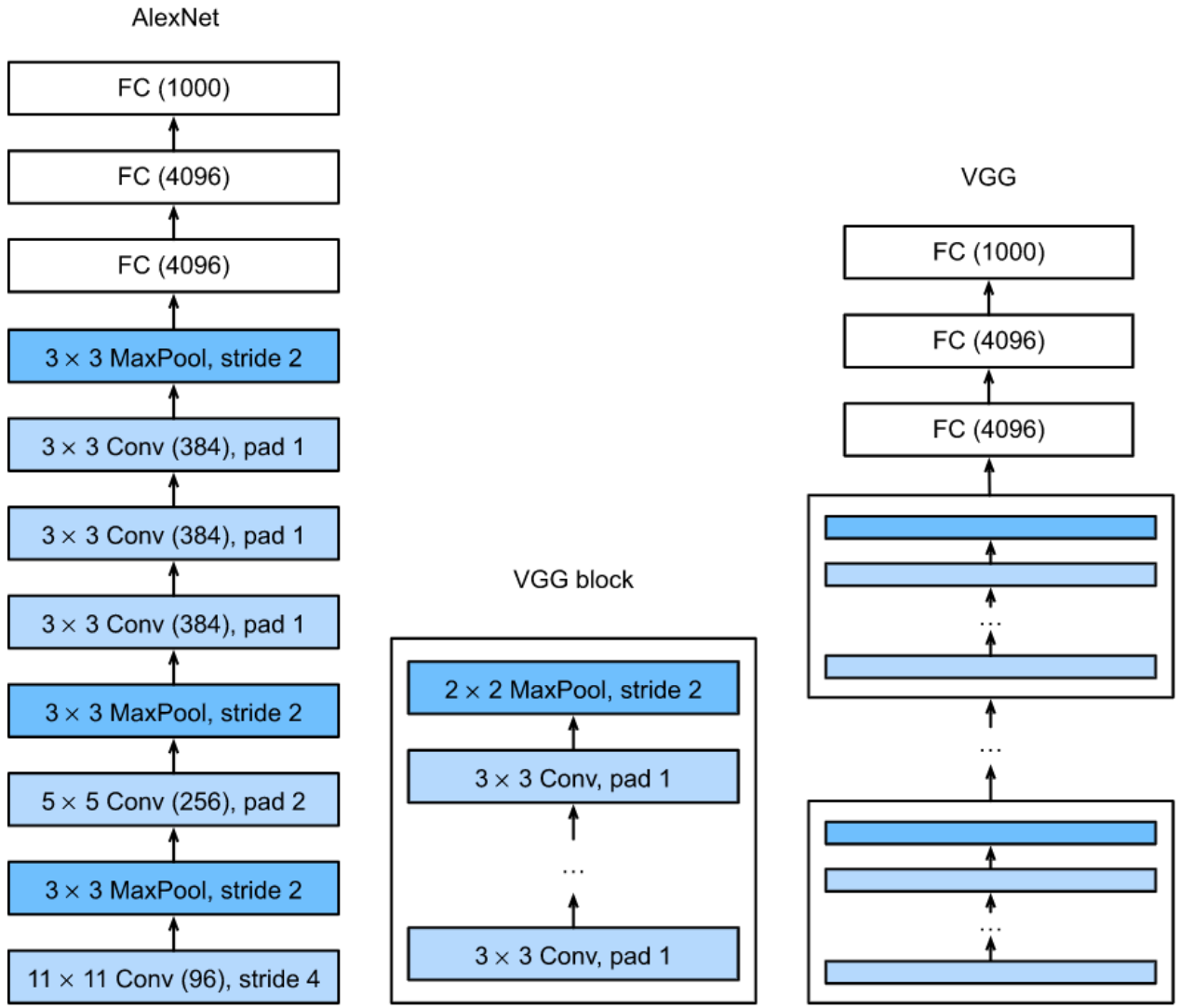
http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

http://d2l.ai/chapter_convolutional-modern/vgg.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

VGG - 2014



http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

http://d2l.ai/chapter_convolutional-modern/vgg.html

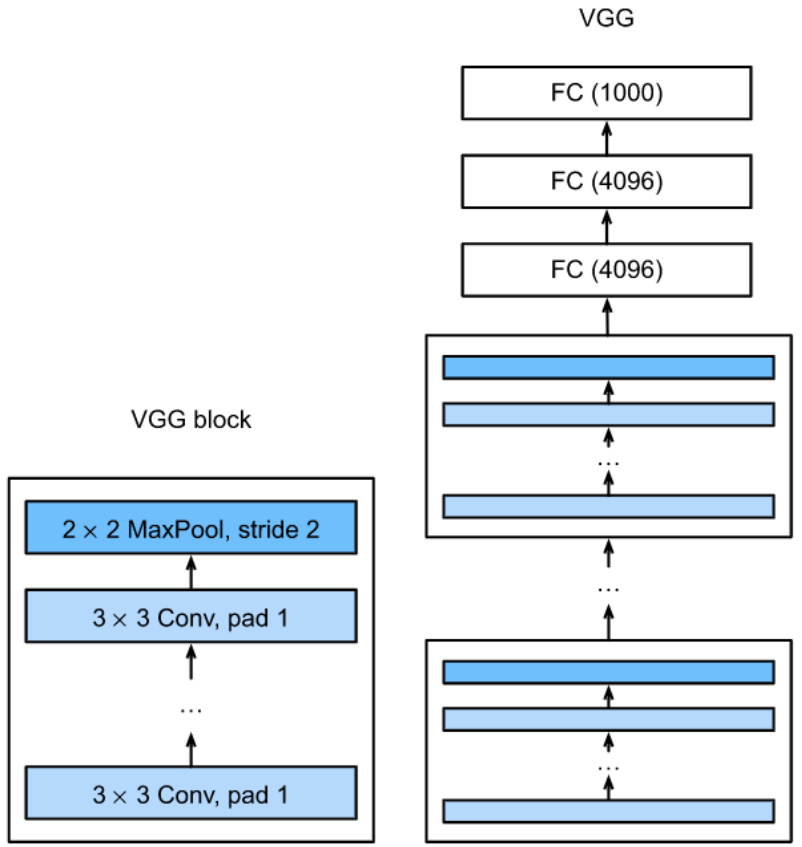
Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

VGG - 2014

```
template <typename SUBNET> using block_1 = max_pool<2,2,2,2, relu<con<8,3,3,1,1, relu<con<8,3,3,1,1, SUBNET>>>>>;
template <typename SUBNET> using block_2 = max_pool<2,2,2,2, relu<con<16,3,3,1,1, relu<con<16,3,3,1,1, SUBNET>>>>>;
template <typename SUBNET> using block_3 = max_pool<2,2,2,2, relu<con<32,3,3,1,1, relu<con<32,3,3,1,1, SUBNET>>>>>;
template <typename SUBNET> using block_4 = max_pool<2,2,2,2, relu<con<64,3,3,1,1, relu<con<64,3,3,1,1, SUBNET>>>>>;
template <typename SUBNET> using block_5 = max_pool<2,2,2,2, relu<con<128,3,3,1,1, relu<con<128,3,3,1,1, SUBNET>>>>>;
```

```
using net_type_vgg_3 = loss_multiclass_log<
                        fc<2,
                        dropout<relu<fc<4096,
                        dropout<relu<fc<4096,
                        block_5<
                        block_4<
                        block_3<
                        block_2<
                        block_1<
                        input<matrix<unsigned char>>
                        >>>>>>>>>>>>>;
```



http://dlib.net/dnn_introduction2_ex.cpp.html

http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

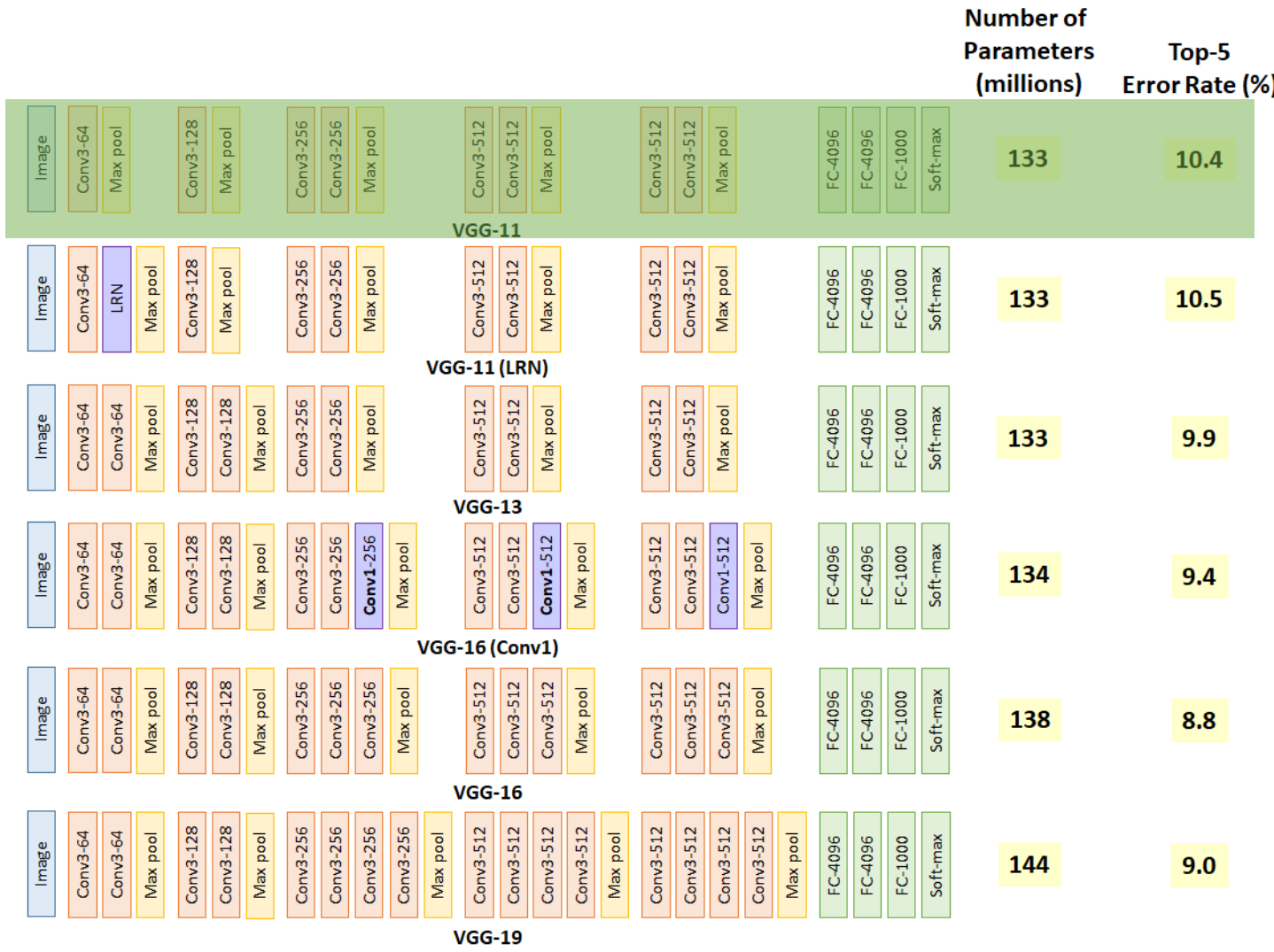
http://d2l.ai/chapter_convolutional-modern/vgg.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

VGG – 2014 Pytorch

```
VGGNet = nn.Sequential(
    nn.Conv2d(3, 64, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=2, stride=2),
    nn.Conv2d(64, 128, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=2, stride=2),
    nn.Conv2d(128, 256, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.Conv2d(256, 256, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=2, stride=2),
    nn.Conv2d(256, 512, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.Conv2d(512, 512, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=2, stride=2),
    nn.Conv2d(512, 512, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.Conv2d(512, 512, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(kernel_size=2, stride=2),
    nn.Flatten(),
    nn.LazyLinear(4096),
    nn.ReLU(),
    nn.Dropout2d(0.5),
    nn.LazyLinear(4096),
    nn.ReLU(),
    nn.Dropout2d(0.5),
    nn.LazyLinear(2)
)
```



http://d2l.ai/chapter_convolutional-modern/vgg.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

CovNet Architectures

- LeNet (1990s)
- AlexNet (2012)
- ZF Net (2013)
- VGGNet (2014)
- GoogLeNet (2014)
- ResNets (2015)
- DenseNet (2017)

GoogLeNet - 2014

2014

- ImageNet Challenge http://www.image-net.org/challenges/LSVRC/2014/
- http://www.image-net.org/challenges/LSVRC/2014/results
- **GoogLe**Net > Google/LeNet
- **1 x 1 convolution**
- Network In Network
- **Variously-sized kernels**
- **Inception Blocks**

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). **Going deeper with convolutions**. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).

GoogLeNet - 2014

všechny plakáty (98)

Počátek

Inception
Inception
(další názvy)

koupit
koupit

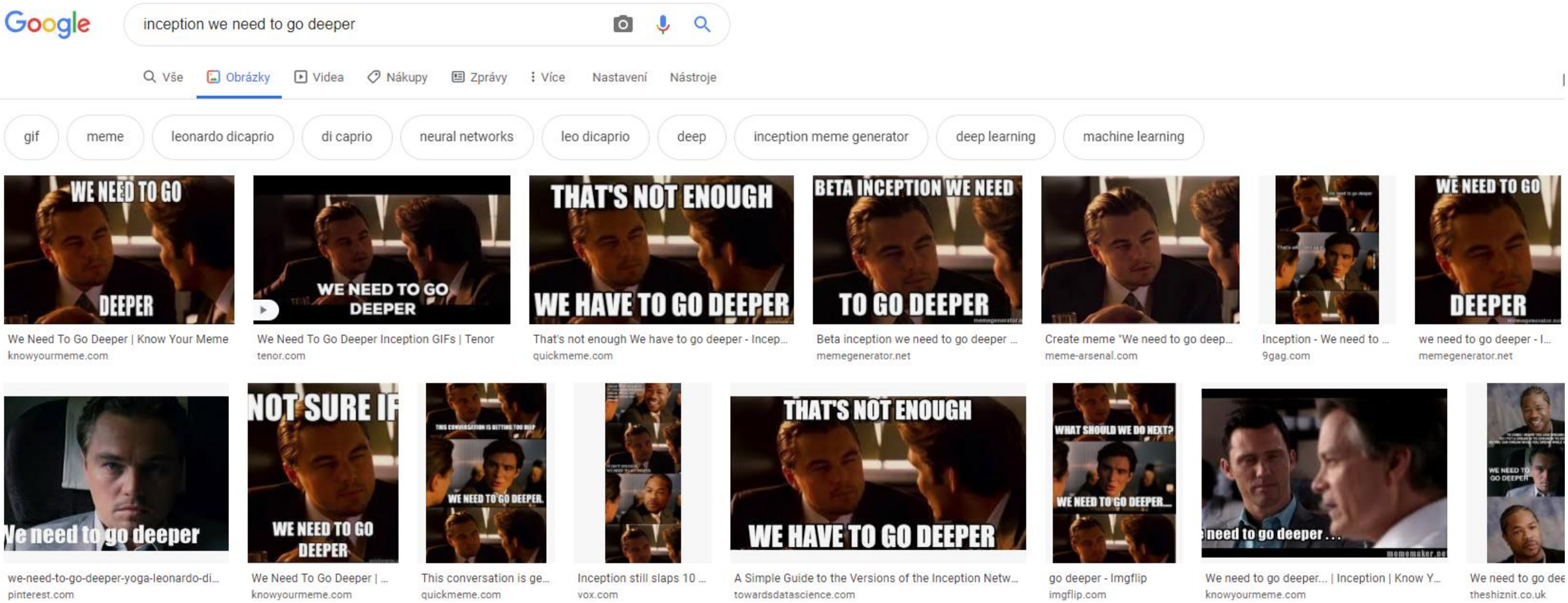
Thriller / Mysteriózní / Akční / Sci-Fi / Dobrodružný
USA / Velká Británie, 2010, 148 min

Režie: Christopher Nolan
Scénář: Christopher Nolan
Kamera: Wally Pfister
Hudba: Hans Zimmer
Hrají: Leonardo DiCaprio, Joseph Gordon-Levitt, Ellen Page, Tom Hardy, Ken Watanabe, Dileep Rao, Cillian Murphy, Tom Berenger, Marion Cotillard, Pete Postlethwaite, Michael Caine, Lukas Haas, Tai-Li Lee, Claire Geare, Taylor Geare, Johnathan Geare, Tohoru Masamune, Júdži Okumoto, Earl Cameron, Ryan Hayward, Tim Kelleher, Talulah Riley, S… (více)
(další profese)

https://www.csfd.cz/film/254156-pocatek/prehled/

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

GoogLeNet - 2014



GoogLeNet - 2014

- **Deep Network > overfitting**
- **What kernel sizes is right?**
  - **3 x 3 (VGGNet)**
  - **5 x 5 (LeNet)**
  - **7 x 7 (ZFNet)**
  - **11 x 11 (AlexNet)**

http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

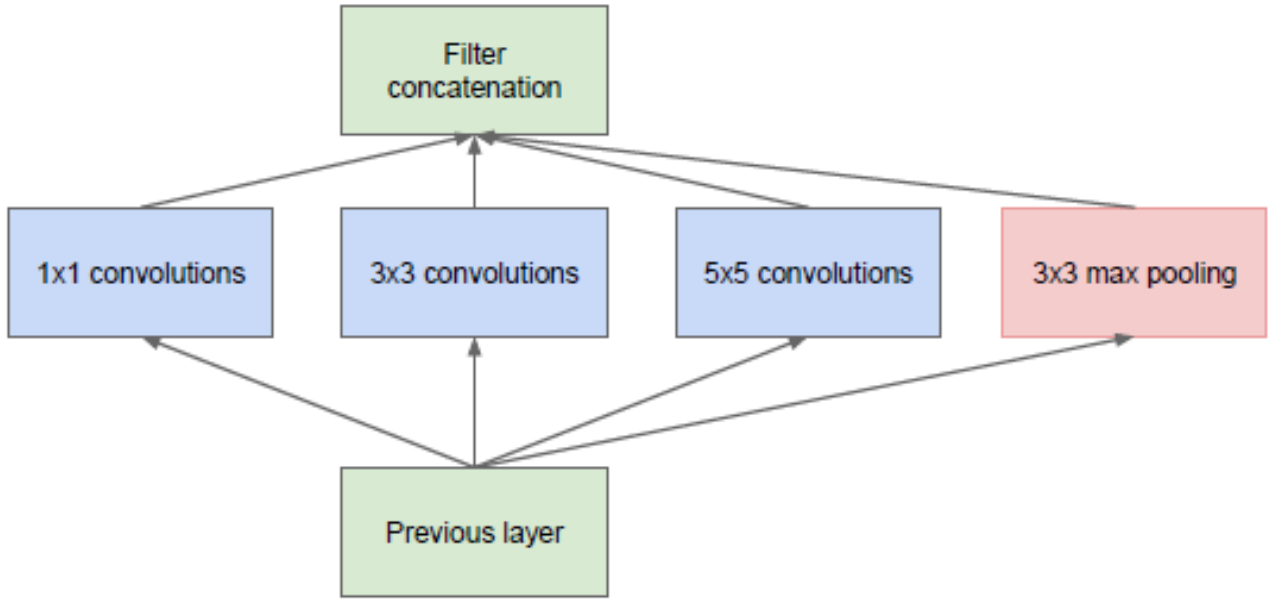
GoogLeNet - 2014

- **Multiple filter sizes** in on the same level
- **“Wider”** rather than “Deeper”

http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

GoogLeNet - 2014

- **Multiple filter sizes** in on the same level
- **“Wider”** rather than “Deeper” - **3 different filters**

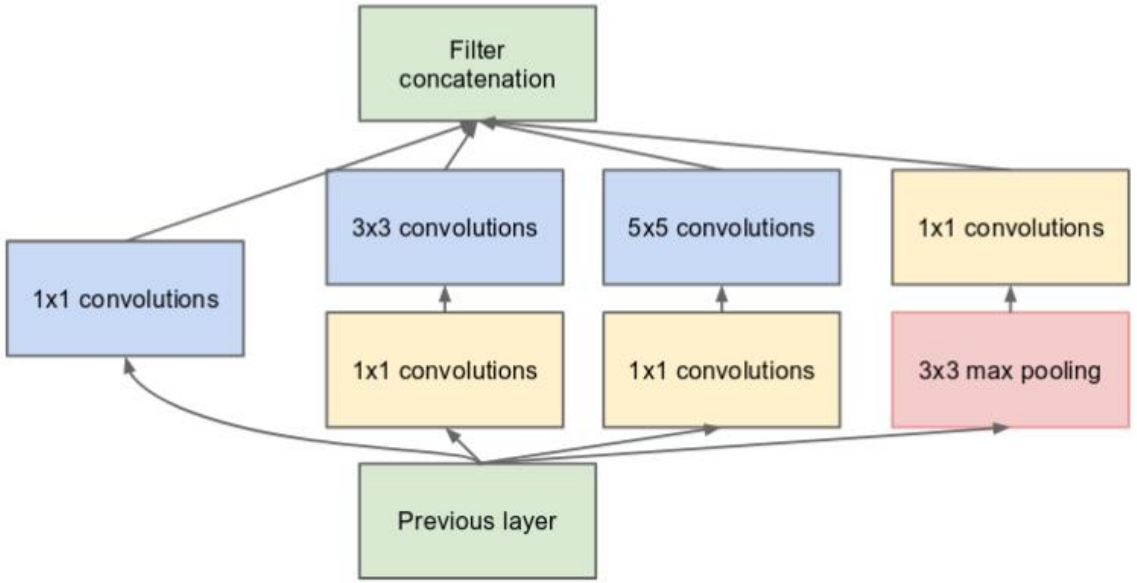


(a) Inception module, naïve version

http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

GoogLeNet - 2014

- **1 x 1 filters for dimension reduction**
- **“Wider”** rather than “Deeper” - **3 different filters**

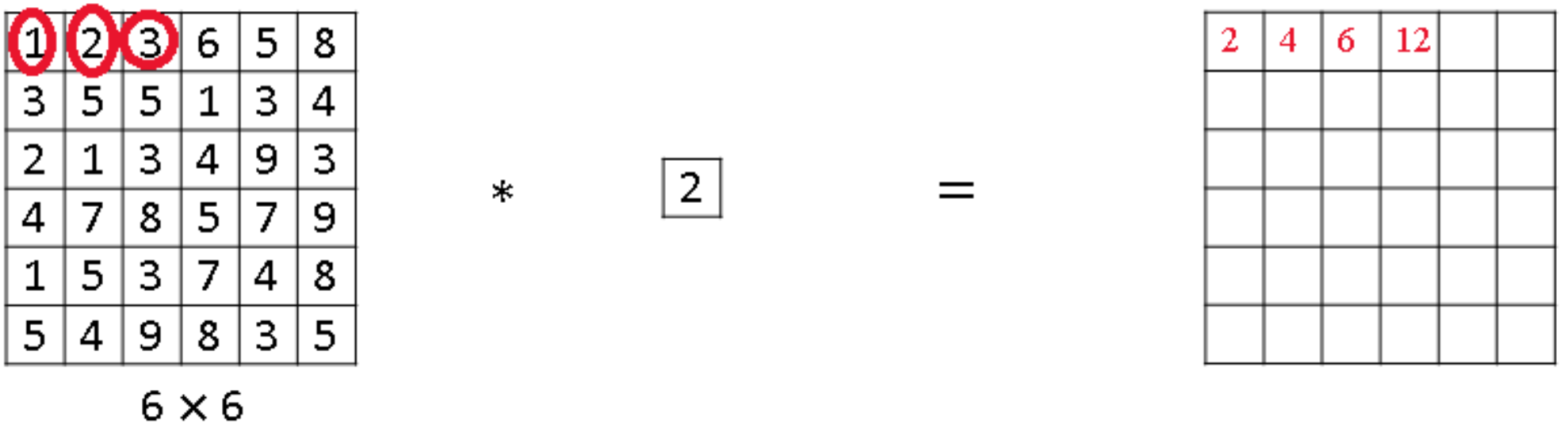


(b) Inception module with dimension reductions

http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

GoogLeNet - 2014

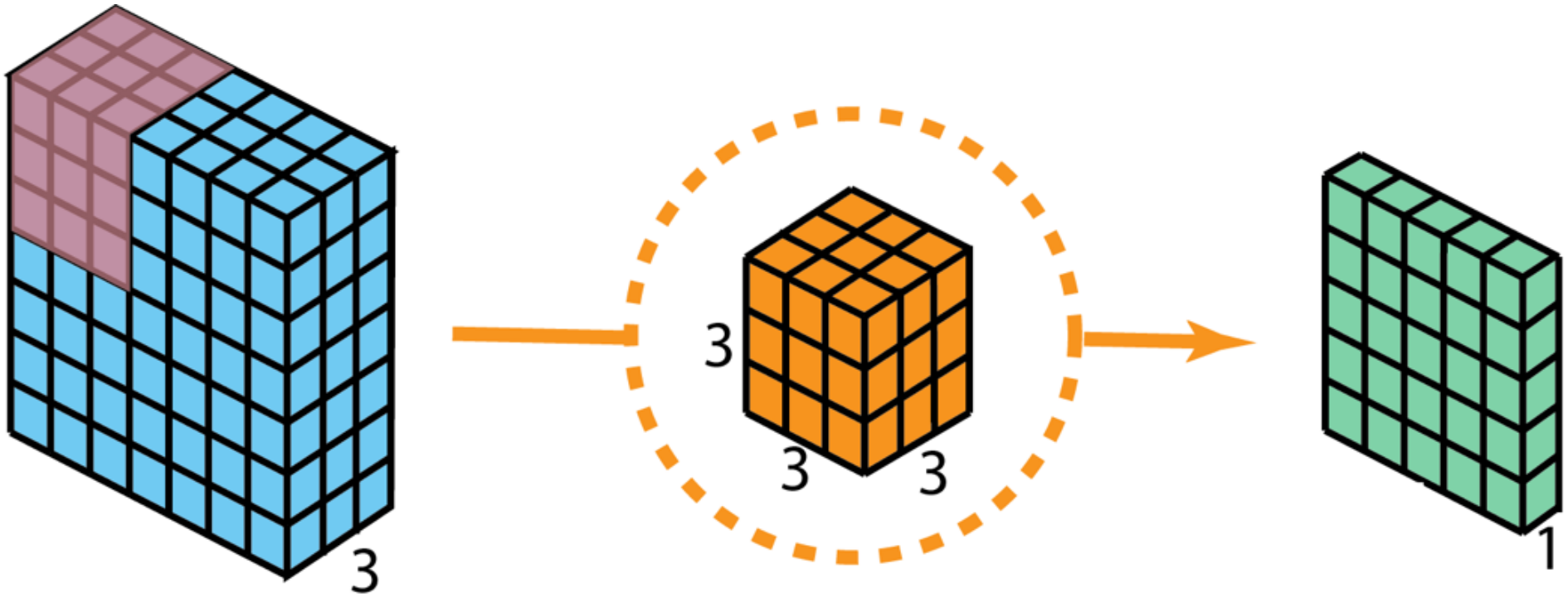
- **1 x 1 filters in 2D**



https://upscfever.com/upsc-fever/en/data/deeplearning4/15.html
http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

GoogLeNet - 2014

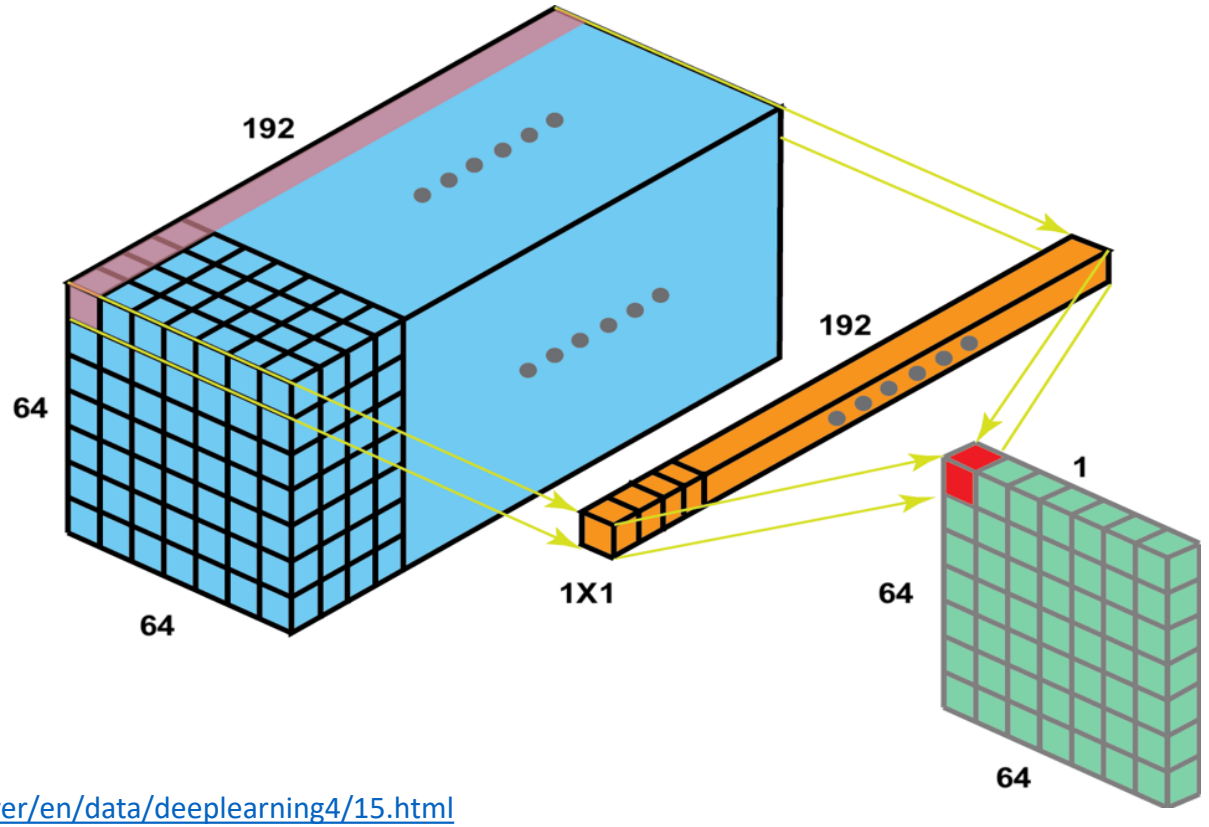
- **1 x 1 filters in 3D**



https://medium.com/analytics-vidhya/talented-mr-1x1-comprehensive-look-at-1x1-convolution-in-deep-learning-f6b355825578
http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

GoogLeNet - 2014

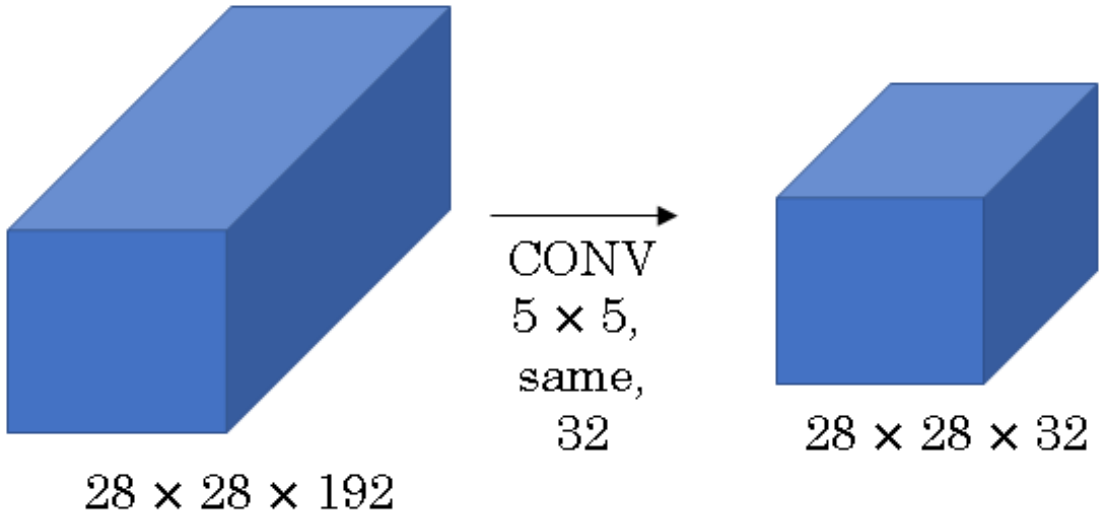
- **1 x 1 filters in** large number of channels - **192**



https://upscfever.com/upsc-fever/en/data/deeplearning4/15.html
https://medium.com/analytics-vidhya/talented-mr-1x1-comprehensive-look-at-1x1-convolution-in-deep-learning-f6b355825578
http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

GoogLeNet - 2014

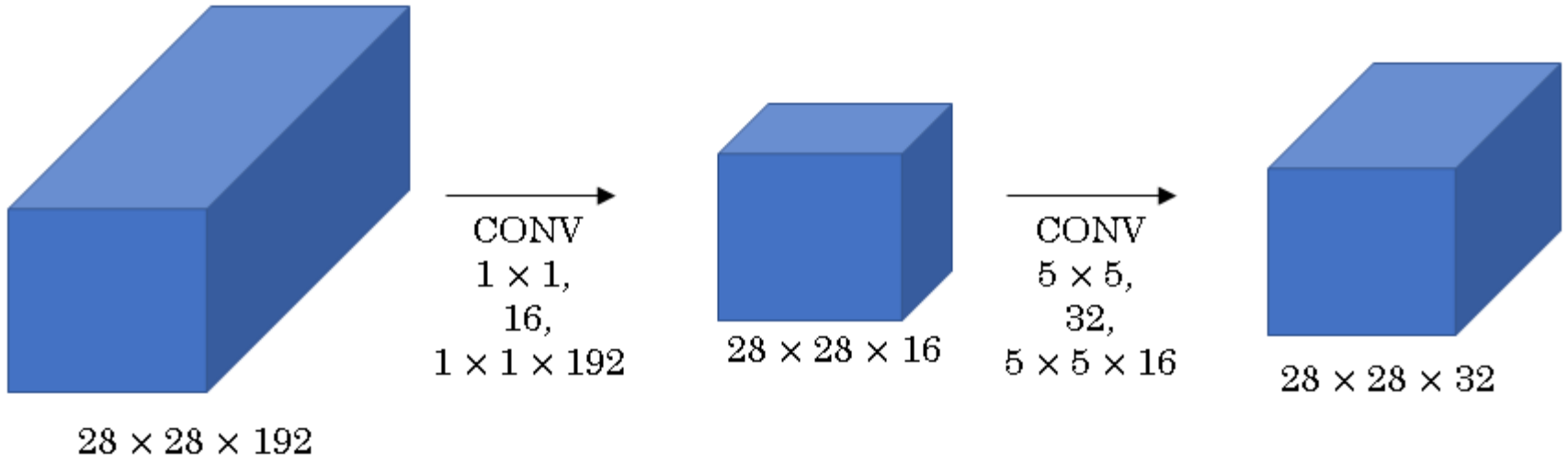
- convolve 28 X 28 X 192 input feature maps with 5 X 5 X 32 filters.
- **This will result in aprox. 120 *Million operations***



https://upscfever.com/upsc-fever/en/data/deeplearning4/15.html
https://medium.com/analytics-vidhya/talented-mr-1x1-comprehensive-look-at-1x1-convolution-in-deep-learning-f6b355825578
http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

GoogLeNet - 2014

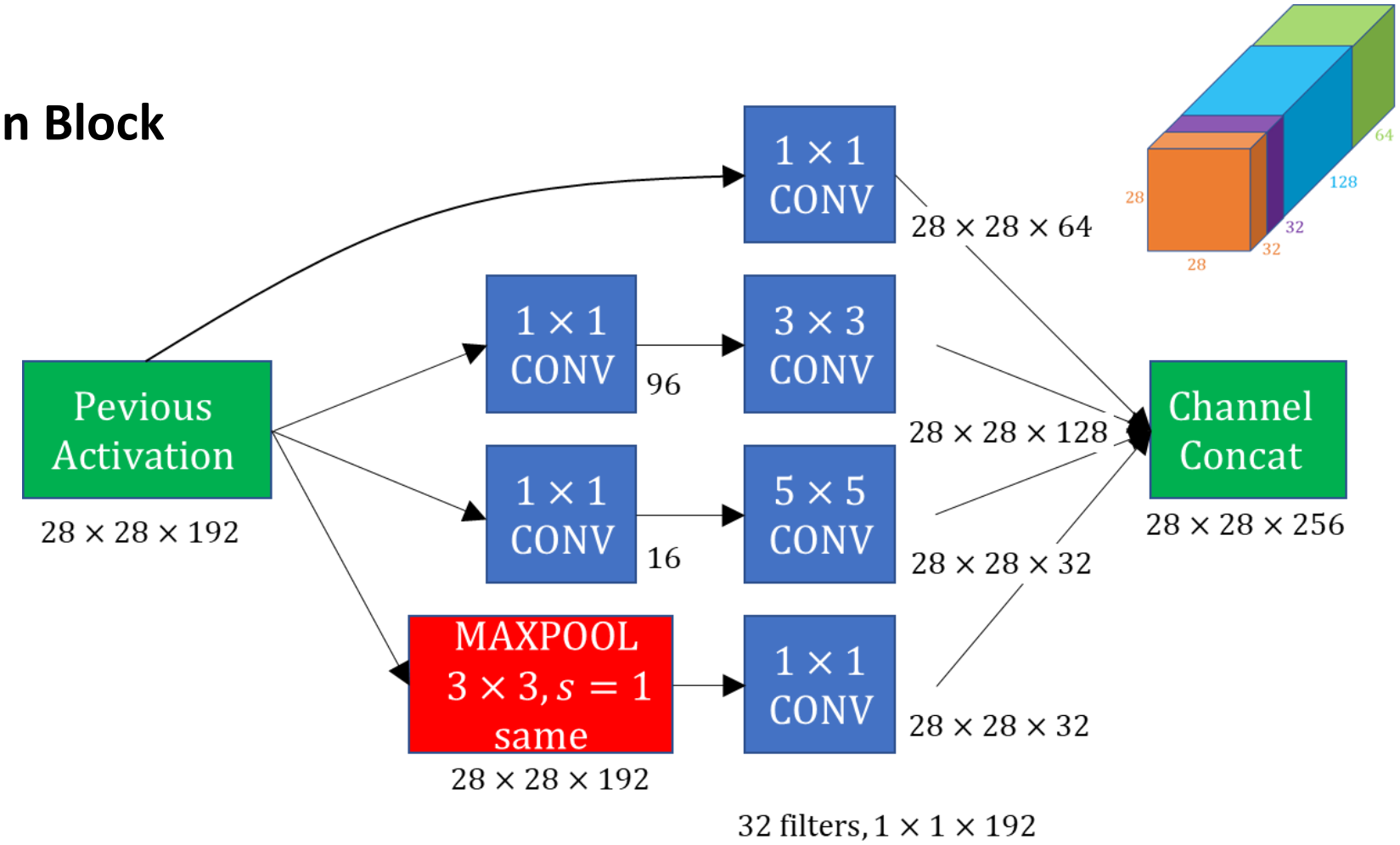
- Using 1 X 1 filters - **aprox. 12 *Million operations***
- **“Bottleneck layer”**



https://upscfever.com/upsc-fever/en/data/deeplearning4/15.html
https://medium.com/analytics-vidhya/talented-mr-1x1-comprehensive-look-at-1x1-convolution-in-deep-learning-f6b355825578
http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

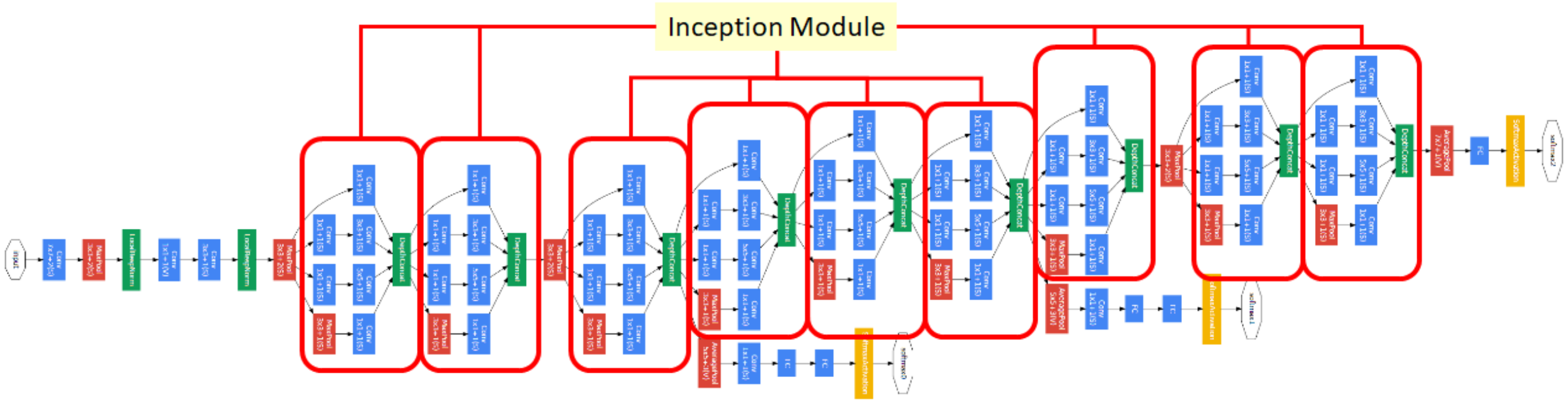
GoogLeNet - 2014

- **Inception Block**



https://upscfever.com/upsc-fever/en/data/deeplearning4/15.html
https://medium.com/analytics-vidhya/talented-mr-1x1-comprehensive-look-at-1x1-convolution-in-deep-learning-f6b355825578
http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

GoogLeNet - 2014



https://towardsdatascience.com/review-inception-v4-evolved-from-googlenet-merged-with-resnet-idea-image-classification-5e8c339d18bc
http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

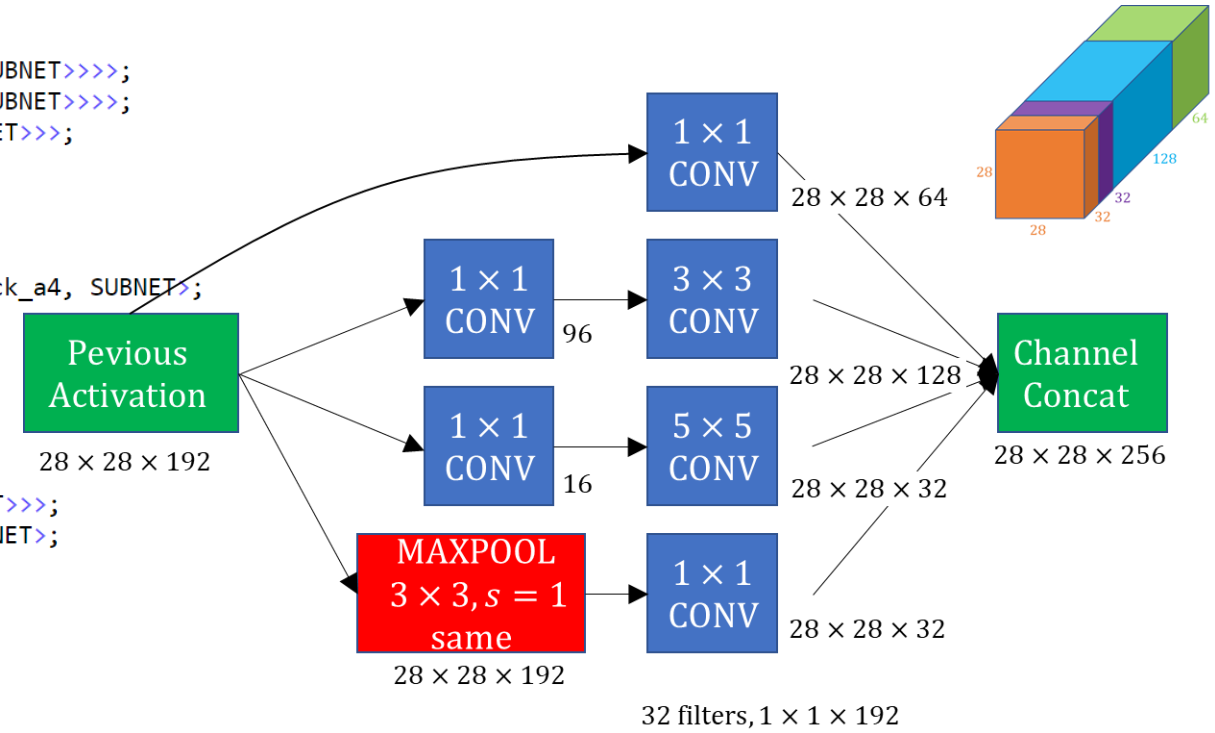
GoogLeNet - 2014

```
// Inception layer has some different convolutions inside.  Here we define
// blocks as convolutions with different kernel size that we will use in
// inception layer block.
template <typename SUBNET> using block_a1 = relu<con<10,1,1,1,1,SUBNET>>;
template <typename SUBNET> using block_a2 = relu<con<10,3,3,1,1,relu<con<16,1,1,1,1,SUBNET>>>>;
template <typename SUBNET> using block_a3 = relu<con<10,5,5,1,1,relu<con<16,1,1,1,1,SUBNET>>>>;
template <typename SUBNET> using block_a4 = relu<con<10,1,1,1,1,max_pool<3,3,1,1,SUBNET>>>;

// Here is inception layer definition. It uses different blocks to process input
// and returns combined output.  Dlib includes a number of these inceptionN
// layer types which are themselves created using concat layers.
template <typename SUBNET> using incept_a = inception4<block_a1,block_a2,block_a3,block_a4, SUBNET>;

// Network can have inception layers of different structure.  It will work
// properly so long as all the sub-blocks inside a particular inception block
// output tensors with the same number of rows and columns.
template <typename SUBNET> using block_b1 = relu<con<4,1,1,1,1,SUBNET>>;
template <typename SUBNET> using block_b2 = relu<con<4,3,3,1,1,SUBNET>>;
template <typename SUBNET> using block_b3 = relu<con<4,1,1,1,1,max_pool<3,3,1,1,SUBNET>>>;
template <typename SUBNET> using incept_b = inception3<block_b1,block_b2,block_b3,SUBNET>;

// Now we can define a simple network for classifying MNIST digits.  We will
// train and test this network in the code below.
using net_type = loss_multiclass_log<
        fc<10,
        relu<fc<32,
        max_pool<2,2,2,2,incept_b<
        max_pool<2,2,2,2,incept_a<
        input<matrix<unsigned char>>
        >>>>>>>>>;
```



http://dlib.net/dnn_inception_ex.cpp.html
http://datahacker.rs/building-inception-network/
http://d2l.ai/chapter_convolutional-modern/googlenet.html
https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

CovNet Architectures

- LeNet (1990s)
- AlexNet (2012)
- ZF Net (2013)
- VGGNet (2014)
- GoogLeNet (2014)
- ResNets (2015)
- DenseNet (2017)

ResNets - 2015

2015

- ImageNet Challenge http://www.image-net.org/challenges/LSVRC/2015/
- http://www.image-net.org/challenges/LSVRC/2015/results
- http://image-net.org/challenges/talks/ILSVRC2015_12_17_15_clsloc.pdf
- **Deep networks are hard to train - Vanishing gradient issue**
- **Residual Blocks**
- **1 x 1 convolutions**

http://datahacker.rs/deep-learning-residual-networks/

He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.

ResNets - 2015

"Is learning better networks as easy as stacking more layers?"

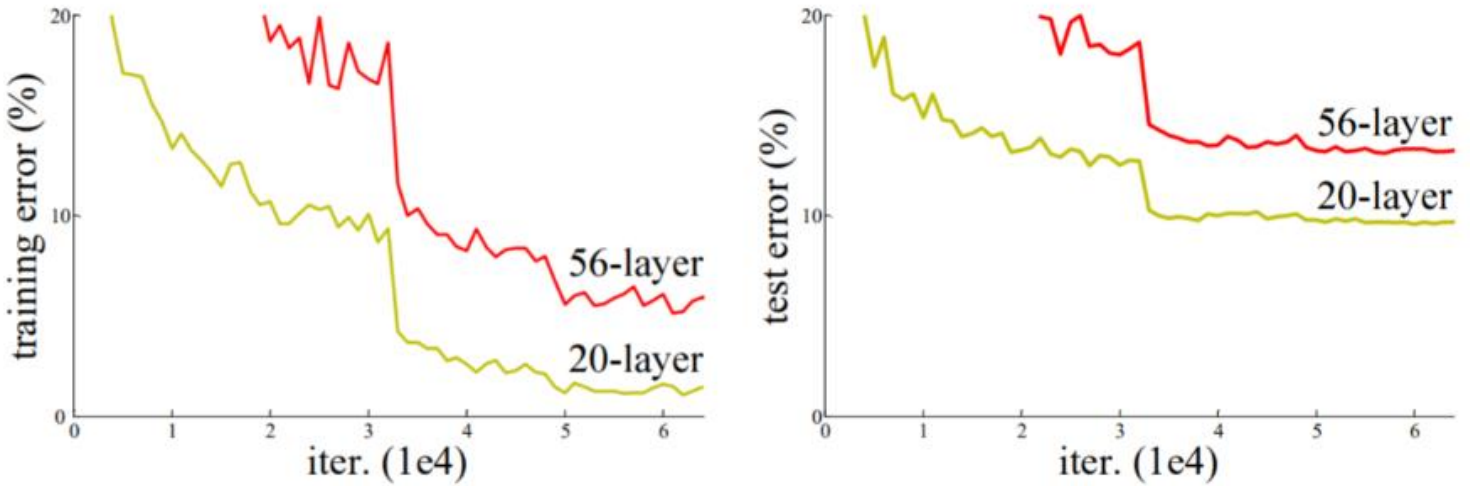


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

ResNets - 2015

- in the **classical** way: the output of layer 1 is pass to layer 2
- in **ResNet**: information can go much deeper into the neural network
  - information (gradient) from higher layer can directly pass to the lower layer
  - via **shortcut** or **skip connection** (identity connection, addition operator)
  - stacking a lot of residual blocks together, we can build much deeper neural networks

"shortcut connections"

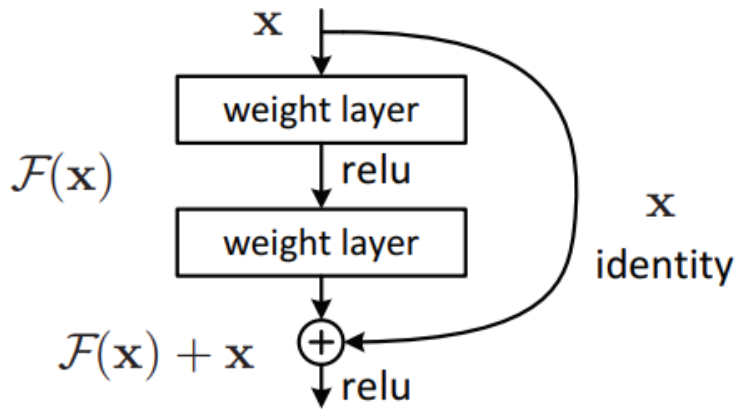


Figure 2. Residual learning: a building block.

He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

ResNets - 2015

“shortcut connections”

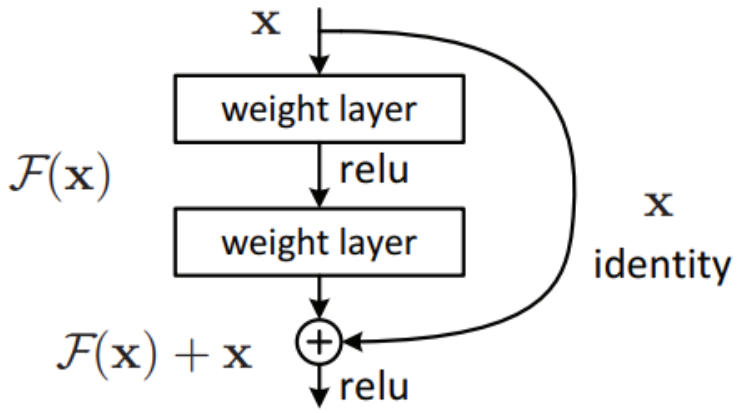


Figure 2. Residual learning: a building block.

“bottleneck”

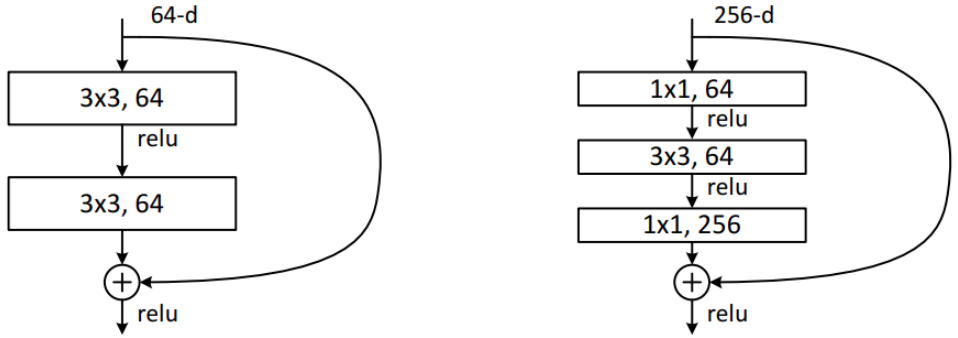
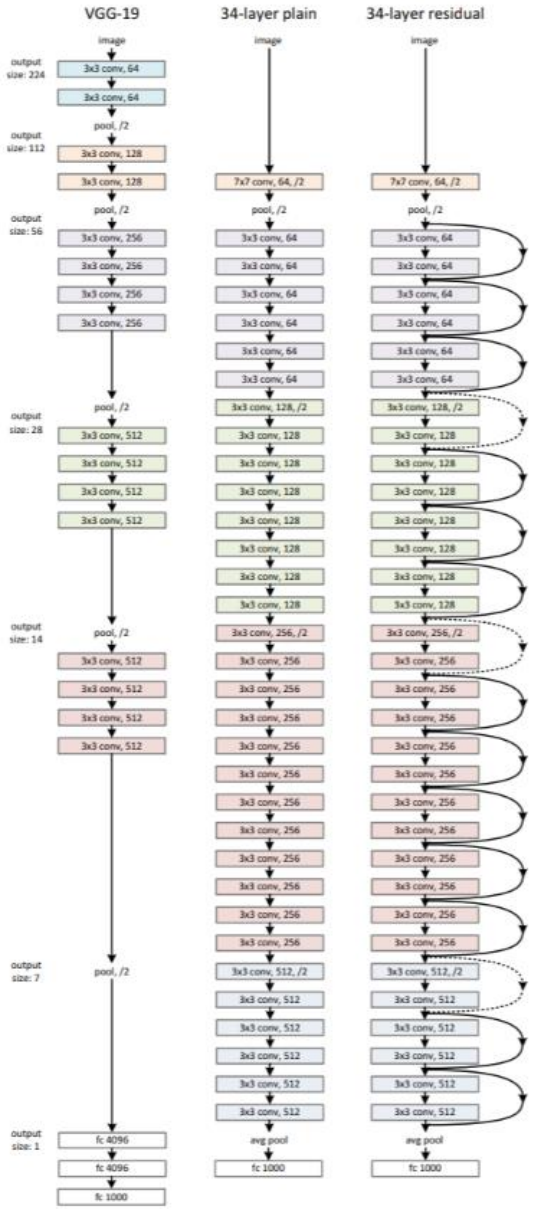


Figure 5. A deeper residual function $\mathcal{F}$ for ImageNet. Left: a building block (on 56×56 feature maps) as in Fig. 3 for ResNet-34. Right: a “bottleneck” building block for ResNet-50/101/152.

He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

ResNets - 2015



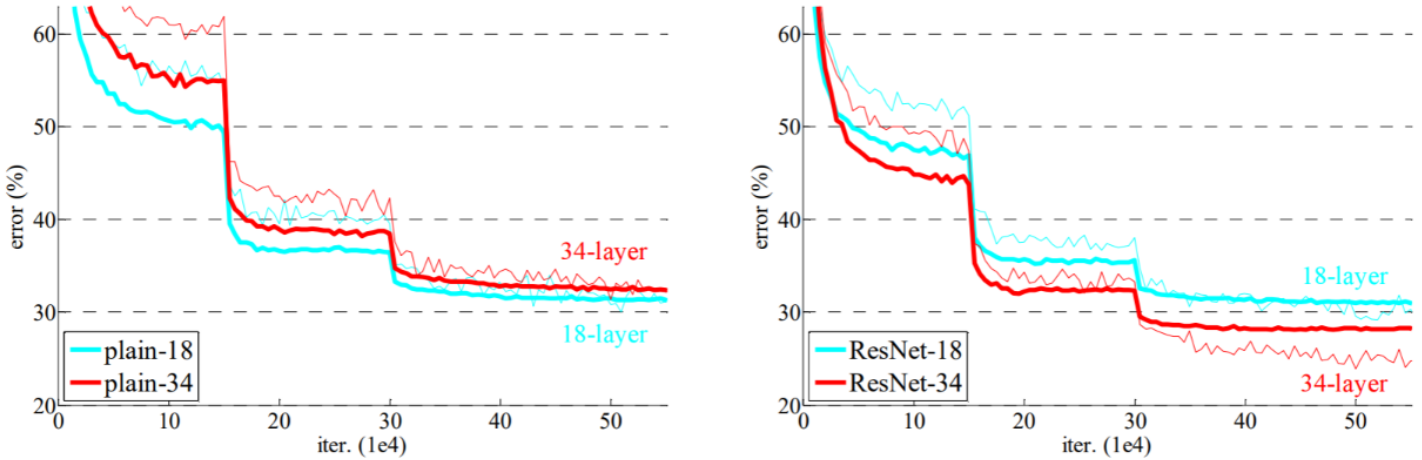


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

ResNets - 2015

Table 2. Classification error (%) on the CIFAR-10 test set using different activation functions.

case	Fig.	ResNet-110	ResNet-164
original Residual Unit [1]	Fig. 4(a)	6.61	5.93
BN after addition	Fig. 4(b)	8.17	6.50
ReLU before addition	Fig. 4(c)	7.84	6.14
ReLU-only pre-activation	Fig. 4(d)	6.71	5.91
full pre-activation	Fig. 4(e)	**6.37**	**5.46**

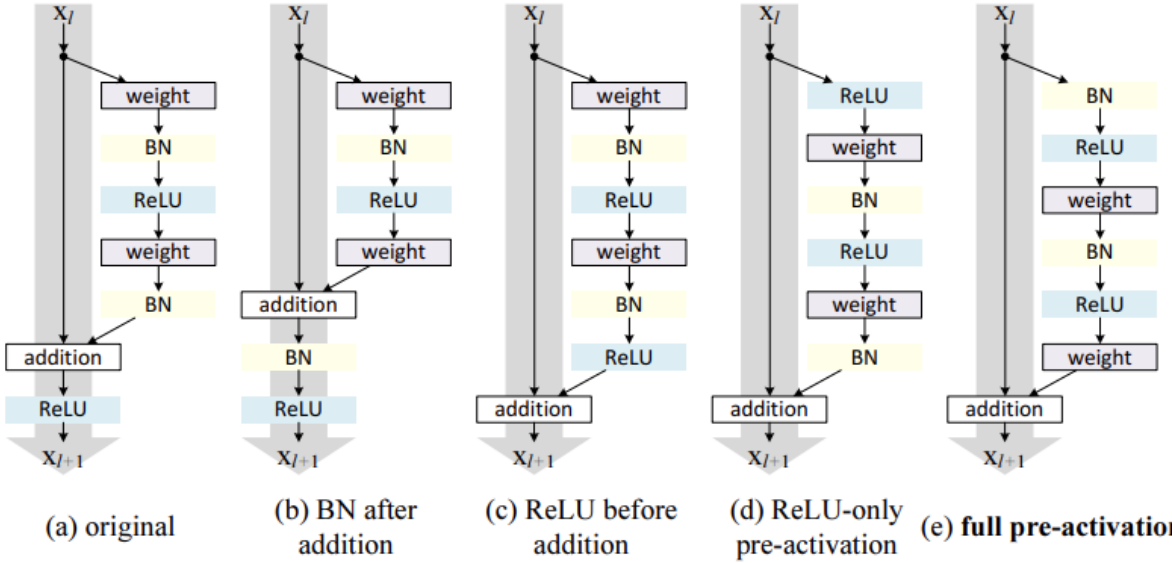


Figure 4. Various usages of activation in Table 2. All these units consist of the same components — only the orders are different.

He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Identity Mappings in Deep Residual Networks. 9908. 630-645. 10.1007/978-3-319-46493-0_38.
He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.

- **Variations of ResNets:**
  - **Inception-ResNet V1**
  - **Inception-ResNet-v2**
  - **ResNeXt - 1st Runner Up of ILSVRC classification task**
    - http://image-net.org/challenges/LSVRC/2016/results

	224×224		320×320 / 299×299	
	top-1 err	top-5 err	top-1 err	top-5 err
ResNet-101 [14]	22.0	6.0	-	-
ResNet-200 [15]	21.7	5.8	20.1	4.8
Inception-v3 [39]	-	-	21.2	5.6
Inception-v4 [37]	-	-	20.0	5.0
Inception-ResNet-v2 [37]	-	-	19.9	4.9
ResNeXt-101 (**64** × **4d**)	20.4	5.3	**19.1**	**4.4**

Winner of ILSVRC 2015
Pre-Activation ResNet
1st Runner-Up of ILSVRC 2015
Better Than Inception-v3
Inception-v4 + ResNet

https://towardsdatascience.com/review-inception-v4-evolved-from-googlenet-merged-with-resnet-idea-image-classification-5e8c339d18bc
https://towardsdatascience.com/review-resnext-1st-runner-up-of-ilsvrc-2016-image-classification-15d7f17b42ac

S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, "Aggregated Residual Transformations for Deep Neural Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 5987-5995, doi: 10.1109/CVPR.2017.634.

He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Identity Mappings in Deep Residual Networks. 9908. 630-645. 10.1007/978-3-319-46493-0_38.
He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

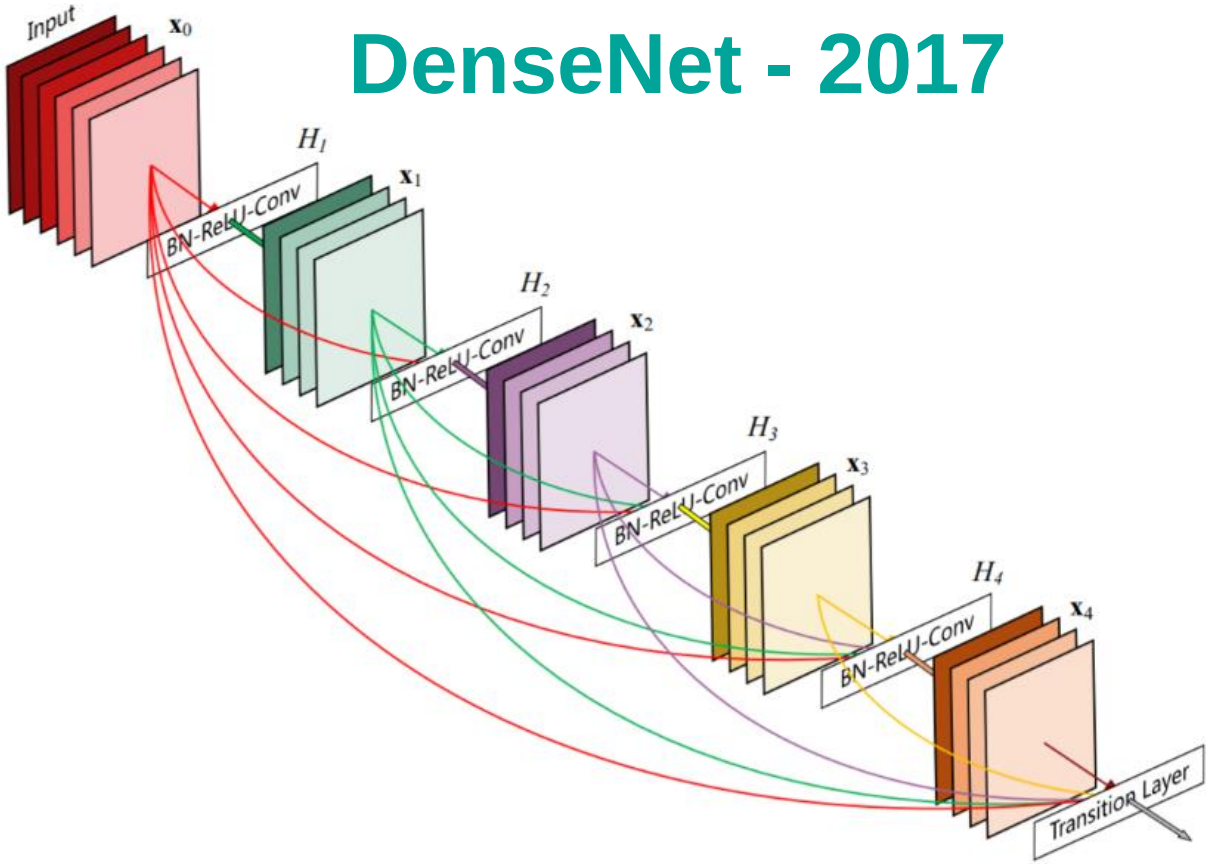
CovNet Architectures

- LeNet (1990s)
- AlexNet (2012)
- ZF Net (2013)
- VGGNet (2014)
- GoogLeNet (2014)
- ResNets (2015)
- DenseNet (2017)

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

DenseNet - 2017

- connects all layers directly with each other
- handle vanishing gradients problem
- addition vs. concatenation (ResNet vs. DenseNet)
- dense blocks



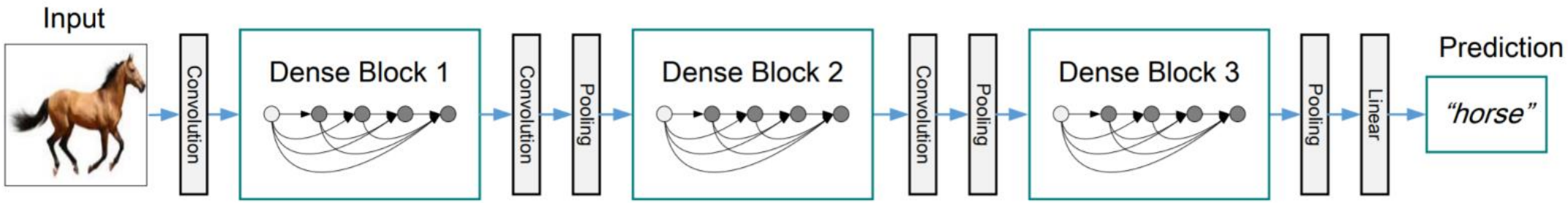
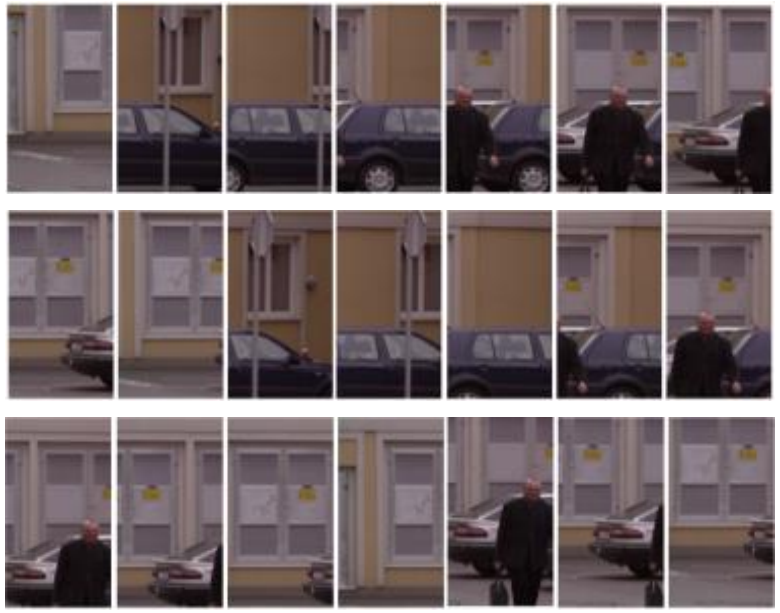


Figure 2: A deep DenseNet with three dense blocks. The layers between two adjacent blocks are referred to as transition layers and change feature-map sizes via convolution and pooling.

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

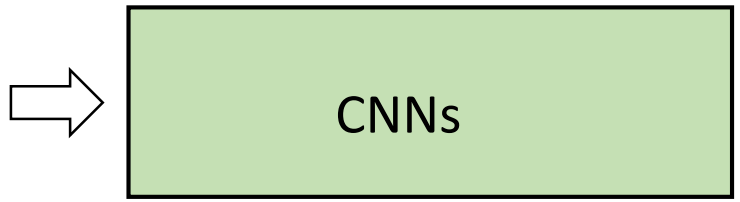
Region-Based CNNs (R-CNNs)

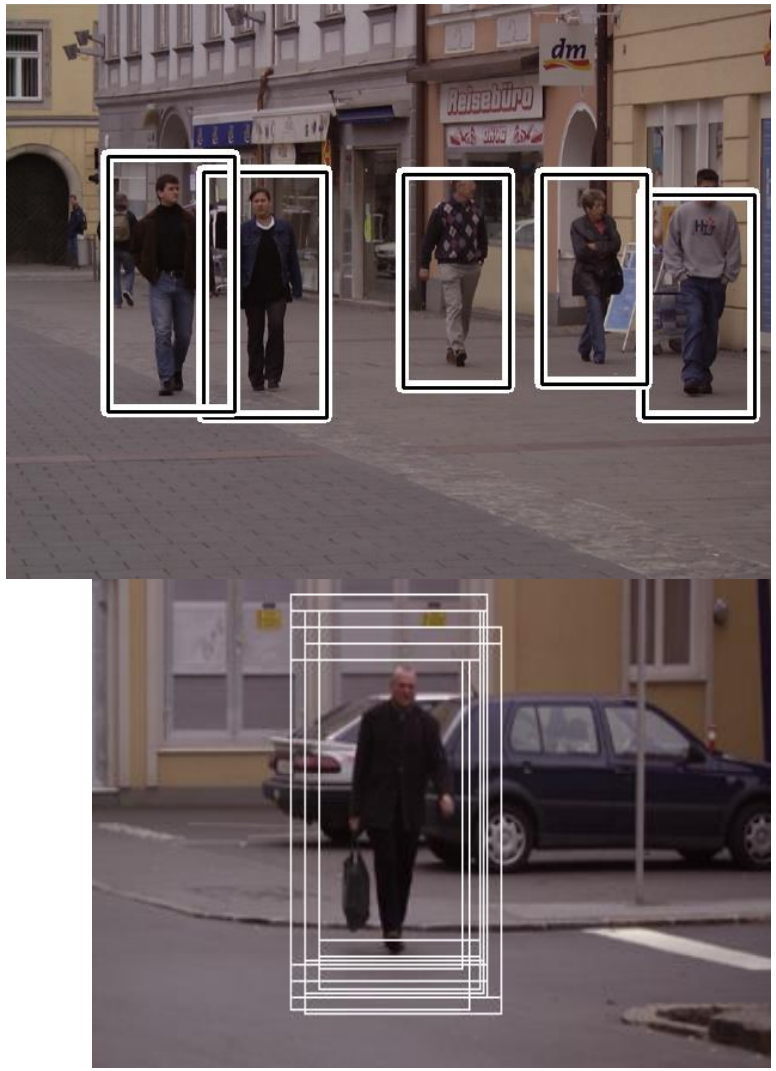
- Classical way (how to localize/detect object) is based on sliding window technique



Region-Based CNNs (R-CNNs)

- Disadvantages of sliding window with the use off very deep CNNs for object detection
  - many different image regions
  - each region is used as an input for CNNs
    - computational cost – overlapping regions (stride parameters)
    - duplicated operations





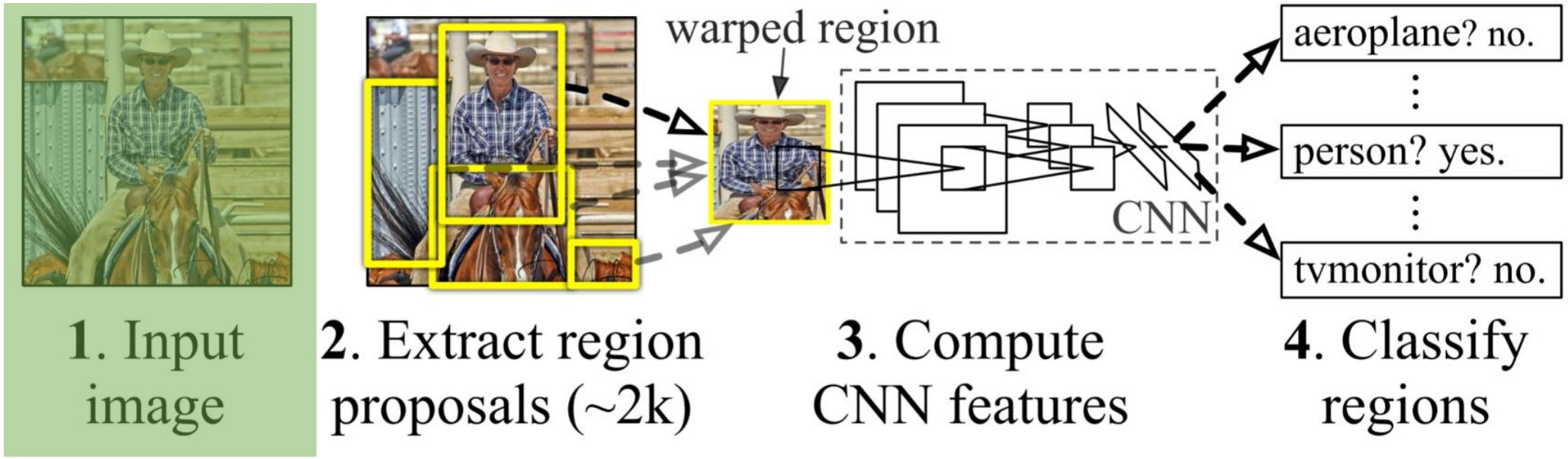
Region-Based CNNs (R-CNNs)

- **R-CNN**
  - Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580–587).
- **Fast R-CNN**
  - Girshick, R. (2015). Fast r-cnn. *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448).
- **Faster R-CNN**
  - Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. *Advances in neural information processing systems* (pp. 91–99).
- **Mask R-CNN**
  - He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Region-Based CNNs (R-CNNs)

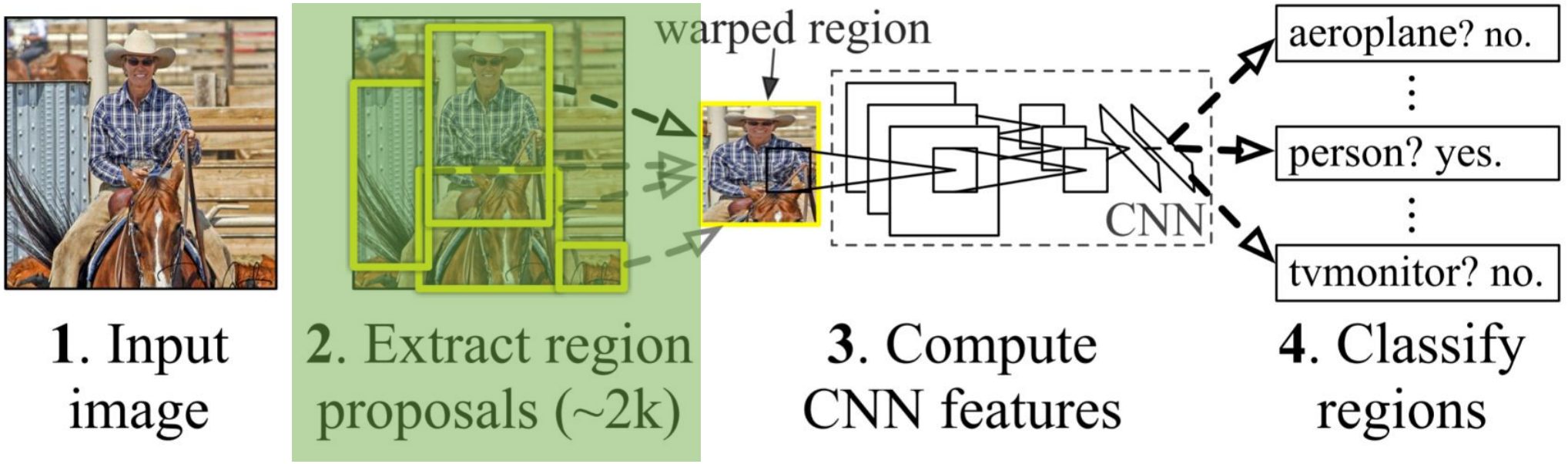
- **R-CNN - 2014**
- **(1) takes an input image**



Region-Based CNNs (R-CNNs)

- **R-CNN - 2014**

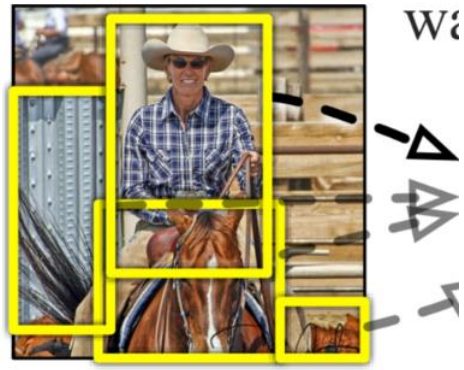
- (1) takes an input image
- **(2) extracts around 2000 bottom-up regions using selective search**
  - J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013



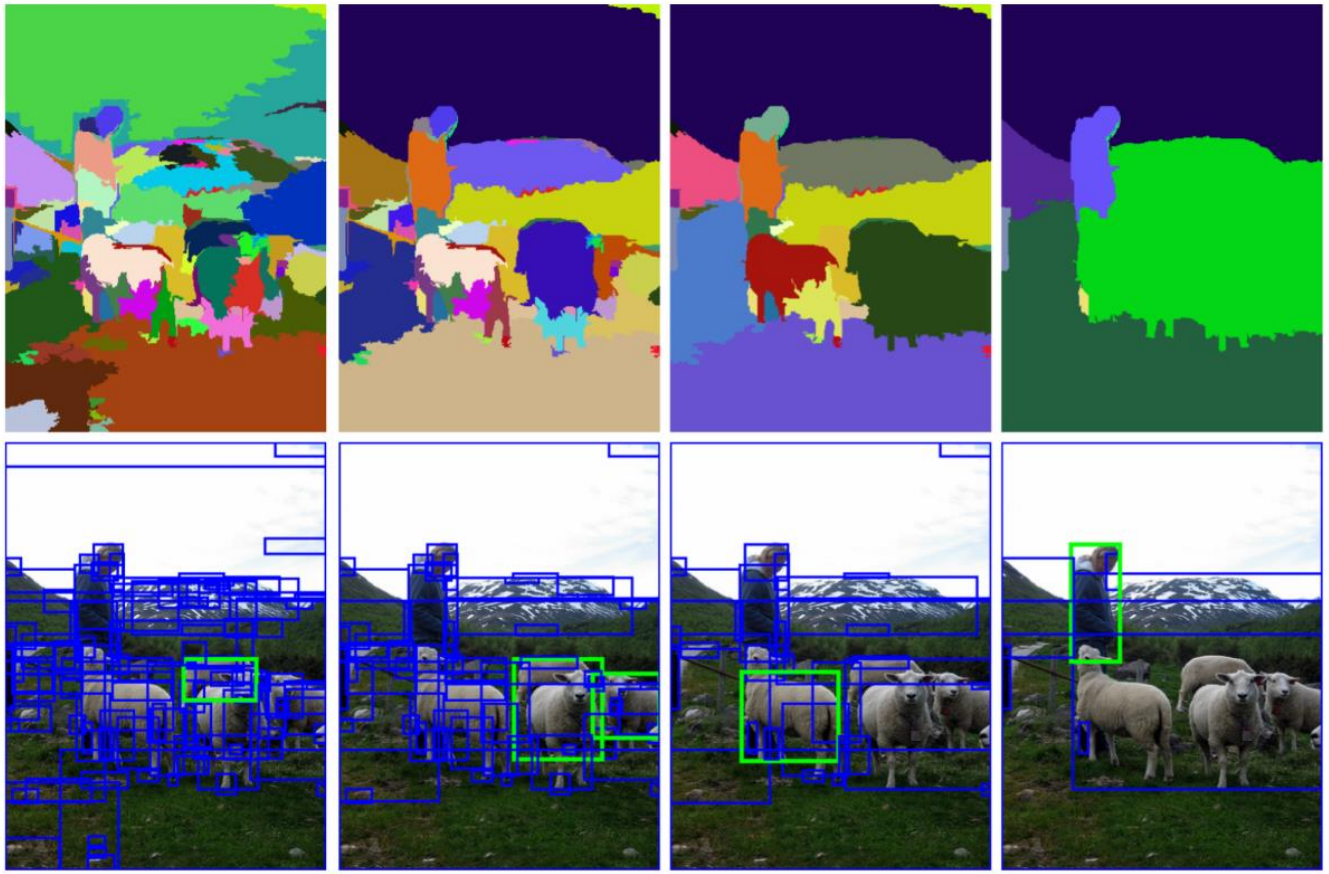
Region-Based CNNs (R-CNNs)

- **R-CNN - 2014**

- (1) takes an input image
- **(2) extracts around 2000 bottom-up regions using selective search**
  - J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013



2. Extract region proposals (~2k)

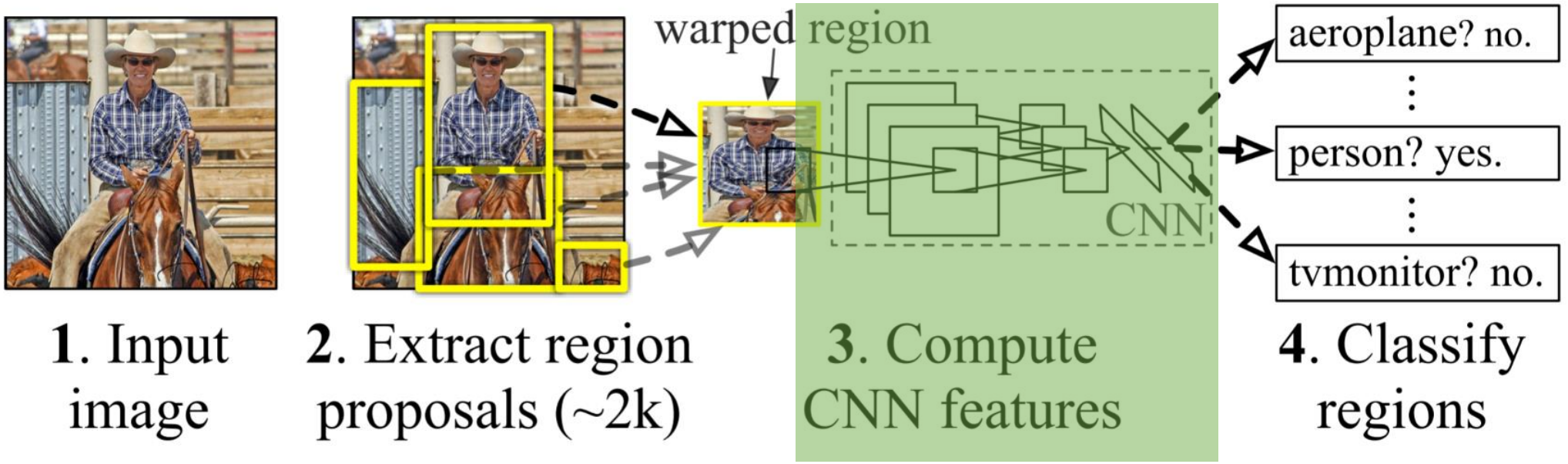


EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Region-Based CNNs (R-CNNs)

- **R-CNN - 2014**

- (1) takes an input image
- (2) extracts around 2000 bottom-up regions using selective search
  - J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013
- **(3) computes features for each region using a large convolutional neural network (CNN)**
  - AlexNet is used to compute the features

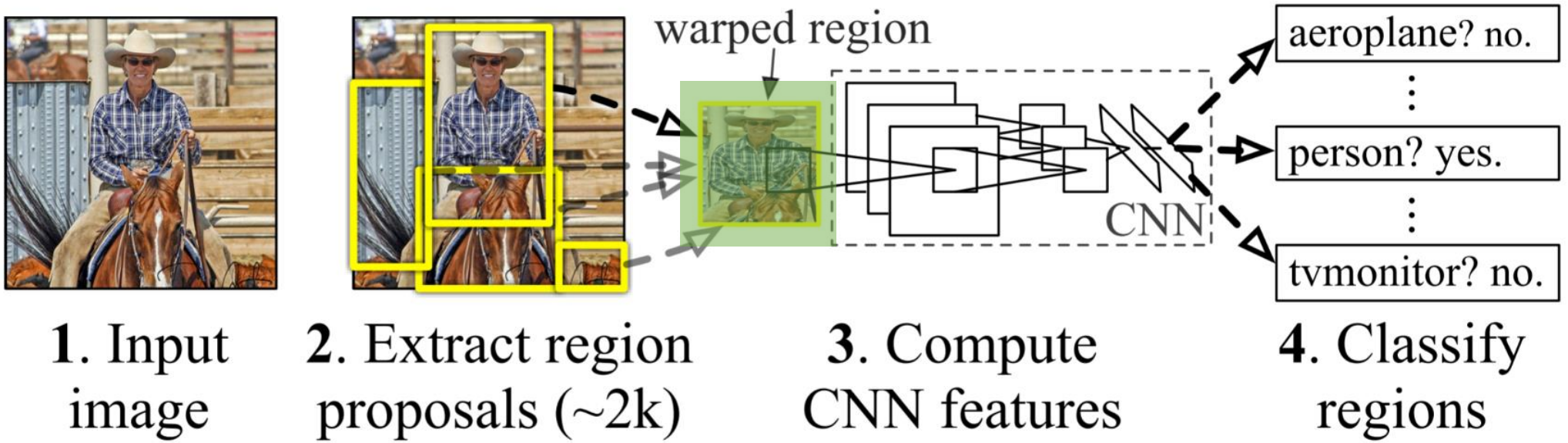


EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Region-Based CNNs (R-CNNs)

- **R-CNN - 2014**

- (1) takes an input image
- (2) extracts around 2000 bottom-up regions using selective search
  - J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013
- **(3) computes features for each region using a large convolutional neural network (CNN)**
  - AlexNet is used to compute the features **(227×227 pixels)**

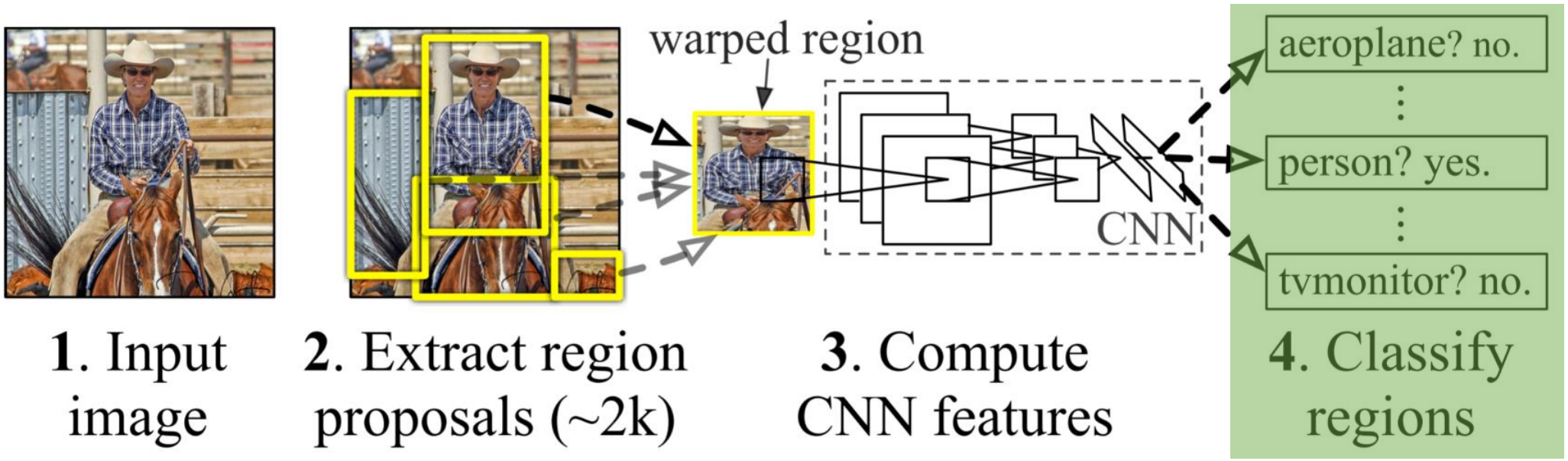


EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Region-Based CNNs (R-CNNs)

- **R-CNN - 2014**

- (1) takes an input image
- (2) extracts around 2000 bottom-up regions using selective search
  - J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013
- (3) computes features for each region using a large convolutional neural network (CNN)
  - AlexNet is used to compute the features (227×227 pixels)
- **(4) classifies each region using class-specific linear SVMs**



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Region-Based CNNs (R-CNNs)

- **R-CNN - 2014**

- (1) takes an input image
- (2) extracts around 2000 bottom-up regions using selective search
  - J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013
- (3) computes features for each region using a large convolutional neural network (CNN)
  - AlexNet is used to compute the features (227×227 pixels)
- **(4) classifies each region using class-specific linear SVMs**