Convolutional neural networks
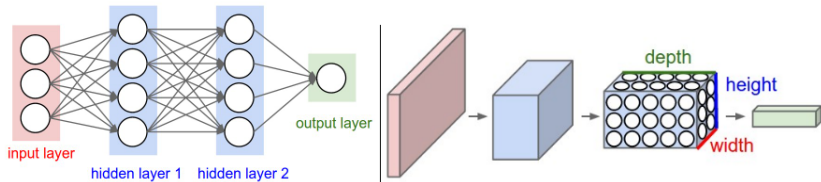
Mikhail Yurushkin
m.yurushkin@broutonlab.com

October 19, 2022
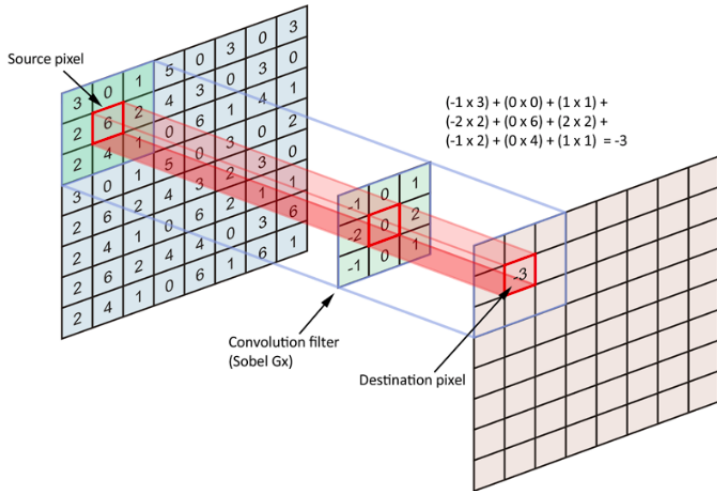

https://broutonlab.com

Regular neural networks and ConvNets

- ▶ Regular Neural Nets don't scale well to full images.
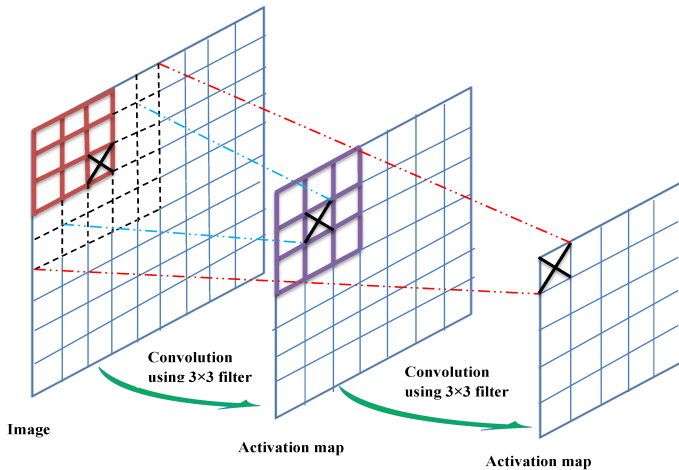- ▶ ConvNet have neurons arranged in 3 dimensions: width, height, depth

Convolutional Layer

Source pixel

$(-1 \times 3) + (0 \times 0) + (1 \times 1) +$
$(-2 \times 2) + (0 \times 6) + (2 \times 2) +$
$(-1 \times 2) + (0 \times 4) + (1 \times 1) = -3$
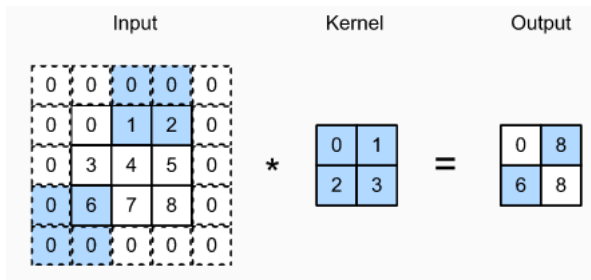
Convolution filter
(Sobel Gx)

Destination pixel

Parameters

- ▶ Kernel Size - is a filter applied in slidding window.
- ▶ Stride - controls how the filter convolves around the input volume
- ▶ Depth - controls number of neurons in a layer that connect to the same region of the input volume
- ▶ Padding - controls the output volume spatial size

Receptive field

Image

Convolution using 3×3 filter

Activation map

Convolution using 3×3 filter

Activation map

Striding

Input   Kernel   Output

▶ strides of 3 and 2 for height and width, respectively.

Padding

- ▶ helps to preserve size of image
- ▶ can be used for segmentation

Spatial Pooling layers (Max/Average)

Max(1, 1, 5, 6) = 6
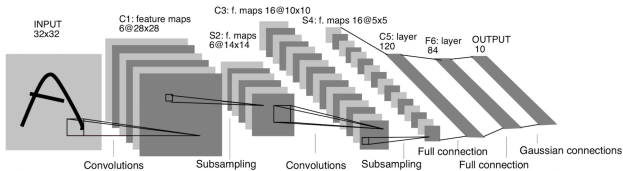
max pool with 2x2 filters
and stride 2

Rectified Feature Map
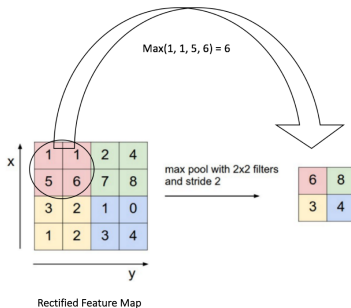
▶ Reduces the dimensionality of each feature map
▶ Spatial Pooling can be of different types: Max, Average, Sum etc.
▶ Seems to be will be discarded in the future (see Striving for Simplicity: The All Convolutional Net'14).

Speed & size bottlenecks

conv (180w + 5b)
maxpool
conv (450w + 10b)
non-linear
non-linear
maxpool
non-linear
fully-connected
(1600w + 10b)
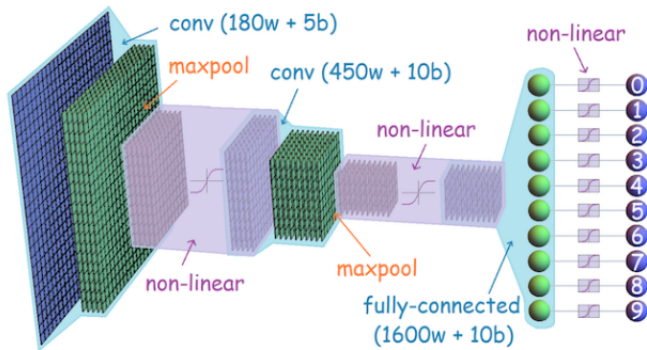
1. Backbone - speed issues
2. head contains 99% of weights
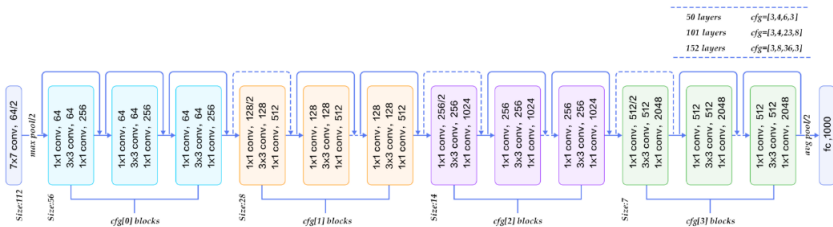
LeNet (1990s)

1. Very first convolutional neural network
2. Classifies images of 10 classes (dog, cat, bird etc)
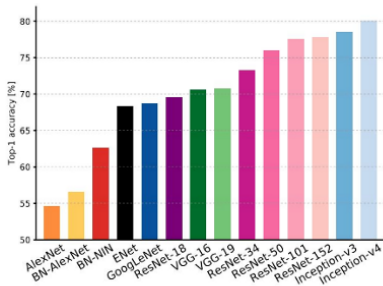
AlexNet (2012)
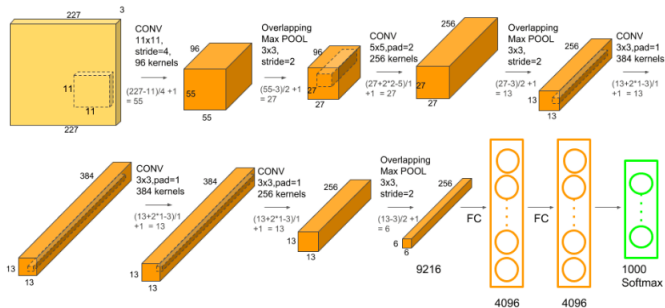
1. won the ImageNet challenge in 2012
2. deep CNN trained on ImageNet and outperformed all the entries that year
3. Compared to modern architectures, a relatively simple layout

ResNet (2015)

1. residual blocks usage
2. does not overfit with increasing of layers number
3. took three weeks to train it on an 8 GPU machine

Comparison of state of art architectures
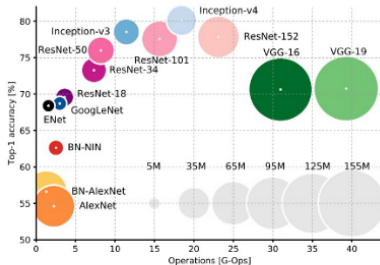
An Analysis of Deep Neural Network Models for Practical Applications, 2017.

Comparison of state of art architectures

Year	CNN	Developed by	Place	Top-5 error rate	No. of parameters
1998	LeNet(8)	Yann LeCun et al			60 thousand
2012	AlexNet(7)	Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever	1st	15.3%	60 million
2013	ZFNet()	Matthew Zeiler and Rob Fergus	1st	14.8%	
2014	GoogLeNet(19)	Google	1st	6.67%	4 million
2014	VGG Net(16)	Simonyan, Zisserman	2nd	7.3%	138 million
2015	ResNet(152)	Kaiming He	1st	3.6%	

Trainable filters
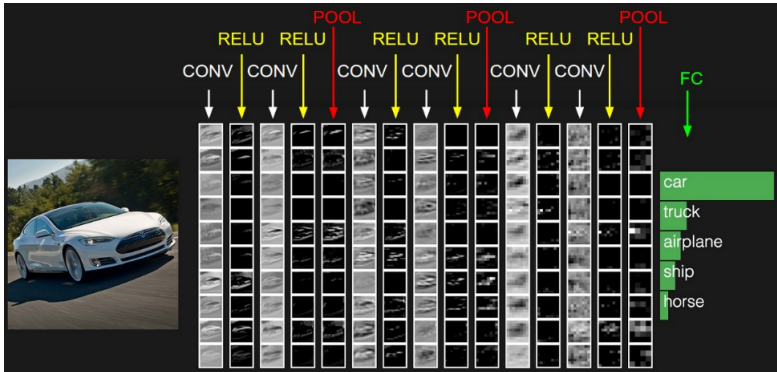
Benchmarks and datasets
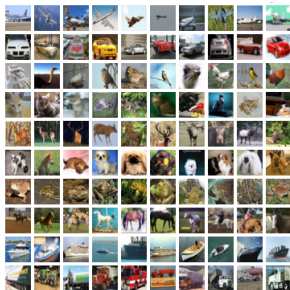
- ▶ PASCAL VOC, COCO, ImageNet, CIFAR (2D classification, object detection)
- ▶ KITTI Vision Benchmark (tereo, optical flow, visual odometry, 3D object detection and 3D tracking)
- ▶ CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.