Implementing Artificial Neural Networks (ANNs) with TensorFlow

Session 1: Introduction

University of Osnabrück
Institute of Cognitive Science

WS 2016 / 17

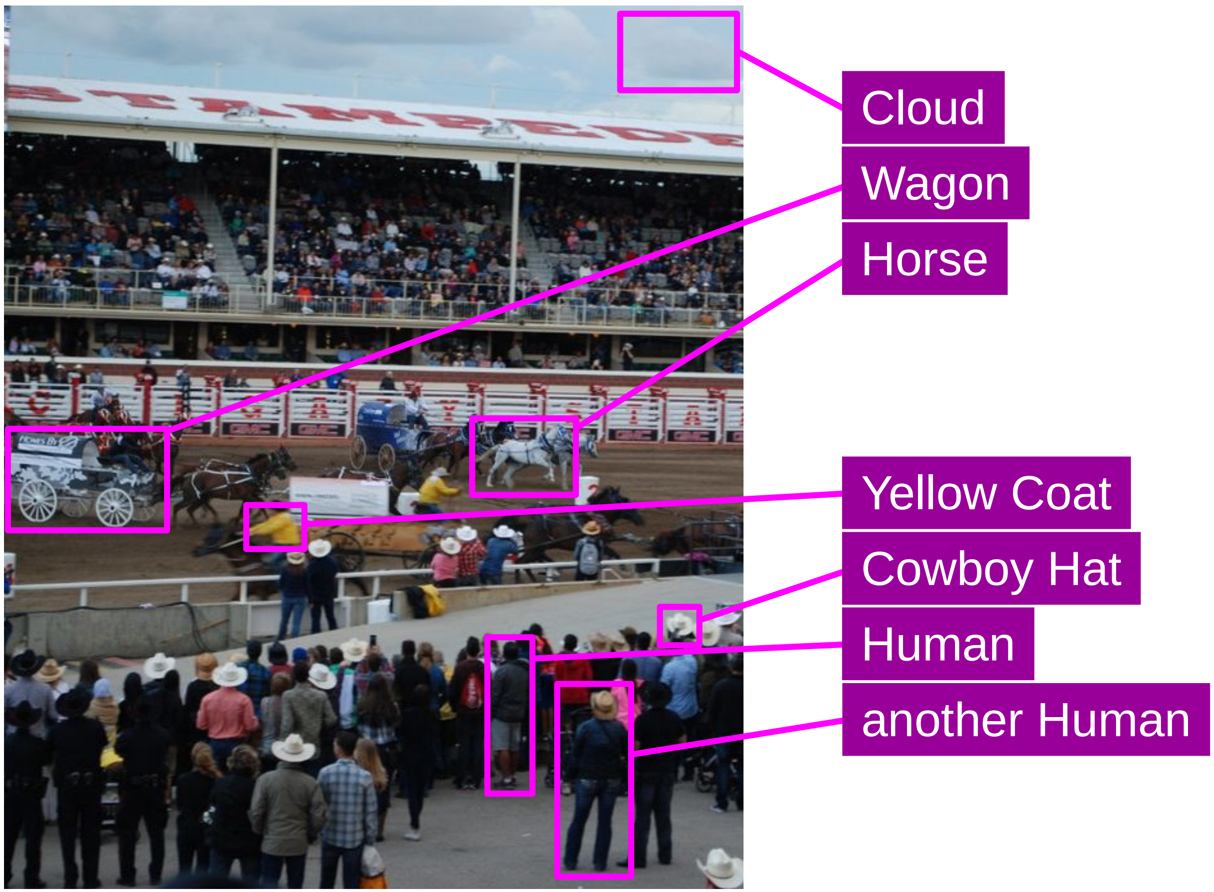
Lukas Braun, lbraun@uos.de

The Human Brain

The human brain

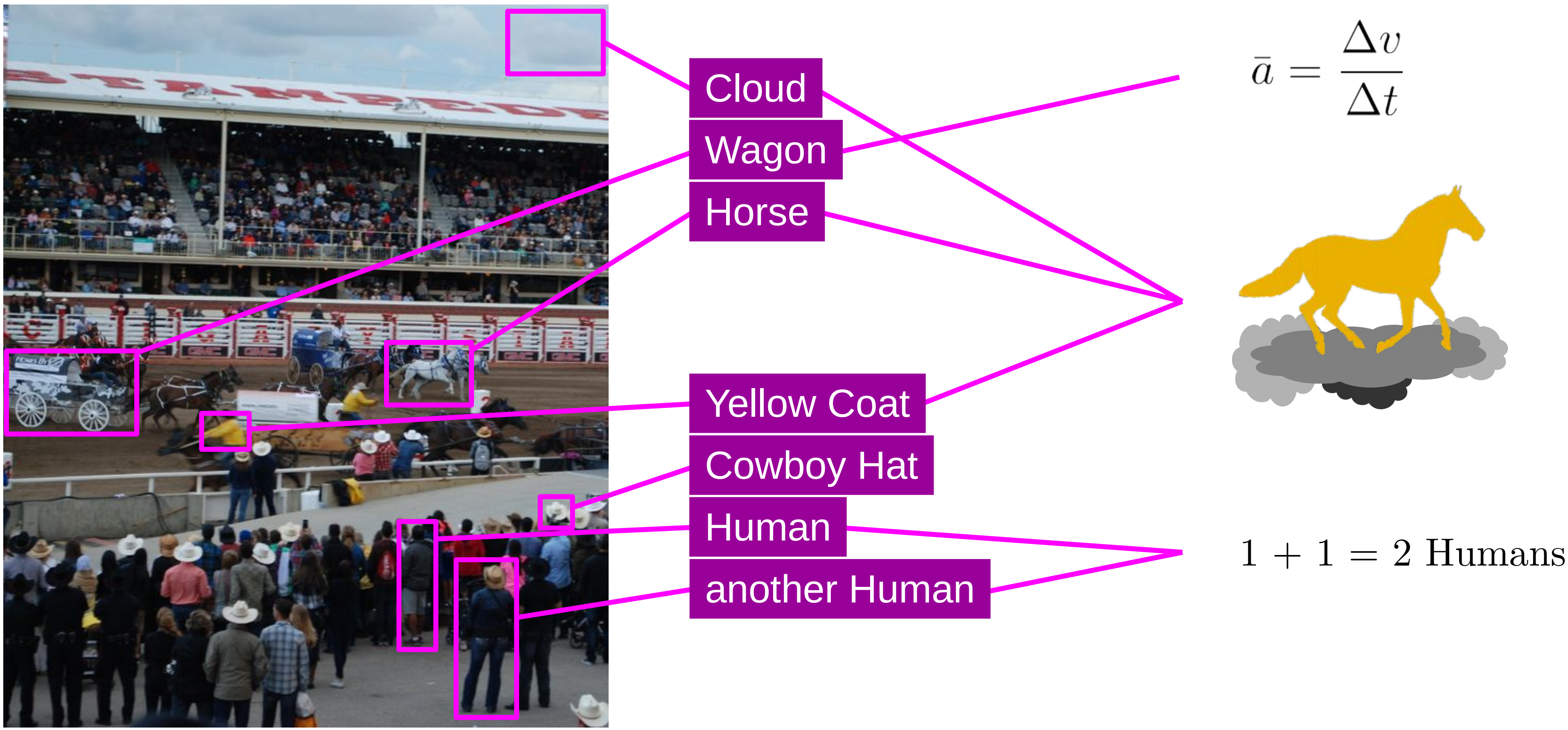
Incredible capabilities: detect, categorize, learn, memorize, navigate, imagine, predict, plan, speech ...

The human brain



"A crowd is watching the start of a wagon-race. It seems to be cold and cloudy.

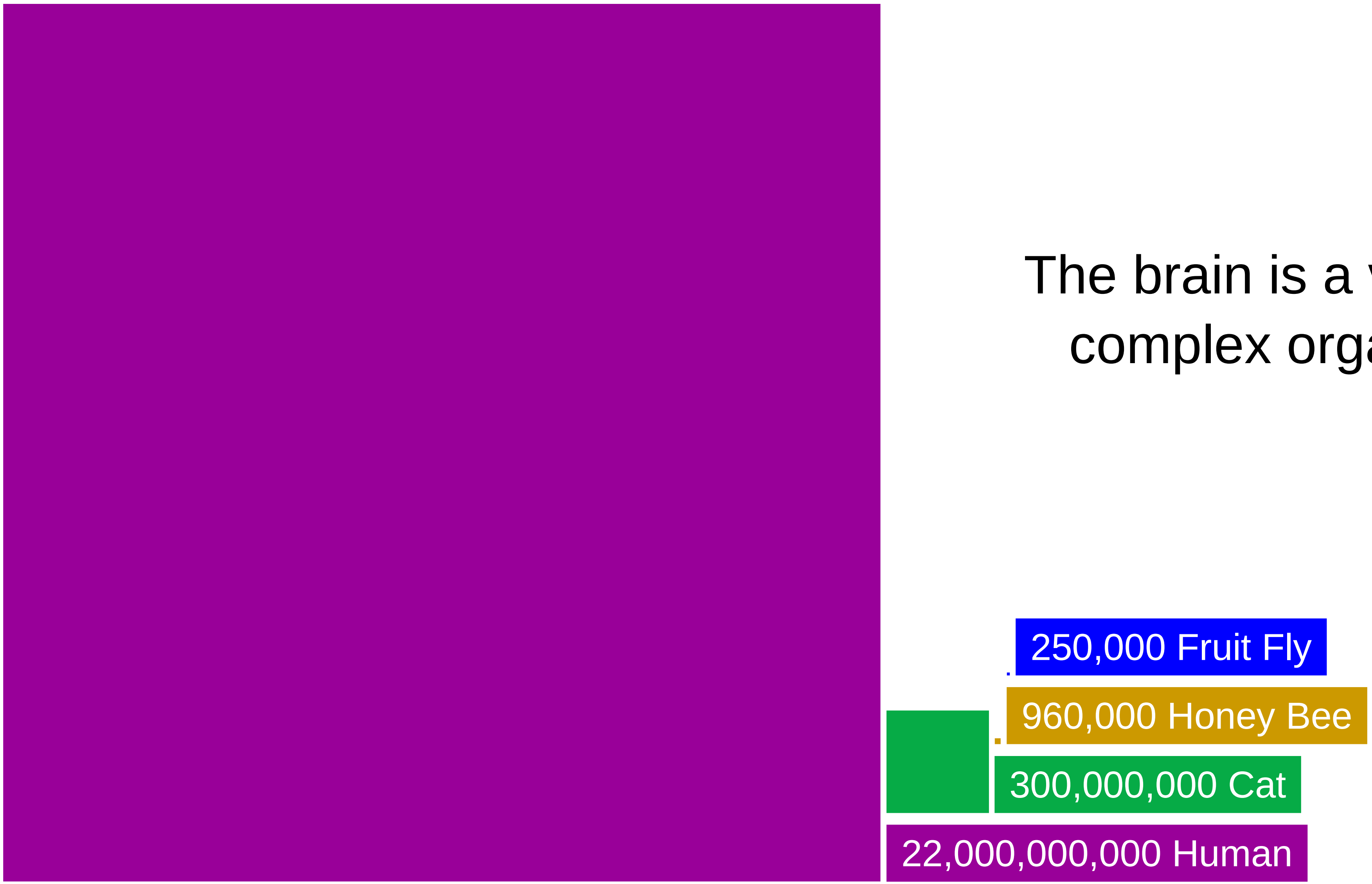
The human brain



Biological Neural Networks

Neurons in the cortex

The brain is a very complex organ



Fruit fly

250,000 Neurons

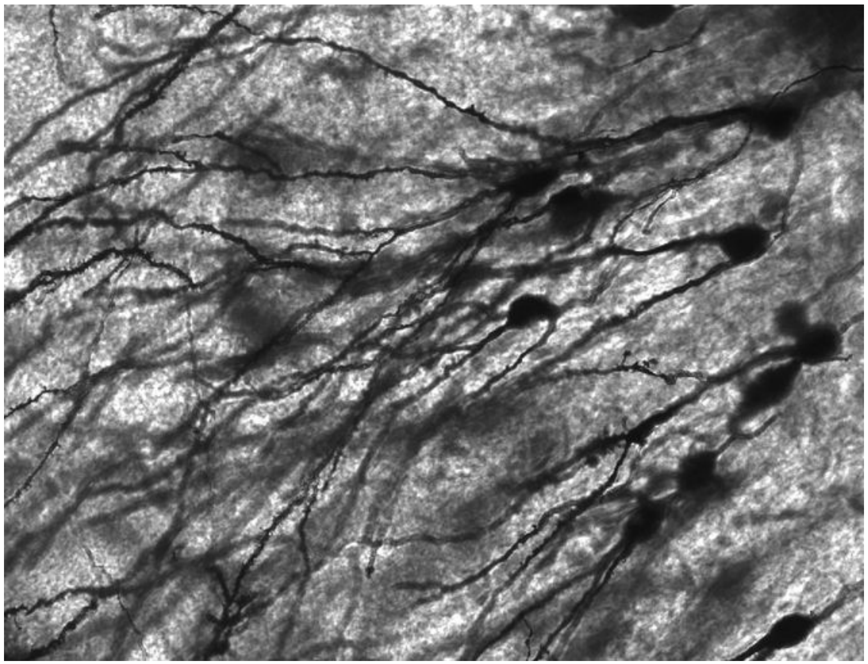
- Navigate by walking and flying
- Recognize and find food
- Find mates and reproduce
- Learn ...

[1]

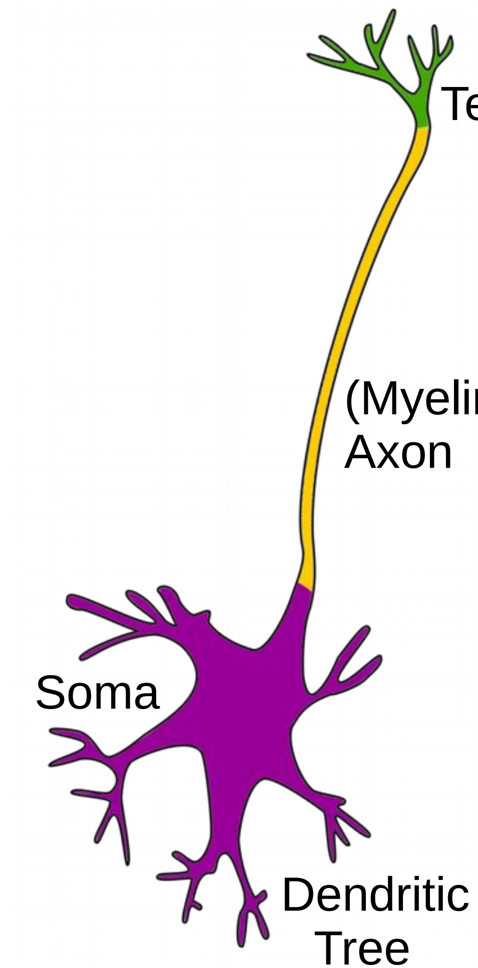
DARPA Robot Challenge 2015

[1] youtube.com

Neurons

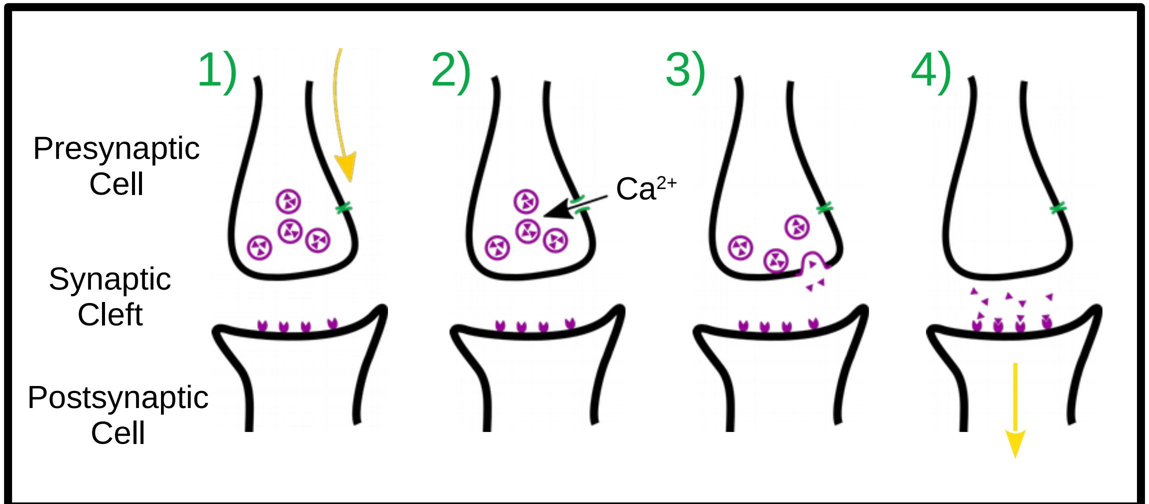
 [2]

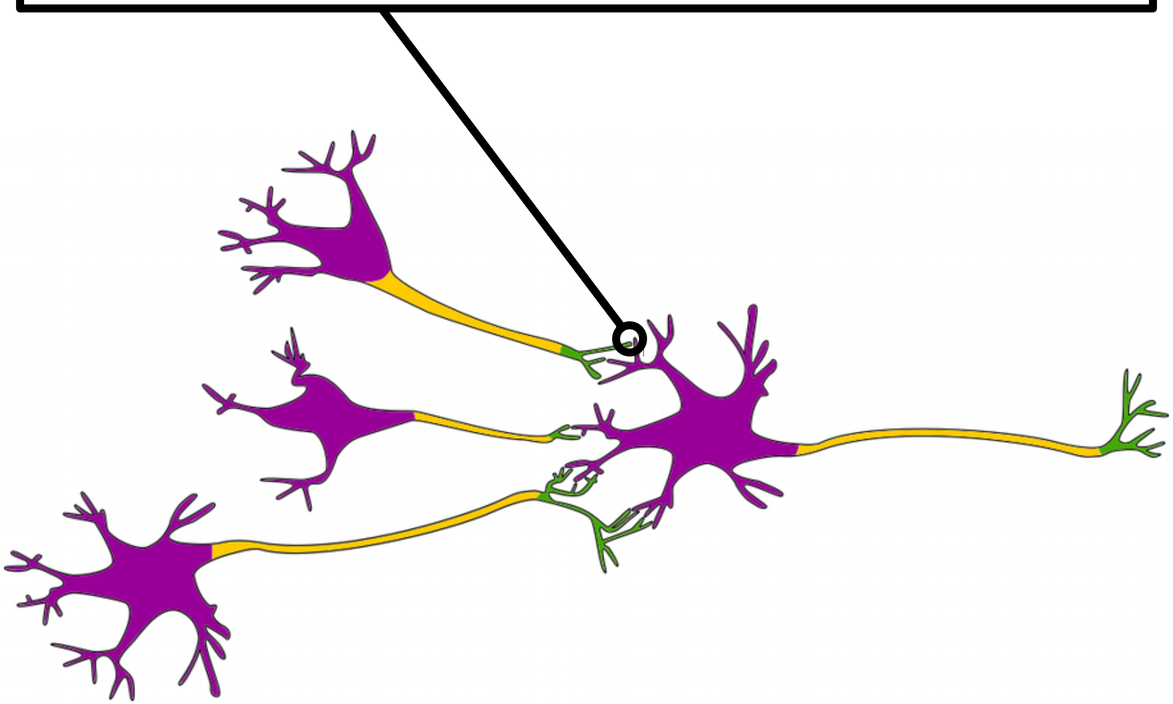
Golgi-stained neurons from human hippocampus



[2] *commons.wikipedia.org*

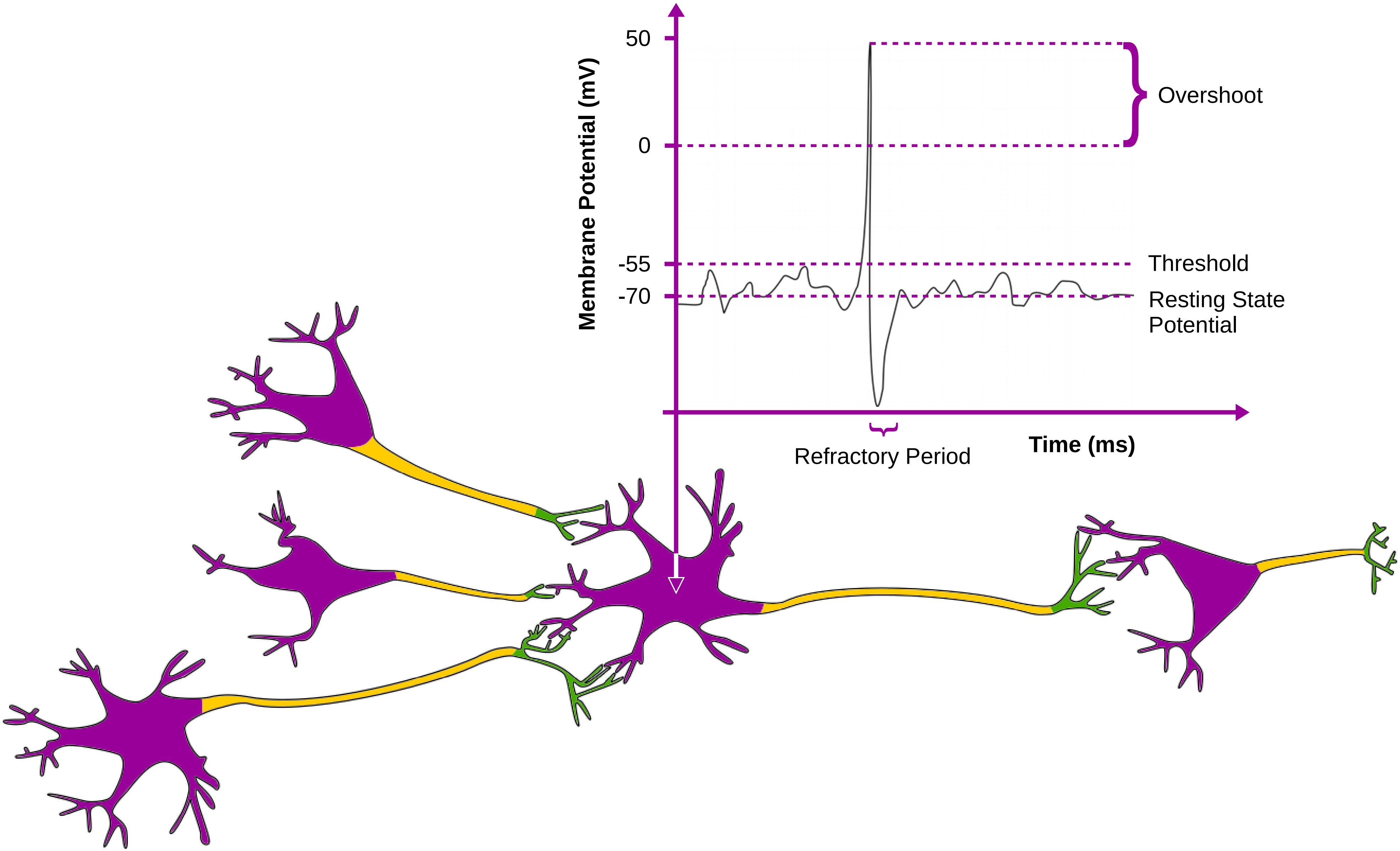
Neurons





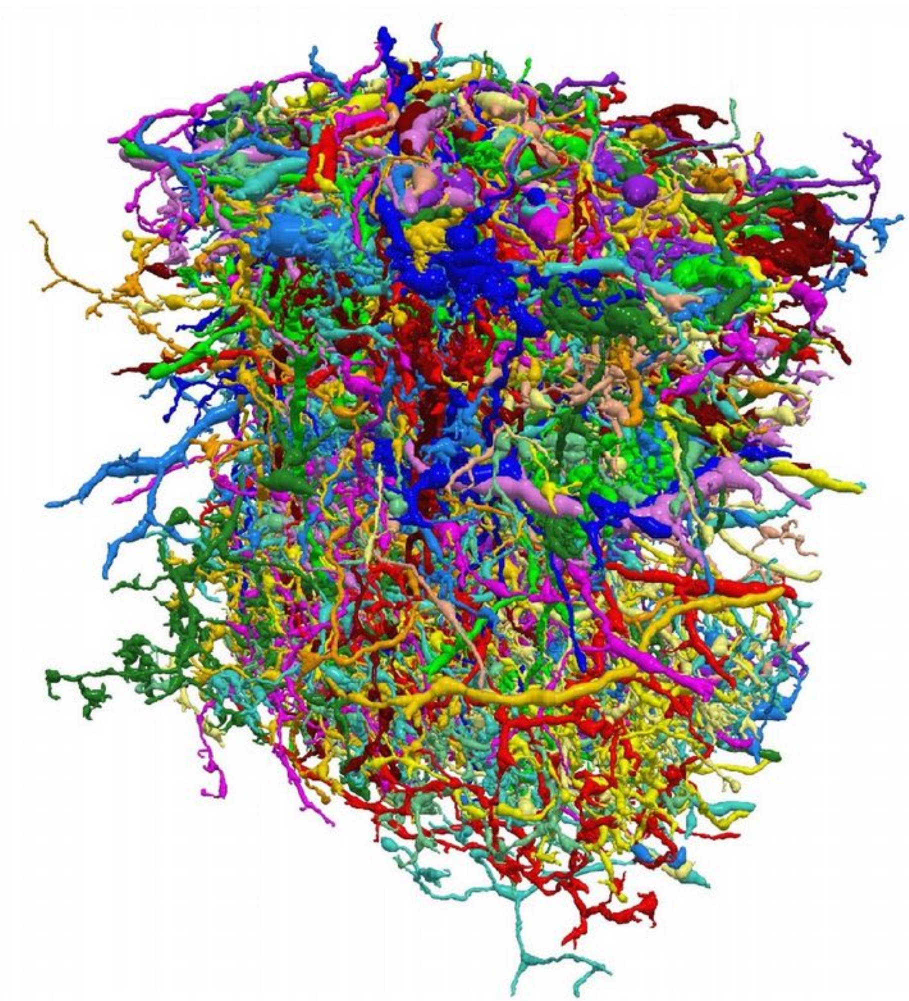
1) Action potential arrives at axon terminal
2) Voltage-gated Ca^{2+} channel open
3) Ca^{2+} signals to vesicles. Vesicles move to membrane and release neuro-transmitter
4) Neurotransmitter dock at receptor and induce an excitatory / inhibitory postsynaptic potential

Neurons



Membrane Potential (mV)
50
0
-55
-70
Overshoot
Threshold
Resting State Potential
Refractory Period
Time (ms)

Biological Neural Networks



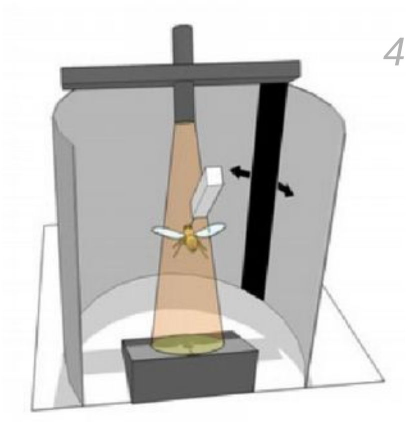
3

Biological neural networks are incredibly complex!

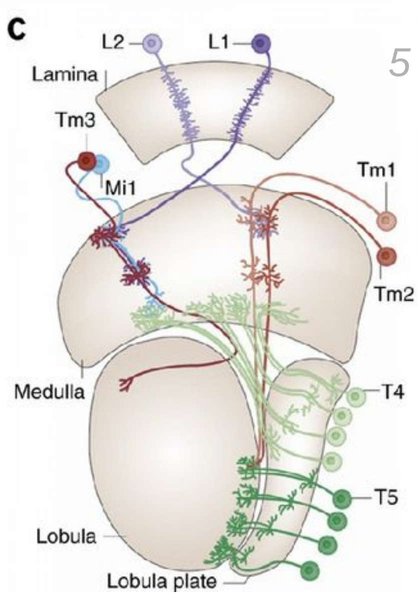
Reconstruction of 379 neurons of the optic lobe in the fruit fly

[3] *technologyreview.com*

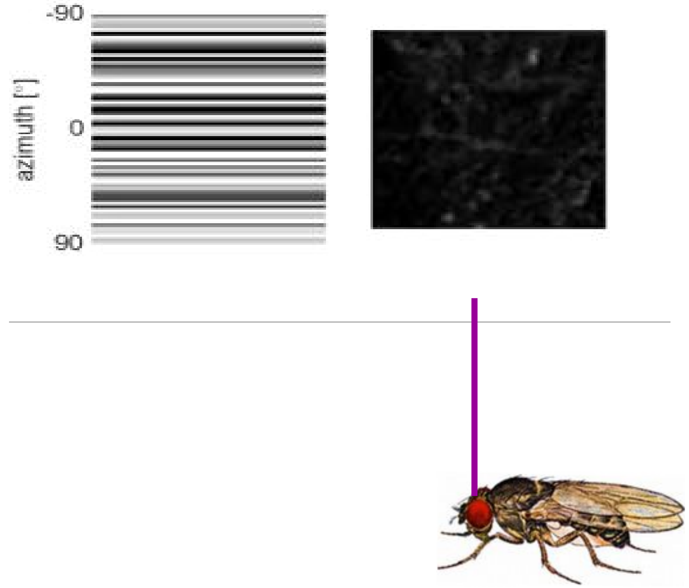
Motion vision in the fruit fly



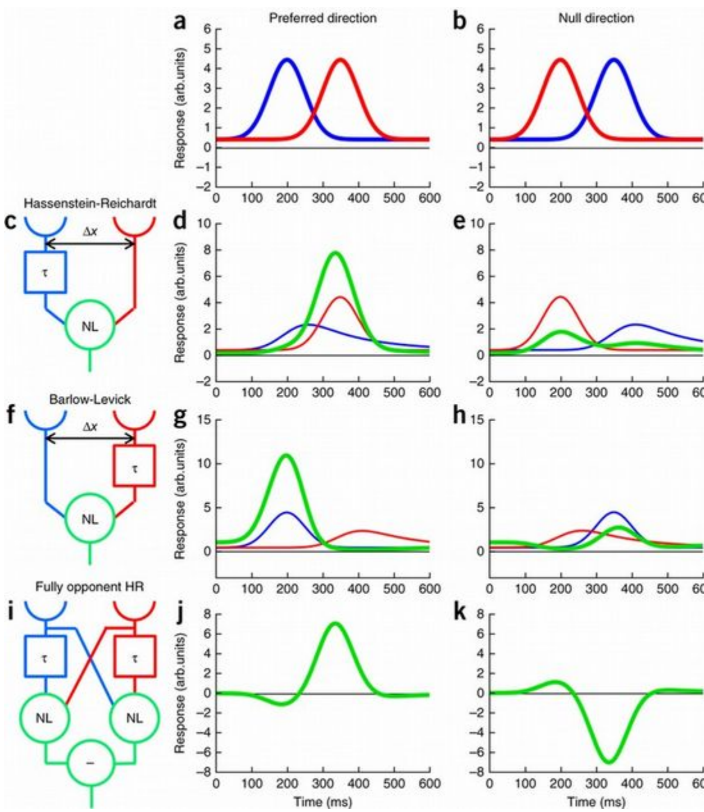
[4]



[5]



[6]

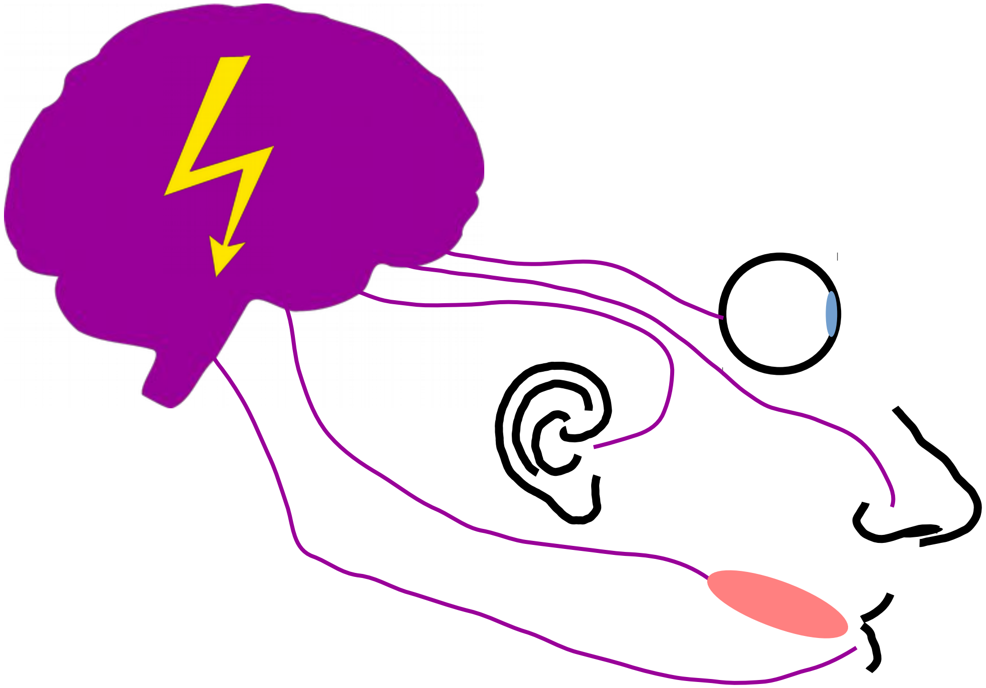


[5]

[4] *Schilling et al. 2015*
[5] *Borst et al. 2015*
[6] *Drews 2016*

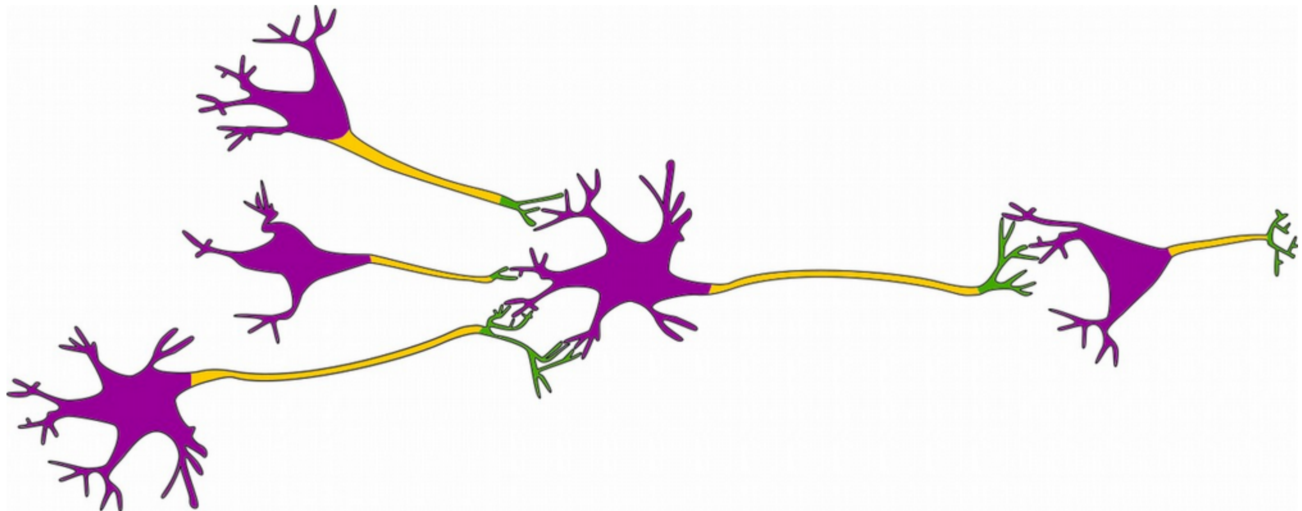
The human brain

In the brain it is all the same: electrical signals



Different modalities are encoded similarly

The Promise



Model

Speech
Text
Images
Videos
Labels
Entities
Words
Audio
Features

Input

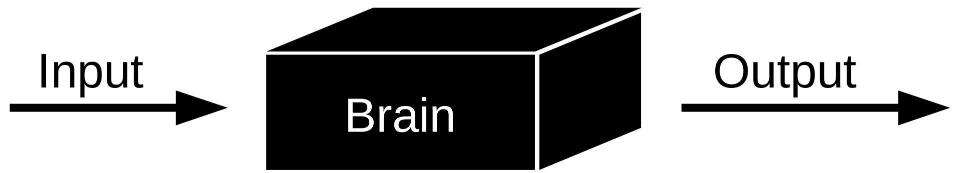
Simple and reconfigurable building blocks

Output

Speech
Text
Images
Videos
Labels
Entities
Words
Audio
Features

From Biological Neural Networks to Artificial Neural Networks

Blckbox model



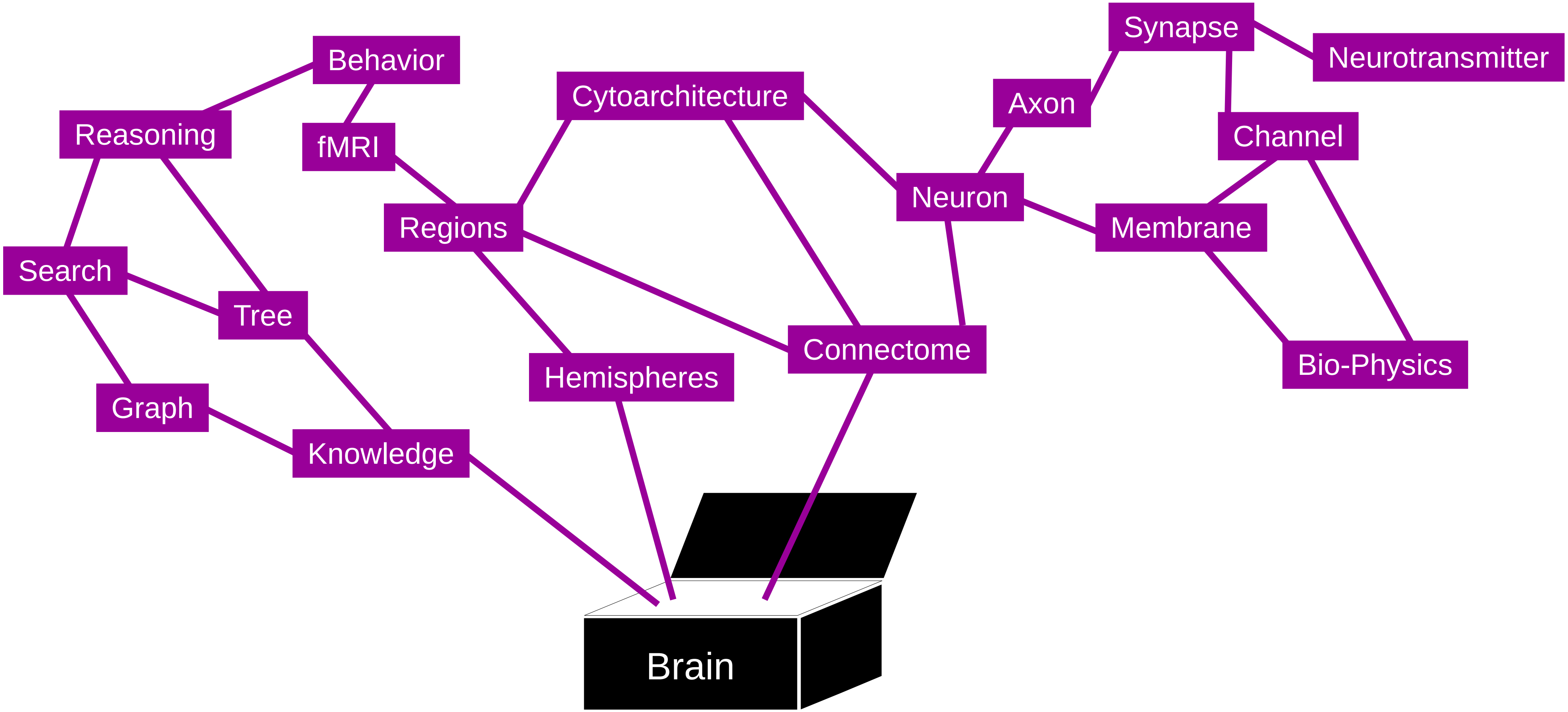
“A cute yawning cat!”

The brain as a function which maps inputs to outputs

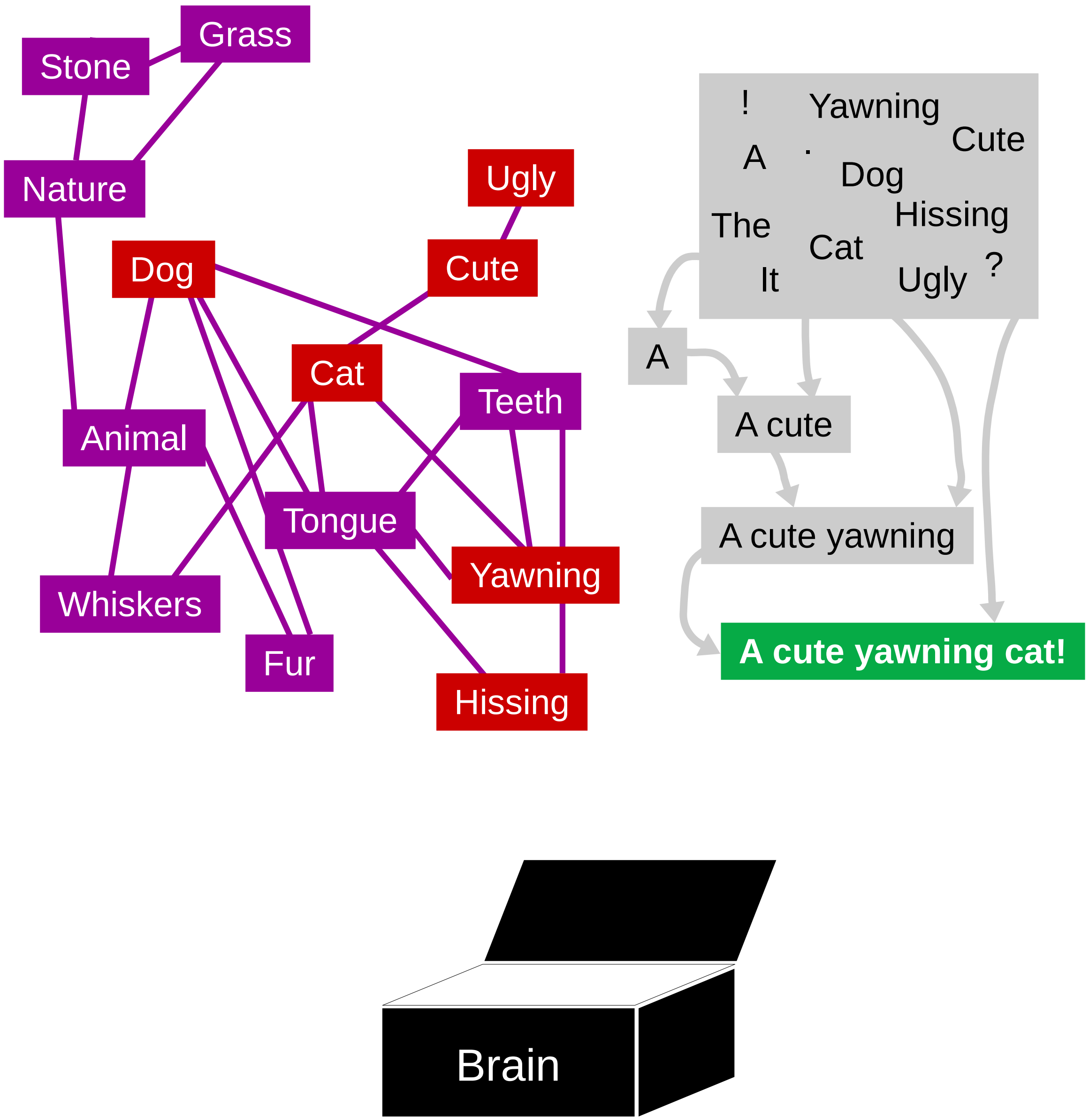
$$\text{“A Cute yawning cat!”} = \mathrm{f}(\quad)$$

Modeling

How to find this function and its parameters?



Modeling

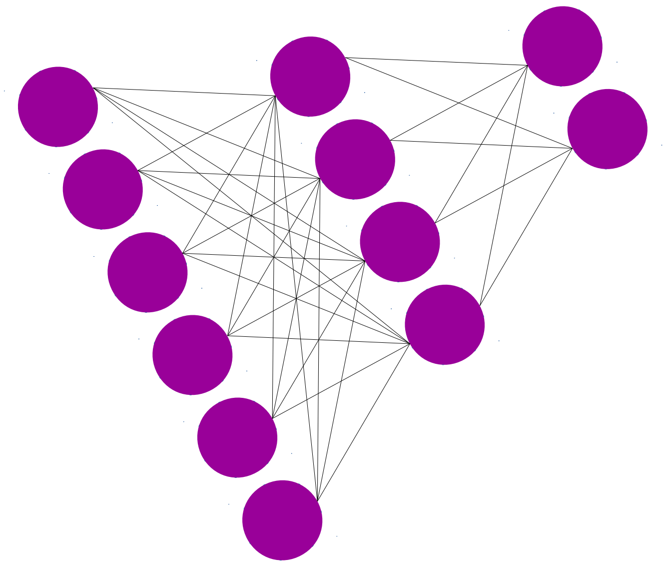


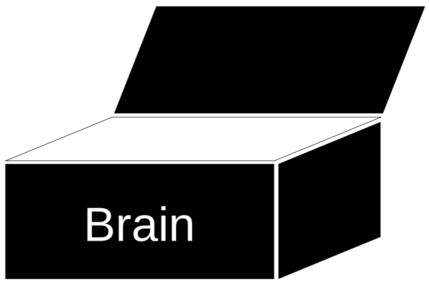
Classical AI: The black box as rule based symbol manipulation procedure

1) Convert input into symbolic representations
2) Search for and select candidates
3) Rule based production of output

Lukas Braun ❈ lbraun@uos.de

Modeling

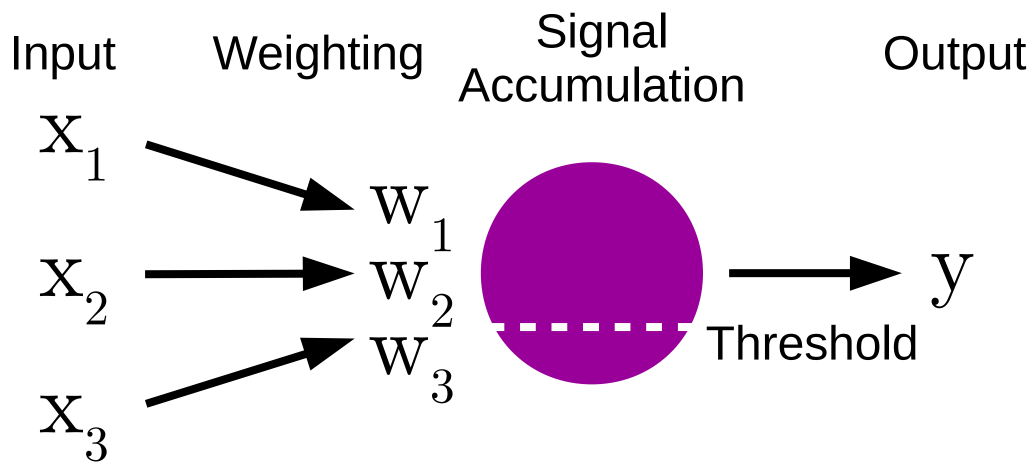




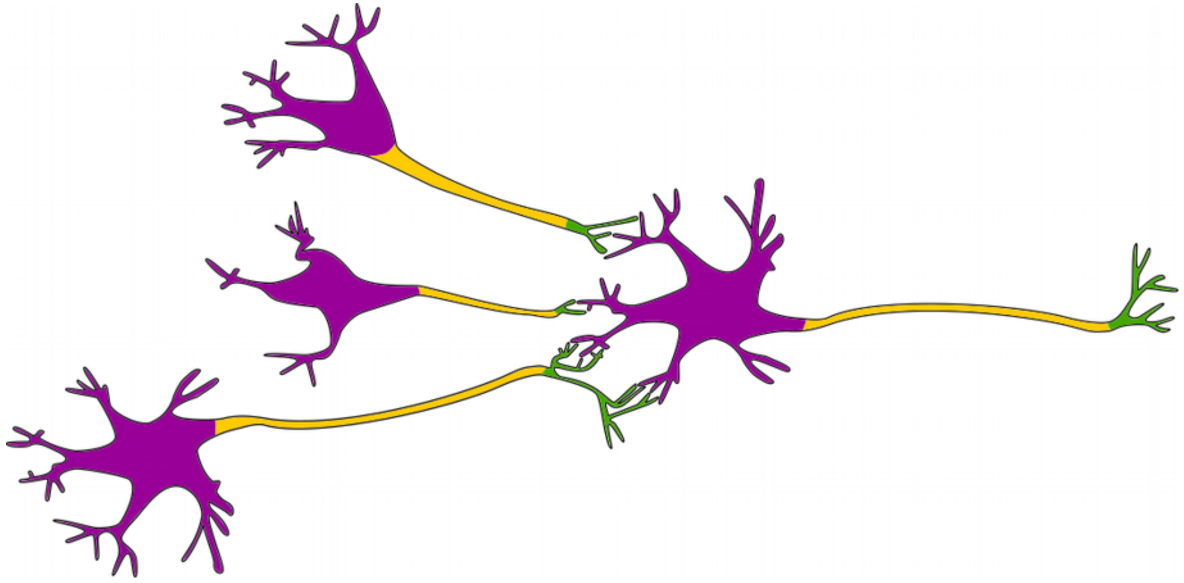
ANNs: The blackbox as abstraction of biological neural networks

1) “One kind of neuron”
2) Extremely simplified connectome
3) No bio-physical model (reservoir computing)
4) Stepwise processing

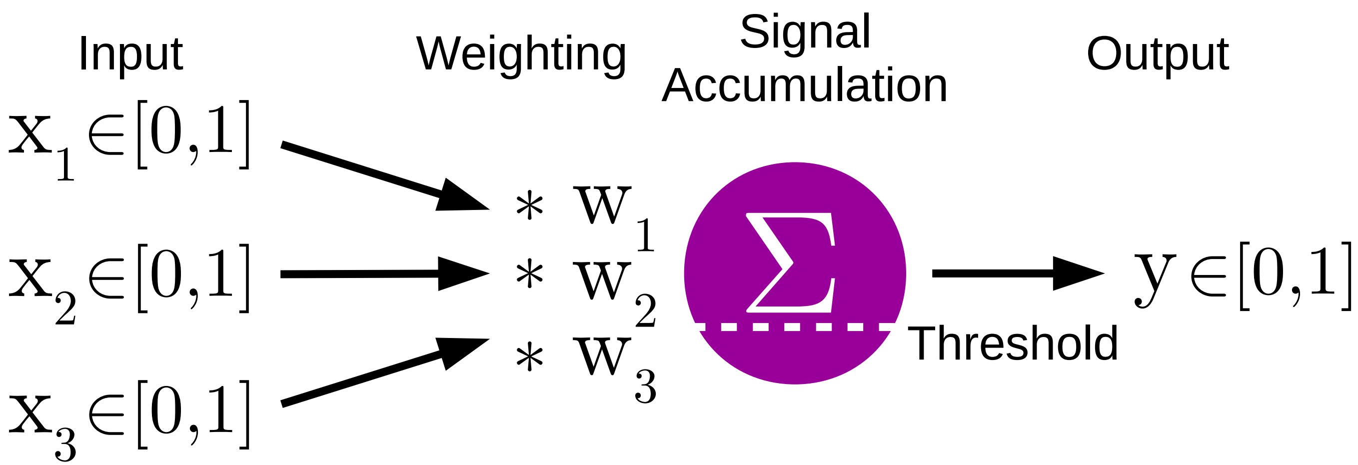
Perceptron



Input
Weighting
Signal Accumulation
Output
x_1
x_2
x_3
w_1
w_2
w_3
y
Threshold

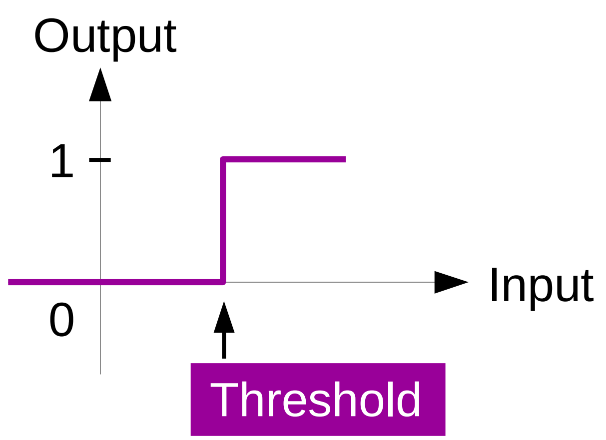


Perceptron



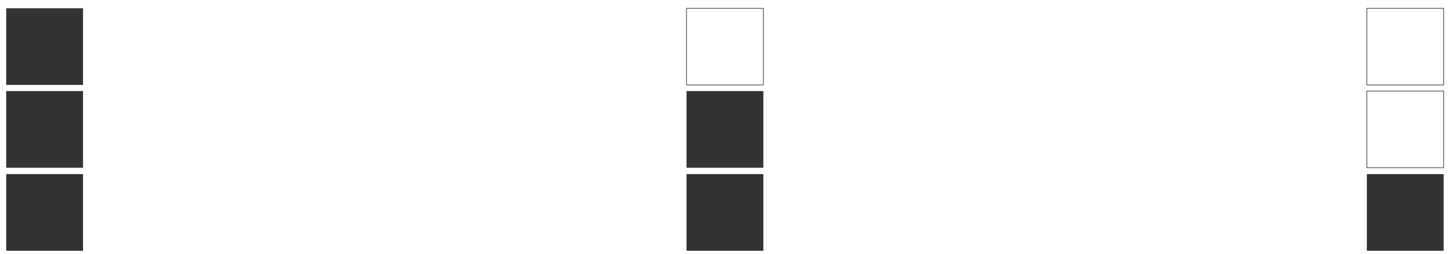
$$y = \sigma\left(\sum_i x_i w_i\right)$$

Step function



Example

Giraffe shade | Tiger shade | Frog shade



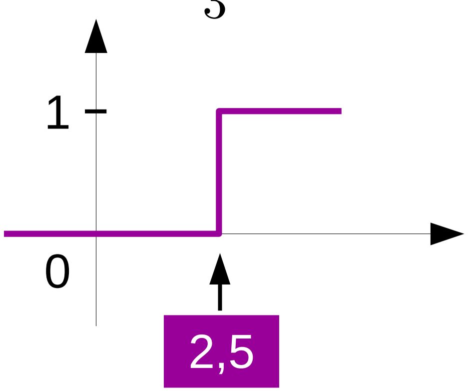
1*-1 + 1*2 + 1*1 = 2

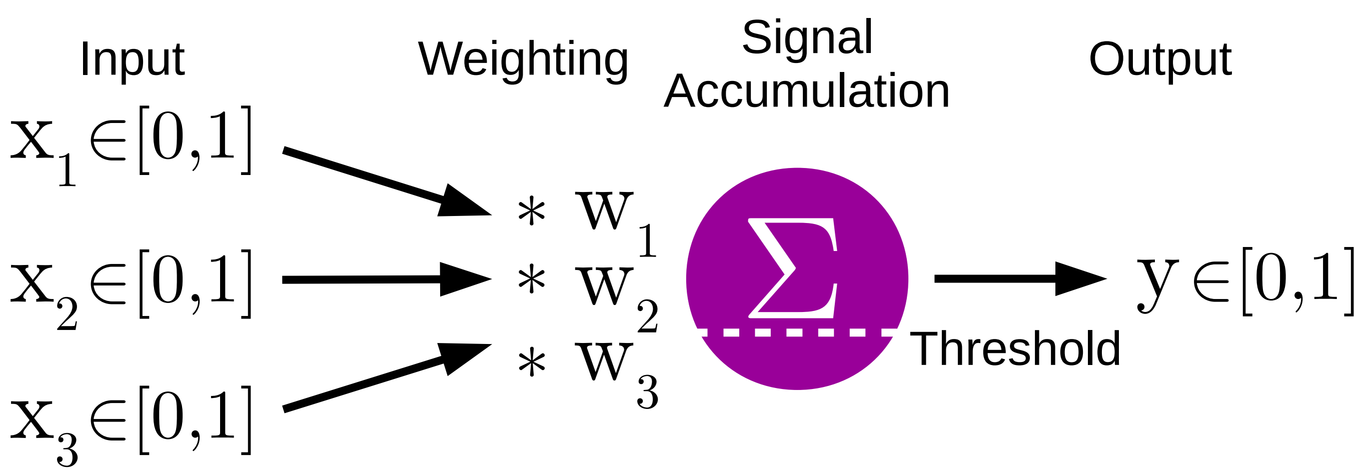
0*-1 + 1*2 + 1*1 = 3

0*-1 + 0*2 + 1*1 = 1

Solution

$$w_1 = -1$$
$$w_2 = 2$$
$$w_3 = 1$$

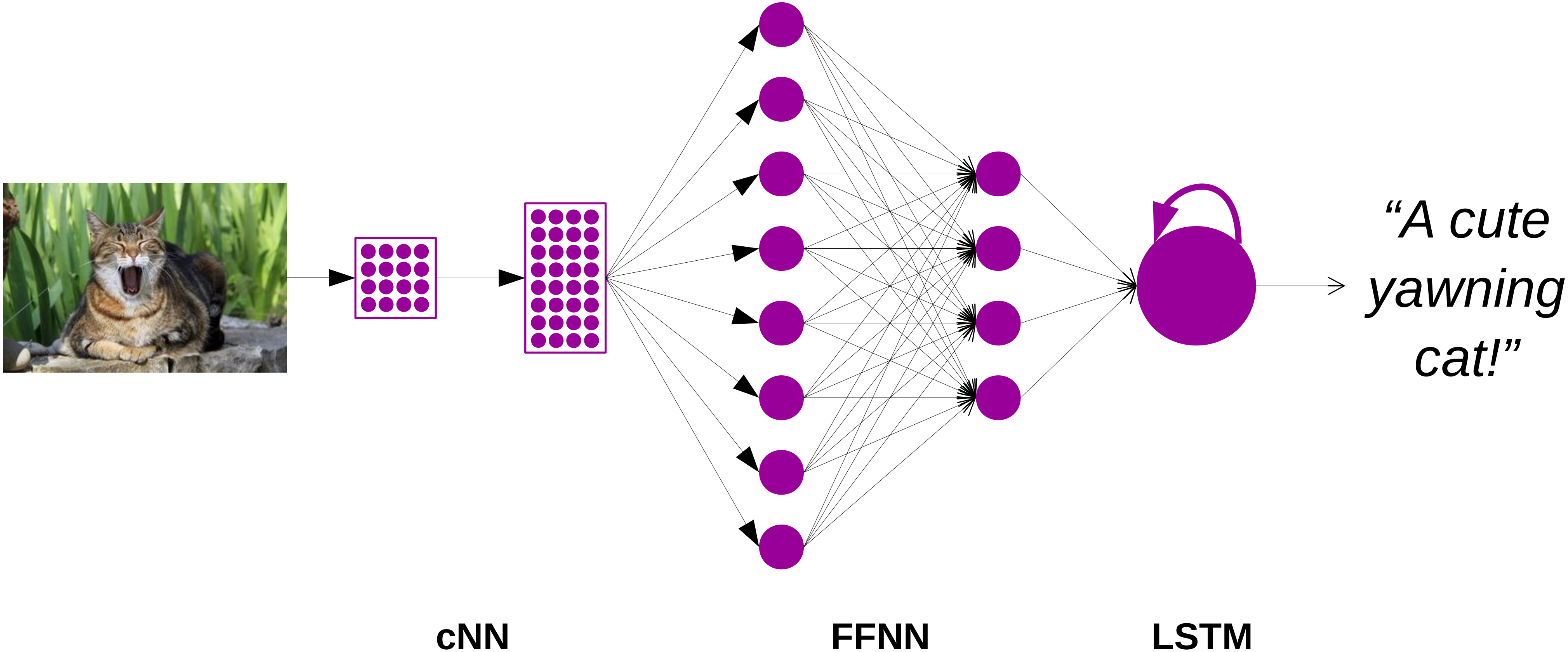




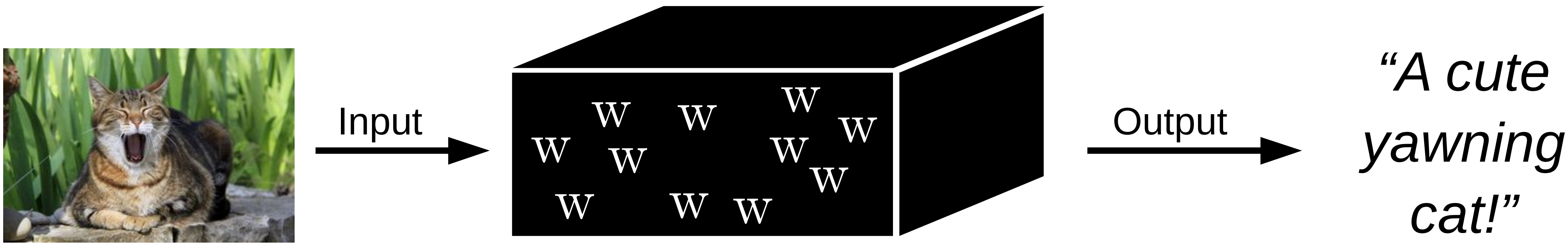
Deep Learning

Why is it deep?

Stacking of layers and eventually different network types



Blackbox model



$$y = f(x,w)$$

By finding the right weights we can find the function which maps the input to the output

However: The ANN will still be a blackbox

Backpropagation

- Artificial neural networks are function approximators
- Inputs are determined by environment
- Compare output to desired output and slowly adjust weights in order to decrease the error
- Backpropagation algorithm and Gradient Descent are used to find the right weights
- For supervised learning: We need labeled samples

Examples

7

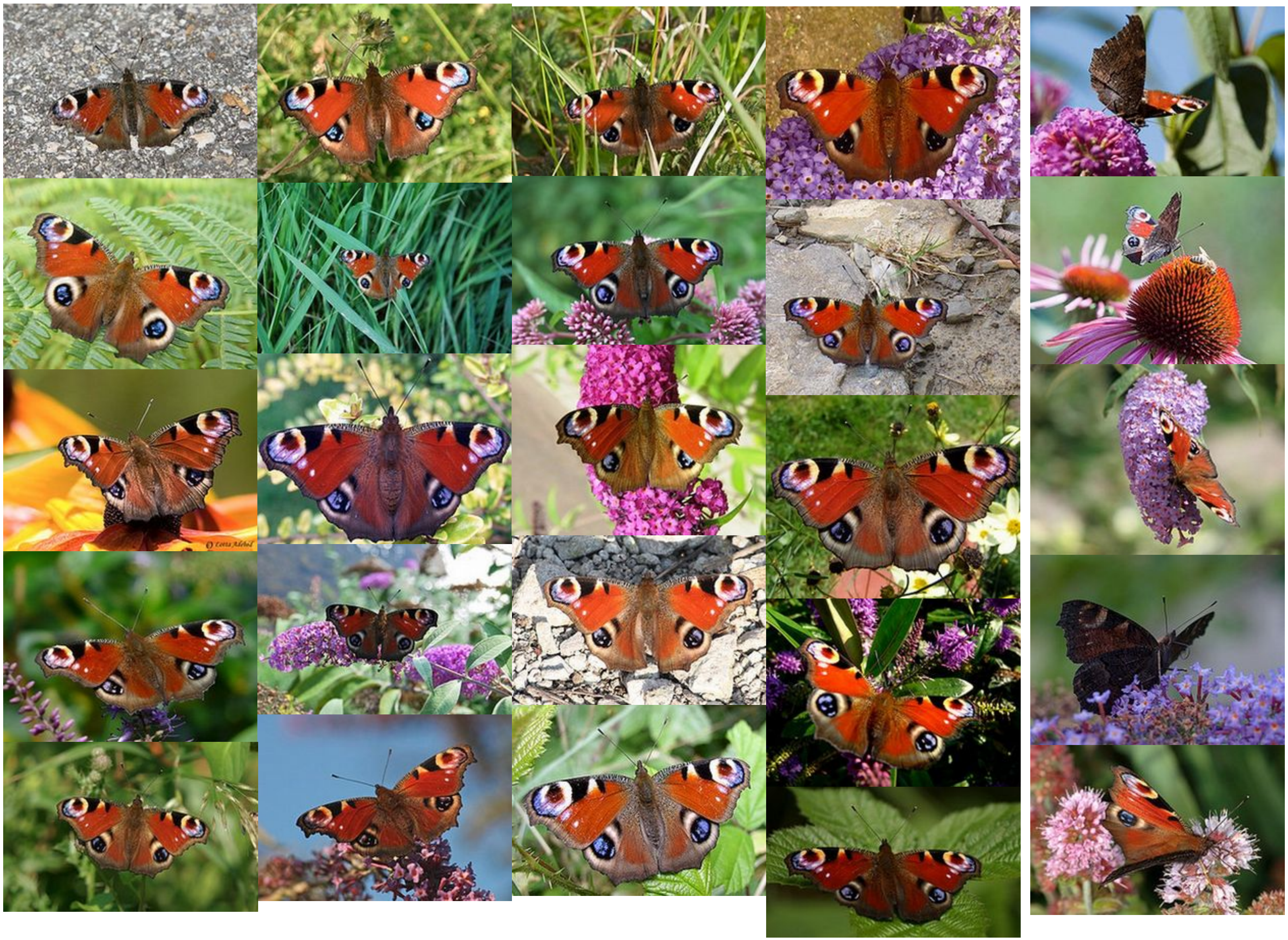
AlexNet

8

Neural Image Caption Generator

[7] Krizhevsky et al. 2012
[8] Vinyals et al. 2015

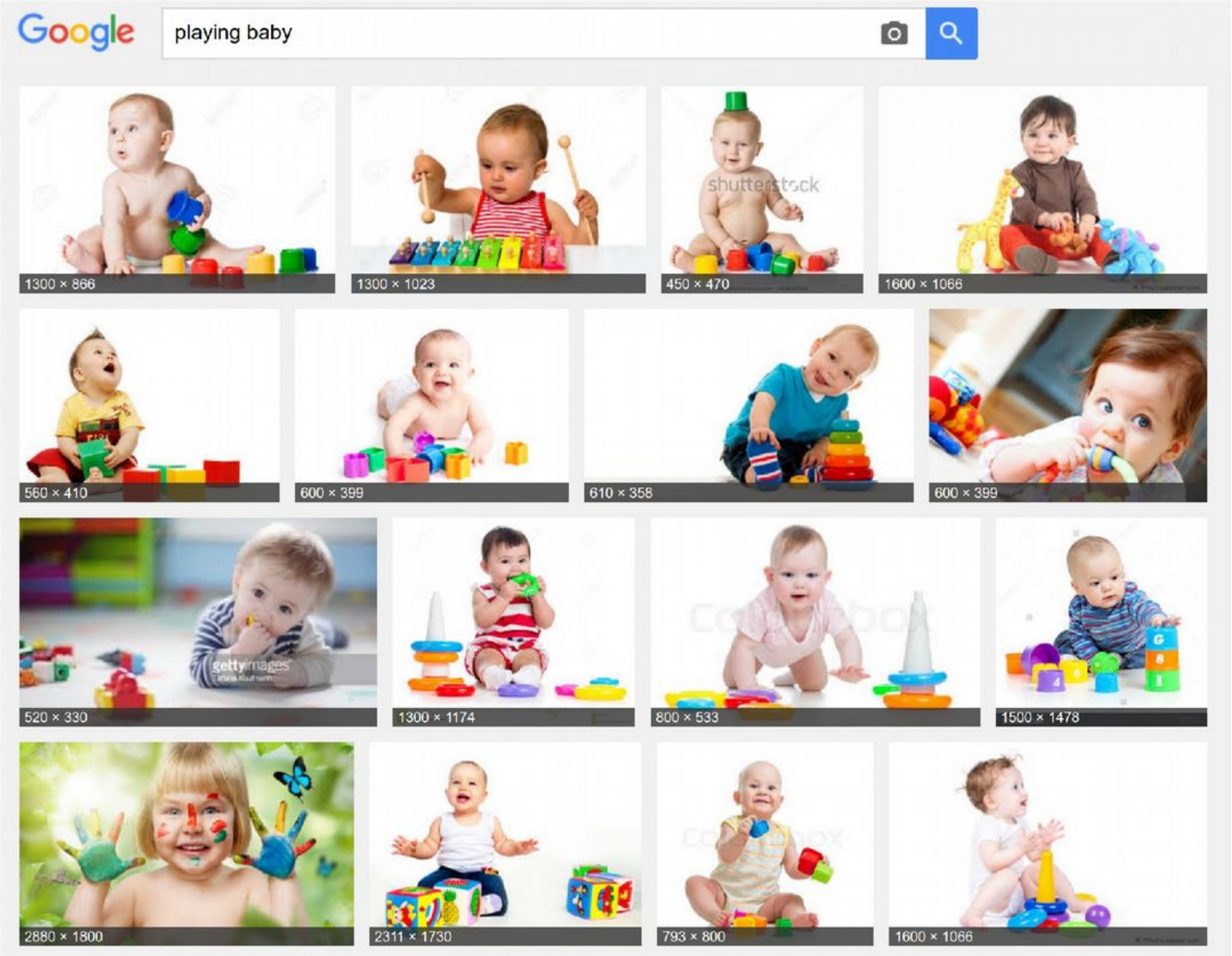
Big data



9

[9] image-net.org

Big data



10

[10] google.com

Big data

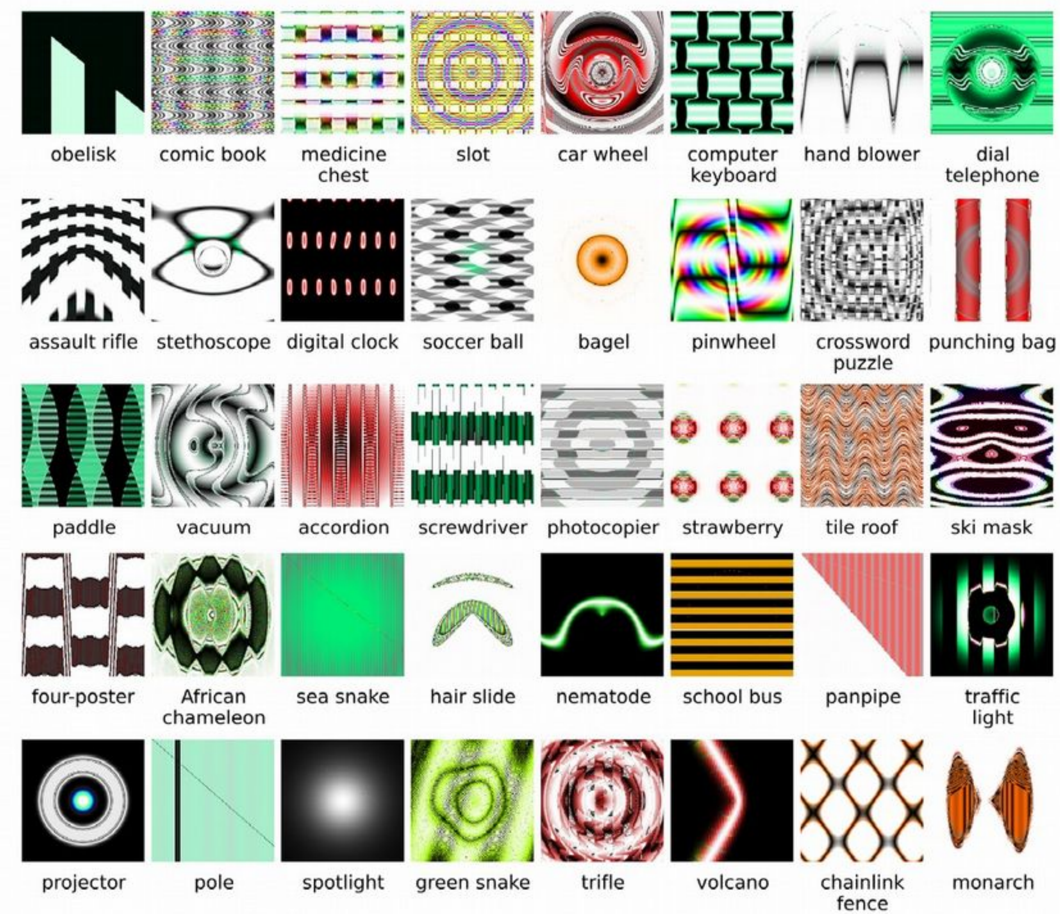
- We can only learn and generalize from given data
- Diversity within data is required for generalization
- We need a lot of data
- Quality of data matters
- Quality of labels matters
- Be aware of biases

Limits of deep learning

> *Deep learning is really just a fancy automatic way of doing pattern recognition*[11]
>
> *- Gary Marcus*

[8]

[12]



[8] *Vinyals et al. 2015*
[11] *oreily.com*
[12] *Nguyen et al. 2015*

Deep Learning Frameworks

The deep learning framework zoo

DL4J
DEEPLEARNING4J

Caffe

theano

K

Caffe2

Microsoft
CNTK

mxnet

PYTORCH

TensorFlow

torch

intel + nervana

Keras

- Object oriented high level API for fast experimentation with deep neural networks
- Build on top of Theano, TensorFlow, MXNet, Deeplearning4j or CNTK
- Difficult to expand

Model generation

Configuration file

```
layers {
  bottom: "data"
  top: "conv1"
  name: "conv1"
  type: CONVOLUTION
  convolution_param {
    num_output: 96
    kernel_size: 7
    stride: 2
  }
}
layers {
  bottom: "conv1"
  top: "conv1"
  name: "relu1"
  type: RELU
}
layers {
  bottom: "conv1"
  top: "norm1"
  name: "norm1"
  type: LRN
  lrn_param {
    local_size: 5
    alpha: 0.0005
    beta: 0.75
    k: 2
  }
}
```

Programmatic generation

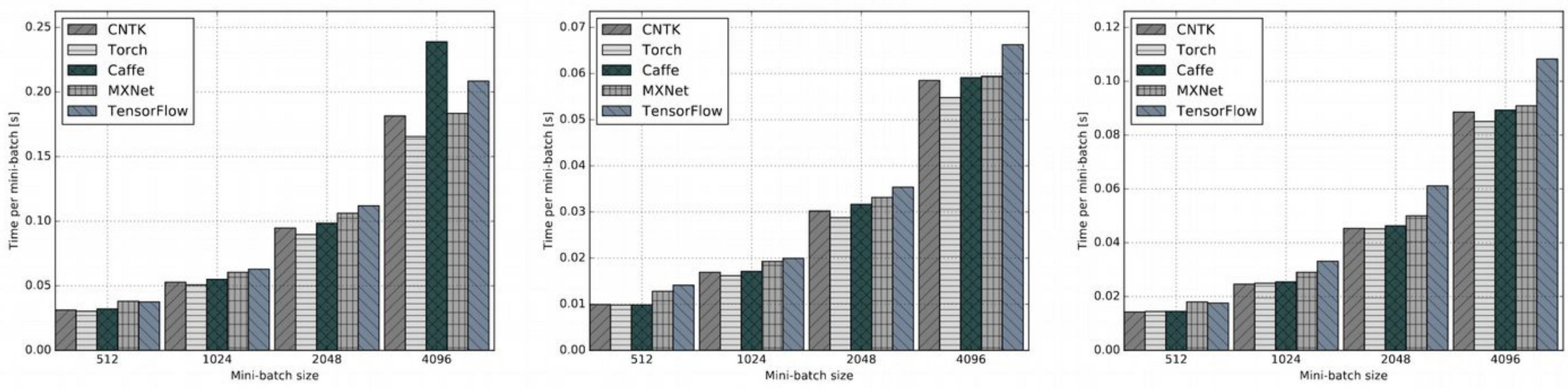
```
x = tf.placeholder(tf.float32, shape = [None, 784])

# reshape flat vectors into 2d images again
x_image = tf.reshape(x, [-1, 28, 28, 1])

# first conv layer
W_conv1 = weight_variable([5, 5, 1, 32])
b_conv1 = bias_variable([32])

h_conv1 = tf.nn.relu(conv2d(x_image, W_conv1) + b_conv1)
h_pool1 = max_pool_2x2(h_conv1)

# second conv layer
W_conv2 = weight_variable([5, 5, 32, 64])
b_conv2 = bias_variable([64])

h_conv2 = tf.nn.relu(conv2d(h_pool1, W_conv2) + b_conv2)
h_pool2 = max_pool_2x2(h_conv2)
```

Deep learning frameworks

Framework	Written in	Interface	Initial Release	Multi GPU	Maintainer
MxNet	C++	C++, Python, Julia, Matlabl JS, Go, R, Scala, Perl	December 2015	Yes	Community Based (Apache Incubator)
CNTK (Microsoft Cognitive Toolkit)	C++	C++, Python, Command Line	Januar 2016	Yes	Microsoft Research
TensorFlow	C++	Python, C++	November 2015	Yes	Google
Caffe	C++	Command Line, Python, C++	March 2014	Yes	Berkeley Vision and Learning Center
Caffe2	C++	C++, Python	April 2017	Yes	Facebook
Torch	C, Lua	Lua	October 2002	Yes	Community based
PyTorch (Beta)	Python	Python	January 2018	Yes	Facebook
Theano	Python	Python	August 2011	Yes	Université de Montréal
DeepLearning4j	C++, Java	Java	February 2014	Yes	Skymind

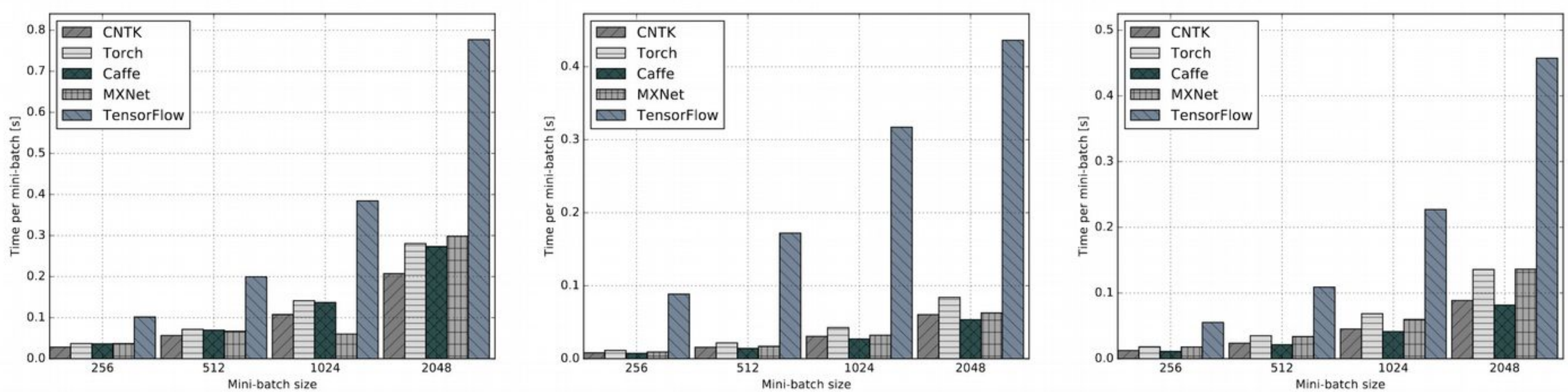
[14] wikipedia.org

Speed



(a) Results on Tesla K80. (b) Results on GTX1080. (c) Results on GTX980.

Figure 12. The performance comparison of FCN-R on GPU platforms.

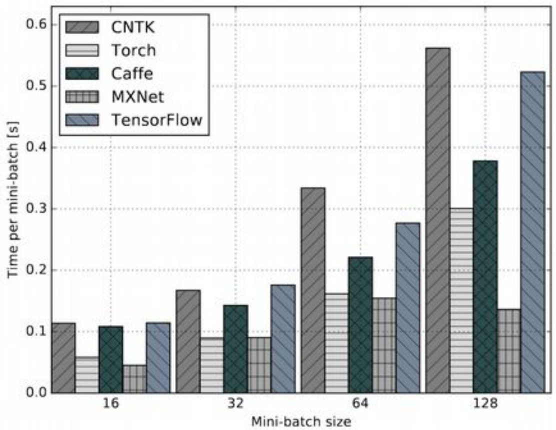


(a) Results on Tesla K80. (b) Results on GTX1080. (c) Results on GTX980.

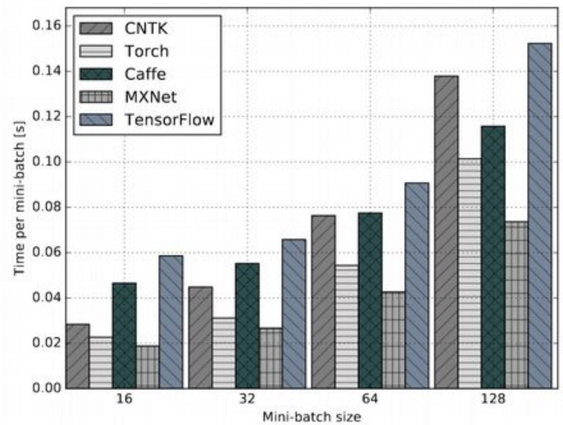
Figure 13. The performance comparison of AlexNet-R on GPU platforms.

[15] *Shi, Shaohuai et al. 2015*

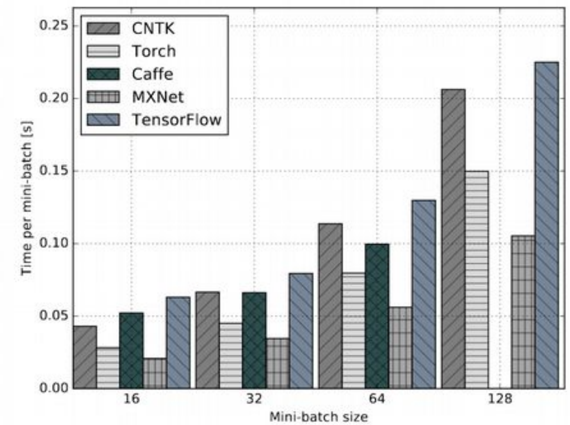
Speed



(a) Results on Tesla K80.

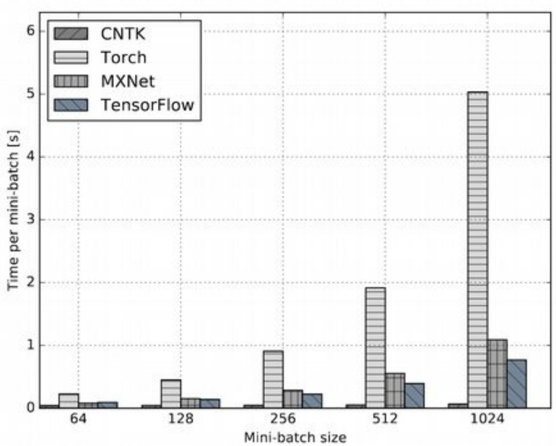


(b) Results on GTX1080.

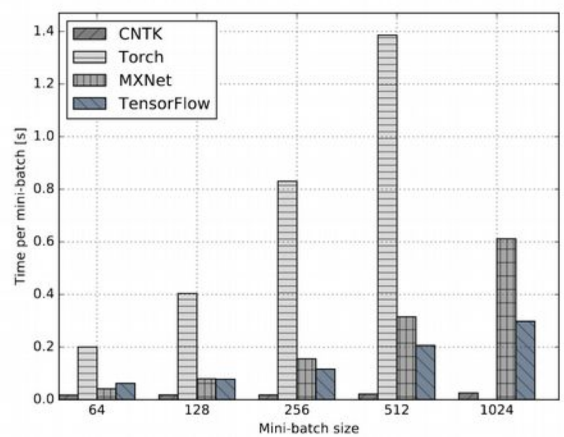


(c) Results on GTX980.

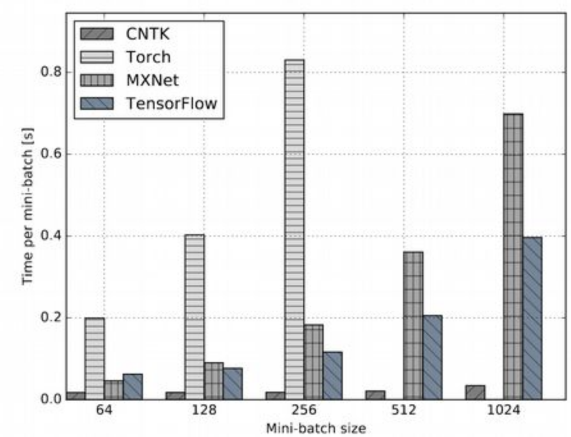
Figure 14. The performance comparison of ResNet-56 on GPU platforms.



(a) Results on Tesla K80.



(b) Results on GTX1080.

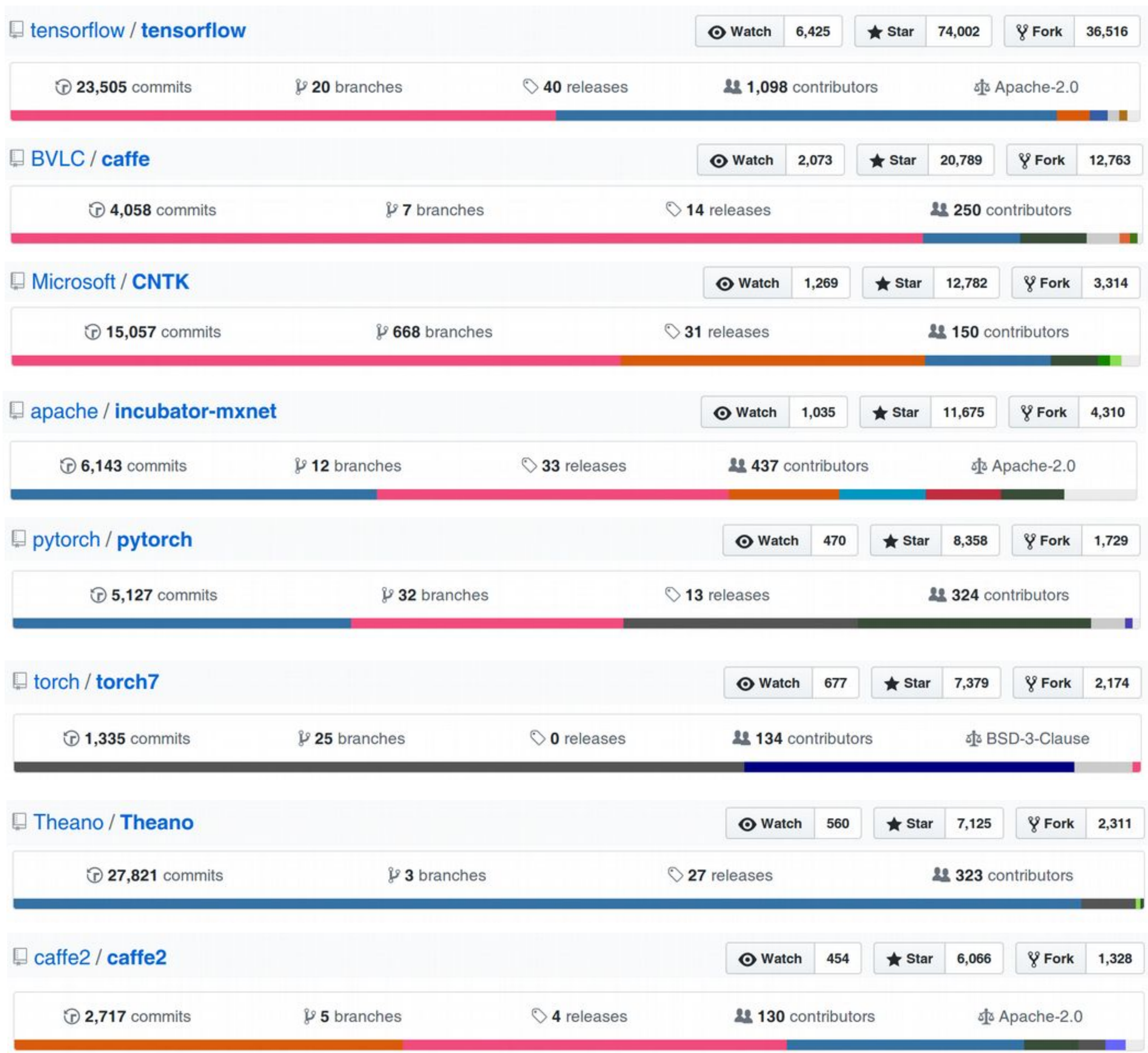


(c) Results on GTX980.

Figure 15. The performance comparison of LSTM on GPU platforms.

[15] Shi, Shaohuai et al. 2015

Popularity



tensorflow / tensorflow
Watch 6,425
Star 74,002
Fork 36,516
23,505 commits
20 branches
40 releases
1,098 contributors
Apache-2.0
BVLC / caffe
Watch 2,073
Star 20,789
Fork 12,763
4,058 commits
7 branches
14 releases
250 contributors
Microsoft / CNTK
Watch 1,269
Star 12,782
Fork 3,314
15,057 commits
668 branches
31 releases
150 contributors
apache / incubator-mxnet
Watch 1,035
Star 11,675
Fork 4,310
6,143 commits
12 branches
33 releases
437 contributors
Apache-2.0
pytorch / pytorch
Watch 470
Star 8,358
Fork 1,729
5,127 commits
32 branches
13 releases
324 contributors
torch / torch7
Watch 677
Star 7,379
Fork 2,174
1,335 commits
25 branches
0 releases
134 contributors
BSD-3-Clause
Theano / Theano
Watch 560
Star 7,125
Fork 2,311
27,821 commits
3 branches
27 releases
323 contributors
caffe2 / caffe2
Watch 454
Star 6,066
Fork 1,328
2,717 commits
5 branches
4 releases
130 contributors
Apache-2.0

Why do we use TensorFlow?

- Python
- Popularity: Fast growing, frequent updates and improvements
- Low level and hence extendable
- Supports huge variety of ANNs (cNN, RNN, DBN ...)
- Build in support for distributed computation
- **At the end, frameworks are extremely similar**

TensorFlow

What is TenorFlow?

> “ *TensorFlow is an open source software library for numerical computation using data flow graphs*[1]

[1] *tensorflow.org*

What is a Data Flow Graph?

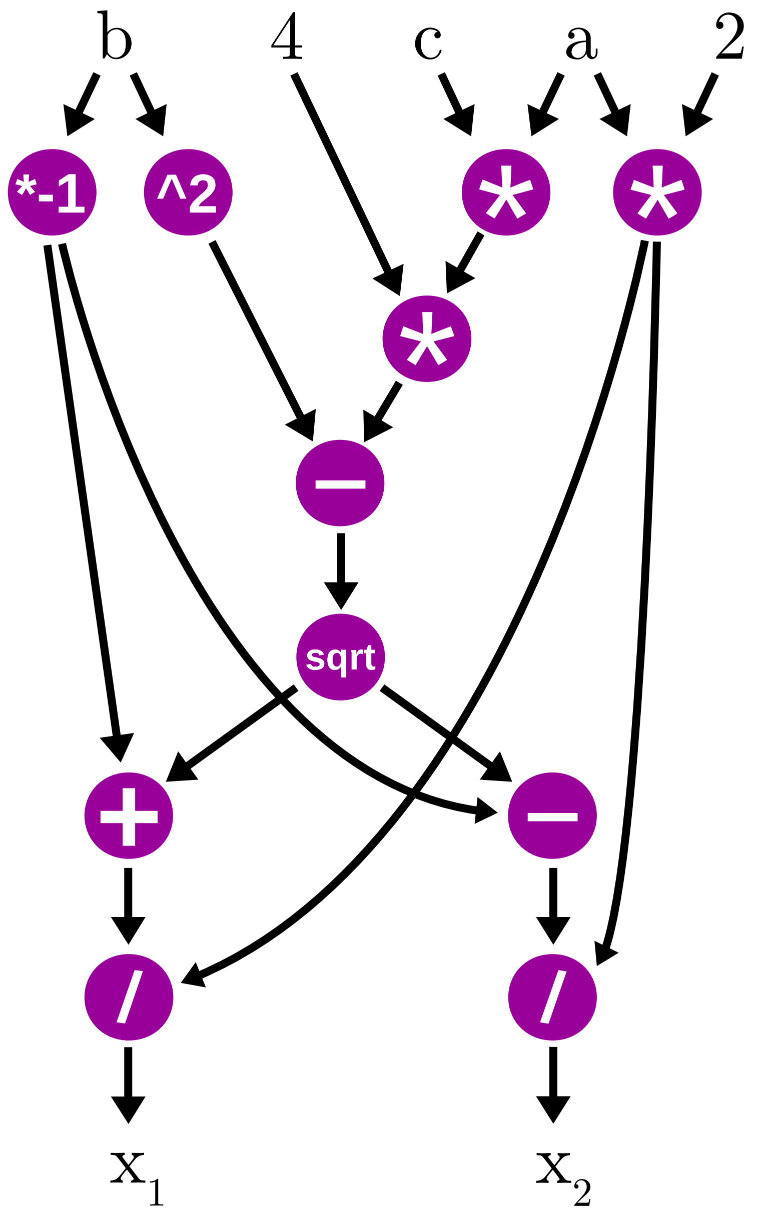
Mathematical computation as a directed graph

Nodes Mathematical operations

Edges Flow of numbers

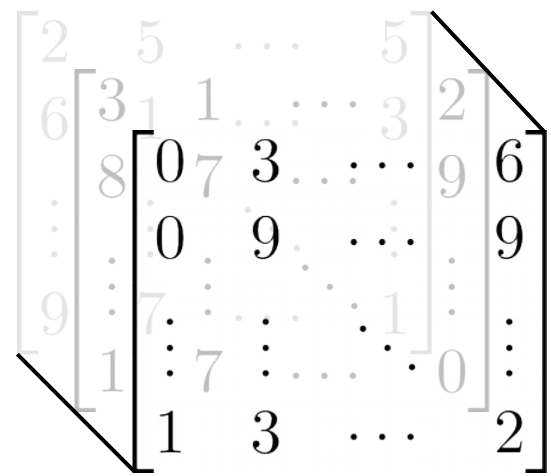
Data Flow Graph for:

$$x_1, x_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

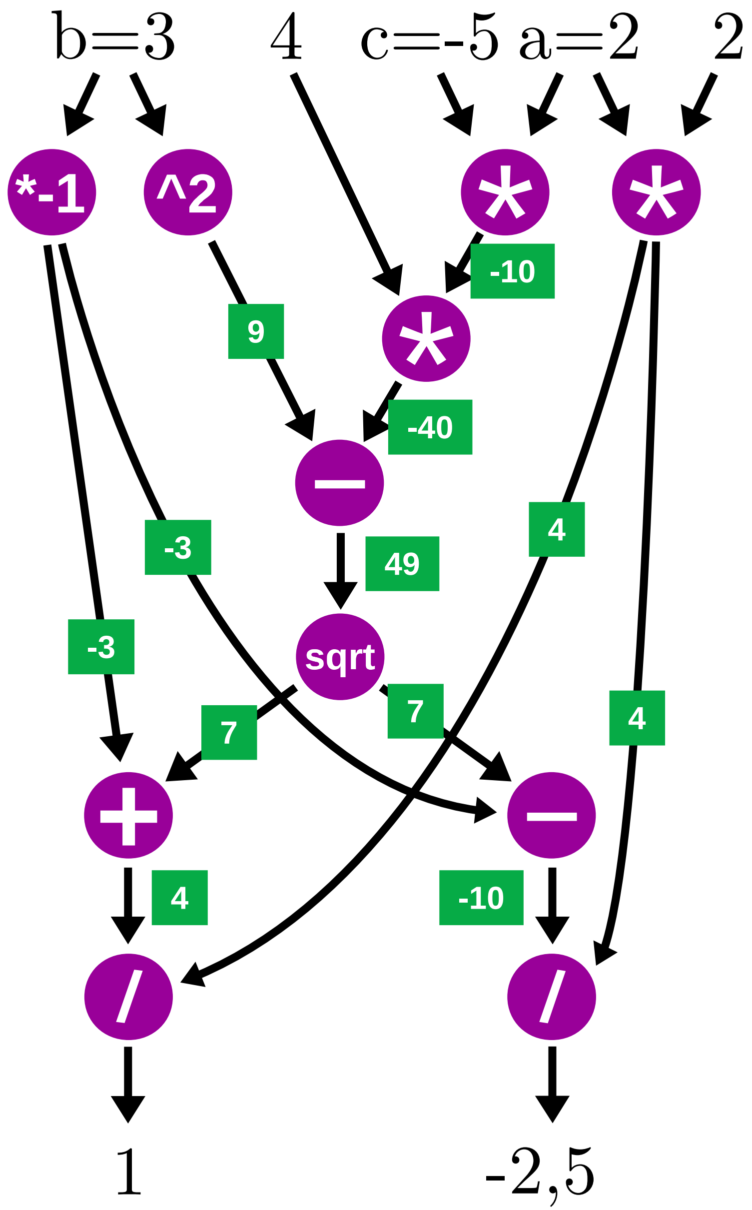


What is flowing?

Tensors are arbitrary dimensional arrays

Scalar	Vector	Matrix	3D Matrix	*N*D Matrix
$[4]$	$\begin{bmatrix}5\\2\\8\\1\\\vdots\\9\\3\end{bmatrix}$	$\begin{bmatrix}3 & 1 & \cdots & 2\\8 & 7 & \cdots & 9\\\vdots & \vdots & \ddots & \vdots\\1 & 7 & \cdots & 0\end{bmatrix}$		...
0-order Tensor	1st-order Tensor	2nd-order Tensor	3rd-order Tensor	*Nth-order* Tensor

Example



DFG for quadratic formula

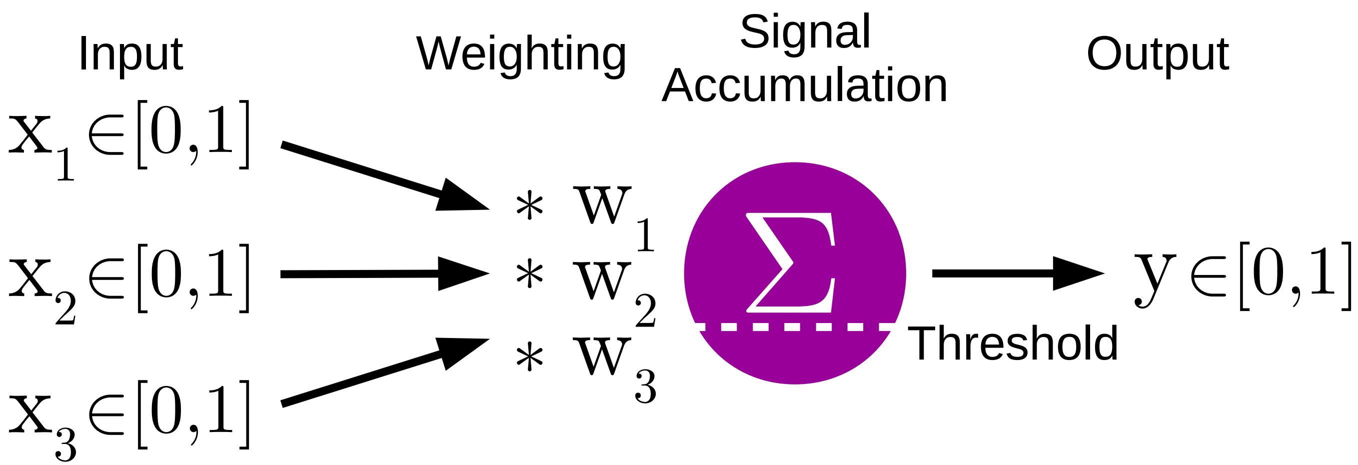
$$x_1, x_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

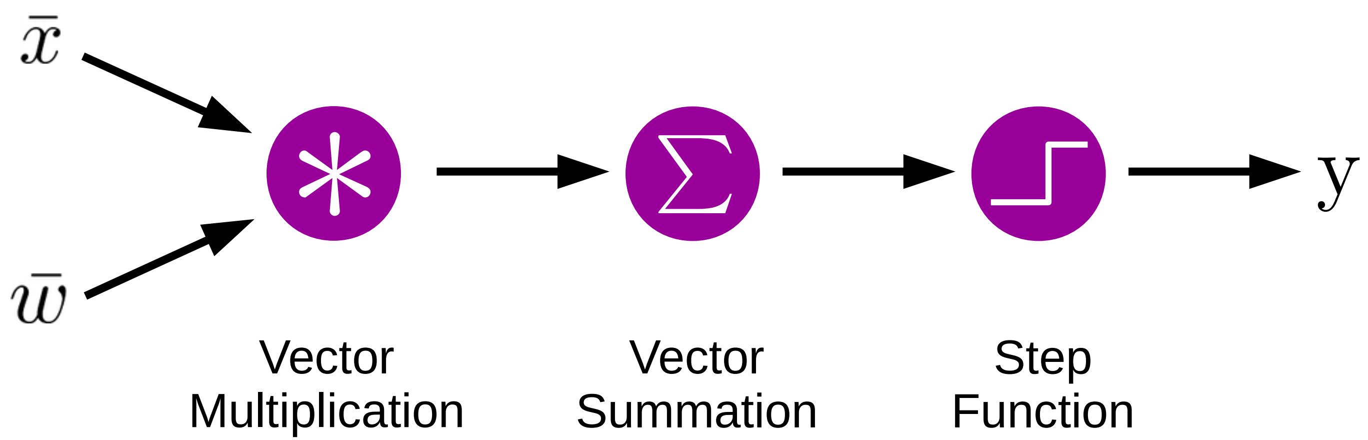
Example:

$$2x^2 + 3x - 5 = 0$$

Great potential for parallelization!

ANNs as Data Flow Graph





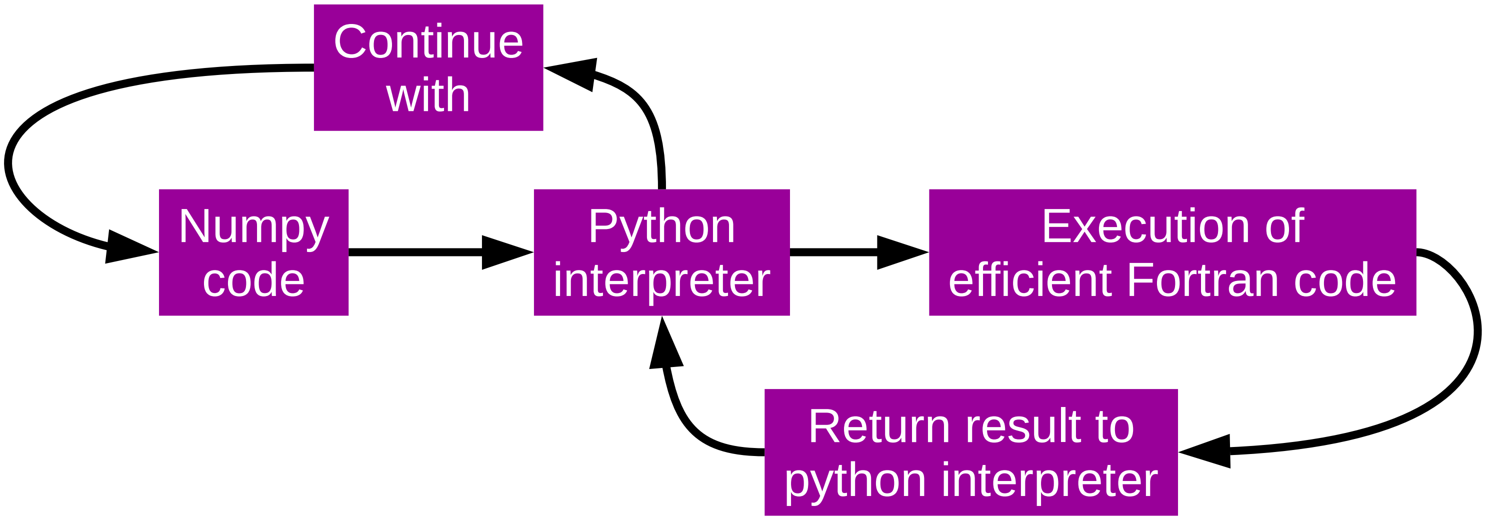
PyTorch

- Still in “early-release” Beta
- Creates Dynamic Computational Graphs which allow dynamic data structures (i.e. text) inside the network
- Other frameworks are based on Static Computational Graphs
- Perfect for dynamically changing systems like recurrent neural networks
- Faster prototyping

Architecture

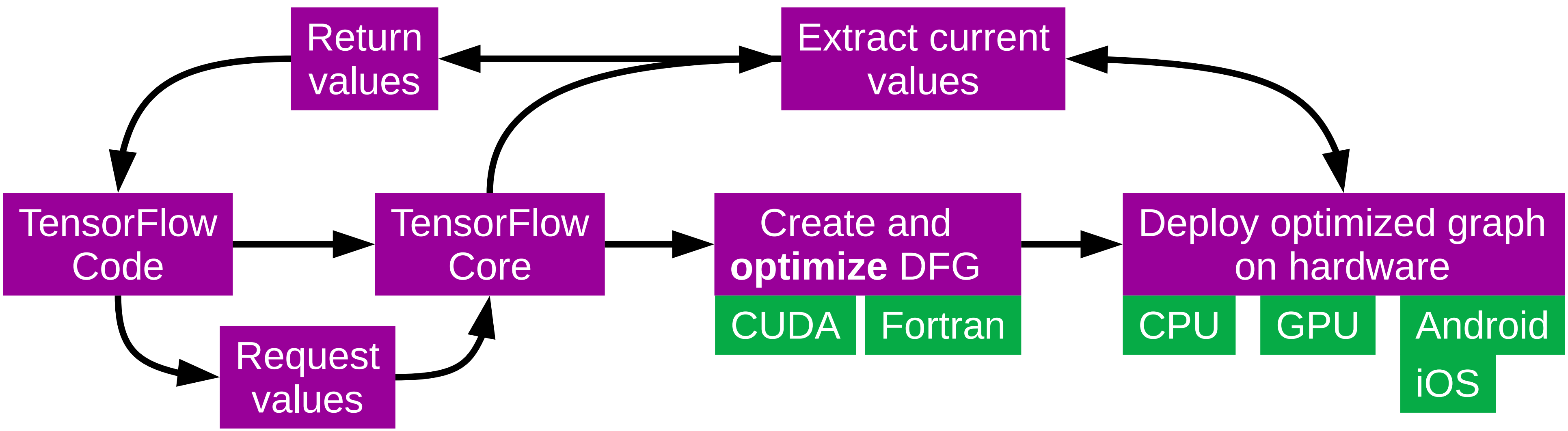
Architecture

Why do we use a framework at all?



Architecture

Why do we use a framework at all?



Easy to share models and to execute very efficient code on different Hardware

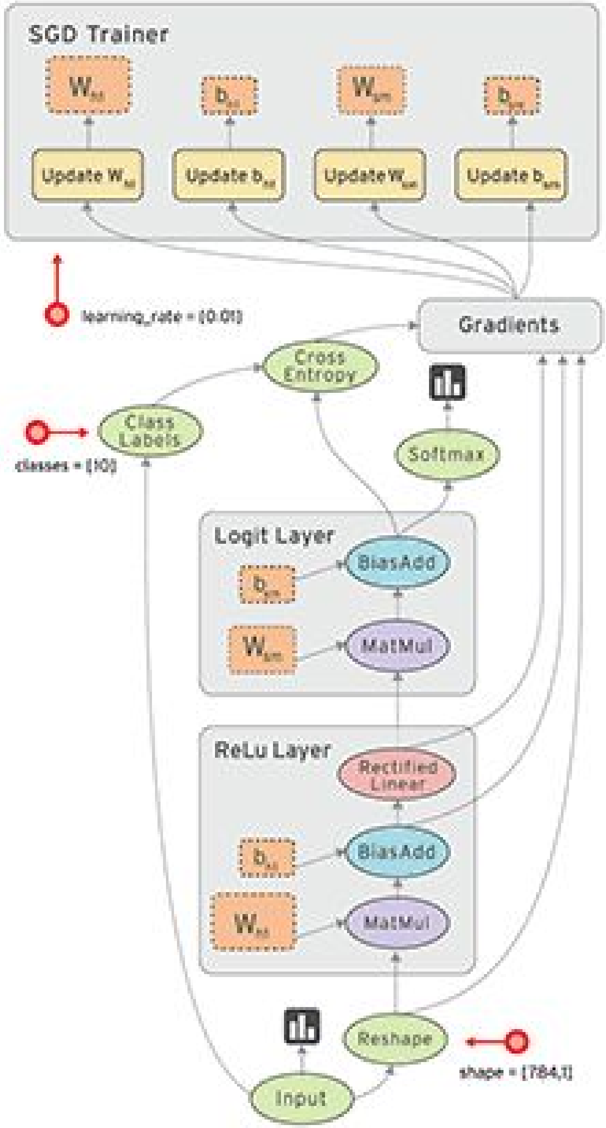
For more technical details see the TensorFlow white paper

TensorFlow workflow

How does coding with TensorFlow look like?

1) Describe the data flow graph with python
2) Define the desired learning algorithm (e.g. gradient descent)
3) Tell TensorFlow to construct the graph
4) Deploy the graph on some hardware in a *session*
5) Feed data into the graph and let TensorFlow **optimize the weights automatically**
6) Save trained network for later use

TensorFlow at work



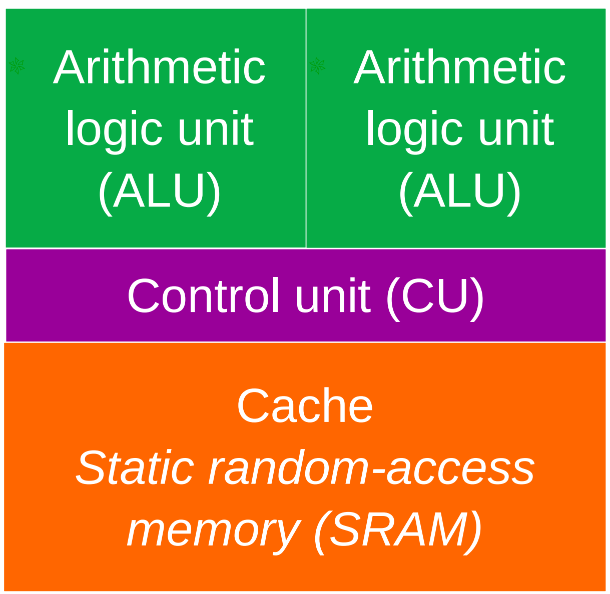
[13]

[13] tensorflow.org

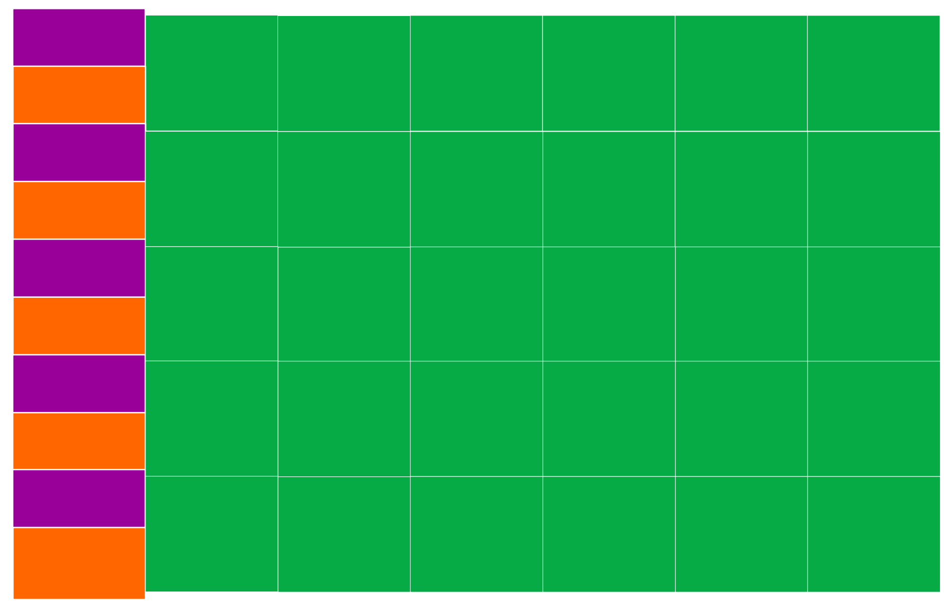
GPU Acceleration and Distributed Computing

CPU vs. GPU – Computing

Central processing unit (CPU)



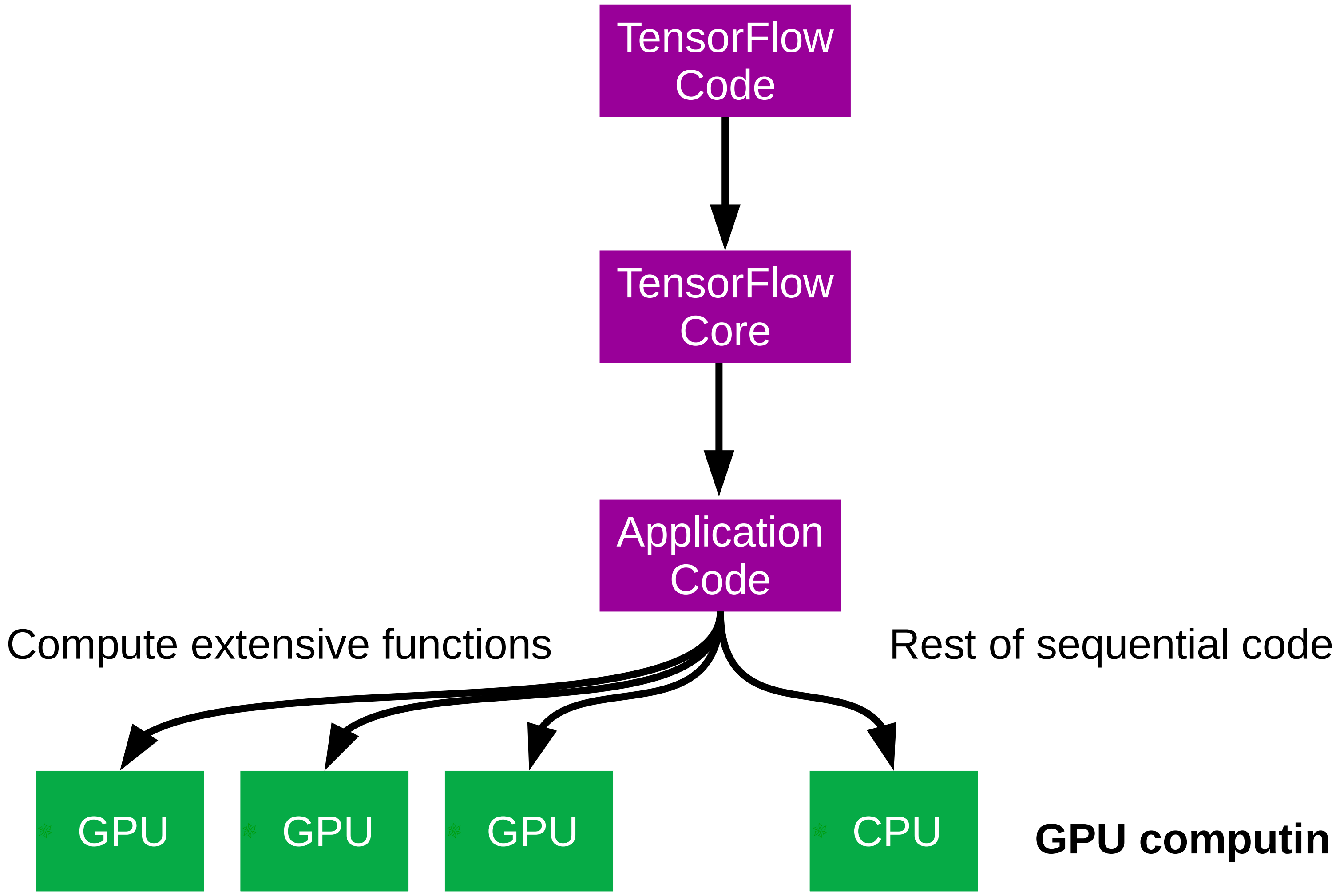
Graphics processing unit (GPU)



CPU vs. GPU – Computing

CPUs	GPU
Multi-core processor	Many-core processor
Very precise operations	Less precise operations
Very high clock rate	Lower clock rate
Huge control unit and cache for general process management	Ideal for parallel execution of heavy calculations

CPU vs. GPU – Computing



GPU computing is >10x faster

Final Notes

How to learn python, numpy and pyplotlib?

Codecademy's python course

1) Interactive online framework
2) Little tasks and exercises
3) No python environment necessary

Quickstart numpy tutorial

Quickstart matplotlib.pyplot tutorial

Send me your questions!

- About ANNs in general
- About specific kinds of networks
- About specific (e.g. optimization) algorithms
- About papers and news that caught your attention
- What do you hope to learn in this course?
- …?

Criticism is always welcome!

This Week’s Task

Installing TensorFlow

- Sign up to groups
- Please download the installing instructions for Windows, Mac or Linux from studip
- We are going to use TensorFlow starting from 6th of November
- Homework assignments are designed such that you are able to solve the tasks with a consumer notebook’s CPU

THANK YOU!

Sources

[1] Retrieved September 6th, 2016 from https://www.youtube.com/watch?v=g0TaYhjpOfo

[2] By MethoxyRoxy - Own work, CC BY-SA 2.5, https://commons.wikimedia.org/w/index.php?curid=1325252 Retrieved September 10th, 2016

[3] Retrieved October 15th, 2016 from https://www.technologyreview.com/s/518006/how-a-fly-brain-detects-motion/

[4] Schilling, Tabea, and Alexander Borst. "Local motion detectors are required for the computation of expansion flow-fields." Biology open (2015): bio-012690.

[5] Borst, Alexander, and Moritz Helmstaedter. "Common circuit design in fly and mammalian motion vision." Nature neuroscience 18.8 (2015): 1067-1076.

[6] Drews, Michael. *Unpublished*

[7] *Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.*

[8] *Vinyals, Oriol, et al. "Show and tell: A neural image caption generator." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.*

[9] *Retrieved October 20th, 2016, http://image-net.org/synset?wnid=n02274259*

[10] *Retrieved October 20th, 2016, https://www.google.de/search?tbm=isch&q=playing+baby*

[11] *Retrieved October 1st, 2016, https://www.oreilly.com/ideas/from-big-data-to-human-level-artificial-intelligence*

[12] *Nguyen, Anh, Jason Yosinski, and Jeff Clune. "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images." 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.*

[13] Retrieved September 9th, 2016 from https://www.tensorflow.org/

[14] Retrieved October 9th, 2017 from https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software

[15] *Shi, Shaohuai, et al. "Benchmarking state-of-the-art deep learning software tools." arXiv preprint arXiv:1608.07249 (2016).*