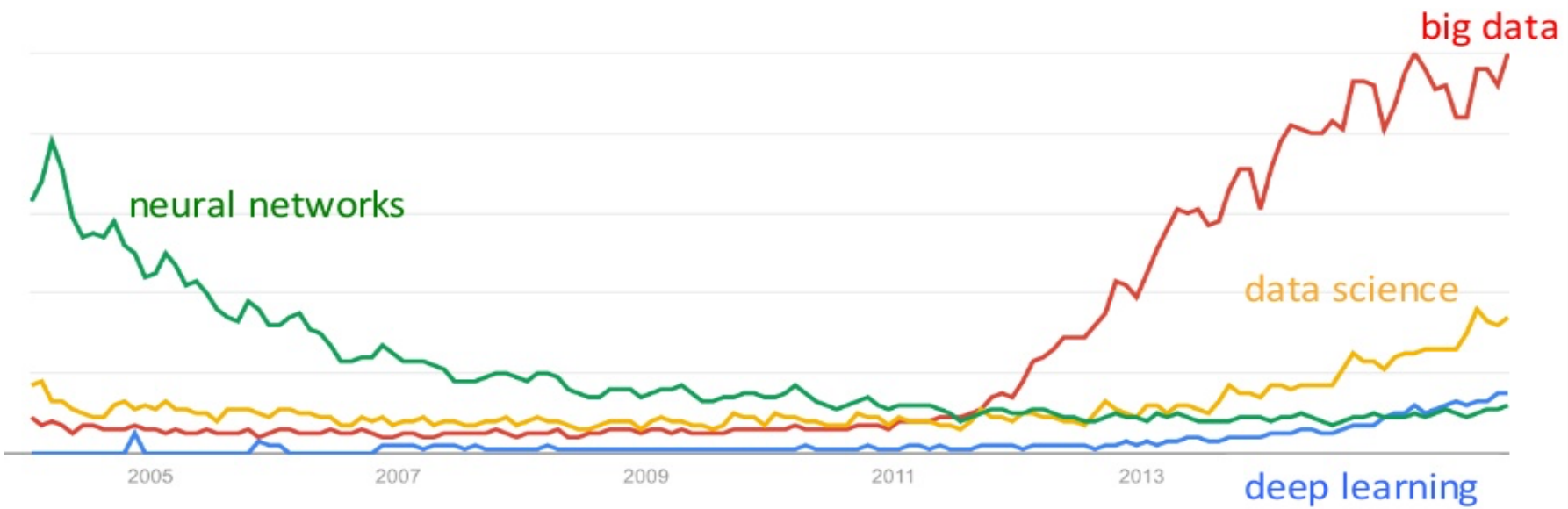
Deep Learning – Möglichkeiten, Chancen und Konflikte

Europäisches Forum Alpbach 2017, Technologiegespräche

Bernhard Nessler
Alpbach, 2017-08-25

BUZZWORD DEEP LEARNING



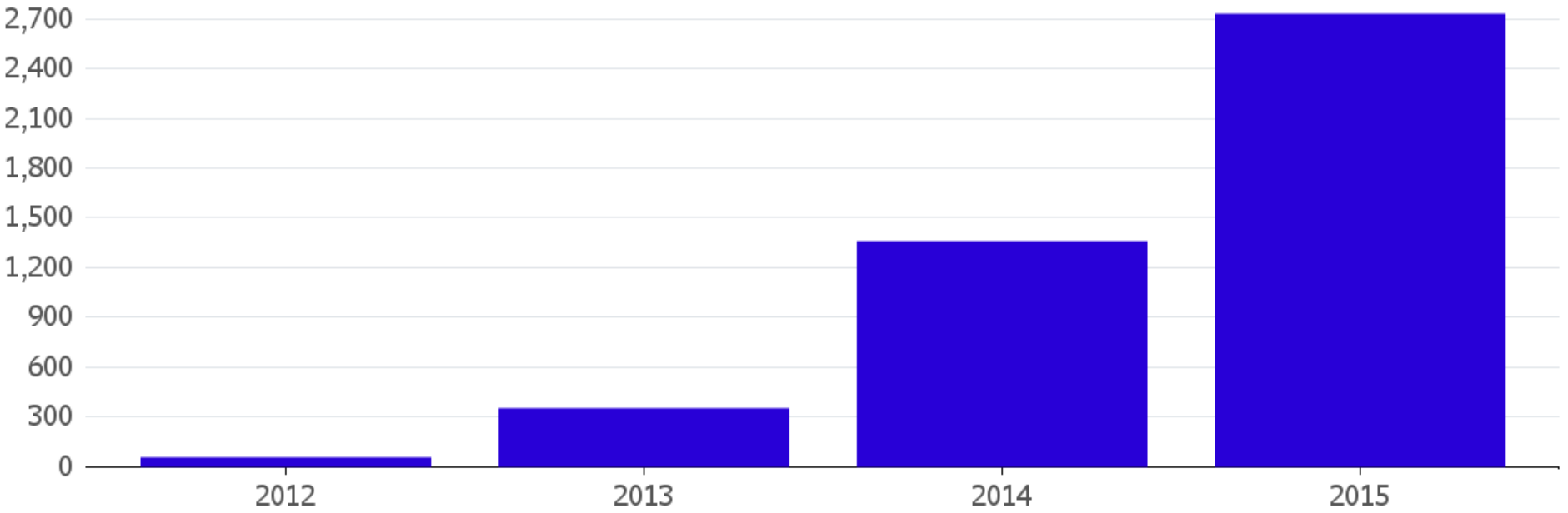
- MIT Technology Review, 2013: *DL among the 10 breakthrought technologies*
- Fortune Magazine, 2016: *Why Deep Learning is suddenly changing your life*

- Google: image recognition/search, speech recognition, automated translation
- Facebook: photo categorization and annotation; Amazon Go stores; Netflix recommender systems ; Apple: Siri Assistant; Merck: drug discovery; Deepmind: Alpha Go,

nessler@bioinf.jku.at

Artificial Intelligence Takes Off at Google

Number of software projects within Google that uses a key AI technology, called Deep Learning.



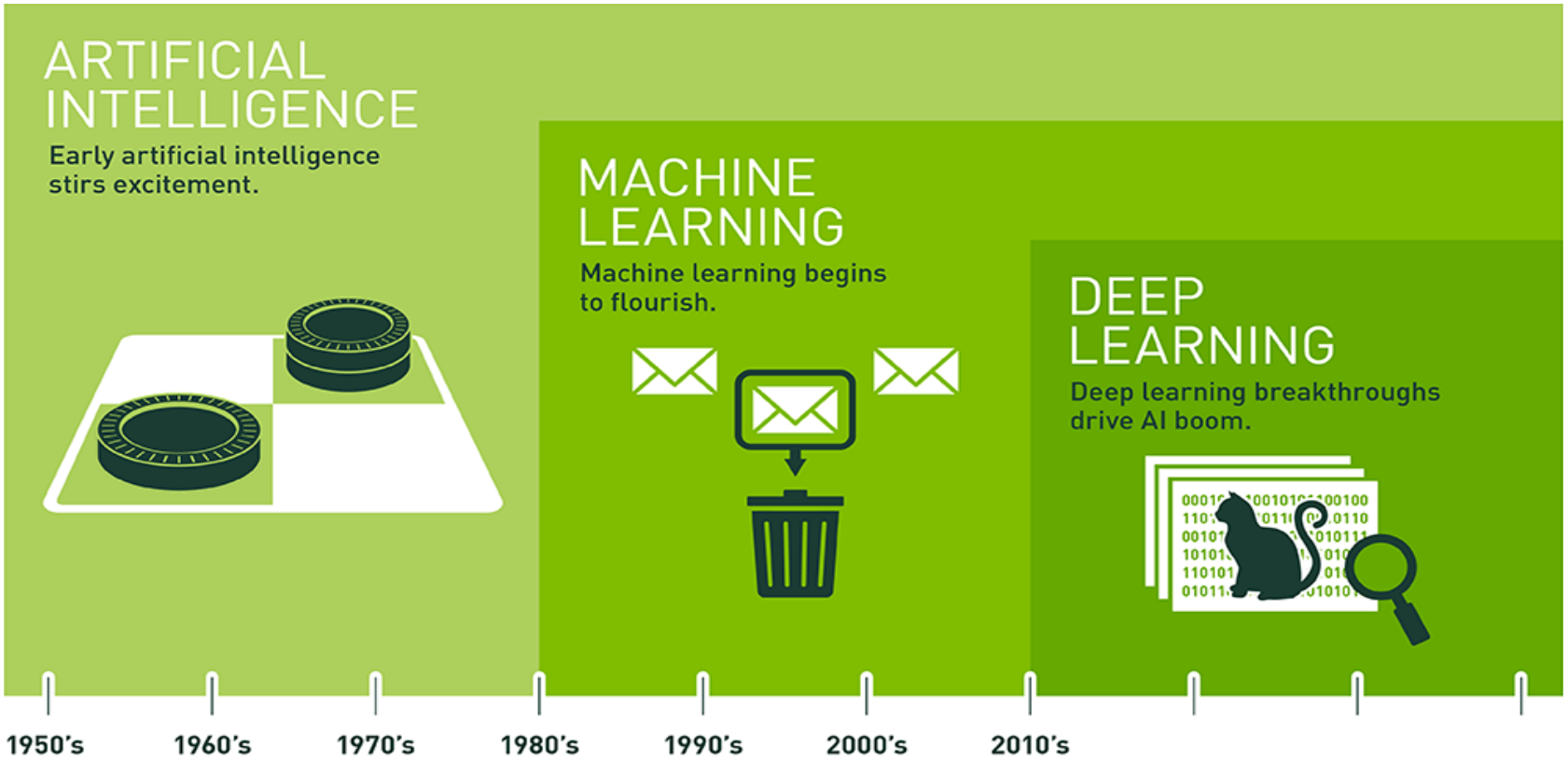
Source: Google

Note: 2015 data does not incorporate data from Q4

Bloomberg

nessler@bioinf.jku.at

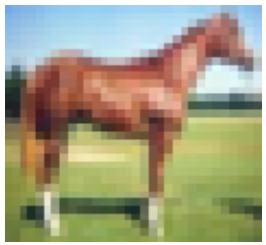
Relation AI / ML / DL



- Turing Test, Perceptron, Eliza
- Classical ML: OCR, Spam Detection, Genomics, statistical predictions, MS Office Assistant, MS Printer Troubleshouter, Old Google Search, Early Recomender Systems, Industrial quality checks
- Deep Learning: ImageNet, Alpha Go, Tox21 Challenge (Winner: bioinf@JKU) , MobilEye , Spotter

nessler@bioinf.jku.at

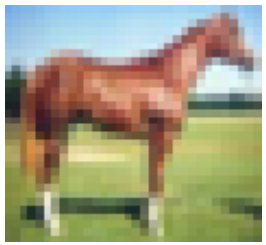
Machine Learning



Horse 33%
Dog 33%
Cat 33%

nessler@bioinf.jku.at

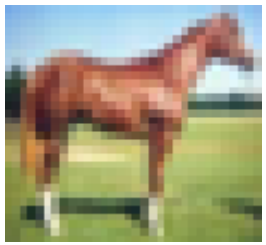
Machine Learning



Horse	**33%**
Dog	33%
Cat	33%

nessler@bioinf.jku.at

Machine Learning - Training

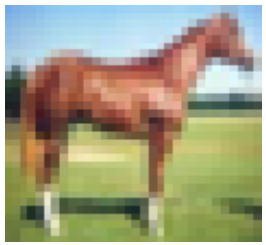


	Target values	
Horse	**33%**	**100%**
Dog	33%	0%
Cat	33%	0%

Error = 3*0.66 = 2

nessler@bioinf.jku.at

Machine Learning – Training Step



Target values

Horse	**33%**	**100%**
Dog	33%	0%
Cat	33%	0%

Error = 3*0.66 = 2

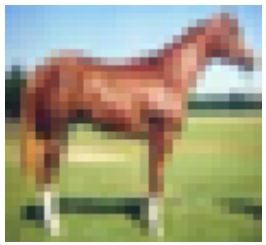
Teacher Feedback, Supervision

Very small training step, iterated many times over many training images

JKU

nessler@bioinf.jku.at

Machine Learning – Training Error

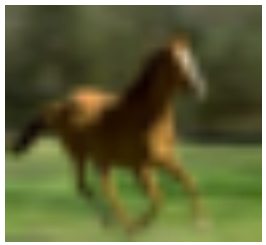


		Target values
Horse	90%	100%
Dog	4%	0%
Cat	6%	0%

Error = 0.2

nessler@bioinf.jku.at

Machine Learning – Testing Error Generalization



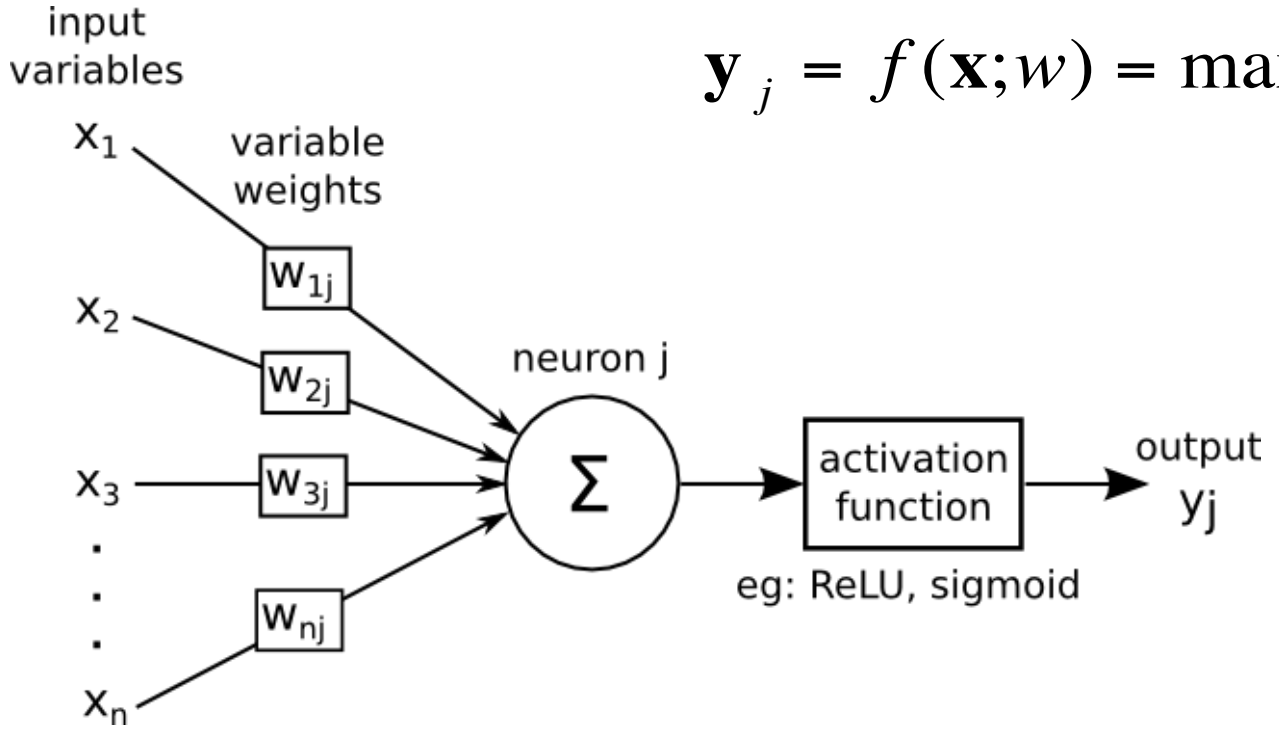
Horse	**60%**
Dog	27%
Cat	13%

Error = 0.8

nessler@bioinf.jku.at

Neural Networks – single neuron

- Neural Networks (ANNs) are a method how to build a trainable program f (there are many others).



$$\mathbf{y}_j = f(\mathbf{x}; w) = \max(0, \sum_i x_i w_{ij})$$

$$\Delta y_j = x_i \Delta w_i \quad \text{oder} \quad 0$$

$$\frac{df}{dw_i} = x_i \quad \text{oder} \quad 0$$

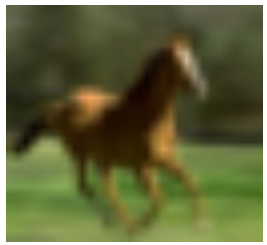
$$\frac{df}{dx_i} = w_i \quad \text{oder} \quad 0$$

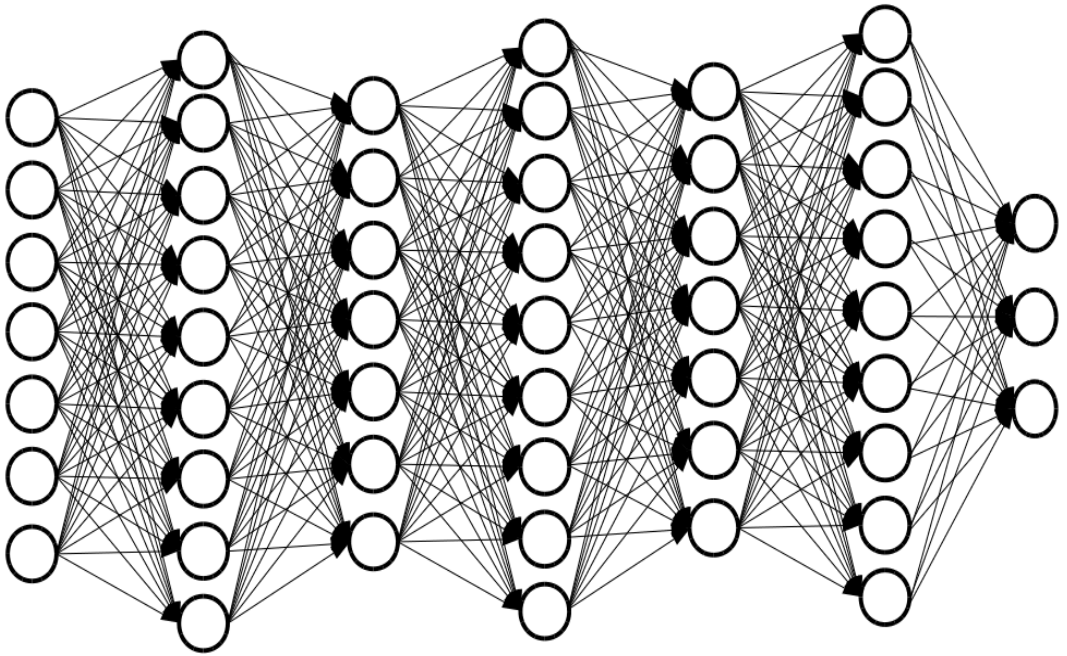
$$\text{Learning}: \quad w \leftarrow w - \gamma \left. \frac{dE}{dw} \right|_w$$

JKU

pessler@bioinf.jku.at

Deep Neural Network

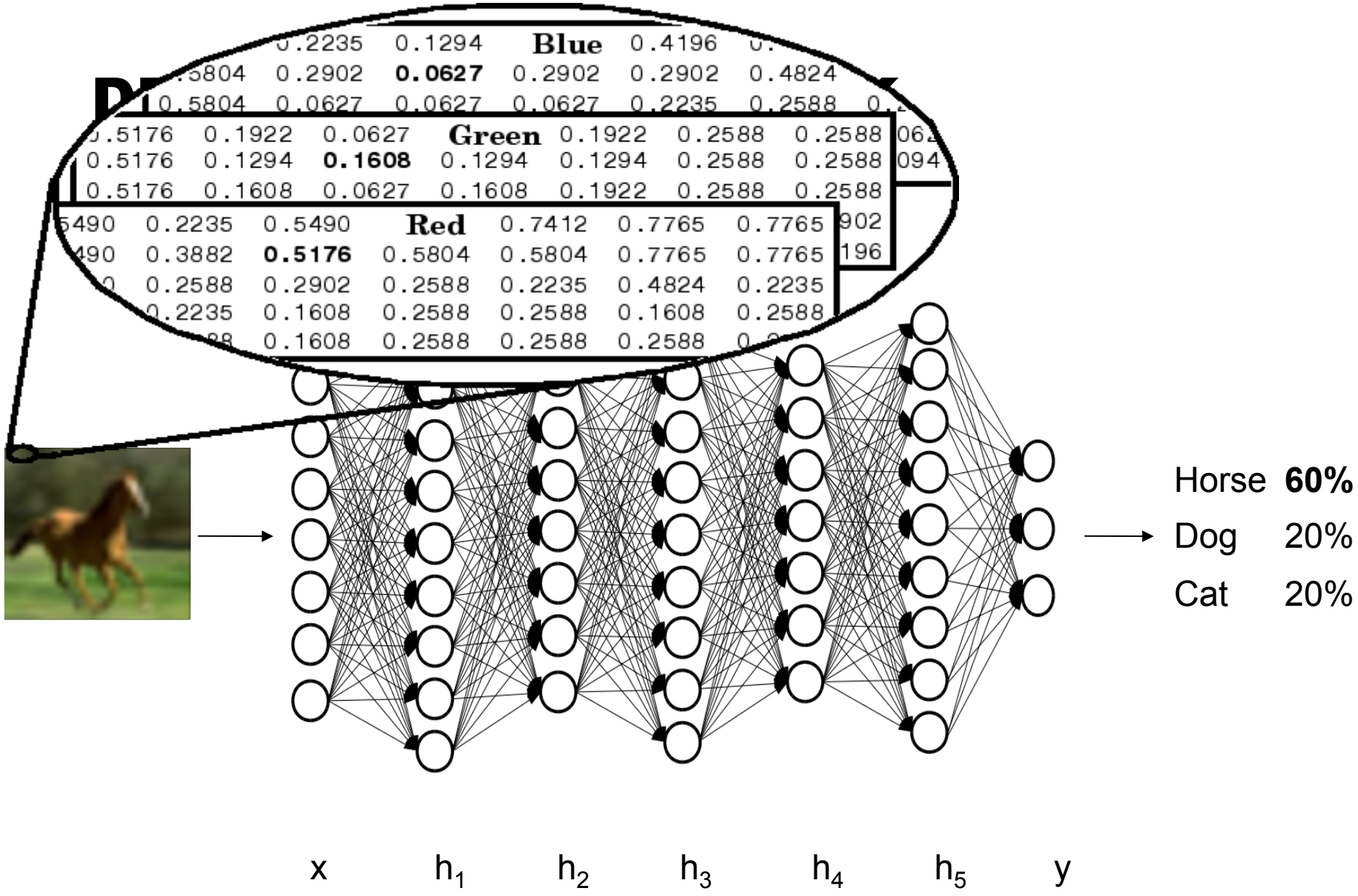




Horse **60%**
Dog 20%
Cat 20%

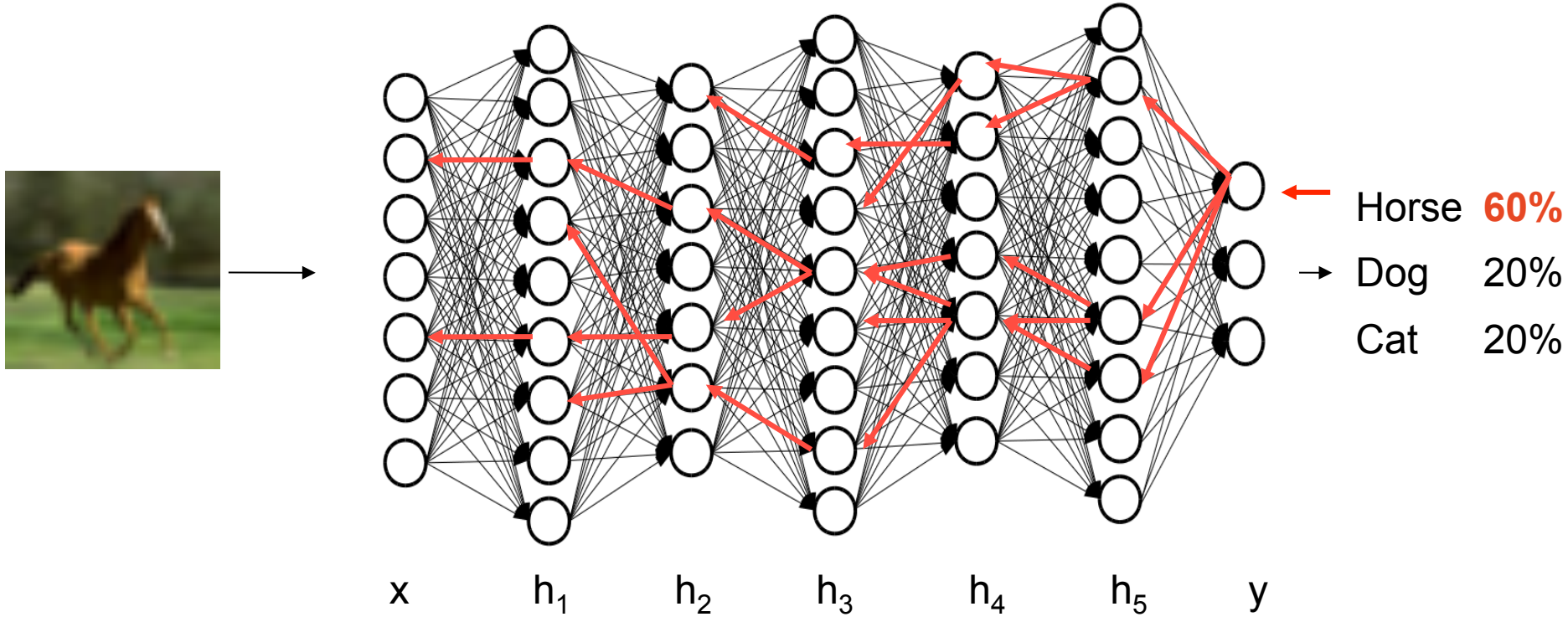
x h_1 h_2 h_3 h_4 h_5 y

nessler@bioinf.jku.at



nessler@bioinf.jku.at

Deep Learning, Backpropagation



$$y = f\Big(h_5\Big(h_4\Big(h_3\Big(h_2\Big(h_1\big(x;w_1\big);w_2\Big);w_3\Big);w_4\Big);w_5\Big);w_6\Big)$$

nessler@bioinf.jku.at

Deep Learning, vanishing gradient

- Function f has a nested structure

$$y = f\Big(h_5\Big(h_4\Big(h_3\Big(h_2\Big(h_1\big(x;w_1\big);w_2\Big);w_3\Big);w_4\Big);w_5\Big);w_6\Big)$$

- Gradients multiply according to the chain-rule

$$w_6 \leftarrow w_6 - \gamma \frac{dE}{dw_6}$$

$$w_5 \leftarrow w_5 - \gamma \frac{dE}{dh_5}\frac{dh_5}{dw_5}$$

$$w_4 \leftarrow w_4 - \gamma \frac{dE}{dh_5}\frac{dh_5}{dh_4}\frac{dh_4}{dw_4}$$

$$w_3 \leftarrow w_3 - \gamma \frac{dE}{dh_5}\frac{dh_5}{dh_4}\frac{dh_4}{dh_3}\frac{dh_3}{dw_3}$$

$$w_2 \leftarrow w_2 - \gamma \frac{dE}{dh_5}\frac{dh_5}{dh_4}\frac{dh_4}{dh_3}\frac{dh_3}{dh_2}\frac{dh_2}{dw_2}$$

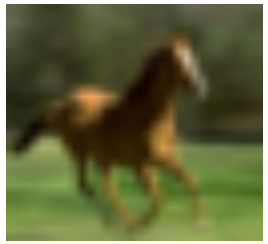
$$w_1 \leftarrow w_1 - \gamma \frac{dE}{dh_5}\frac{dh_5}{dh_4}\frac{dh_4}{dh_3}\frac{dh_3}{dh_2}\frac{dh_2}{dh_1}\frac{dh_1}{dw_1}$$

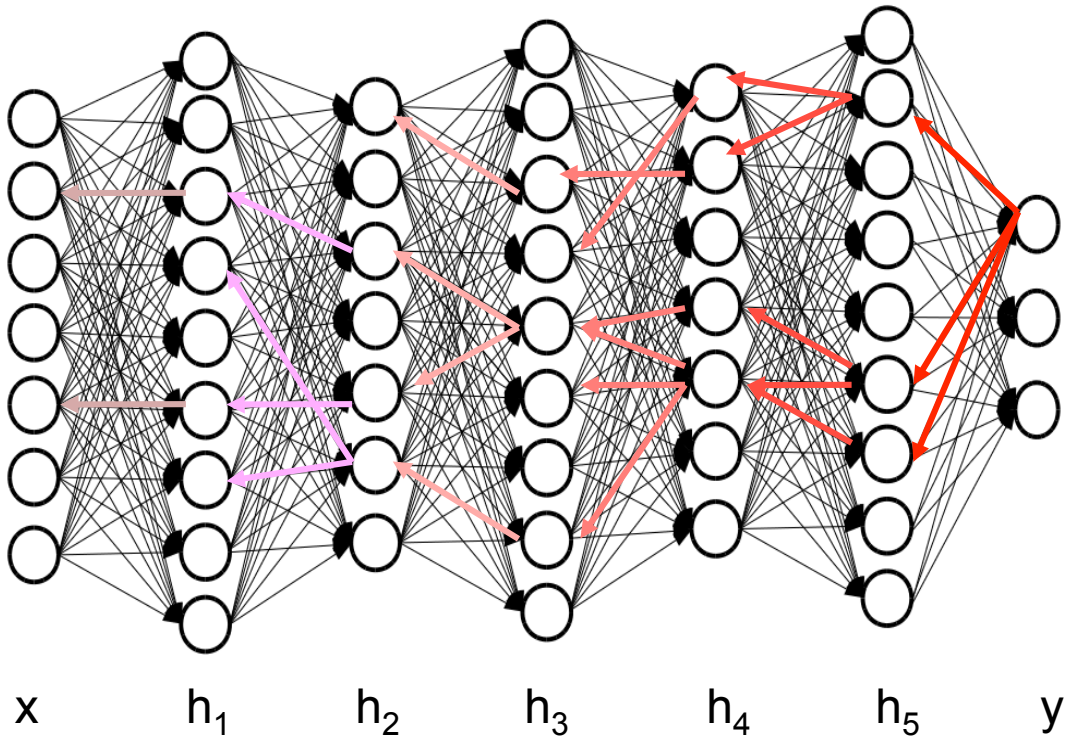
- Gradient signal gets lost in the noise -> Vanishing gradient

- First identified 1991 by Prof. Sepp Hochreiter, JKU Linz

nessler@bioinf.jku.at

Deep Learning, Backpropagation





$$y = f\Big(h_5\Big(h_4\Big(h_3\Big(h_2\Big(h_1\big(x;w_1\big);w_2\Big);w_3\Big);w_4\Big);w_5\Big);w_6\Big)$$

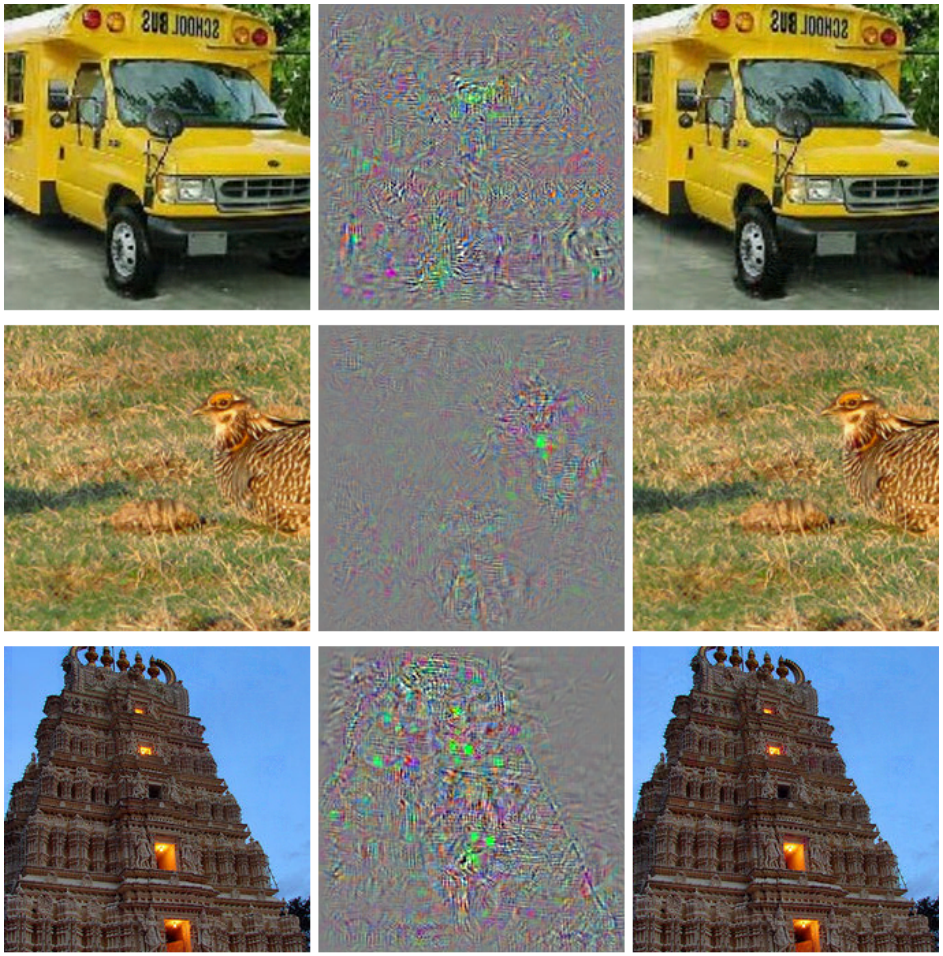
nessler@bioinf.jku.at

Deep Learning, vanishing gradient

- Currently applied methods
  - Gating:
    - try to make most intermediate gradient values close to 0 or 1.
    - Long Short Term Memory (LSTM), (Hochreiter, Schmidhuber, 1997)
    - Residual networks
  - Normalization: distorts gradient, increases noise
  - Clever initialization helps
- New solution:
  - Find clever functions and structures where the problem solves itself by intrinsic regulation.
    - Klambauer, Günter, et al. "Self-Normalizing Neural Networks." *arXiv preprint arXiv:1706.02515 (2017).*

nessler@bioinf.jku.at

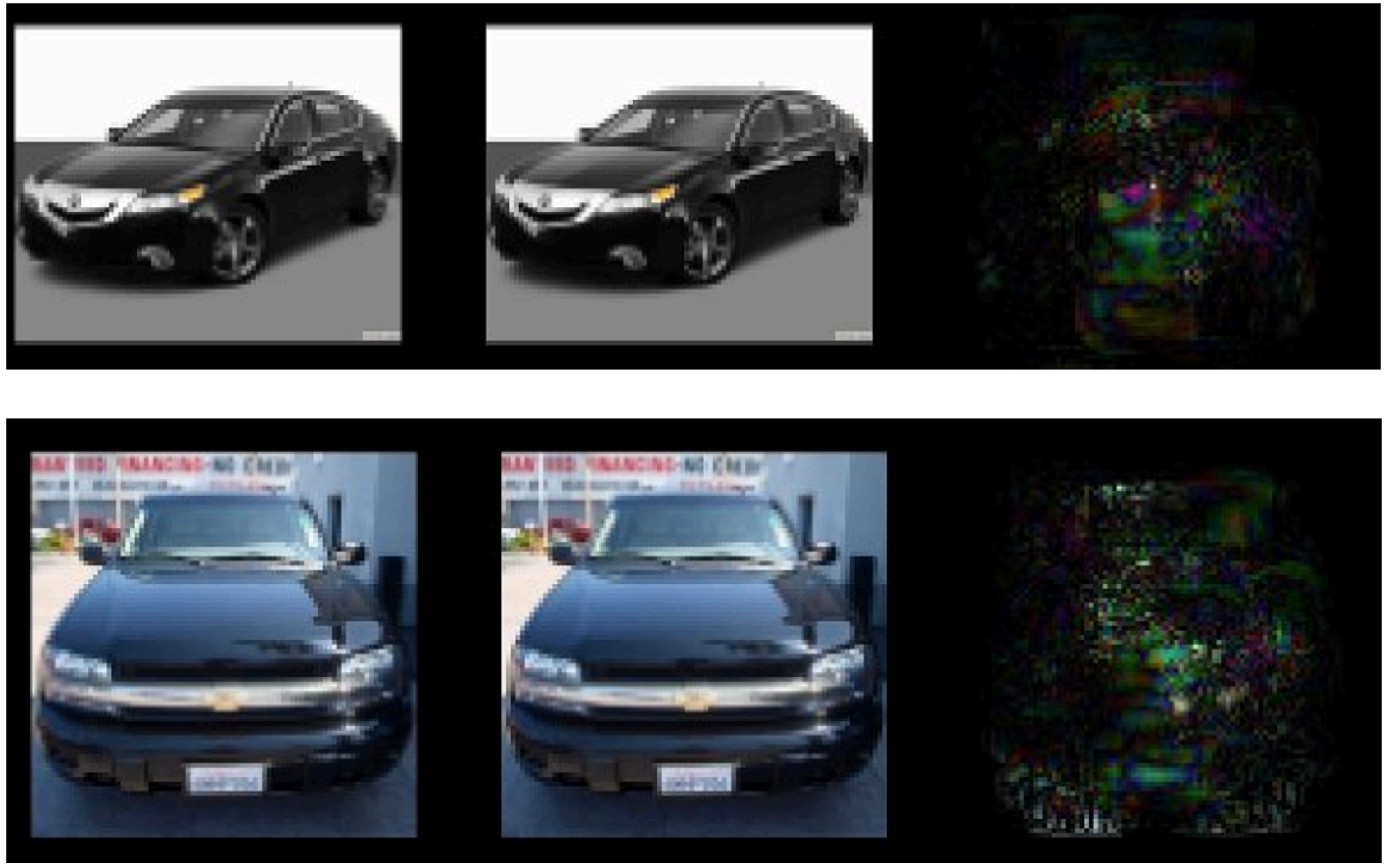
INTRIGUING PROPERTIES OF NEURAL NETWORKS



Szegedy, Christian, et al. "Intriguing properties of neural networks.“ arXiv preprint arXiv: 1312.6199 (2013).

nessler@bioinf.jku.at

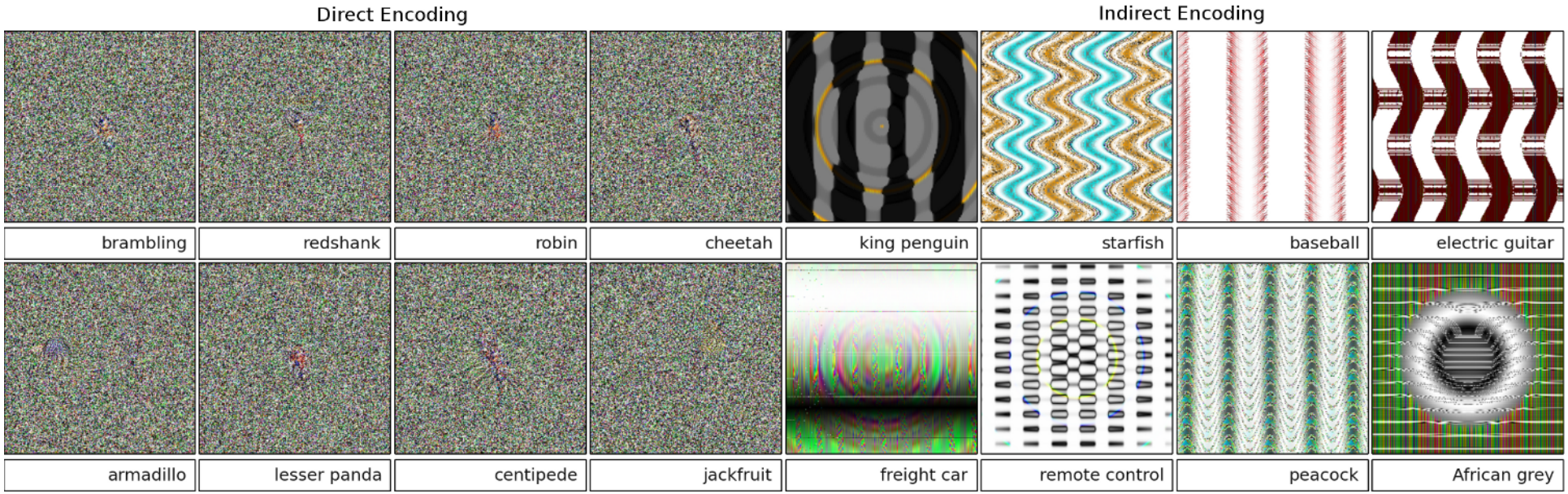
TO CAR OR NOT TO CAR



Szegedy, Christian, et al. "Intriguing properties of neural networks." arXiv preprint arXiv:1312.6199 (2013).

nessler@bioinf.jku.at

DEEP NEURAL NETWORKS ARE EASILY FOOLED



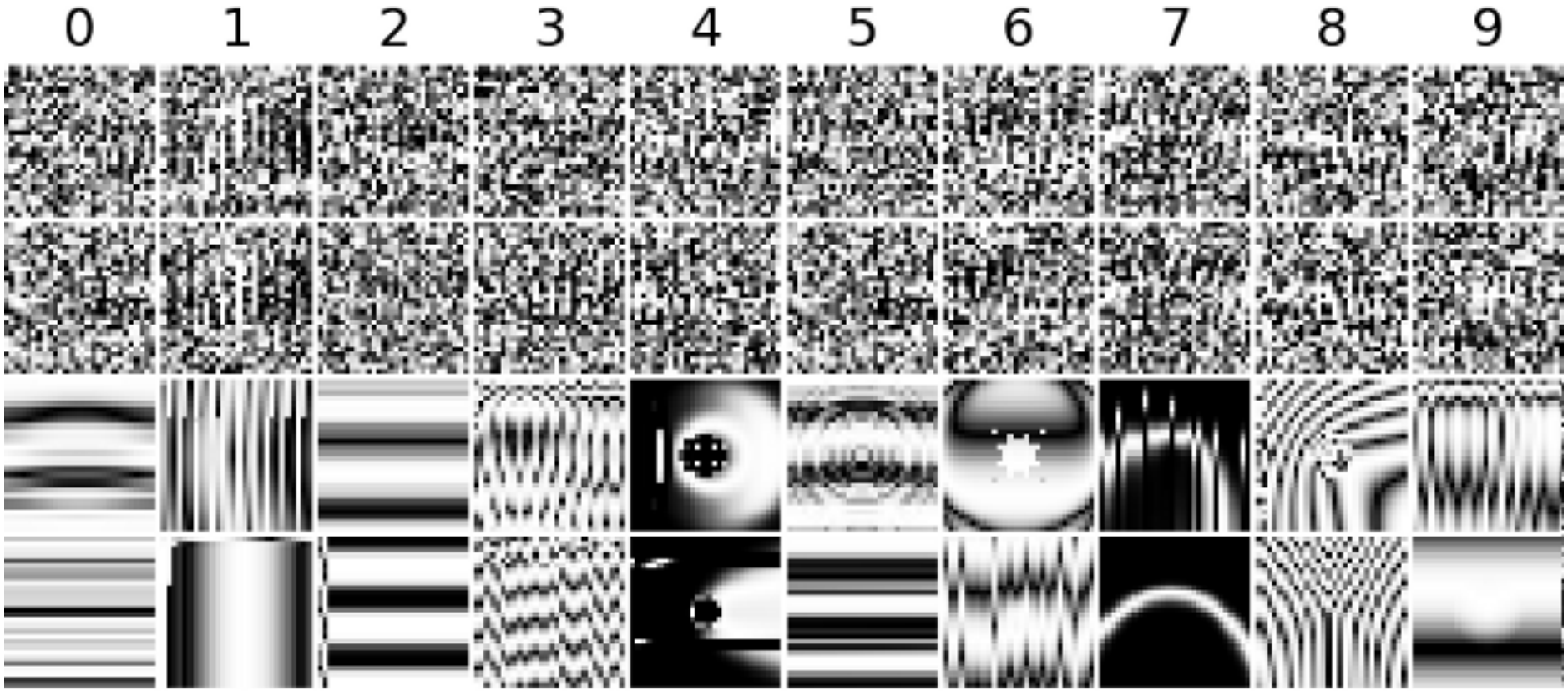
Nguyen etal. CVPR '15, 2015.

nessler@bioinf.jku.at

MNIST DIGITS

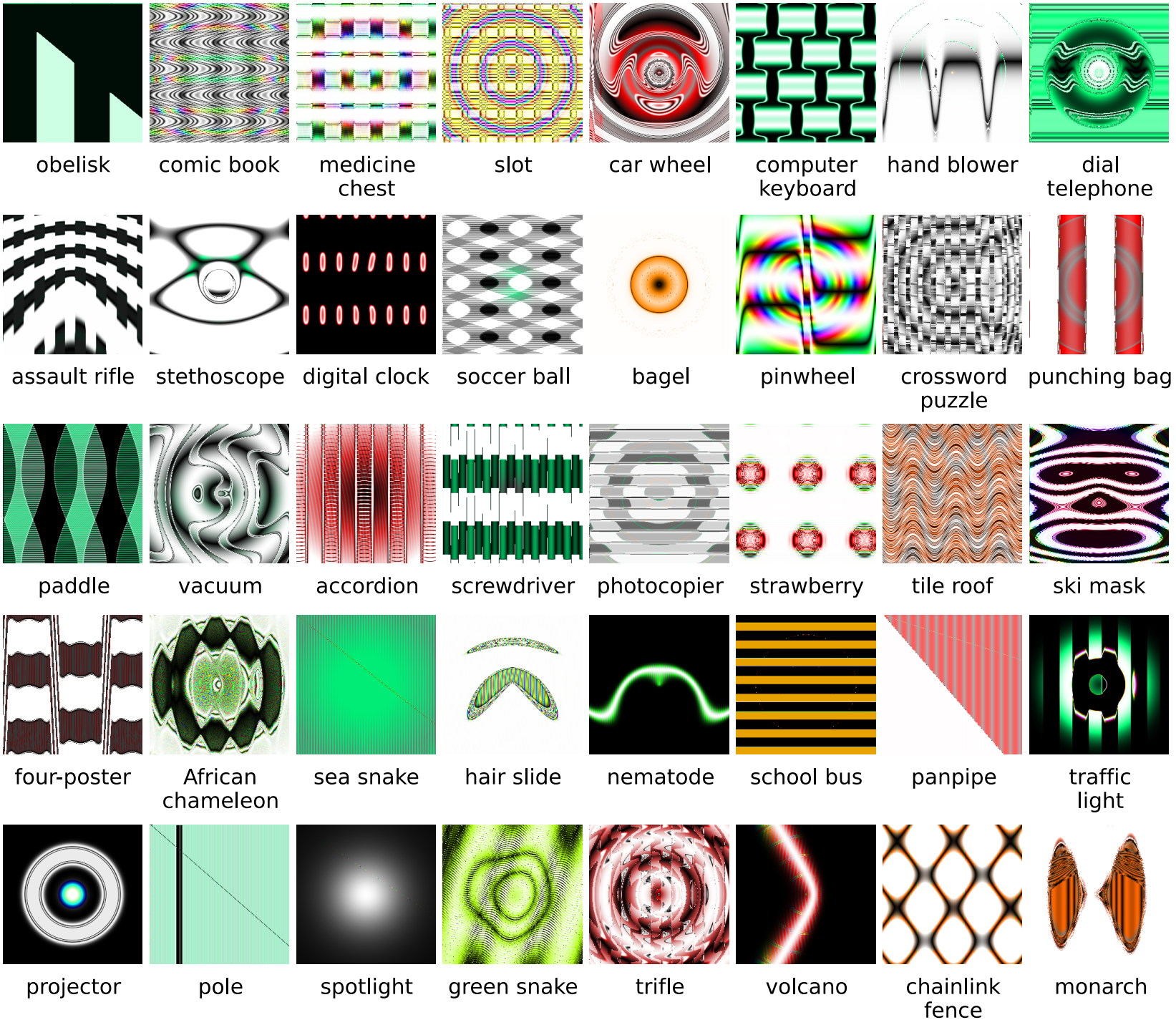
nessler@bioinf.jku.at

MNIST DIGITS



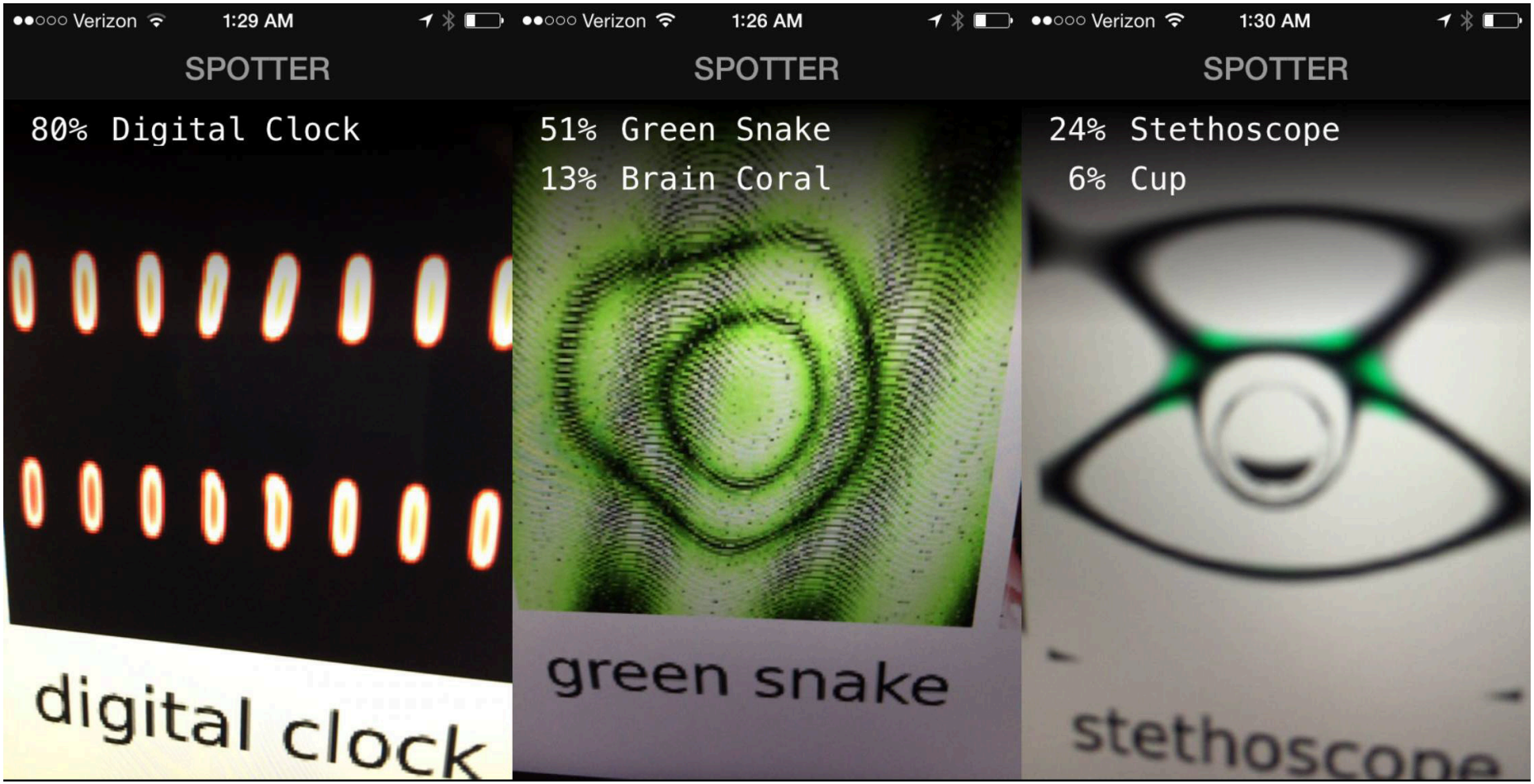
Nguyen etal. CVPR '15, 2015.

nessler@bioinf.jku.at



Nguyen etal. CVPR '15, 2015.

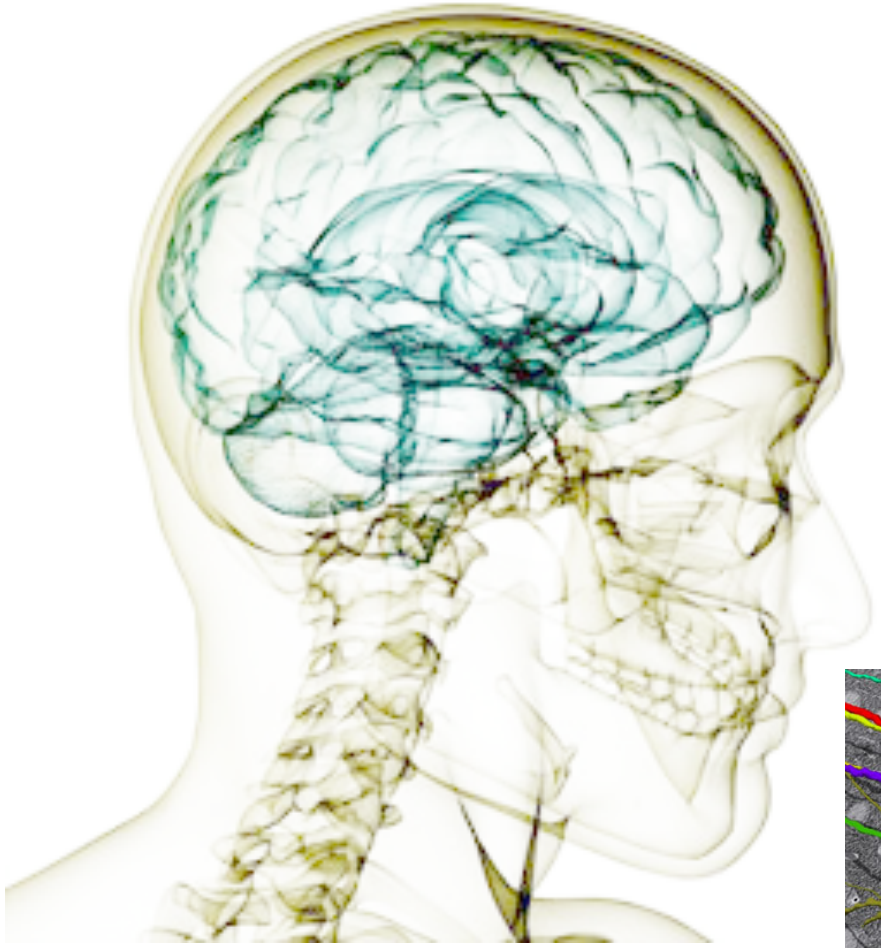
Foto of fooling image



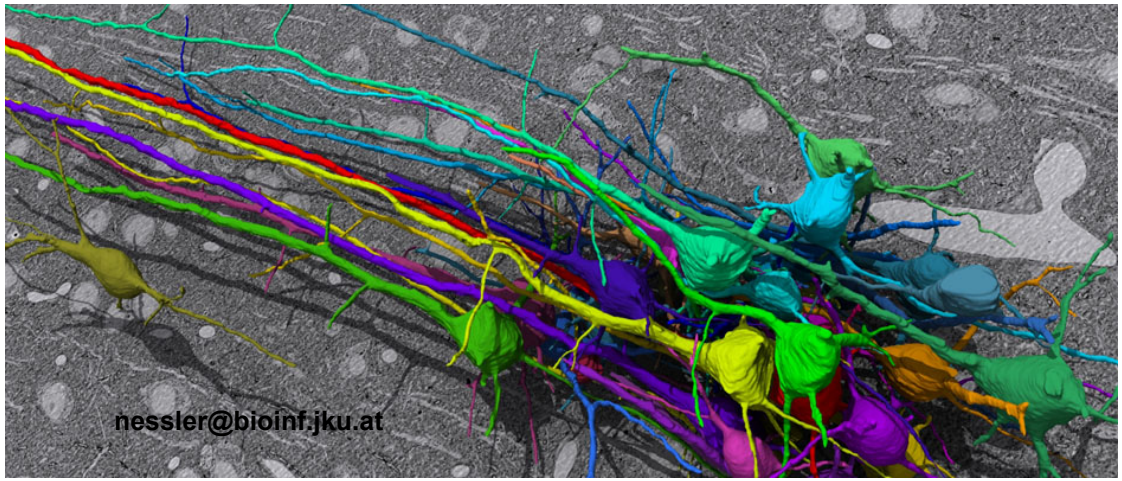
Nguyen etal. CVPR '15, 2015.

nessler@bioinf.jku.at

The Brain

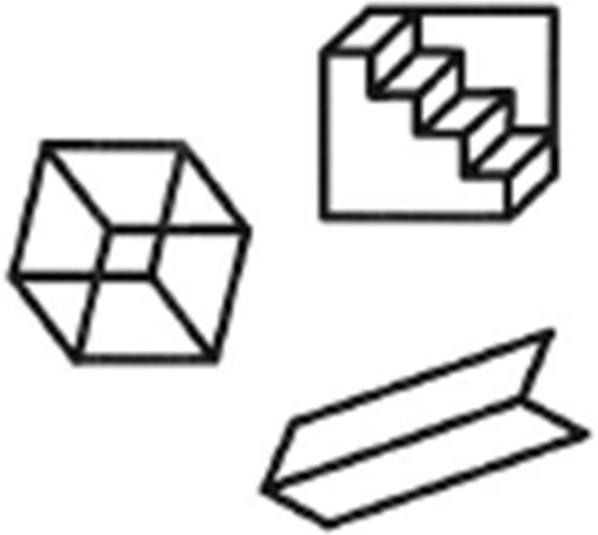


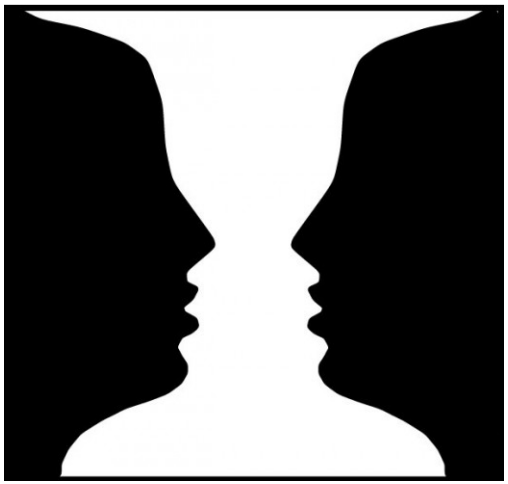
- Weight: 1400 grams
- Neurons: 86 Billion
- Cortex: 19 Billion
- Synapses: ~150 Trillion



JKU

Ambiguity



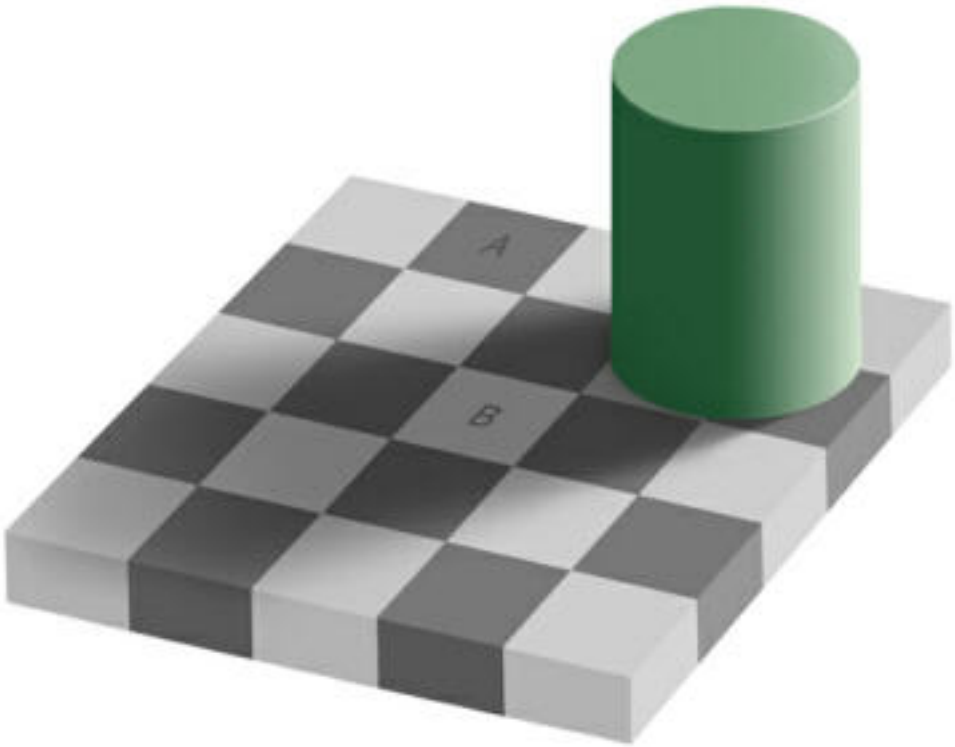


The brain deals with uncertainty in ambiguous visual stimulation (switching percept).

JKU

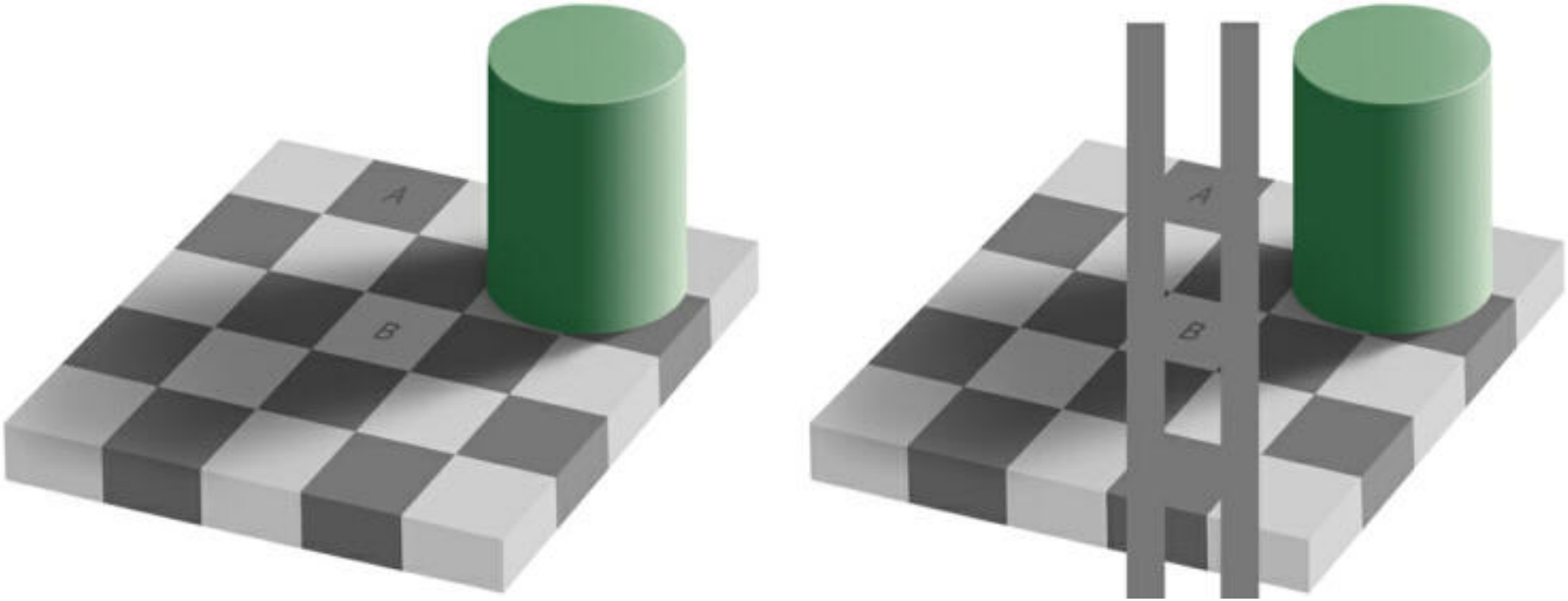
nessler@bioinf.jku.at

Context sensitive Perception



nessler@bioinf.jku.at

Context sensitive Perception



nessler@bioinf.jku.at

Context sensitive Perception

nessler@bioinf.jku.at

PROJEKTS

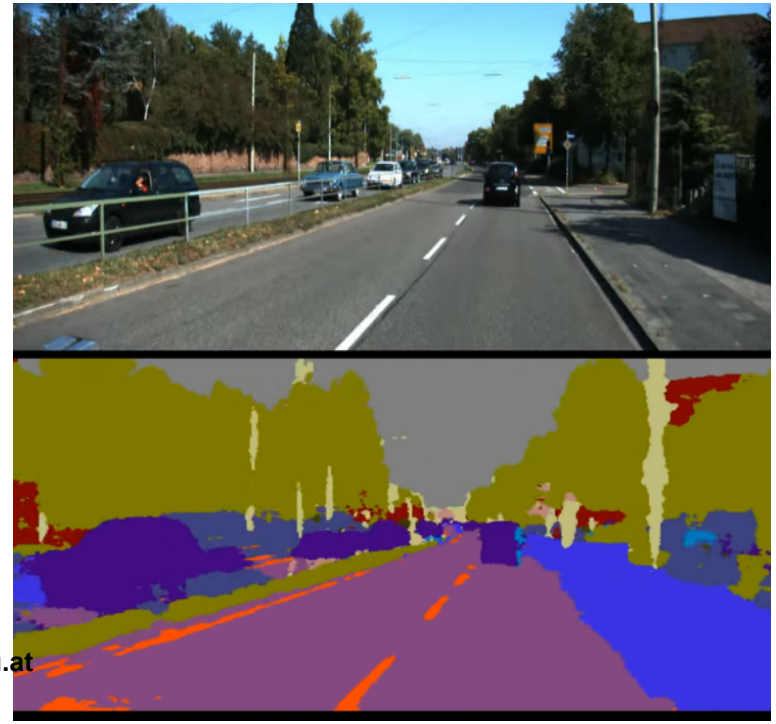
Audi | JKU
deep learning center

LIT AI LAB

NVIDIA

Deep learning for autonomous driving

- Prediction of situations (LSTM)
- Attention (LSTM)
- Sensor (GPS, Radar, Lidar) fusion
- Rare events, dangerous situations



JKU

nessler@bioinf.jku.at

TAKE HOME MESSAGE

- Deep Learning created the revolution in AI. Sepp Hochreiter has pioneered this development (vanishing gradient, LSTMs).
- Deep Learning is a gradual change of parameters (weights) of a neural network according to a gradient upon a teacher signal.

- The internal “understanding” of an AI is (still) completely different from the subjective internal understanding of humans.
- AI-Systems make mistakes that are hard to predict for humans.
- We are working towards a convergence of human and machine understanding of the reality with the hope that this will lead us to the Generalized AI.

nessler@bioinf.jku.at

JOHANNES KEPLER
UNIVERSITY LINZ

THANK YOU

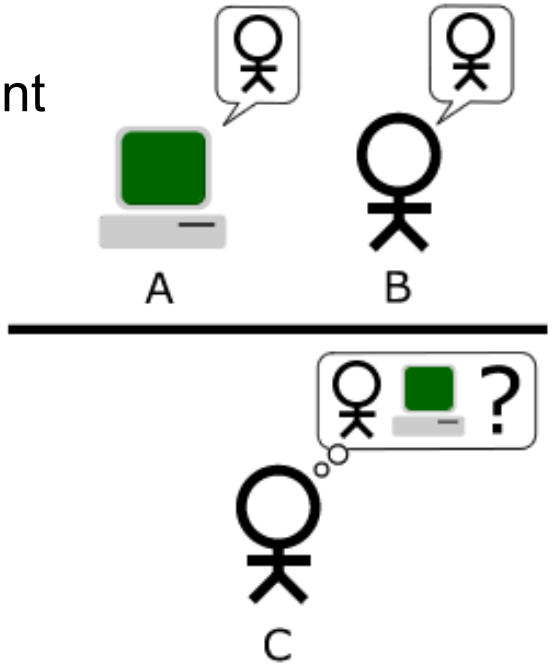
JOHANNES KEPLER
UNIVERSITY LINZ
Altenberger Strasse 69
4040 Linz, Austria
www.jku.at

nessler@bioinf.jku.at

WHAT IS A.I.?

WHAT IS A.I. ?

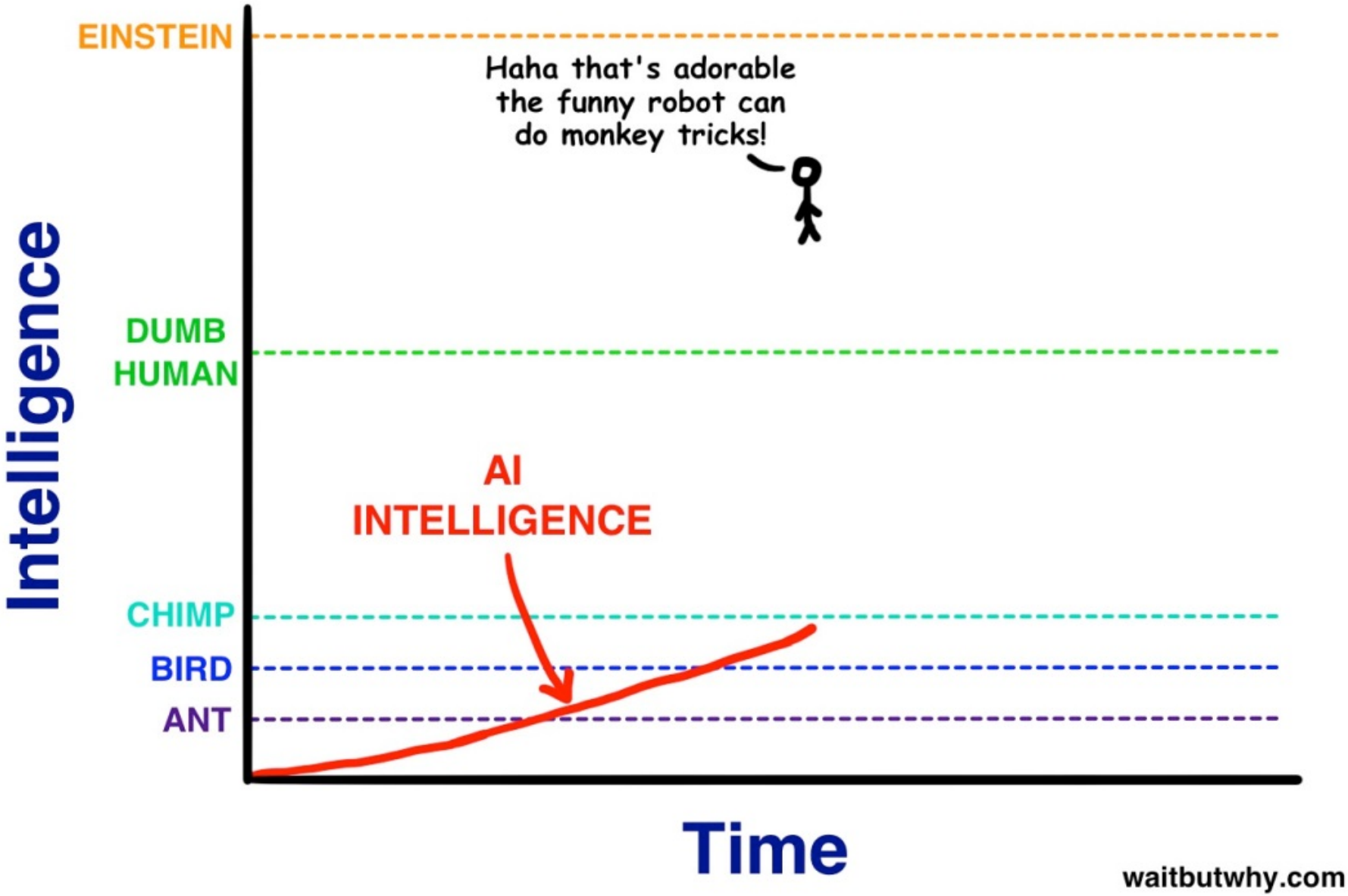
- the capability of a machine to imitate intelligent human behavior (Merriam Webster)
- „I know it, when I see it“
- Turing Test:
  - 1 human (C) has to tell apart an A.I. chatbot (A) from a human chatter (B) while talking to both simultaneously



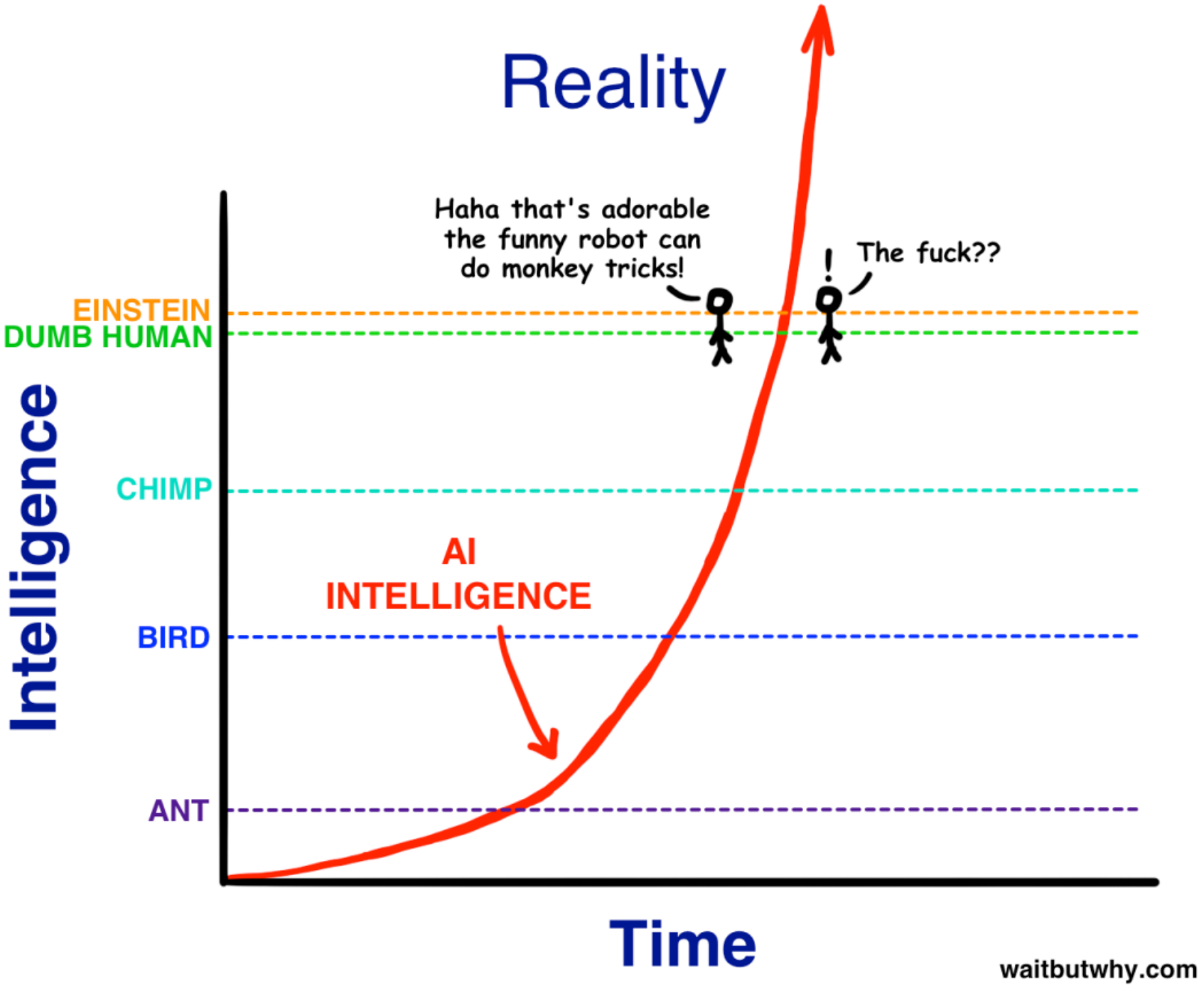
- 1966: Eliza, Joseph Weizenbaum
- 2011: Cleverbot, Rollo Carpenter (unilateral Turing test)
  - Cleverbot got 59,3%, humans achieved 63 %

nessler@bioinf.jku.at

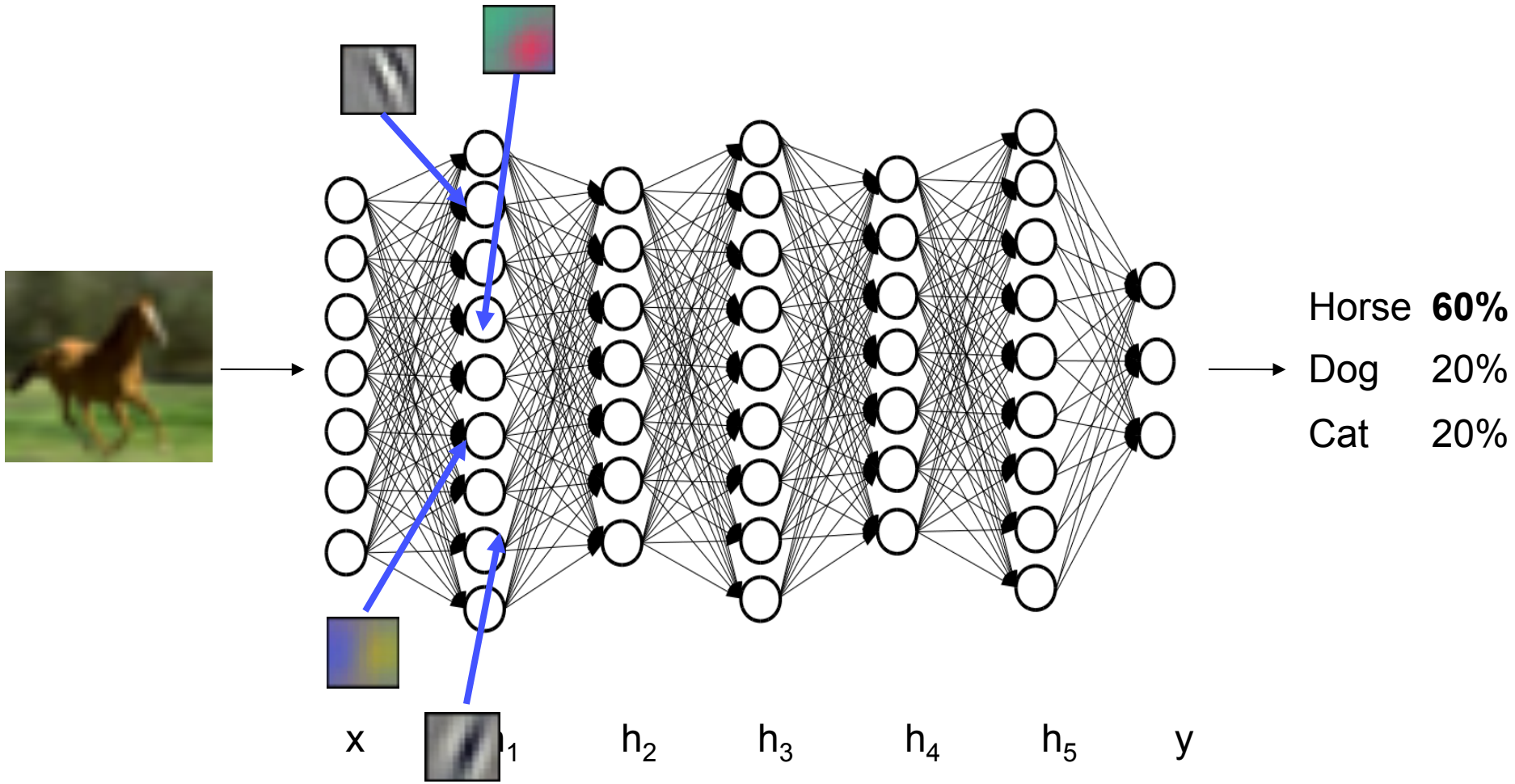
Our Distorted View of Intelligence



waitbutwhy.com

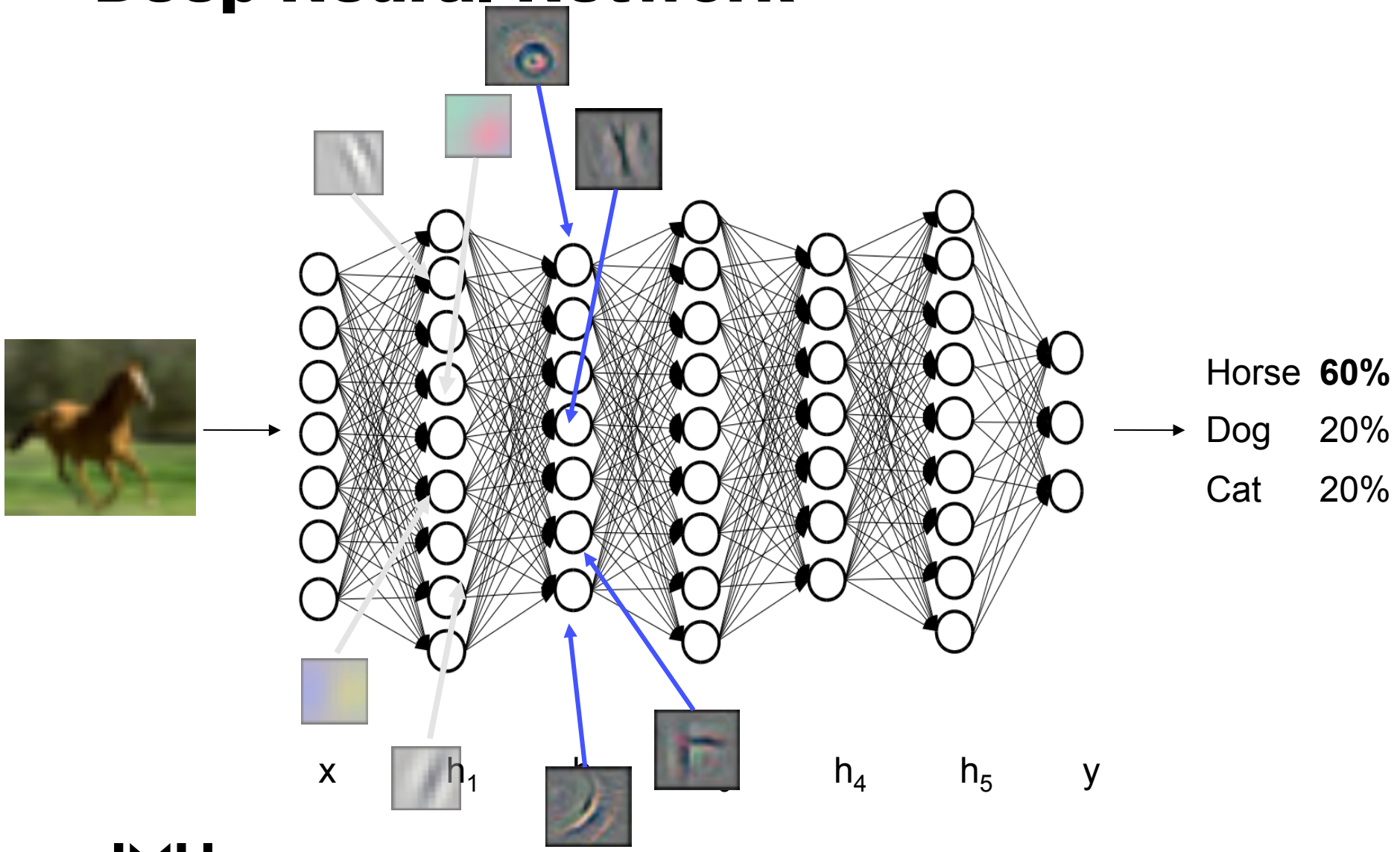

Reality
Haha that's adorable the funny robot can do monkey tricks!
The fuck??
EINSTEIN
DUMB HUMAN
CHIMP
AI INTELLIGENCE
BIRD
ANT
Intelligence
Time
waitbutwhy.com

A closer look



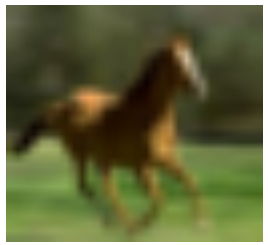
nessler@bioinf.jku.at

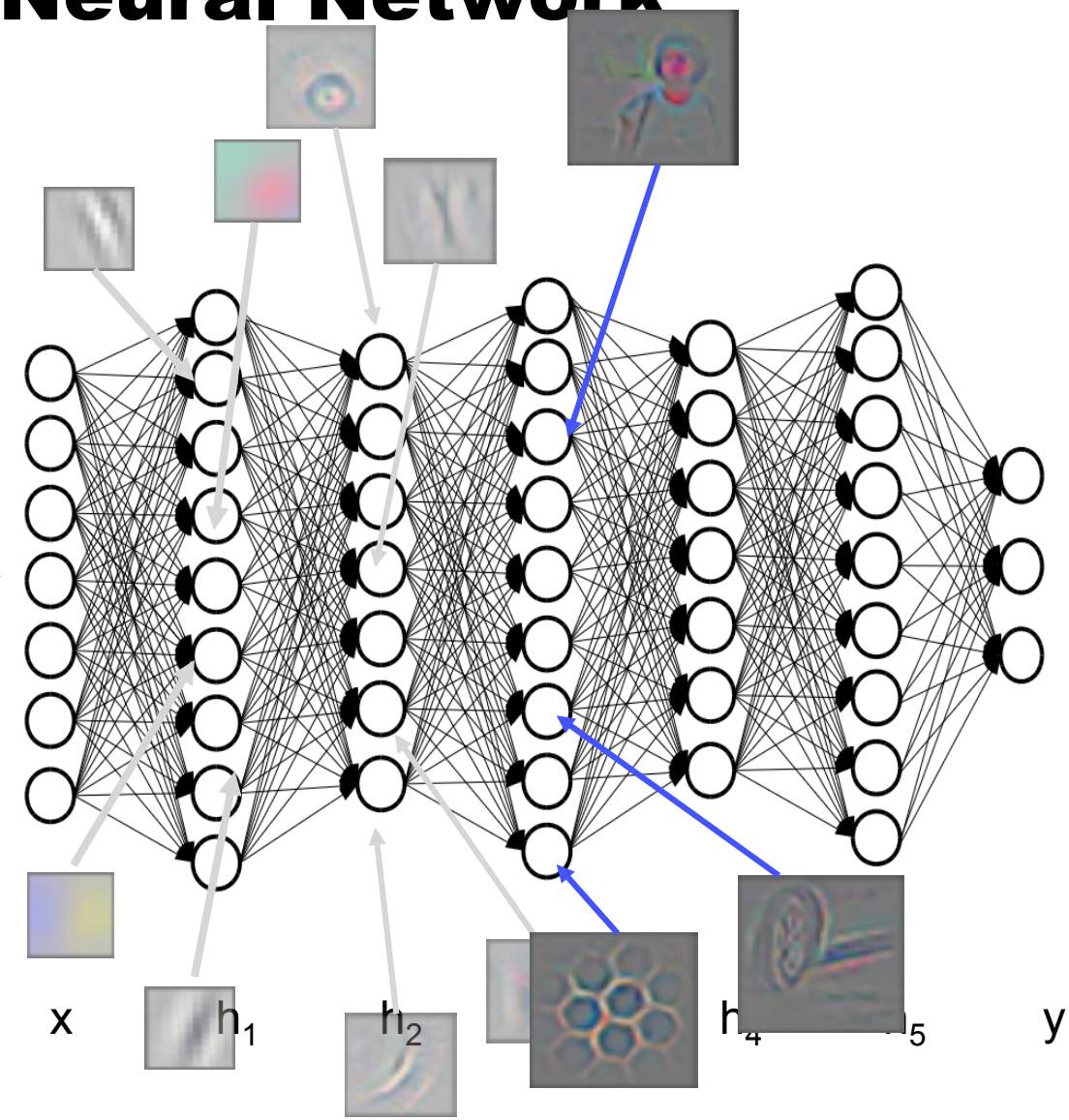
Deep Neural Network



nessler@bioinf.jku.at

Deep Neural Network





Horse **60%**
Dog 20%
Cat 20%

JKU

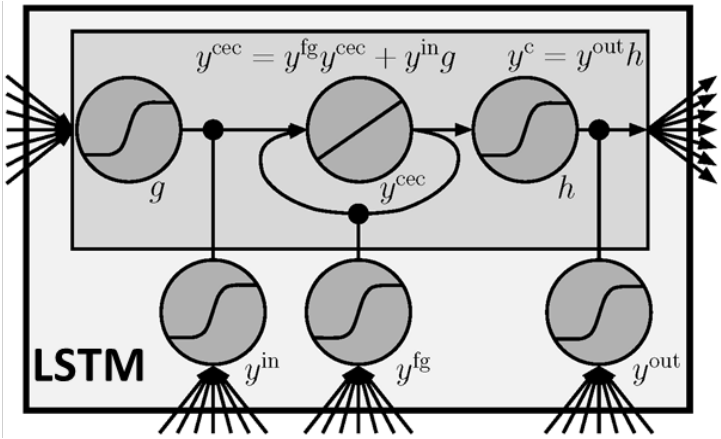
nessler@bioinf.jku.at

WHAT I DID NOT TALK ABOUT:

- Activation function of the neurons
  - ReLU: gradient is always 0 or 1
  - Latest idea from us: scaled Exponential Linear Unit (sELU)
- Stochastic gradient descent, convergence
- Existence or nonexistence of local minima
- initialization of parameters, data augmentation
- overfitting and regularisation, drop-out
- Convolutional structure
- Reinforcment Learning, GANs, …

nessler@bioinf.jku.at

Long Short-Term Memory



- **1991: invented by Hochreiter**
- 1997: major publication Hochreiter&Schmidhuber
- 2009: wins challenges French & Arabic handwriting
- 2011: wins offline Chinese handwriting competition
- 2012: in **Google's Android speech recognizer**
- 2015: in **Google's Voice transcription**
- 2016: in **Apple's iOS 10 → Quicktype**
- 2016: in **Google's Translate**