Deep Learning Revolutionizes Information Technology

Sepp Hochreiter
Institute of Bioinformatics
Johannes Kepler University Linz

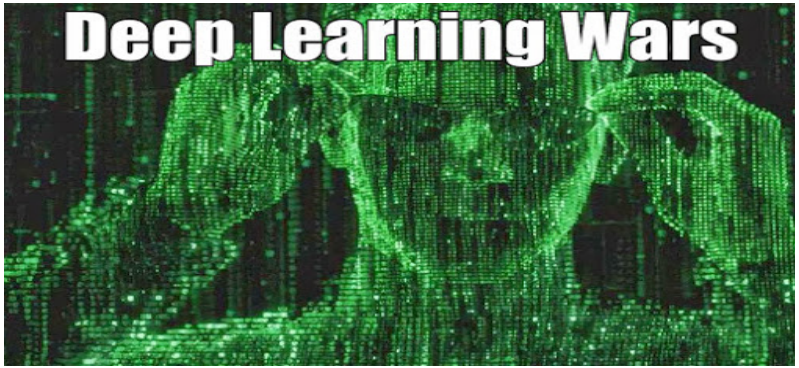
Who are we?


GERMANY
Stockerau
Linz
Vienna
Salzburg
Seefeld
Innsbruck
Tux
AUSTRIA
Graz
HUNGARY
ITALY

Deep Learning

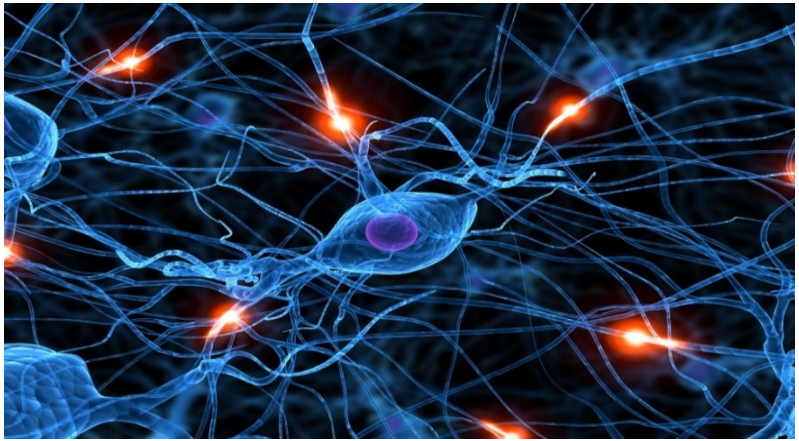
Companies

Google acquired the deep learning start-up DeepMind for $500M, winning bidding against facebook



Google, facebook, Baidu, Amazon, and Microsoft see deep learning as the key technology for

- image processing (object rec.)
- text and language (translation)
- speech



Deep Learning

Media

MIT Technology Review: one of the 10 tech breakthroughs in 2013

The New York Times: two front-page articles

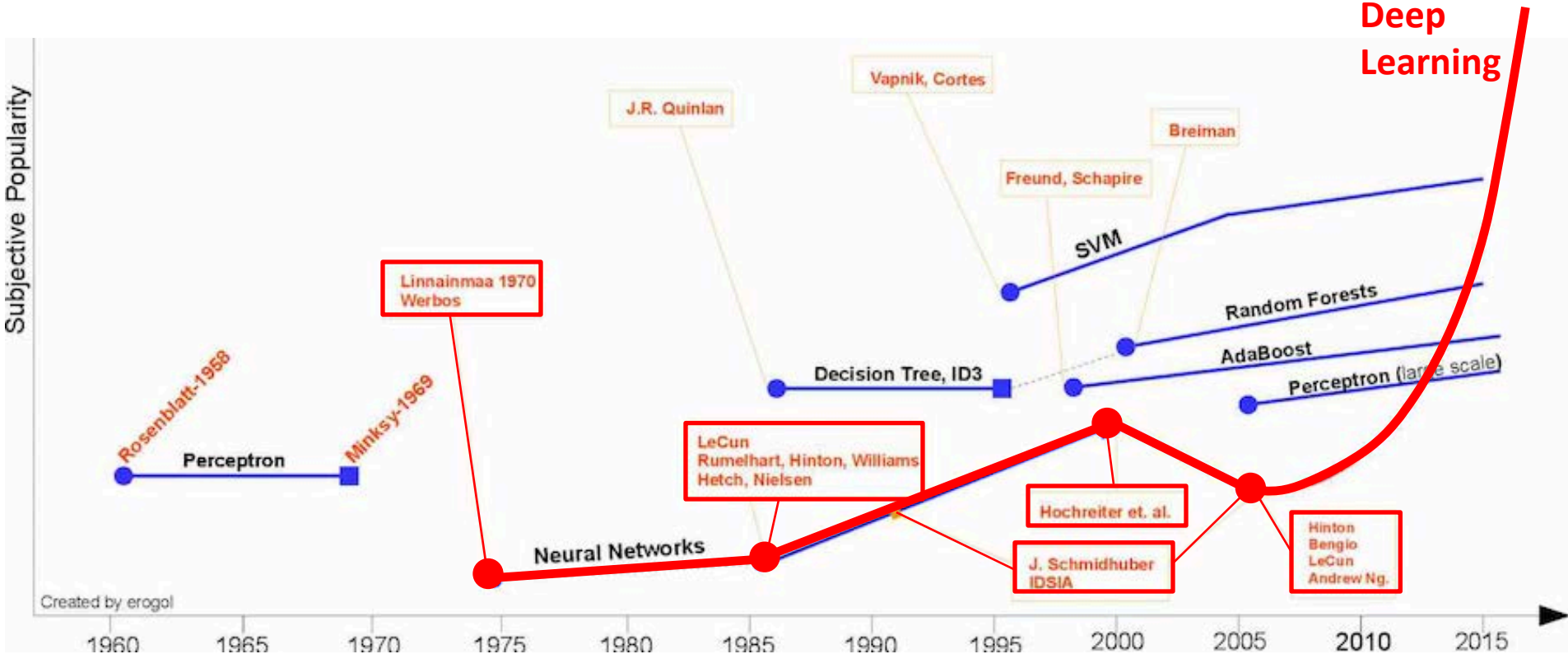
The New York Times

November 23, 2012

Scientists See Promise in Deep-Learning Programs

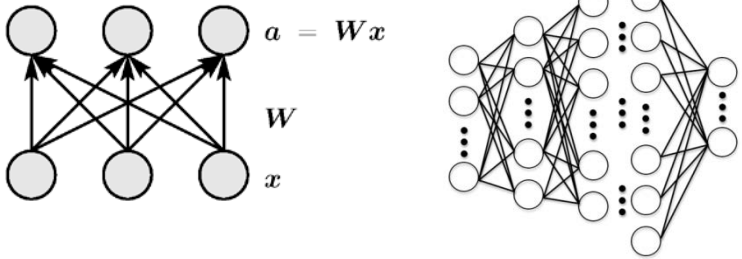
In October, for example, a team of graduate students studying with the University of Toronto computer scientist Geoffrey E. Hinton won the top prize in a contest sponsored by Merck to design software to help find molecules that might lead to new drugs.

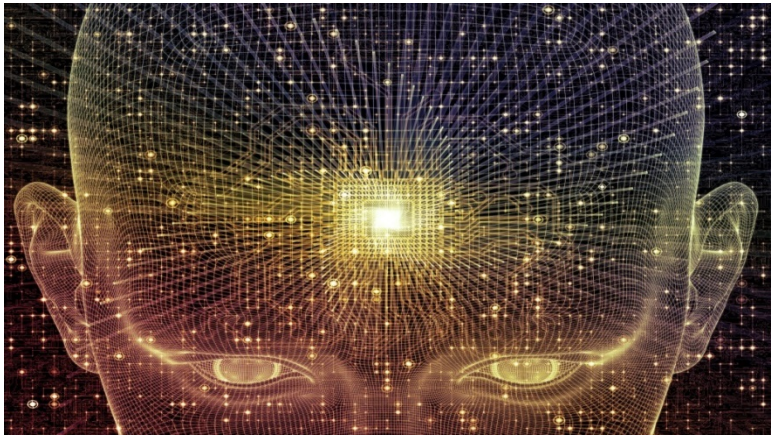
Deep Learning



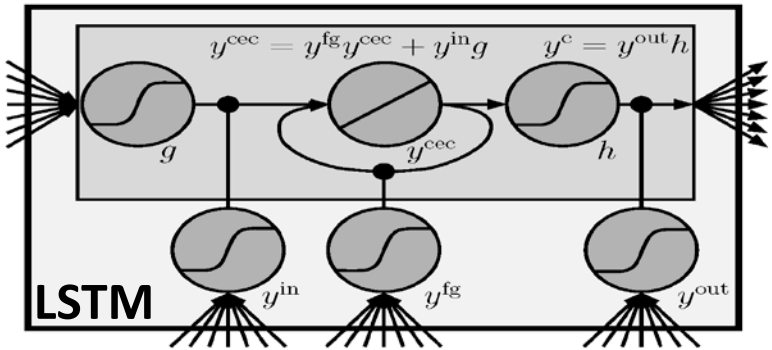
Deep Learning

- Neural networks + very fast computers + massive data sets





- Multiple levels of representations
  → higher levels = abstract concepts



- Started already in 1991 with LSTM by Hochreiter and Schmidhuber

Deep Learning

KEY MOMENTS IN DEEP-LEARNING HISTORY 1989-1997

1989
French researcher Yann LeCun, then at Bell Labs, begins foundational work on a type of neural net that becomes crucial for image recognition.

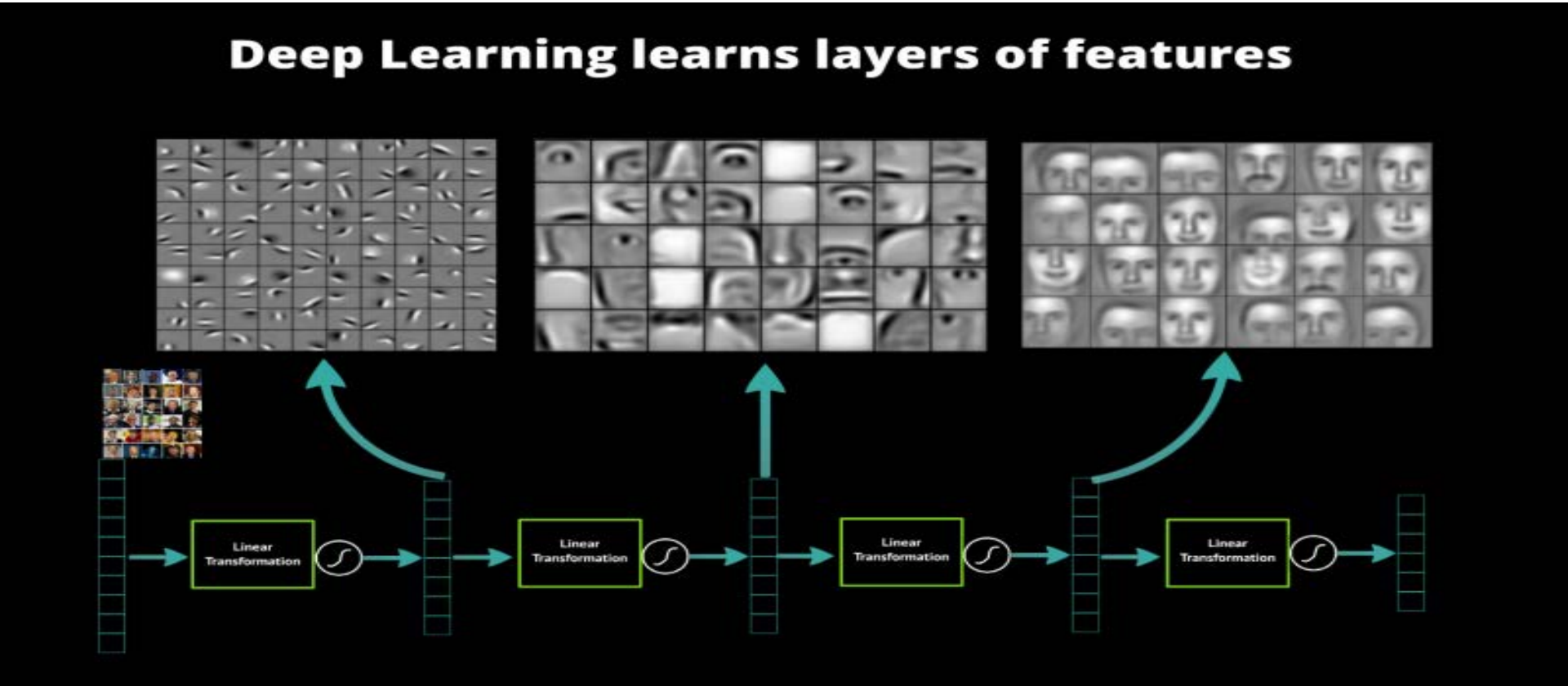
1991
German researchers Sepp Hochreiter and Jürgen Schmidhuber pioneer a neural net with memory features, which eventually proves superior for natural-language processing.

1997
IBM's Deep Blue beats **world champion Garry Kasparov** (right) in chess using traditional AI techniques.

Fortune, 28.09.2016

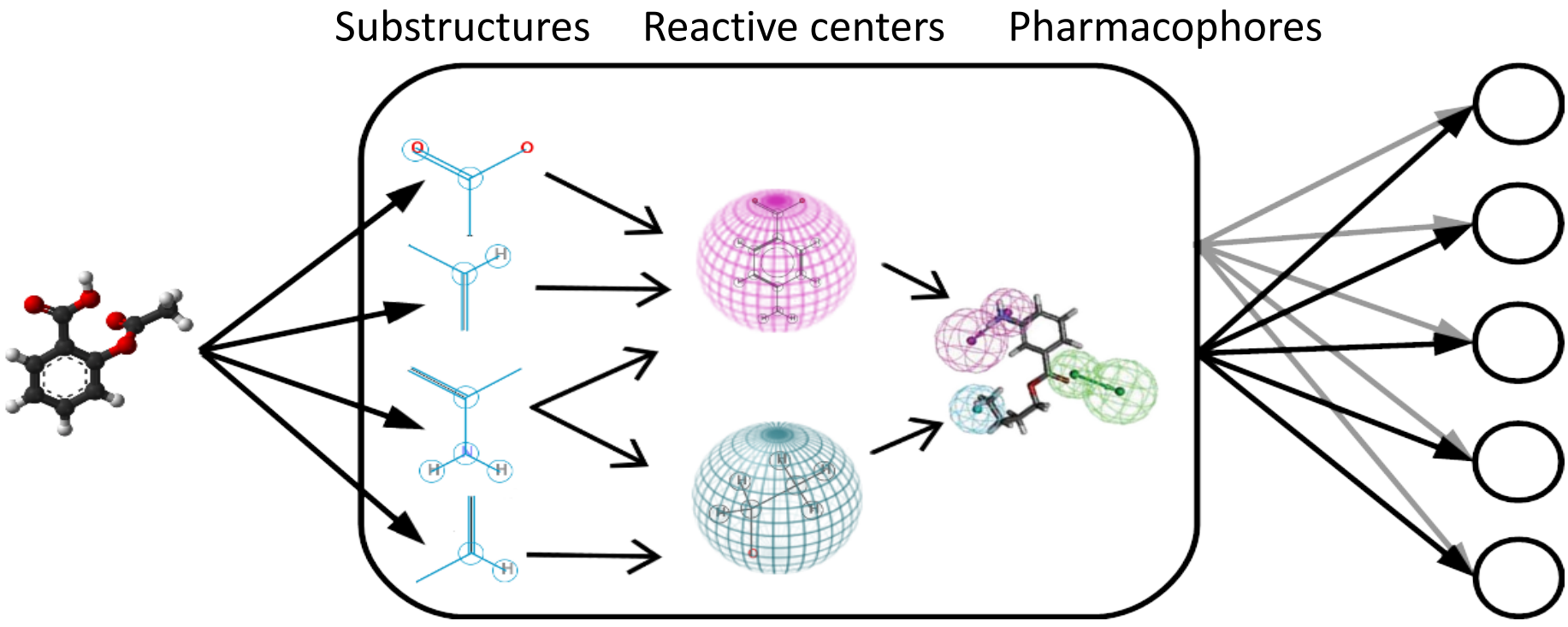
STAN HONDA—AFP/GETTY IMAGES

Deep Learning: Vision



Deep Learning: Drug Design

Abstract representations



Records of Deep Learning

Vision

- 2009: deep networks won French, Arabic, and Chinese handwriting competition (all LSTM nets).
- 2011: deep learning won the German traffic sign recognition contest with **super-human performance:** deep networks 0.56% vs. humans 1.16% error (best non-neural: 3.86%).
- 2012: deep nets won the brain image segmentation challenge with **super-human pixel error rate**.
- 2013: deep networks won the grand challenge on mitosis detection.
- **2013: ImageNet** competition was won by a deep neural network using GPUs. **16%** error at 1.2M images vs. **26%** by its closest competitor.
- 2013: Google researchers obtained a 70% improvement over previous best results on face recognition using 16,000 cores.

Speech

- 2013: Google researchers obtained a 70% improvement over previous best results on face recognition using 16,000 cores.
- **2011**: lowered the error rate of speech recognition on an industry benchmark data set from **24%** to about **16%**. The speech recognition community has been stuck at 24% for more than a decade. Industrial companies which business is related to information technology became interested.

Records of Deep Learning

Audio

TIMIT Phone classification	Accuracy
Prior art (Clarkson et al.,1999)	79.6%
Stanford Feature learning	**80.3%**

TIMIT Speaker identification	Accuracy
Prior art (Reynolds, 1995)	99.7%
Stanford Feature learning	**100.0%**

Images

CIFAR Object classification	Accuracy
Prior art (Krizhevsky, 2010)	78.9%
Stanford Feature learning	**81.5%**

NORB Object classification	Accuracy
Prior art (Ranzato et al., 2009)	94.4%
Stanford Feature learning	**97.3%**

Video

Hollywood2 Classification	Accuracy
Prior art (Laptev et al., 2004)	48%
Stanford Feature learning	**53%**

YouTube	Accuracy
Prior art (Liu et al., 2009)	71.2%
Stanford Feature learning	**75.8%**

KTH	Accuracy
Prior art (Wang et al., 2010)	92.1%
Stanford Feature learning	**93.9%**

UCF	Accuracy
Prior art (Wang et al., 2010)	85.6%
Stanford Feature learning	**86.5%**

Multimodal (audio/video)

AVLetters Lip reading	Accuracy
Prior art (Zhao et al., 2009)	58.9%
Stanford Feature learning	**65.8%**

Other records:

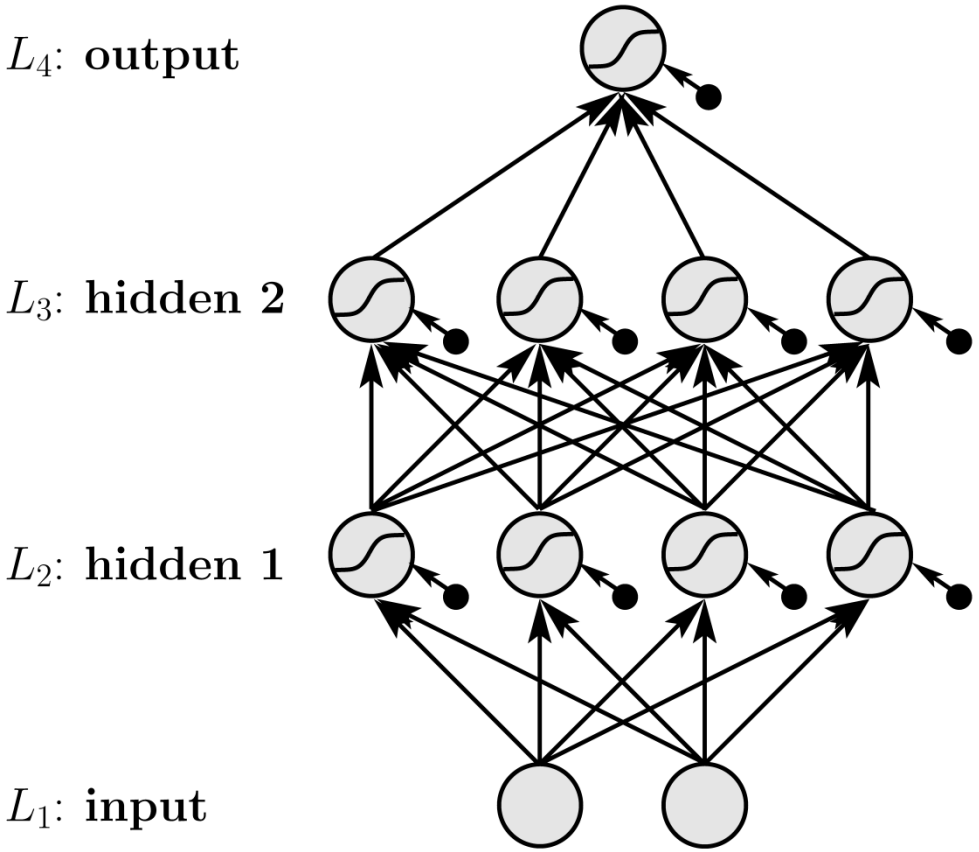
- **Pedestrian detection**
- Different phone recognition task
- PASCAL VOC **object classification**

Vanishing Gradient

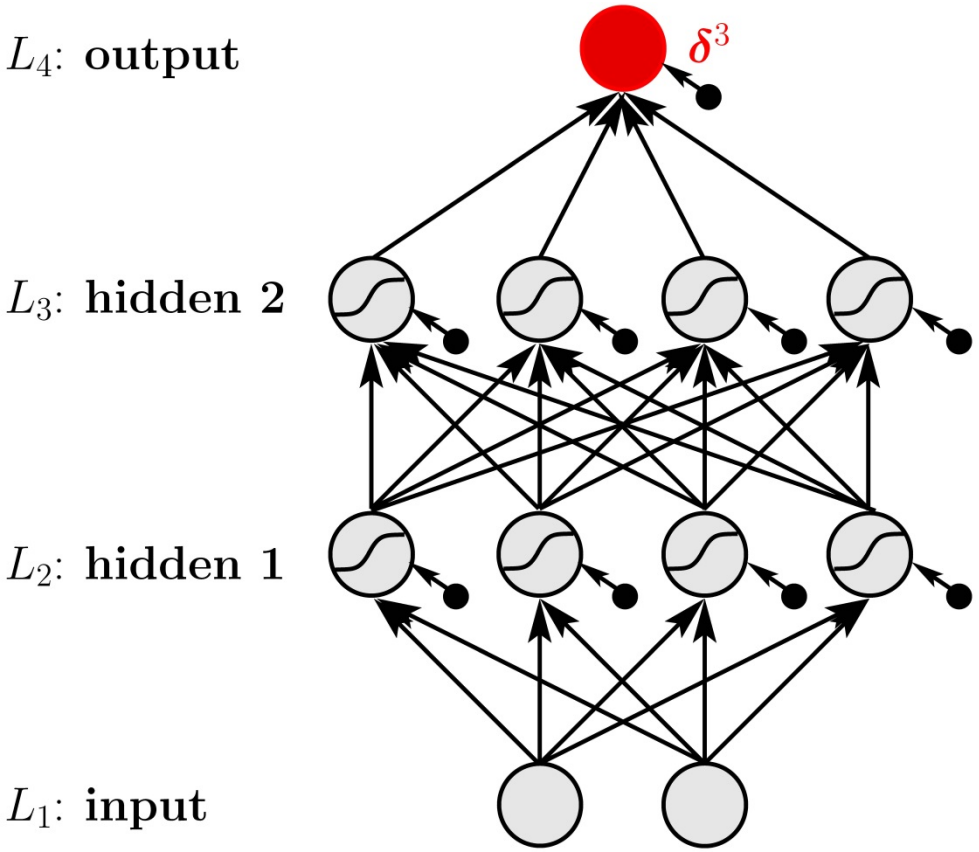
Deep Learning has been enabled by solving the vanishing gradient problem:

- Stacking
- ReLU
- LSTM
- highway nets
- residual nets

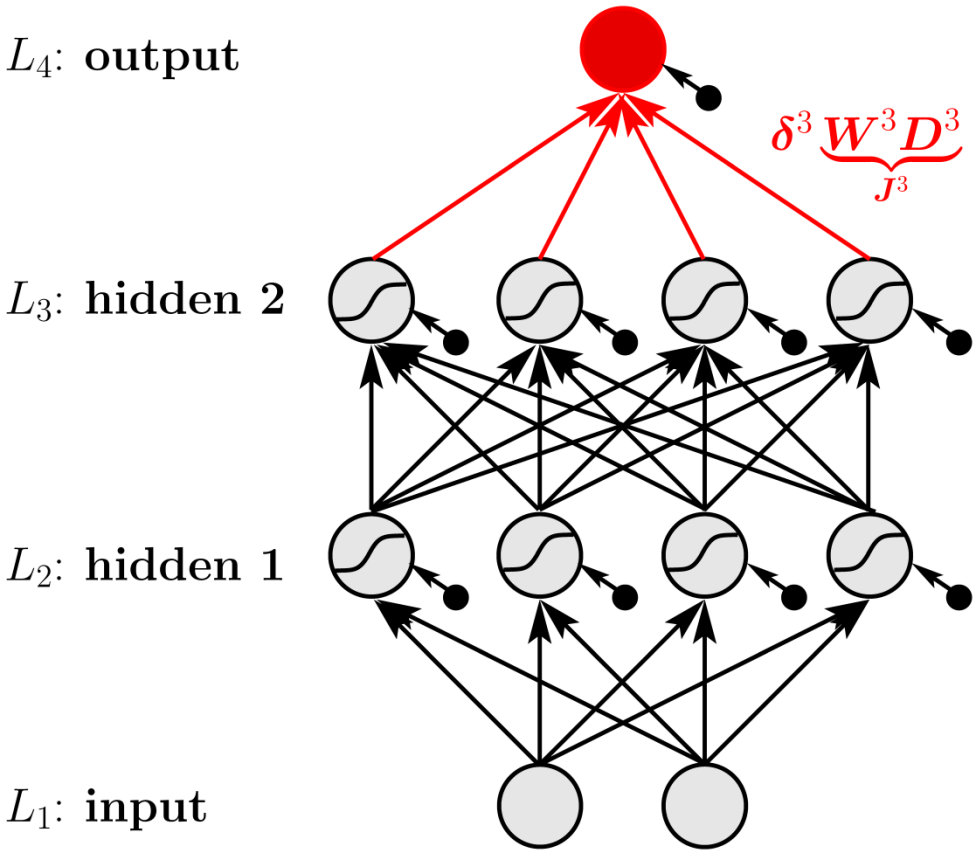
Neural Network



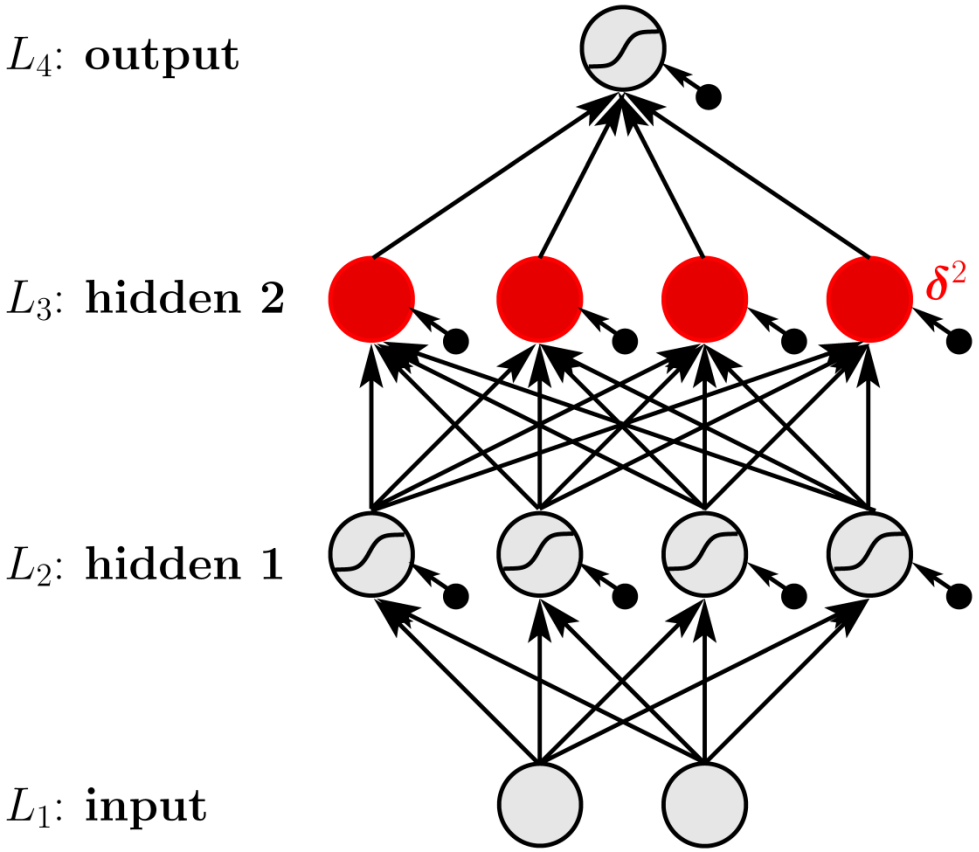
Neural Network



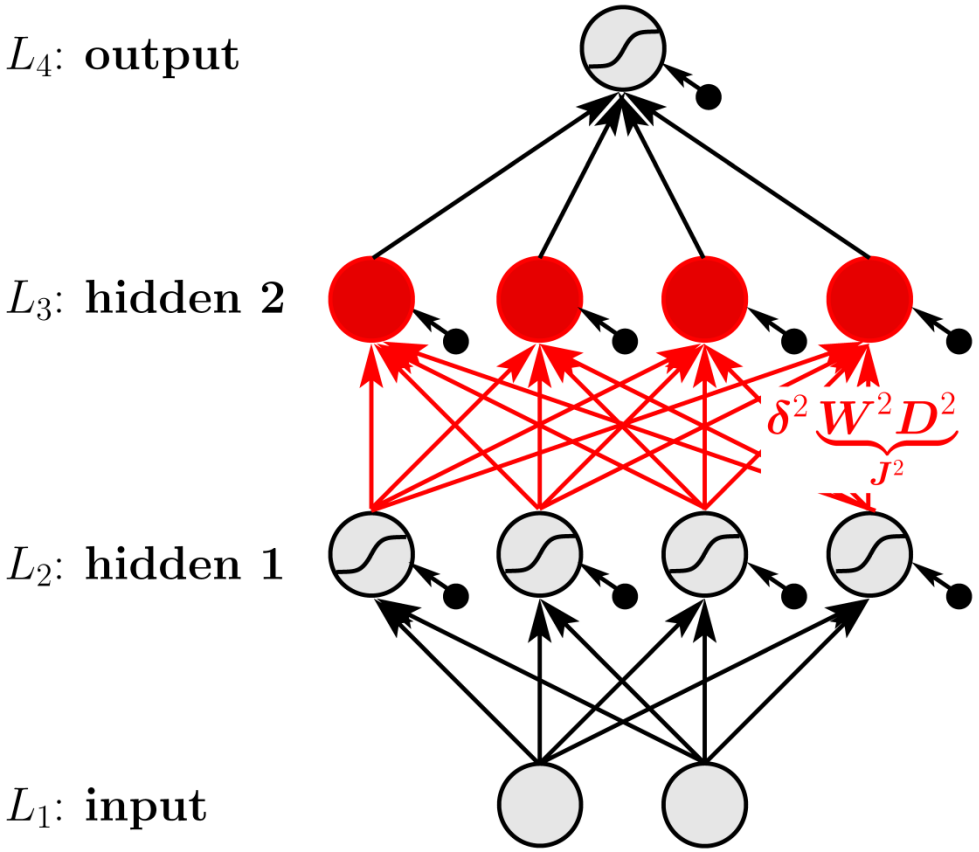
Neural Network



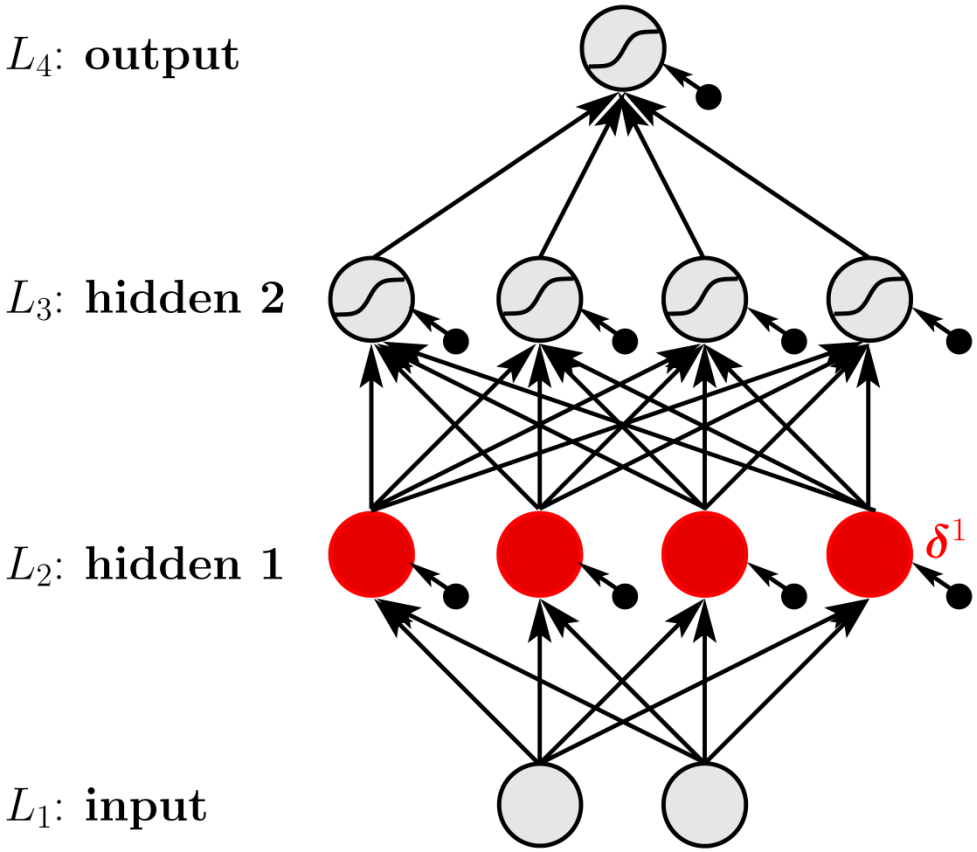
Neural Network



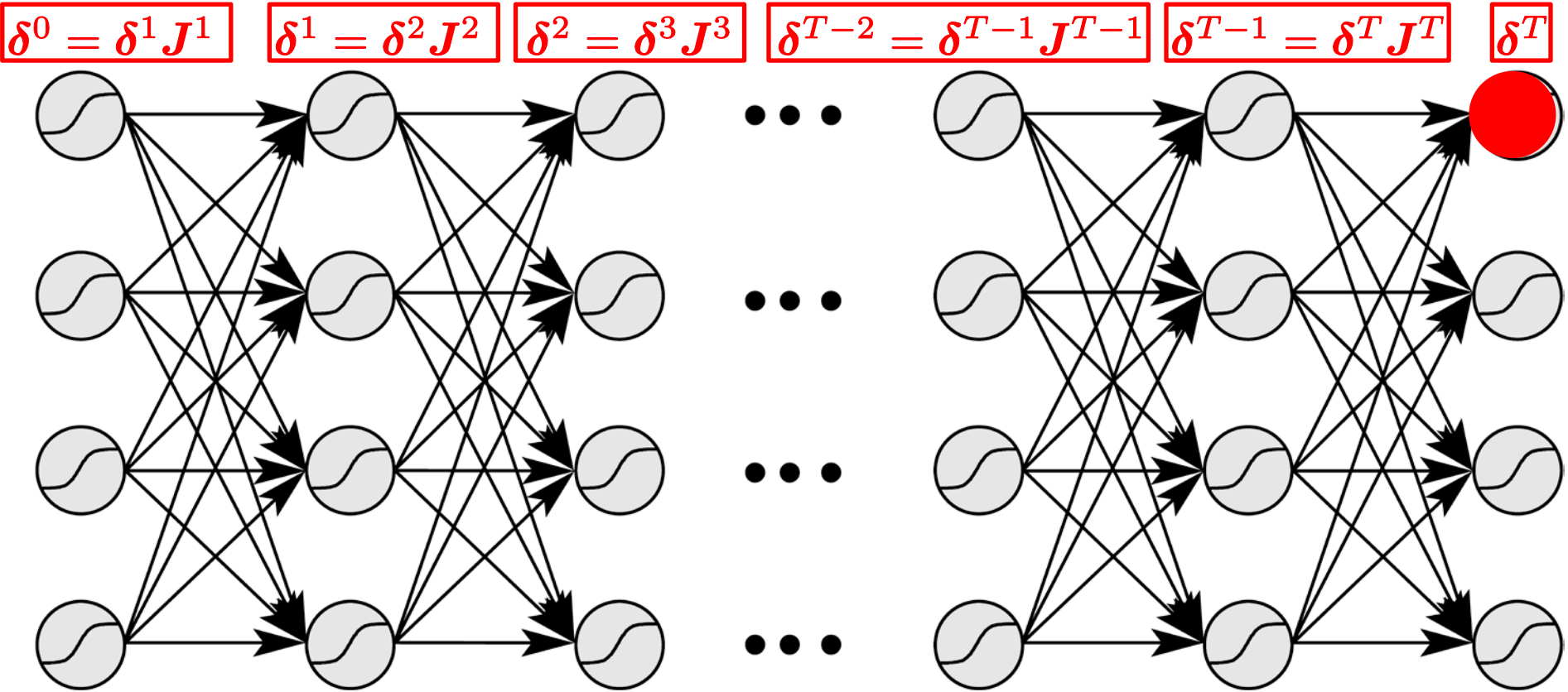
Neural Network



Neural Network



Vanishing Gradient



Vanishing Gradient

δ-propagation: $\boldsymbol{\delta}^{i-1} = \boldsymbol{\delta}^{i} \; J^{i}$ (**backpropagation**)

Stable learning: $\left\| J^{i} \right\| \leqslant k < 1$

$$\left\| \boldsymbol{\delta}^{1} \right\| = \left\| \boldsymbol{\delta}^{T} \prod_{i=T}^{2} J^{i} \right\| \leqslant k^{T-1} \left\| \boldsymbol{\delta}^{T} \right\| \Rightarrow \left\| \boldsymbol{\delta}^{1} \right\| \approx 0$$

Gradient: $\frac{\partial L}{\partial \boldsymbol{W}^{i}} = \boldsymbol{\delta}^{i} \left(\boldsymbol{a}^{i} \right)^{\top}$ → **layer 1 does not learn**

Vanishing Gradient Solutions

- Pre-training (first approach by Hinton 2006)
- Rectified linear units (ReLUs)
  →derivative of 1 and proper initialization
- LSTM (Long Short-Term Memory, recurrent net)
- Highway networks (highlight at NIPS 2015)
- Residual networks (ResNets, winner of Imagenet 2015)

Singular Values Close to One

LSTM, highway network, residual networks

- Singular values close to one
  → no vanishing gradient and volume conserving
- Singular values equal
  → angle preserving

$$\boldsymbol{a}^{i+1} = \boldsymbol{a}^{i} + F(\boldsymbol{a}^{i})$$

Singular Values Close to One

- **LSTM** $a^i = \boxed{a^{i-1} + \underbrace{F(a^{i-1})}_{g(a^{i-1})\ f(a^{i-1})}}$

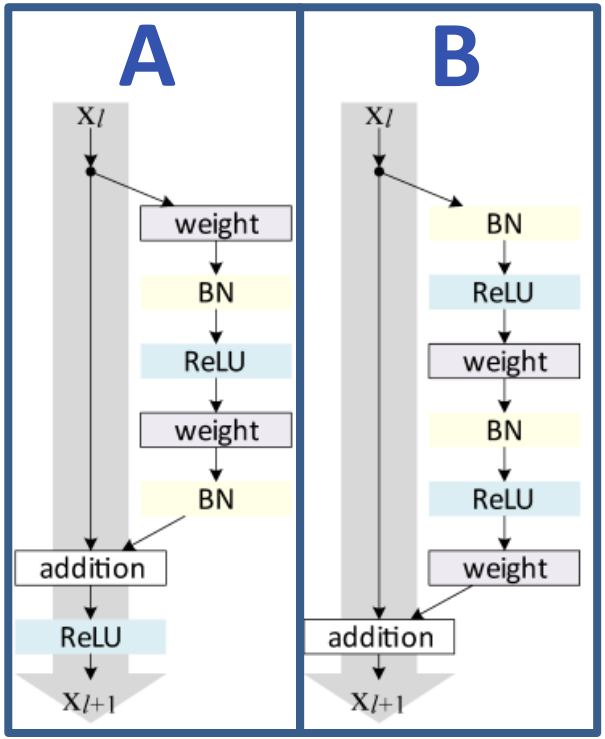
- **Highway networks** (Srivastava et al., 2015)

$$a^i = (1 - g(a^{i-1}))\, a^{i-1} + g(a^{i-1})\, f(a^{i-1}) \approx \boxed{a^{i-1} + F(a^{i-1})} \approx a^{i-1}$$

- **Residual network** (He et al., 2015 and 2016)

$$a^i = g(\boxed{a^{i-1} + F(a^{i-1})}) \quad \textbf{(A)}$$

$$a^i = \boxed{a^{i-1} + F(a^{i-1})} \quad \textbf{(B)}$$



Singular Values Close to One

$$\boldsymbol{a}^i = \boldsymbol{a}^{i-1} + F(\boldsymbol{a}^{i-1})$$

$$\boldsymbol{J}^i = \frac{\partial \boldsymbol{a}^i}{\partial \boldsymbol{a}^{i-1}} = \boldsymbol{I} + \frac{\partial F(\boldsymbol{a}^{i-1})}{\partial \boldsymbol{a}^{i-1}} = \boldsymbol{I} + \boldsymbol{X}$$

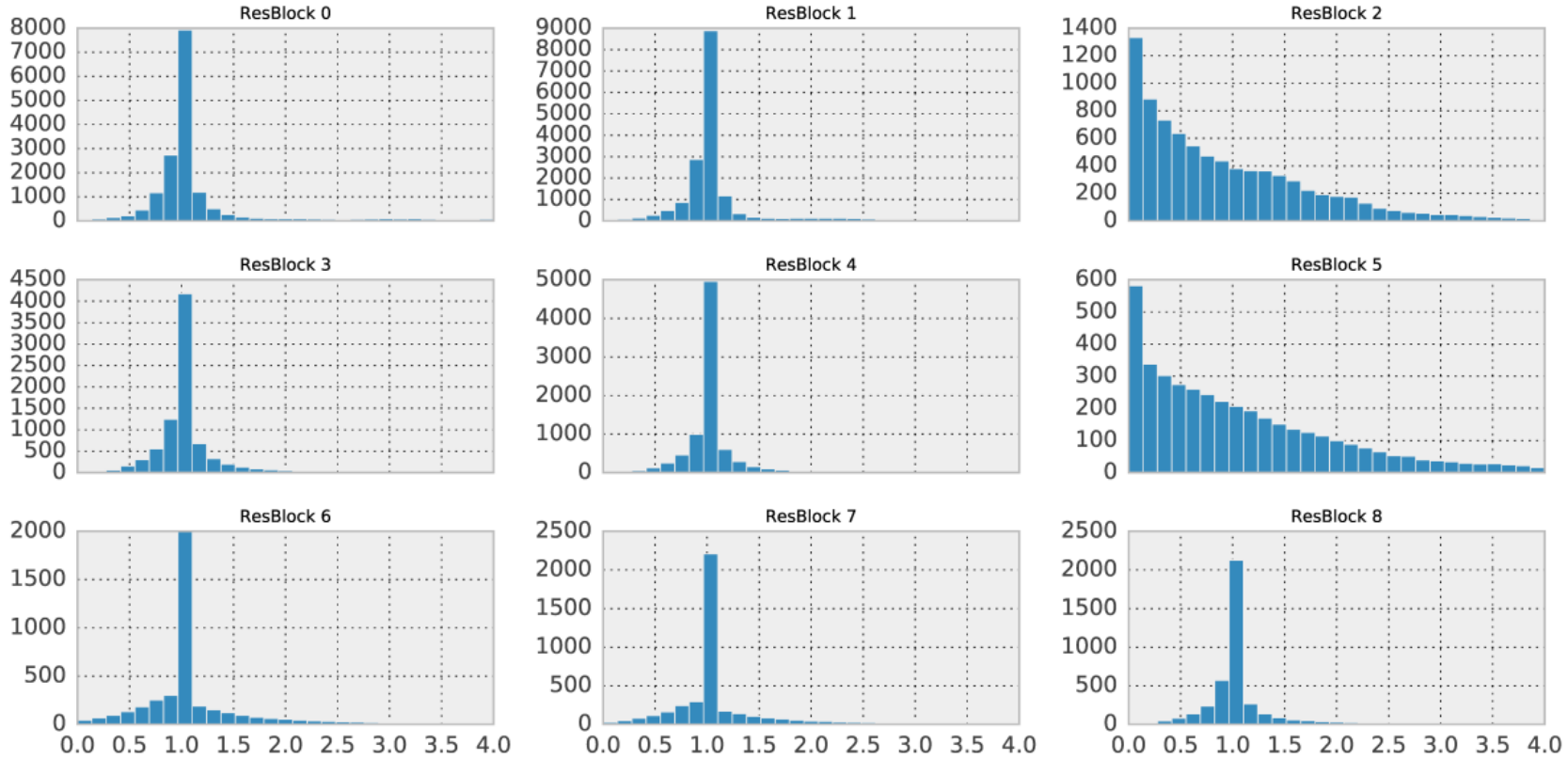
Ky Fan inequalities:

The Ky Fan inequality bounds the singular values of the Jacobian $\boldsymbol{J} = \boldsymbol{I} + \boldsymbol{X}$:

$$1 - s_i(\boldsymbol{X}) \leqslant s_i(\boldsymbol{I} + \boldsymbol{X}) \leqslant 1 + s_i(\boldsymbol{X}),$$

where $s_i(\boldsymbol{X})$ denotes the i-th singular value of $\boldsymbol{X} = \frac{\partial F}{\partial \boldsymbol{a}}$ in decreasing order.

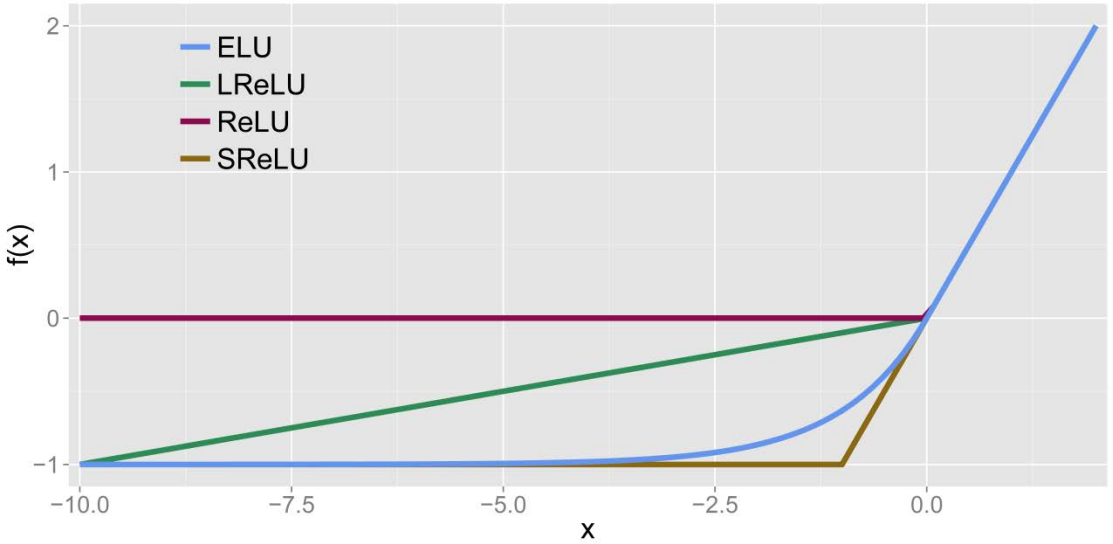
Singular Values: ResNet



Self-Normalizing Networks

Scaled exponential linear units (sELUs)

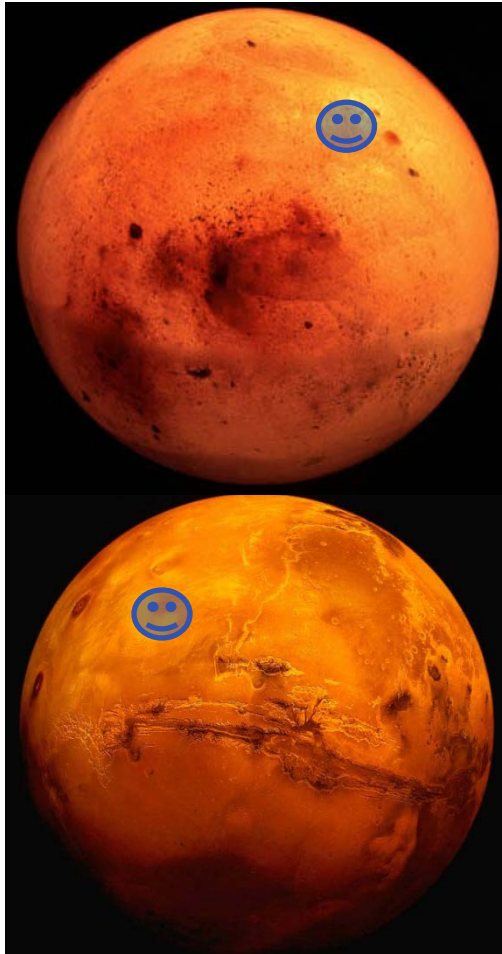
- Moment conserving (→singular values near one)
- Stable fixed point



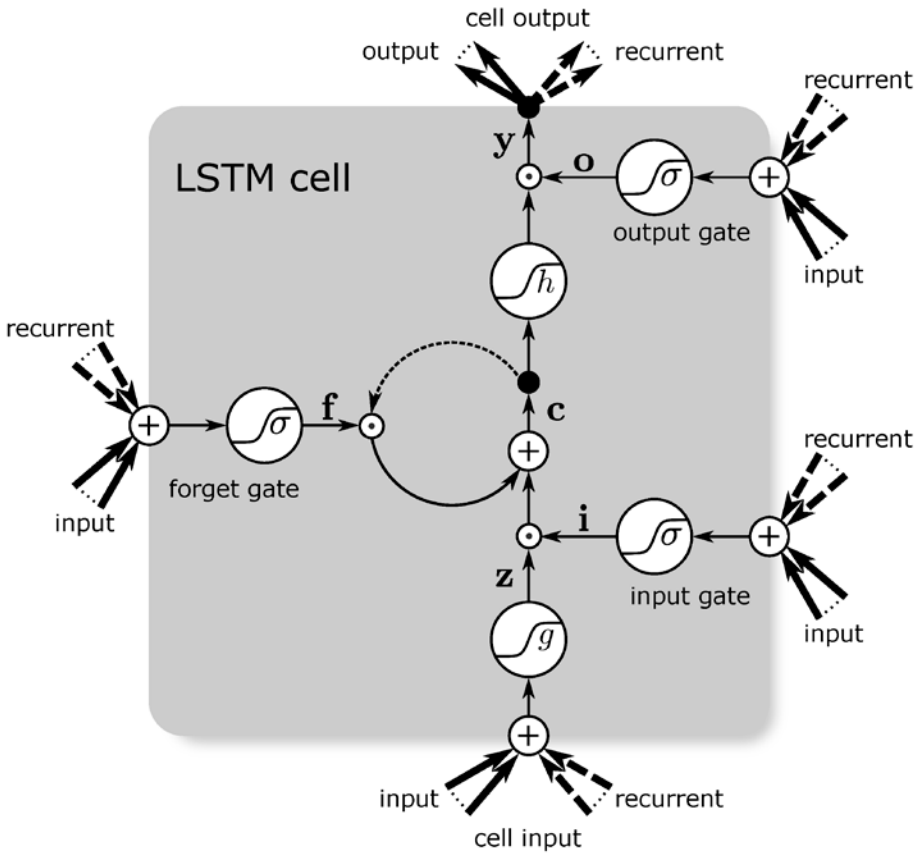
AI to Mars

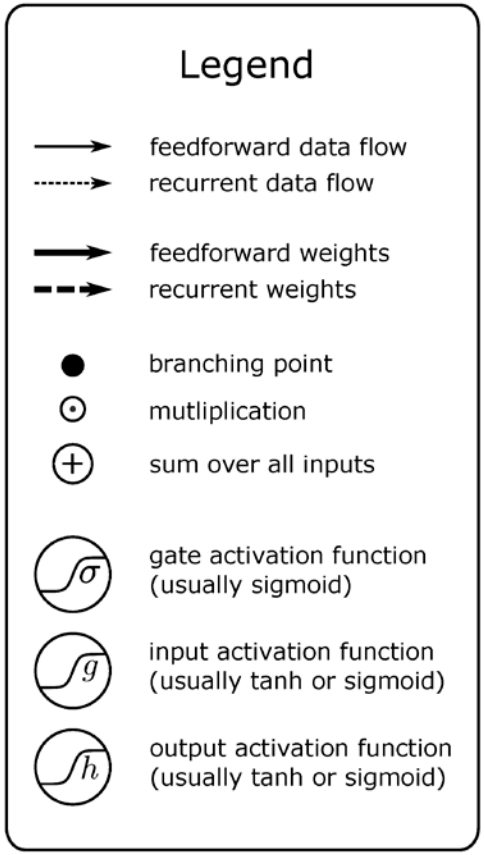
Self-normalizing networks allow
to construct large AI systems

- Personal assistance and friend
- Housekeeping and caretaking systems
- AI to Mars for terraforming

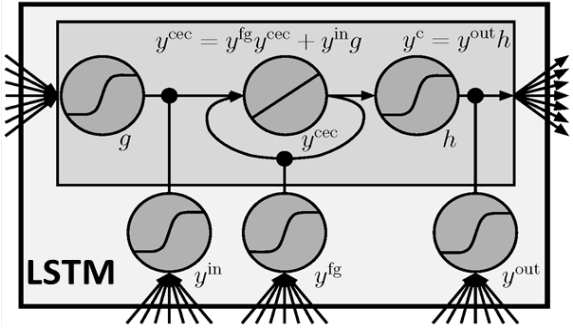


Long Short-Term Memory





Long Short-Term Memory



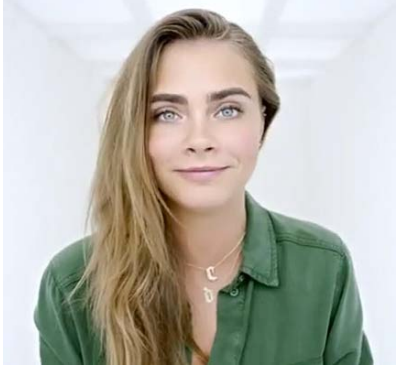
- 1991: invented by Hochreiter
- 1997: major publication Hochreiter&Schmidhuber
- 2009: wins challenges French & Arabic handwriting
- 2011: wins offline Chinese handwriting competition
- 2012: in **Google's Android speech recognizer**
- 2015: in **Google's Voice transcription**
- 2016: in **Apple's iOS 10 → Quicktype**
- 2016: in **Google's Translate (since 28.09.2016)**

Projects

zalando Berlin

Deep learning for fashion images and fashion blogs

- image analysis
- text and language analysis



LSTM: Uniform Credit Assignment

A man wearing a red short shirt and sunglasses is running

A running man with a red short shirt and sunglasses

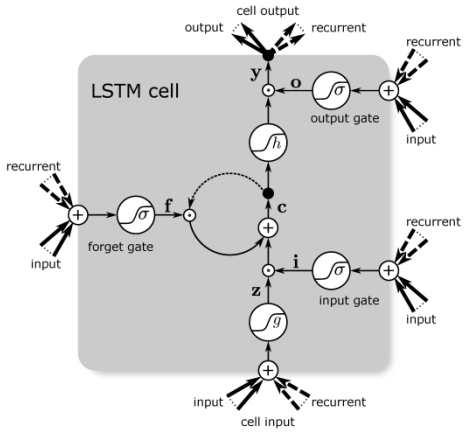
A red short shirt and sunglasses are worn by a running man

LSTM: Uniform Credit Assignment

Storing a thought

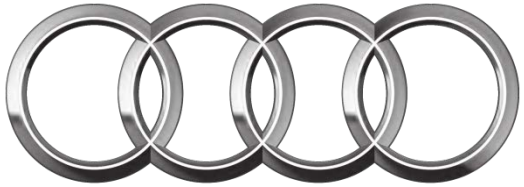
A man is running
A man is racing
A man is sprinting
A man is jogging

A boy is running
A guy is running
A male is running
A husband is running



Translation to german:
„Ein Mann rennt“

Projects



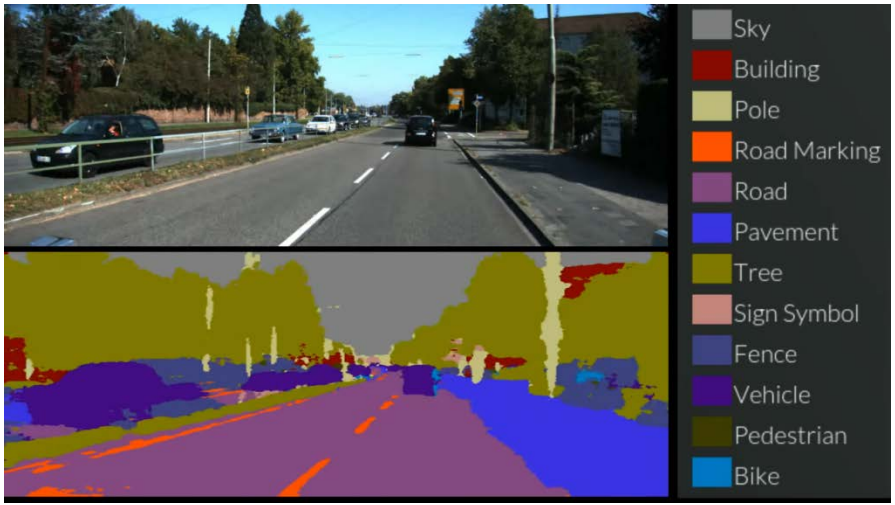
Audi

NVIDIA

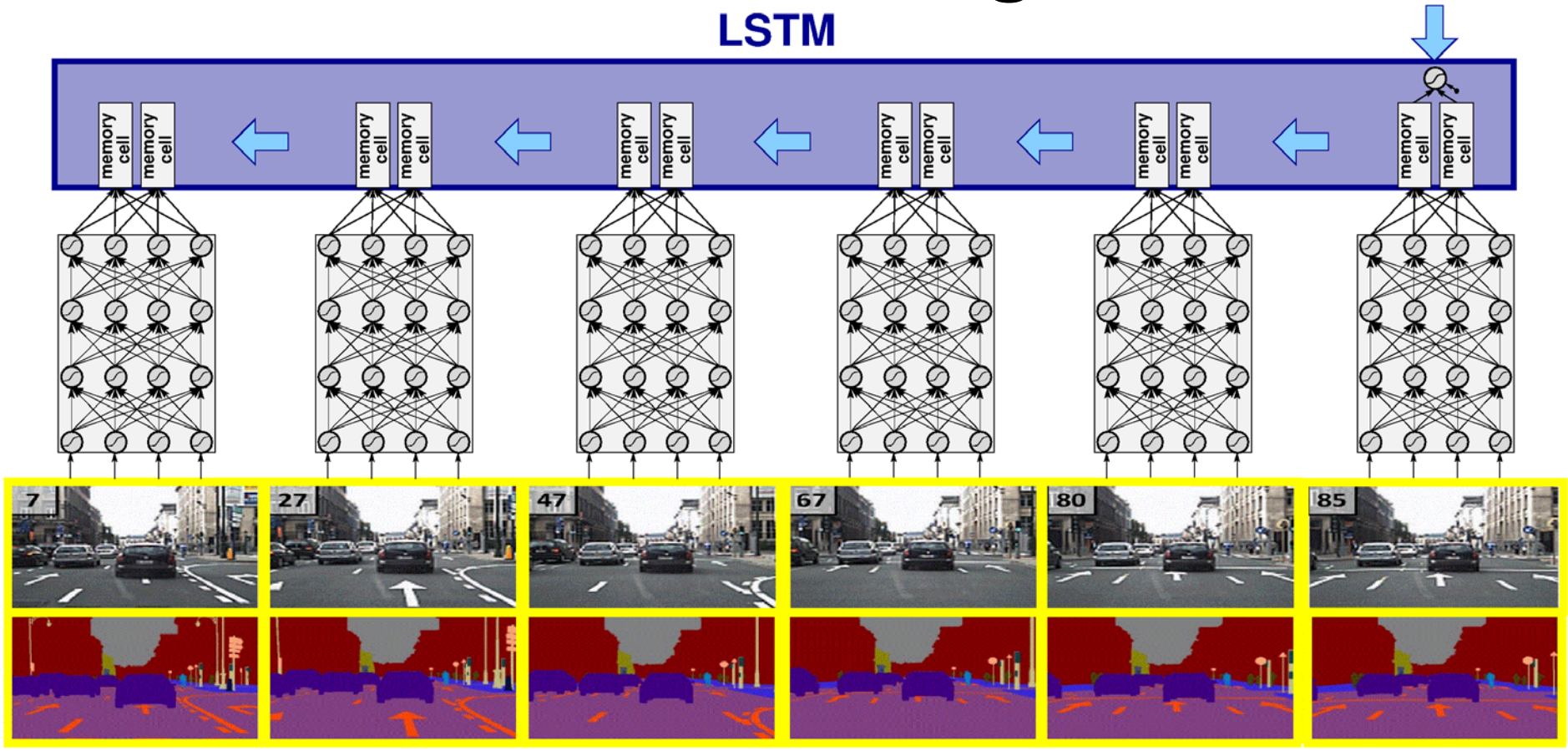
Deep learning for autonomous driving

- Images, radar, laser
- Prediction, attention (LSTM)

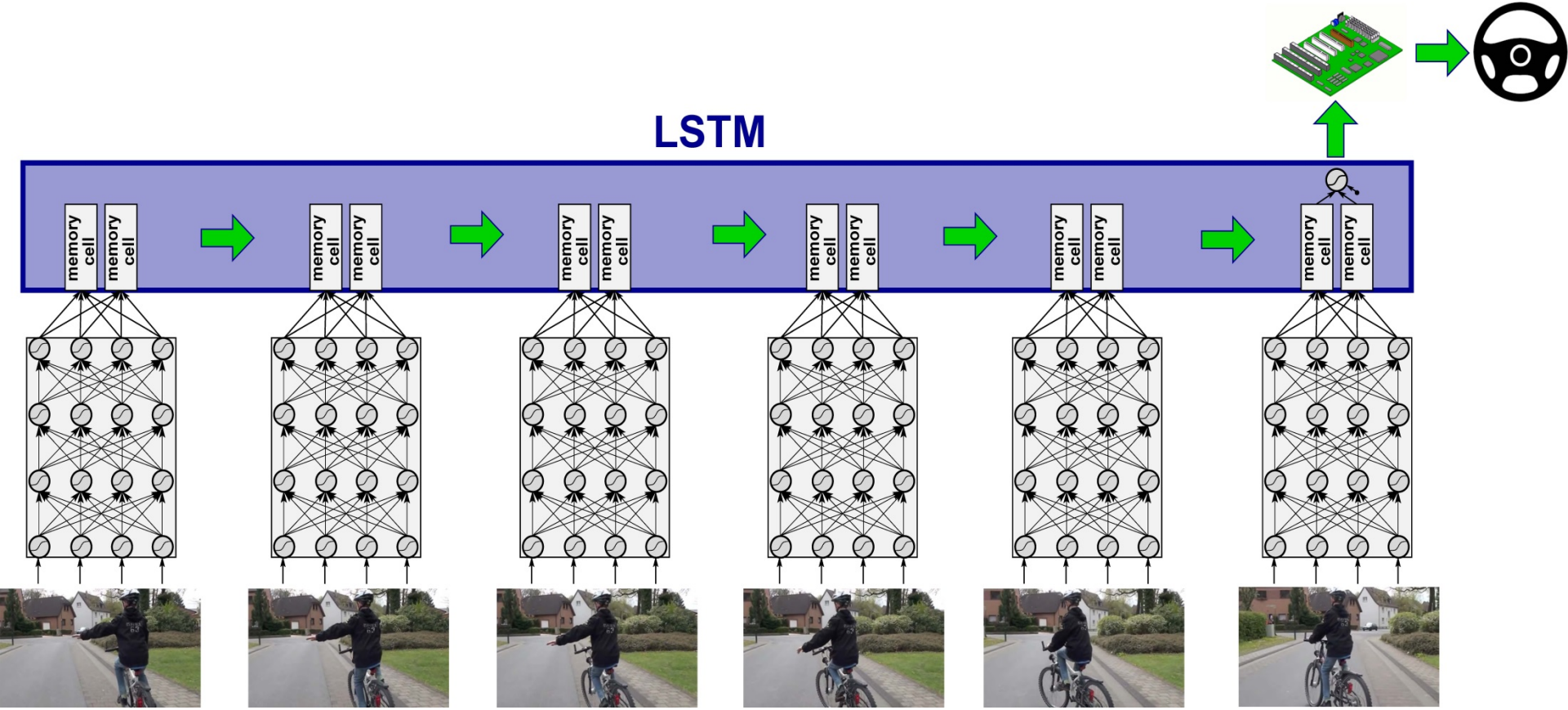
Audi Deep Learning Lab Linz



LSTM: Self-Driving Cars

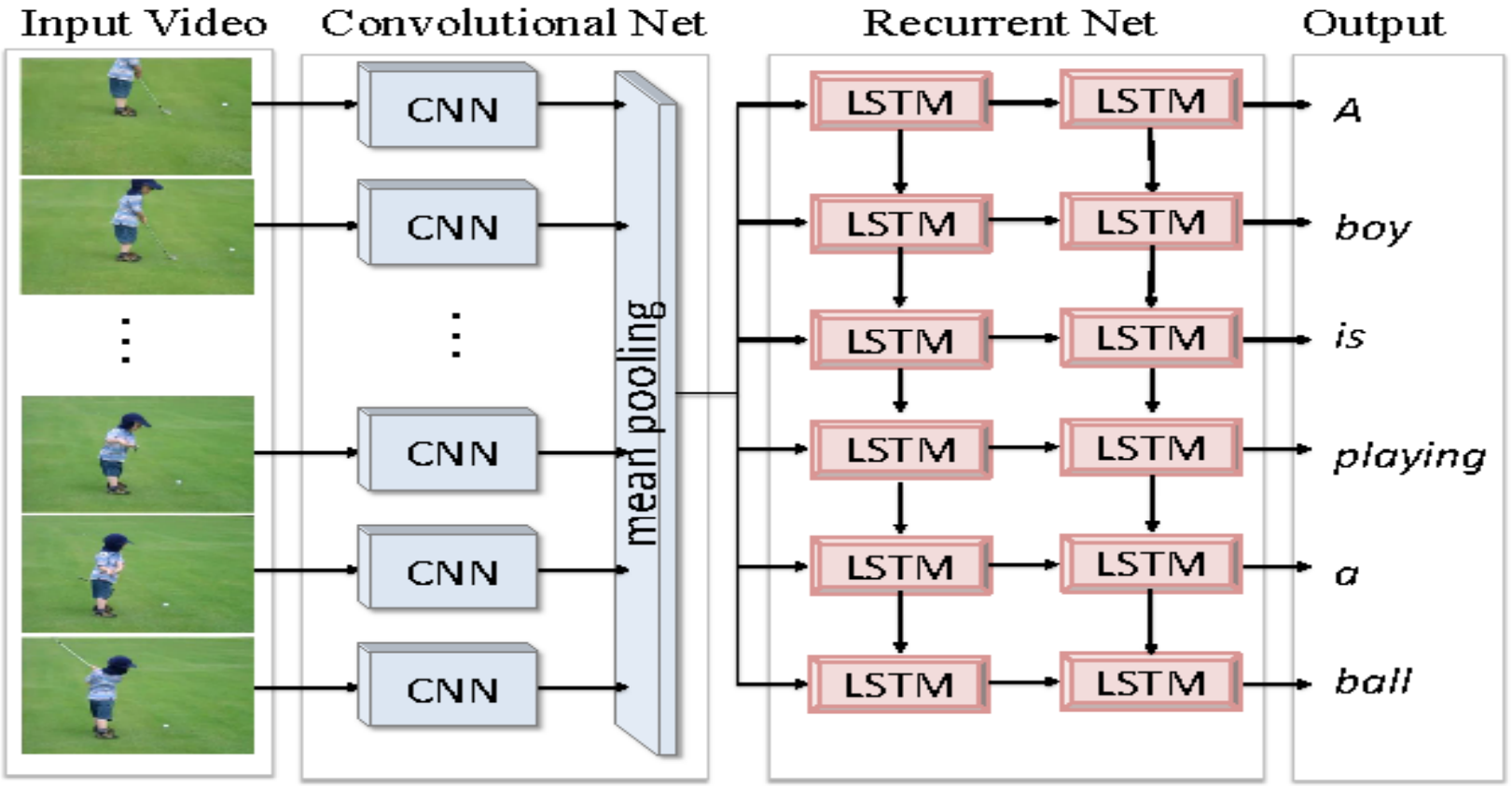


LSTM: Self-Driving Cars



cityscapes

LSTM: Image Captions



a baseball player throwing a ball in front of a field

people are playing a frisbee on grass

there are two men who appear to be practicing martial arts

a pretty young girl holding a wine bottle in her hand

END