

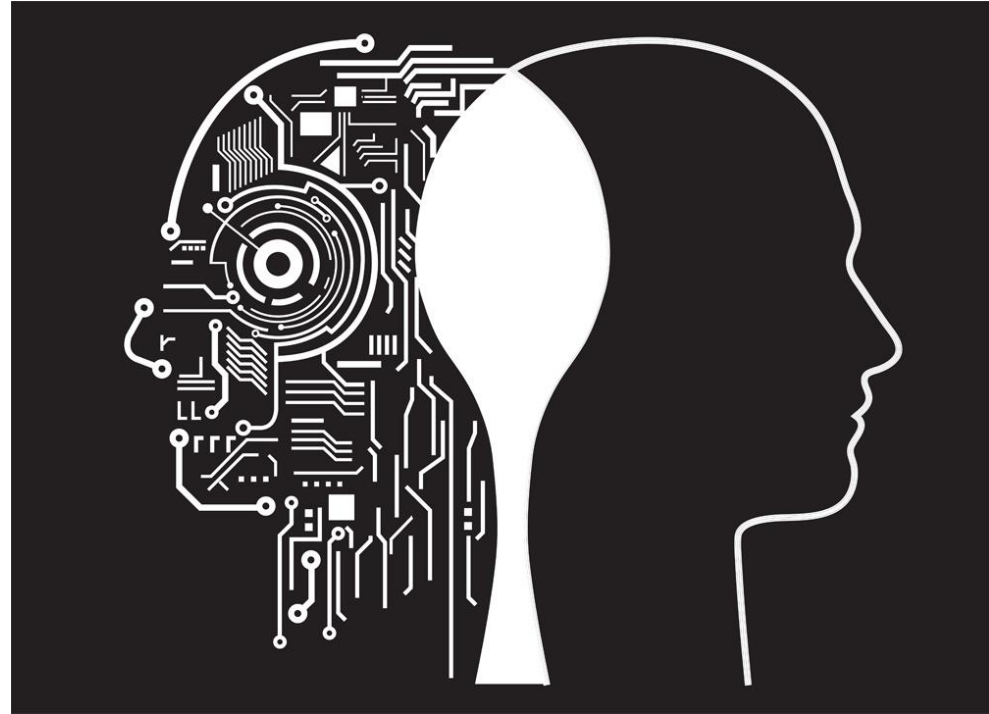
# Vom Gehirn inspiriert: Künstliche Intelligenz durch Deep Learning

**Sven Behnke**



# Künstliche Intelligenz

- Begriff geprägt 1956 (Dartmouth-Konferenz)
- Schwer zu fassen, da Intelligenz vielschichtig
- Verständnis natürlicher Intelligenz
- Entwicklung einer Theorie der Prinzipien intelligenter Systeme
- Schaffung intelligenter Artefakte



[Shutterstock]

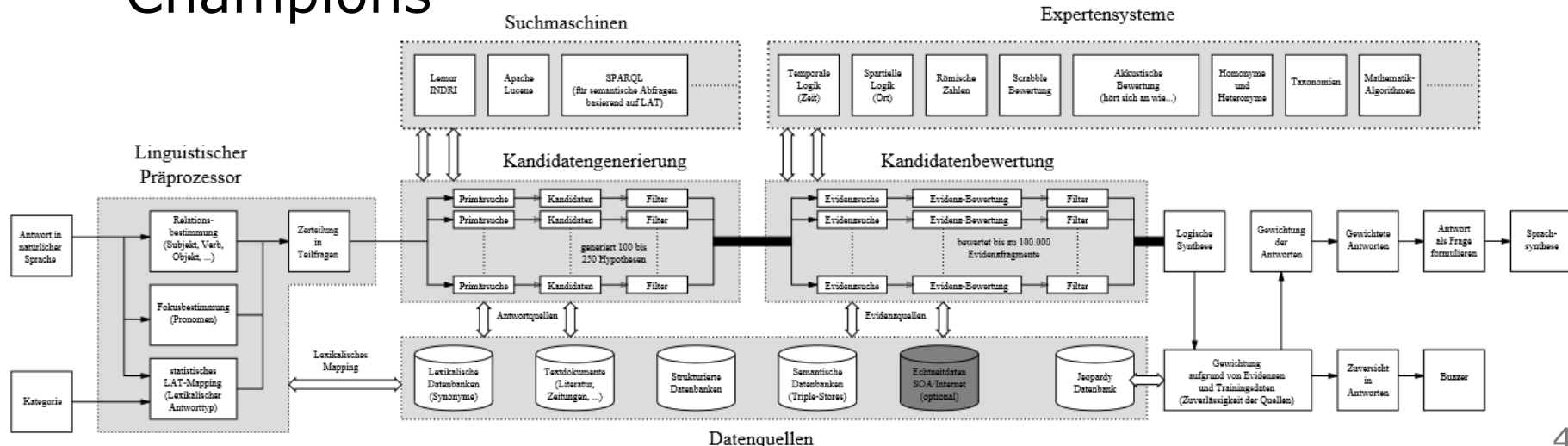
# Schachspiel

- 1997 Kasparov vs. Deep Blue
- IBM Spezialrechner
- Bewertung von bis zu 200 Millionen Stellungen pro Sekunde
- Erstmals Sieg eines Computers unter Wettbewerbsbedingungen



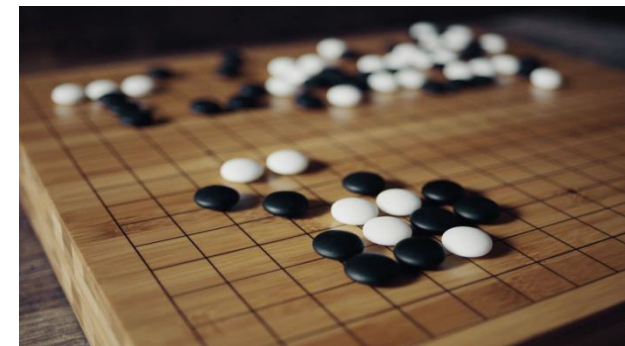
# Spielshow Jeopardy

- Mehrdeutige, natürlichsprachliche Hinweise
- Antwort in Form einer Frage
- 2011 IBM Watson Sieg gegen Champions



# Go-Spiel

- Wesentlich größerer Suchraum als Schach
- 2016 DeepMind AlphaGo Gewinn gegen Meister Lee Sedol
- Spezialrechner zur Bewertung von Stellungen
- Training durch Spiel-Datenbank und Selbstverbesserung



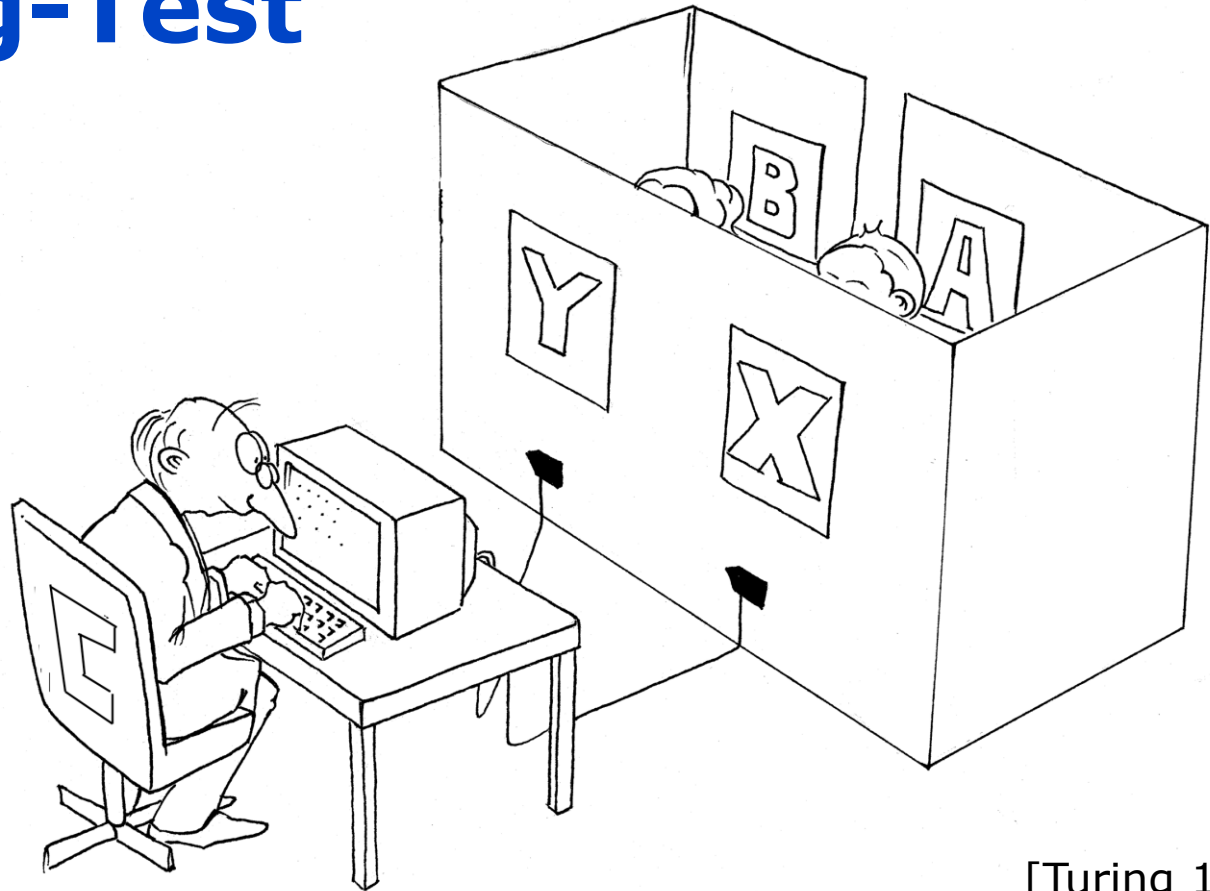
# Physical Symbol System Hypothesis (PSSH)

- "A physical symbol system has the **necessary and sufficient** means for general intelligent action."

[Alan Newell and Herbert Simon, 1976]

- Dies ist die Grundannahme der Good old-fashioned Artificial Intelligence (GOFAI)

# Turing-Test



[Turing 1950]

- Imitationsspiel Mann vs. Frau  
=> Mensch vs. Computer
- Jährlicher Loebner-Preis Wettbewerb,  
Goldmedallie (\$100.000) bislang nicht vergeben

# Zwei Sichtweisen auf Intelligenz

## Klassisch:

- Kognition als Berechnung



## Embodiment:

- Kognition entsteht durch sensomotorische Prozesse und Interaktion





# Sprachassistenten



[Amazon]

# Kommunikationsroboter



[Nieuwenhuisen and Behnke, SORO 2013]

# Serviceroboter



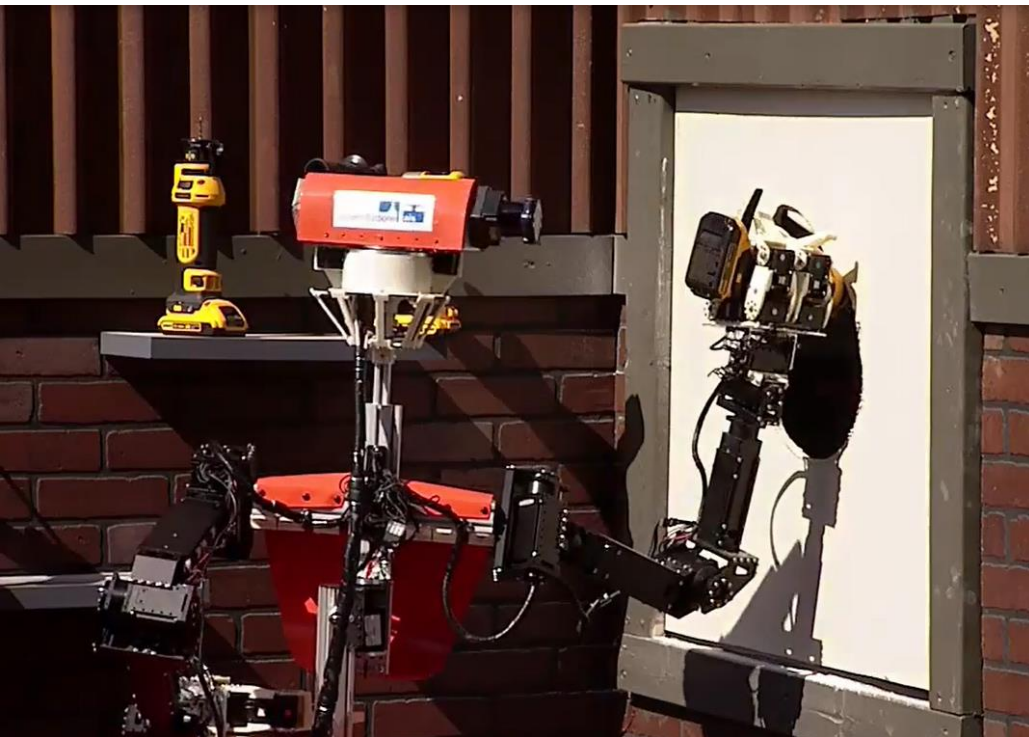
Dynamaid



Cosero

[Stückler et al.: Frontiers in AI and Robotics 2016]

# Rettungs- und Explorationsroboter



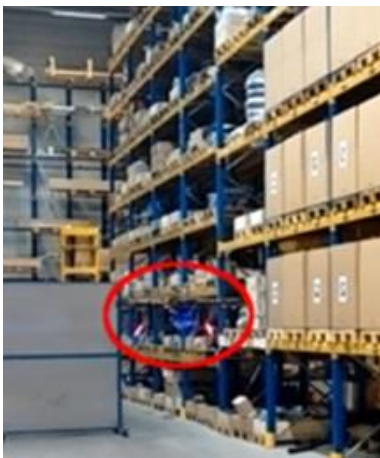
[Schwarz et al.: Frontiers in Robotics and AI 2016, JFR 2017]

# Fußballroboter



[Allgeuer et al.: Humanoids 2015, 2016]

# Flugroboter



[Nieuwenhuisen et al.: JINT 2015, Droschel et al: JFR 2016]

# Roboter für den Griff in die Kiste

ActReMa



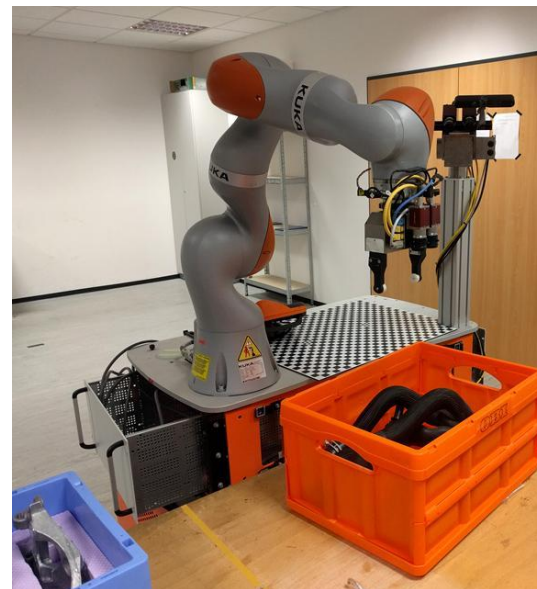
STAMINA



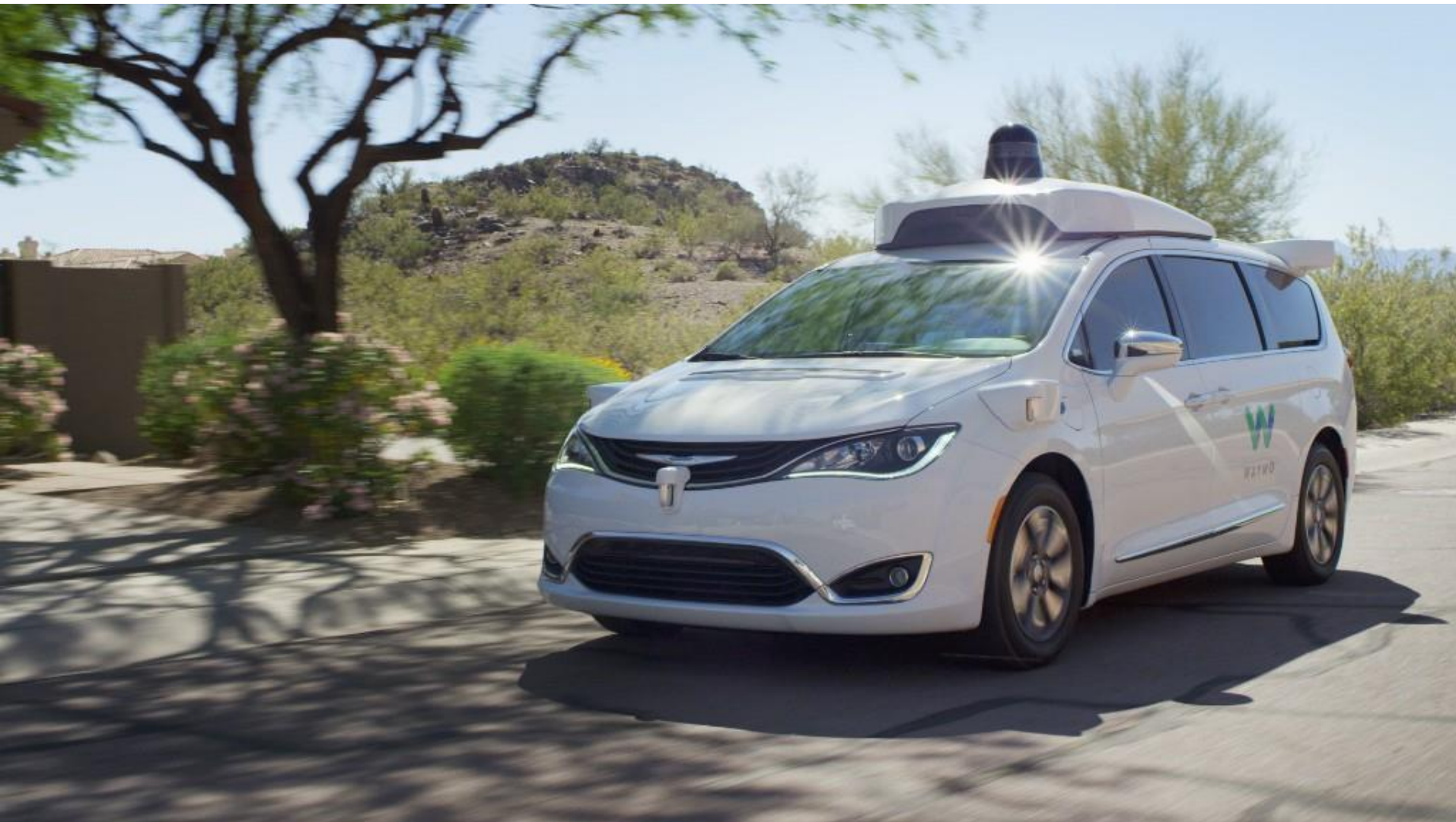
EuRoC  
C1



EuRoC  
C2



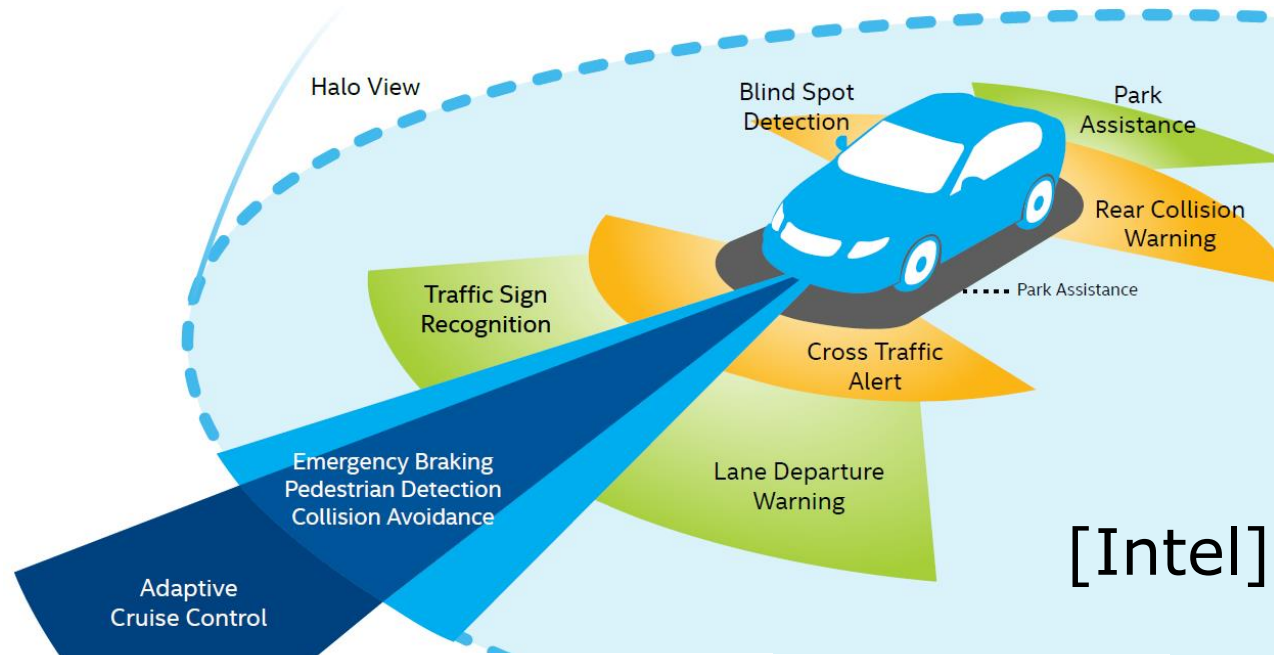
# Selbstfahrende Automobile



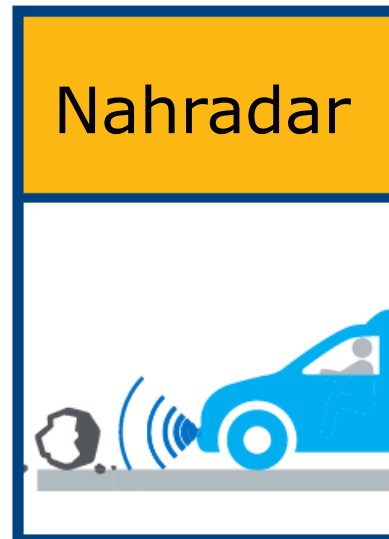
[Waymo]



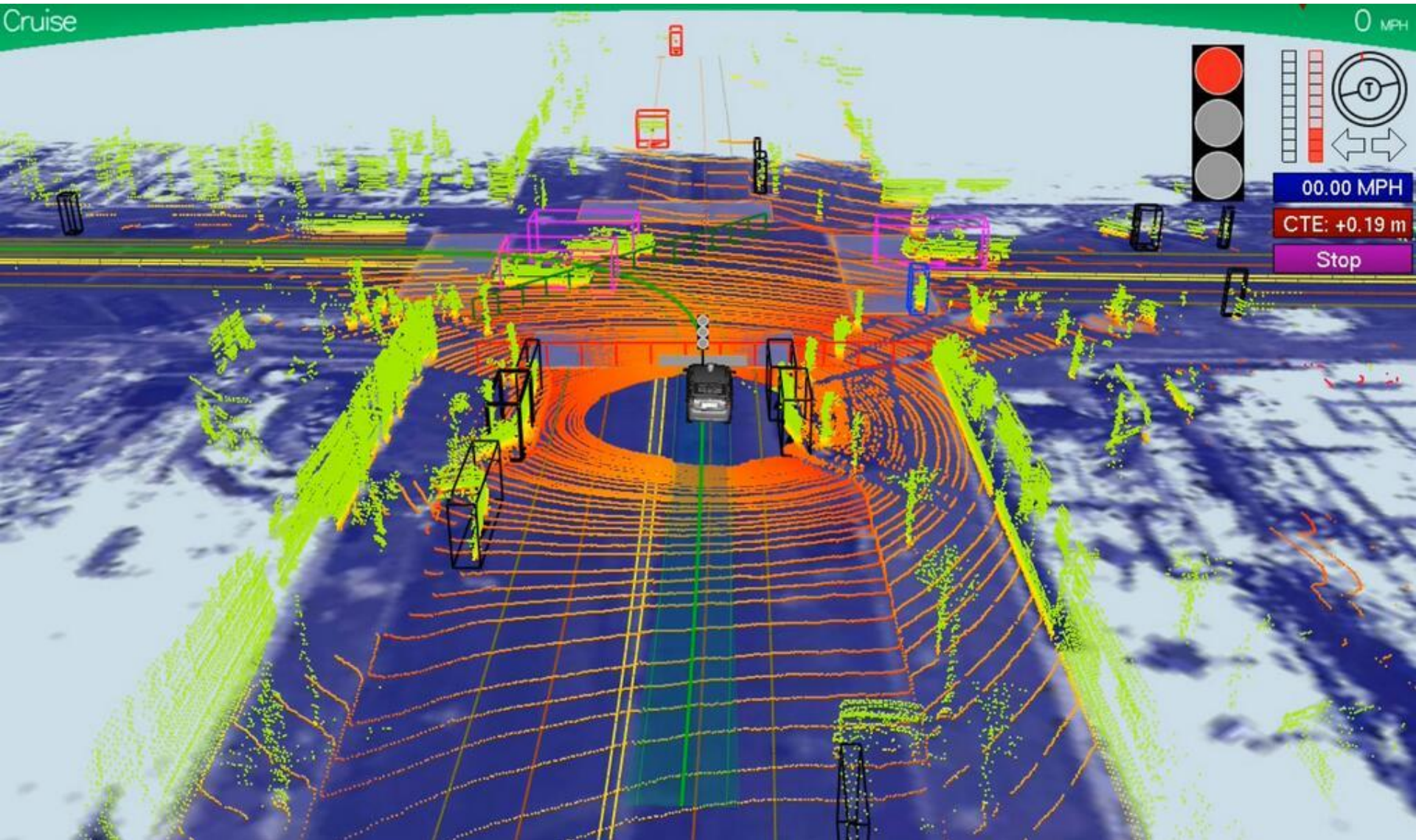
# Sensoren für Automobile



[Intel]

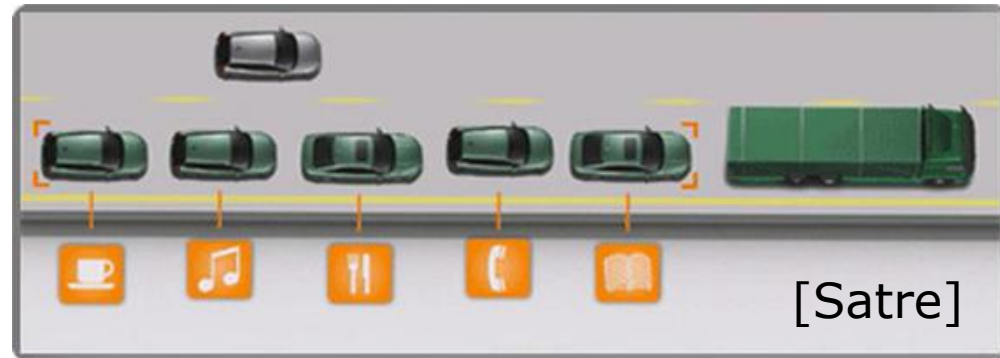


# Umgebungssicht des Google-Autos



# Möglicher Nutzen Selbst-fahrender Automobile

- Höher Sicherheit
- Weniger Energieverbrauch
- Bessere Ausnutzung der Straßenraums
- Fahrer kann Zeit anders nutzen
- Preiswertere Taxis



# Künstliche Neuronale Netze

## Motiviert durch das Gehirn

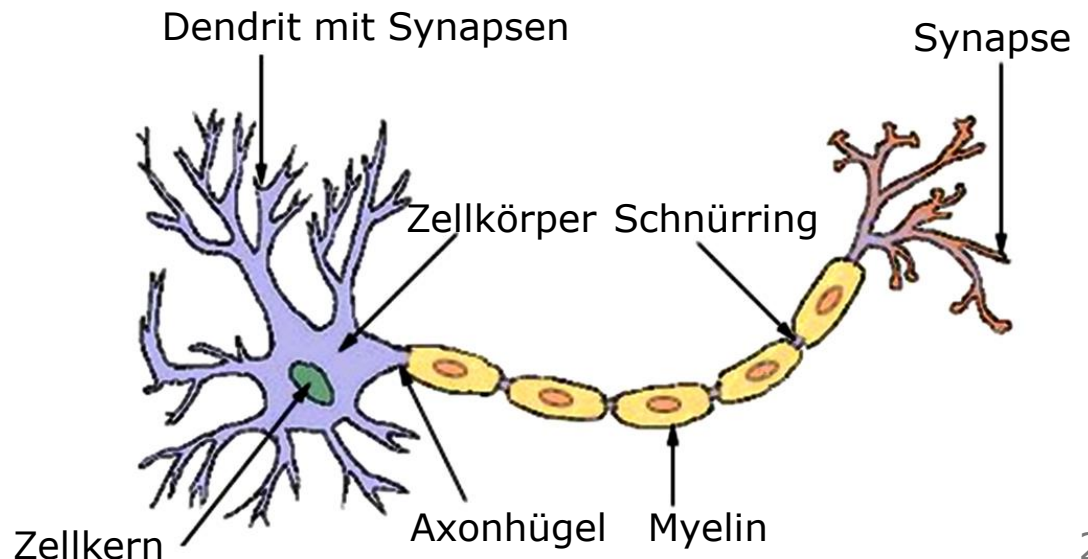
- Ermöglicht anscheinend alle intelligenten Leistungen
- Viele Milliarden hochgradig vernetzter Neuronen

## Prinzipien & Charakteristika

- Die einzelnen Neuronen treffen lokale Entscheidungen
- Viele einfache Neuronen sind vernetzt und arbeiten parallel
- Gesamtnetz kann komplexe Probleme lösen
- Netz ist lernfähig

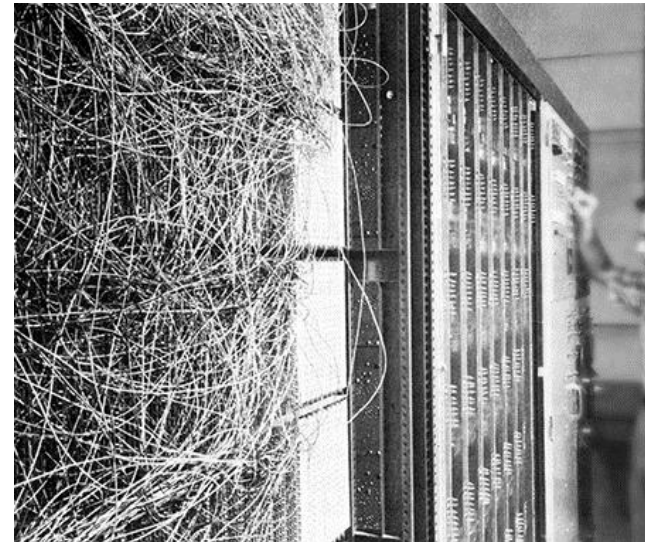
## Anwendungsgebiete

- Mustererkennung / Klassifikation
- Approximation von Funktionen

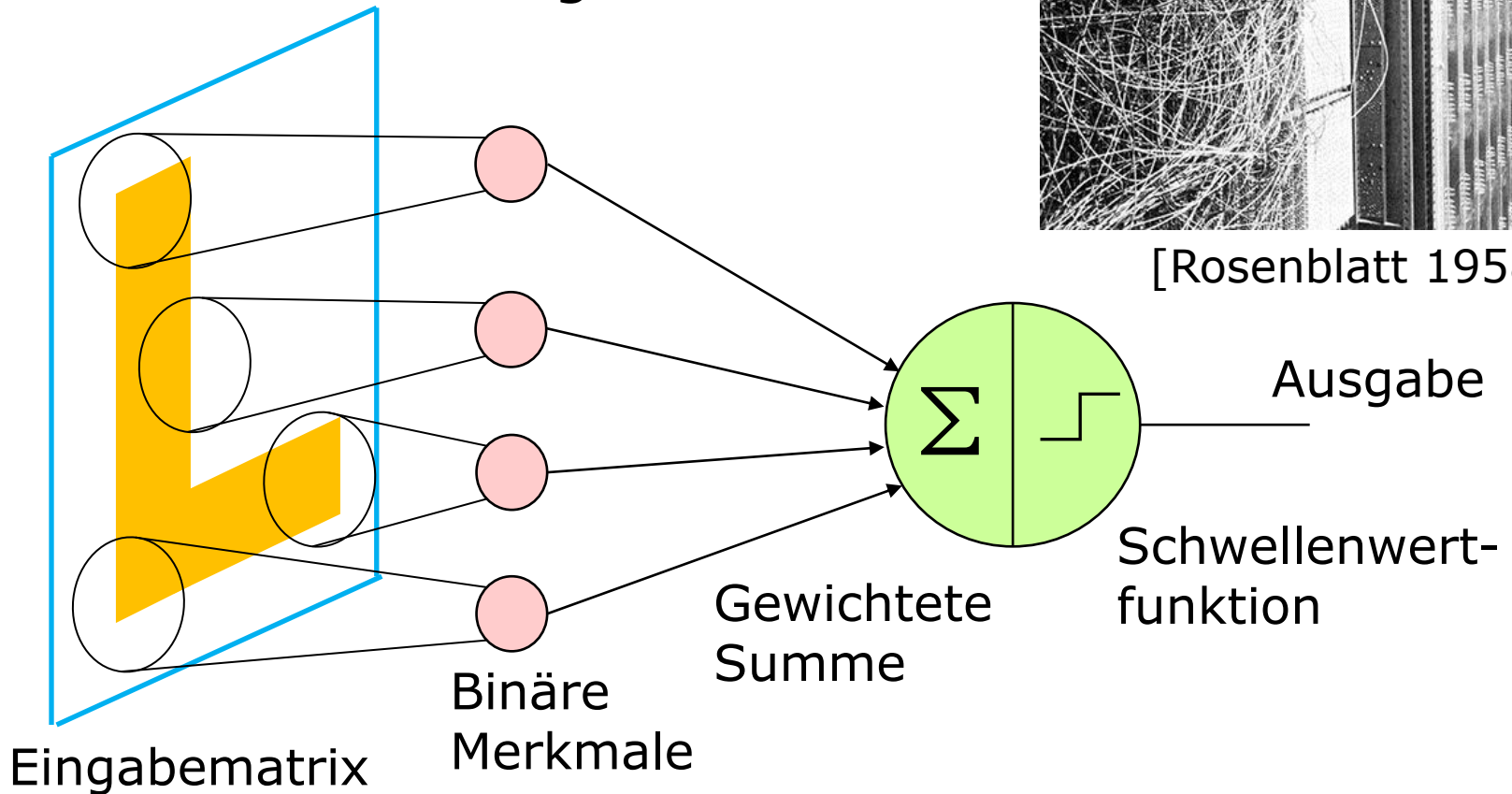


# Perzeptron

- Merkmalsextraktion
- Mustererkennung

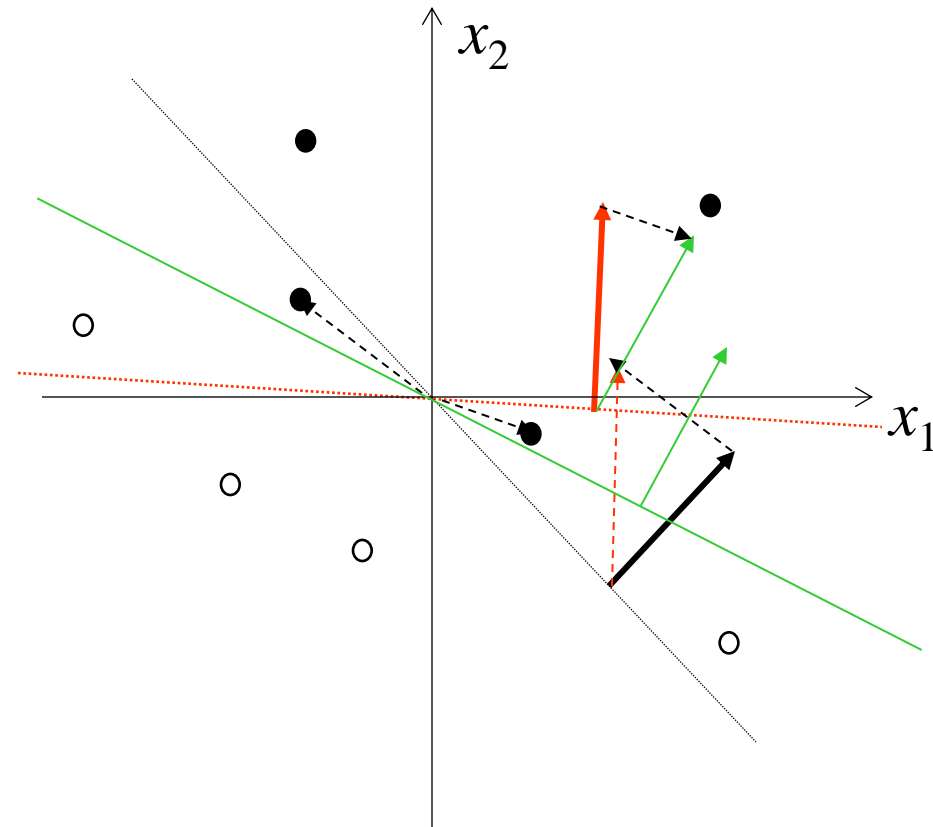


[Rosenblatt 1958]

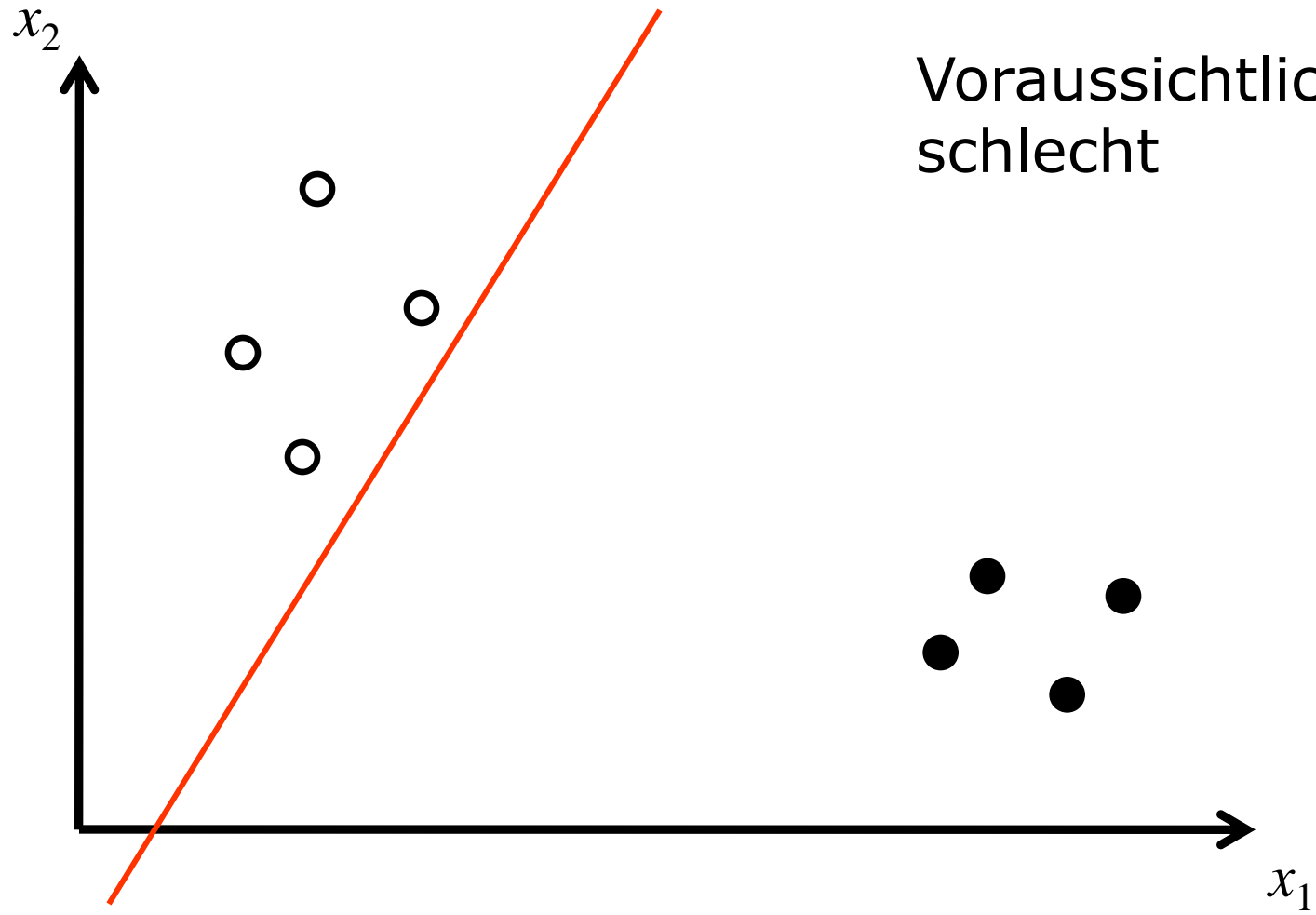


# Perzeptron-Lernalgorithmus

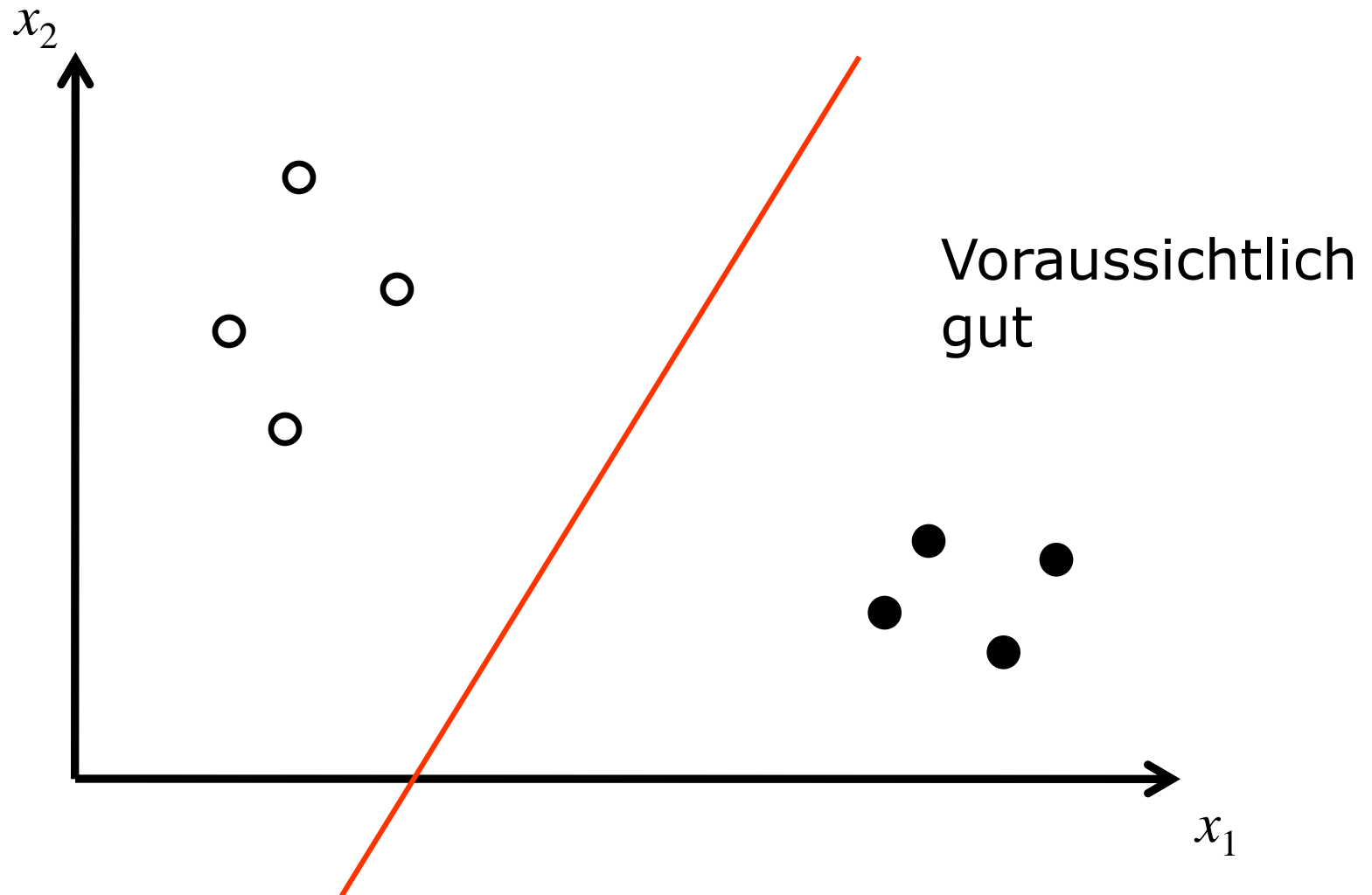
- Eingaben: zwei numerische Werte (gezeichnet in Ebene)
- Aufgabe: Teile Beispiele in zwei Klassen (weiß und schwarz)
- Lerner (Versionsraum): Trenngerade durch den Ursprung
- Lernregel:
  - Nimm Normalvektor
  - Addiere den Punktvektor eines falsch klassifizierten Beispiels
  - Drehe Gerade, sodass neuer Vektor der Normalvektor wird
  - Solange bis alles richtig klassifiziert wird
- Generalisierung: neue Punkte richtig klassifiziert



# Generalisierung



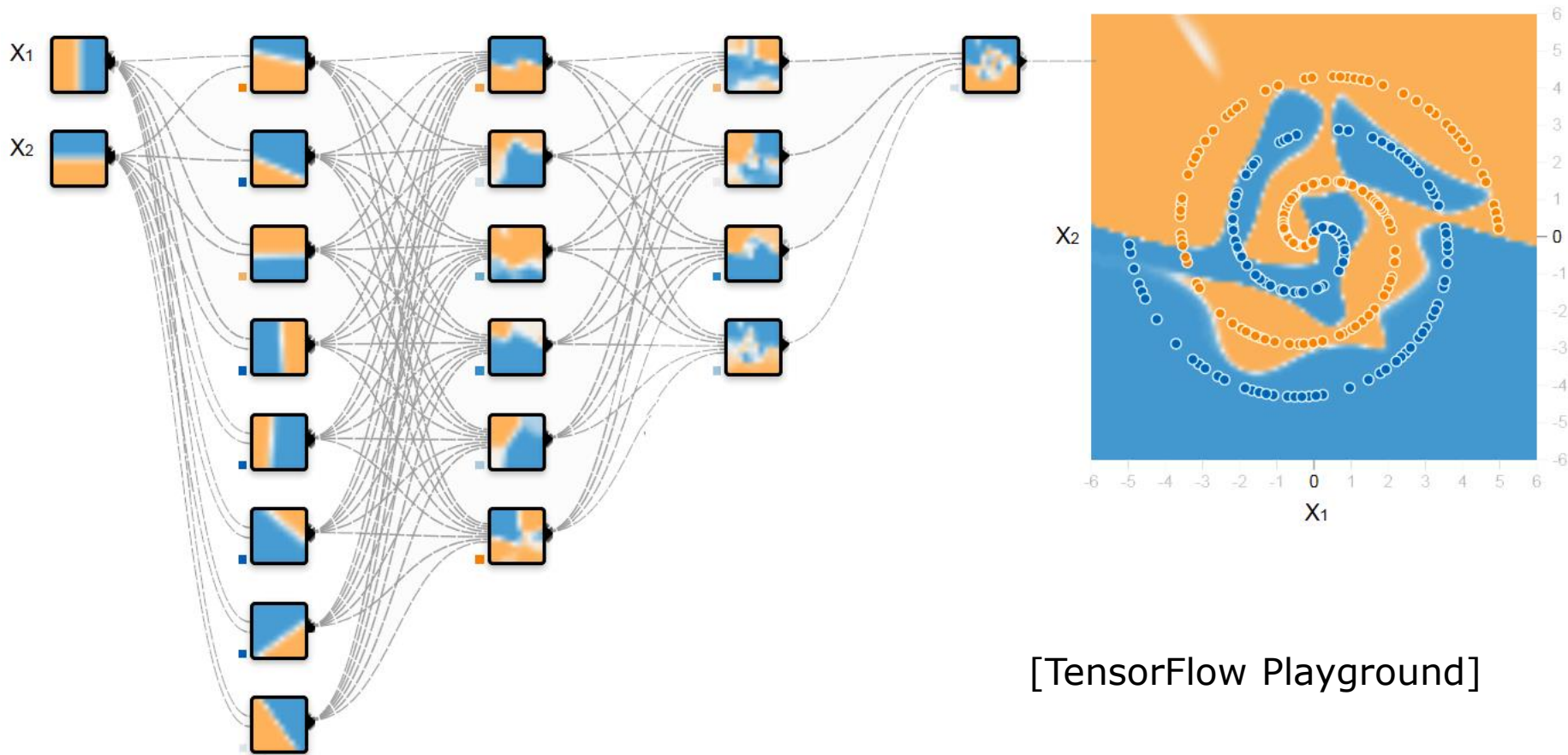
# Generalisierung





# Multi-Layer Perzeptron

- Nichtlineare Trennung des Eingaberaums
- Backpropagation-Algorithmus [Rumelhart et al. 1986]



# Trend seit 2006: Deep Learning



## NEWS DESK

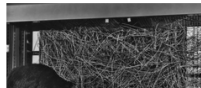
Reporting the latest on Washington and the world.

« How Susan Rice Sees the World | Main | Moral Machines »

### NOVEMBER 25, 2012 IS "DEEP LEARNING" A REVOLUTION IN ARTIFICIAL INTELLIGENCE? POSTED BY GARY MARCUS

f Share 603 Tweet 389 COMMENTS 11

Can a new technique known as deep learning revolutionize artificial intelligence, as yesterday's front-page article at the New York Times suggests? There is good reason to be excited about



## BUSINESS

### Baidu muscles in on Google's turf with Silicon Valley deep learning lab

Chinese search giant beds down next to Apple in Cupertino

By Phil Muncaster, 15th April 2013 Follow 3,371 followers

1

Win a Samsung 40-inch LED HDTV with The Reg and HPI!

#### RELATED STORIES

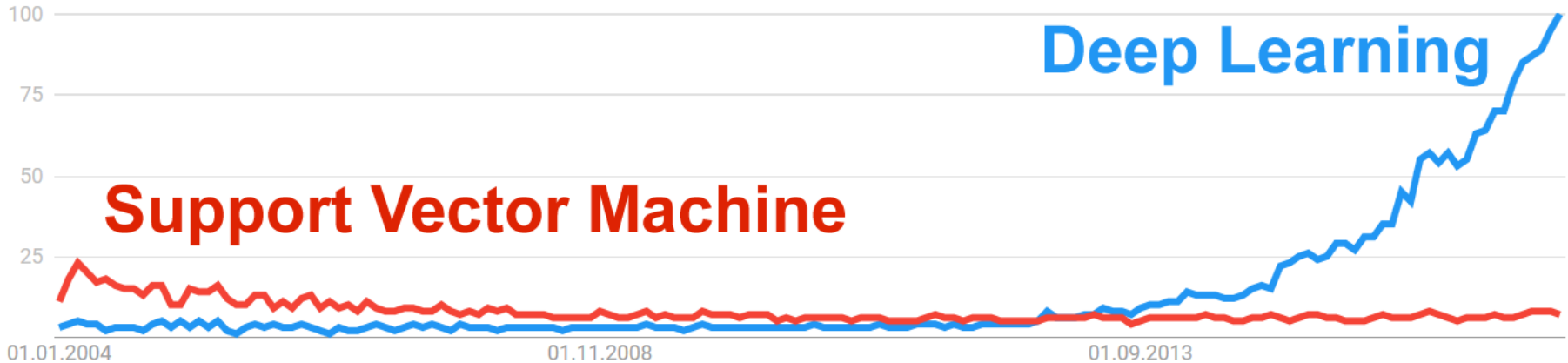
China's Baidu builds new type of App Store

Chinese search giant Baidu has opened the doors to a new research facility in Google's back yard where it's hoping to tap the local talent to consolidate early mover advantage in the burgeoning field of "deep learning".

The Cupertino-based Institute of Deep Learning (IDL) is the Silicon Valley counterpart of another facility back in China dedicated to accelerating research in the emerging machine learning-related discipline.

## Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.



[Google Trends]

# Starkes Interesse der Industrie

- Google
  - DNNresearch (Geoffrey Hinton)
  - DeepMind (Demis Hassabis)
- Baidu
  - Andrew Ng
- Facebook
  - Yann LeCun
- Microsoft
  - Li Deng

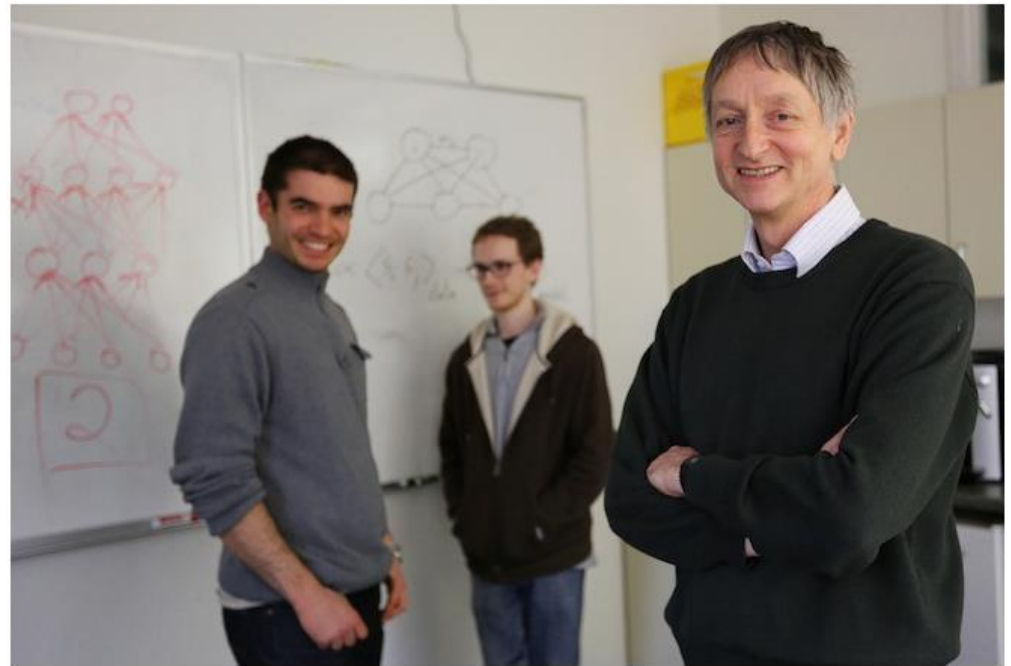


Google Hires Brains that Helped Supercharge Machine Learning

BY ROBERT MCMILLAN 03.13.13 6:30 AM

[Follow @bobmcmillan](#)

Share 672  
Tweet 272  
+1 144  
in 63



Geoffrey Hinton (right) Alex Krizhevsky, and Ilya Sutskever (left) will do machine learning work at Google. Photo: U of T

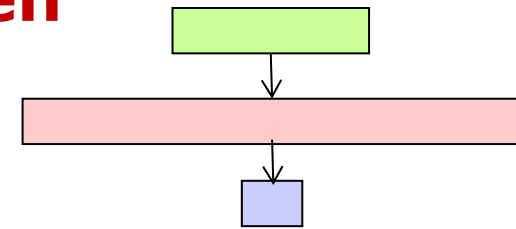
# Definition: Deep Learning

- Deep learning is a set of algorithms in machine learning that attempt to **learn layered models of inputs**, commonly neural networks.
- The layers in such models correspond to **distinct levels of concepts**, where
  - higher-level concepts are defined from lower-level ones, and
  - the same lower-level concepts can help to define many higher-level concepts.

[Bengio 2009]

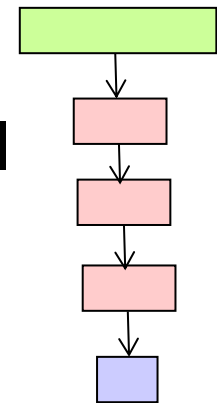
# Flache vs. Tiefe Netzwerke

- Neuronale Netze mit **einer verdeckten Schicht** können jede Funktion berechnen [Cybenko, 1989]



Aber: Manche Funktionen (z.B. Parity) benötigen exponentiell viele Hidden Units

- **Tiefe Netzwerke** (mit mehreren verdeckten Schichten) können exponentiell effizienter sein



Z.B. sequentielle Berechnung der Parity-Funktion

# Ein Bild sagt mehr als tausend Worte



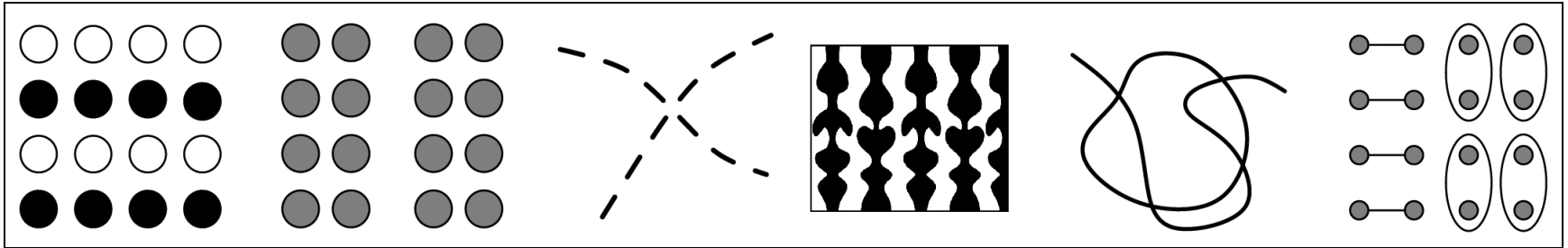
[Vinyals et al. 2014]

# Performanz des Sehsystems

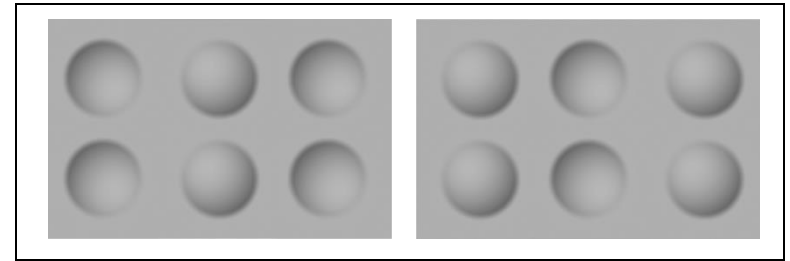


# Psychophysik

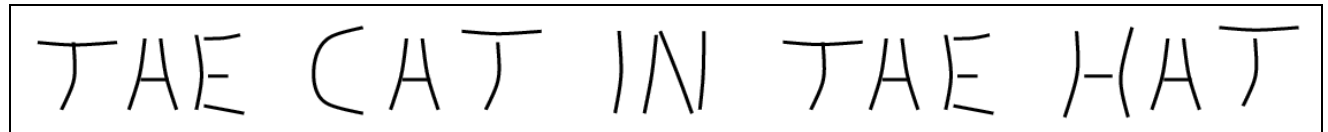
## ■ Gestaltprinzipien



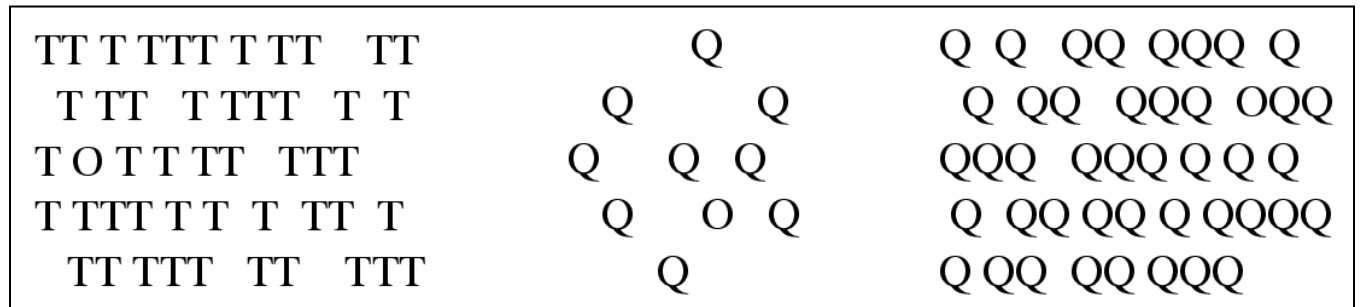
## ■ Heuristiken



## ■ Kontext



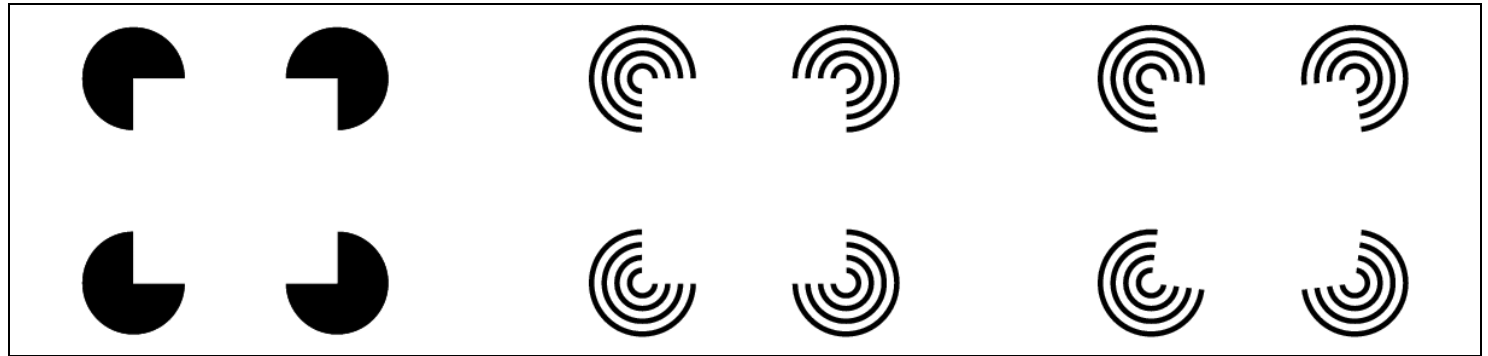
## ■ Aufmerksam- samkeit



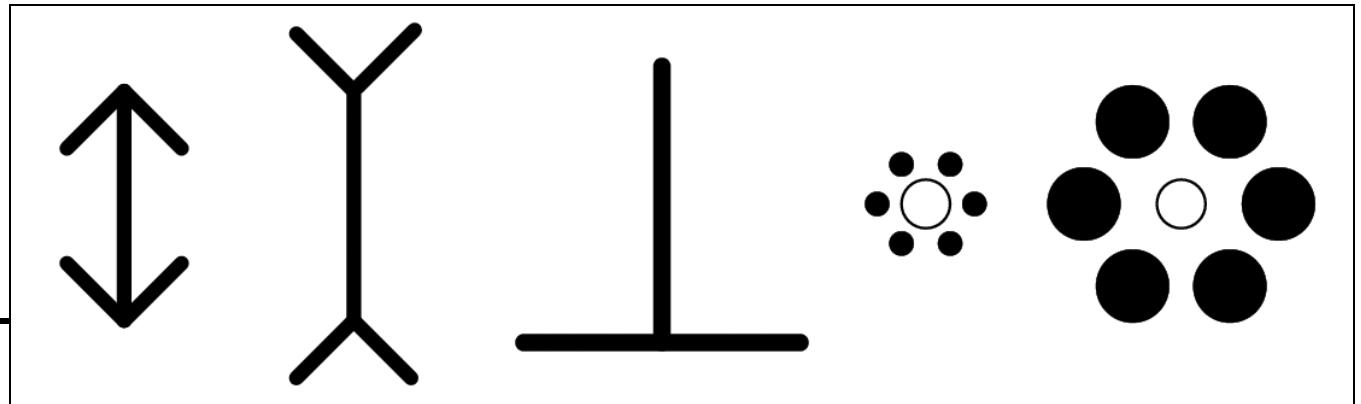


# Visuelle Illusionen

Kanizsa  
Figuren



Müller-Lyer  
horizontal/  
vertikal  
Ebbinghaus  
Titchener

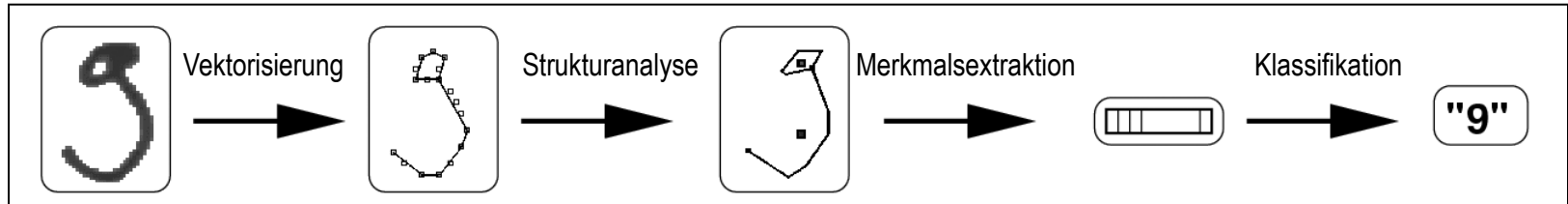


Munker-White

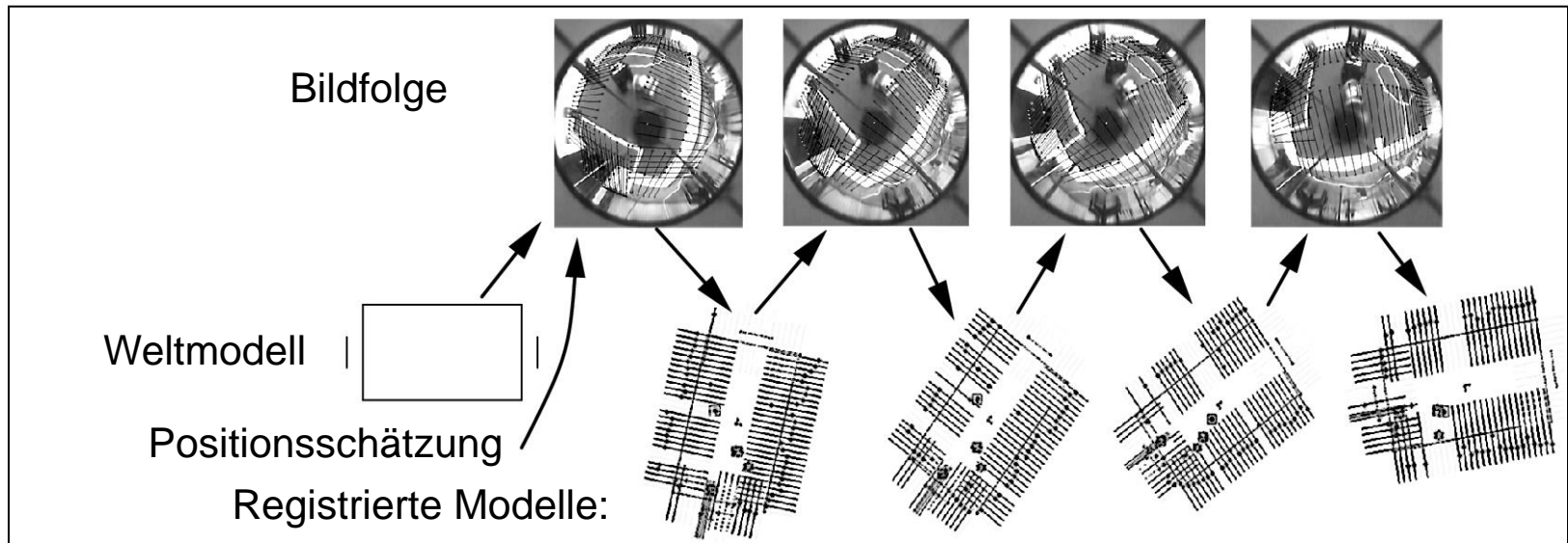


# Computer Vision

## ■ Datengetrieben



## ■ Modellgetrieben



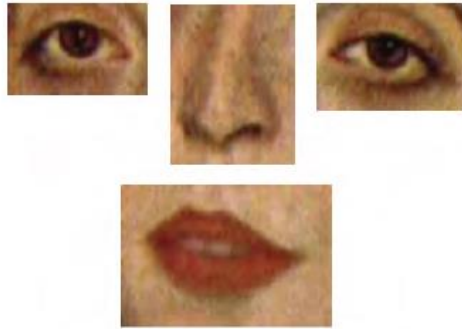
## ■ Interface-Problem

# Beobachtungen

In der Welt trifft meistens zu:

- Benachbarte Dinge haben etwas miteinander zu tun
  - Räumlich
  - Zeitlich
- Es gibt hierarchische Struktur
  - Objekte bestehen aus Teilen
  - Objekte können zu größeren Einheiten zusammengefasst werden, ...

# Räumliche Anordnung von Gesichtsteilen



**A**



**B**



**C**



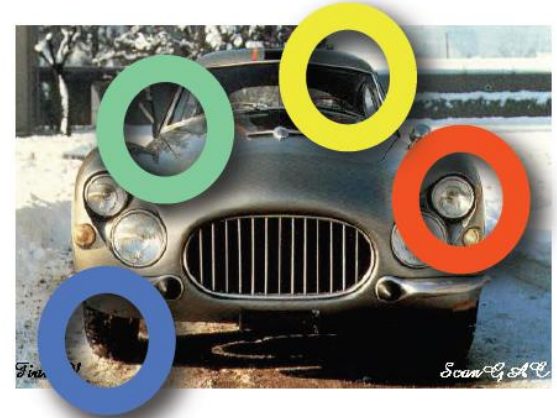
**D**

[Perona]

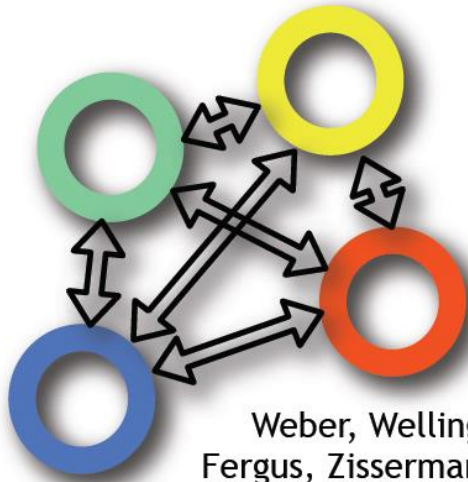
# Gesichtswahrnehmung



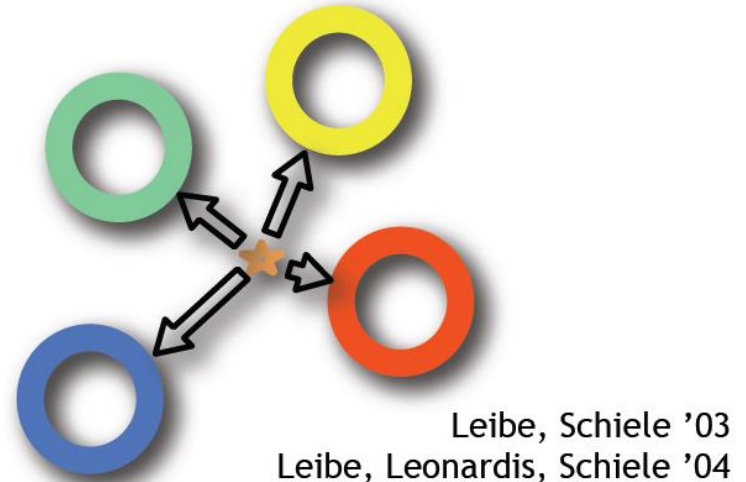
# Horizontale und Vertikale Abhängigkeiten



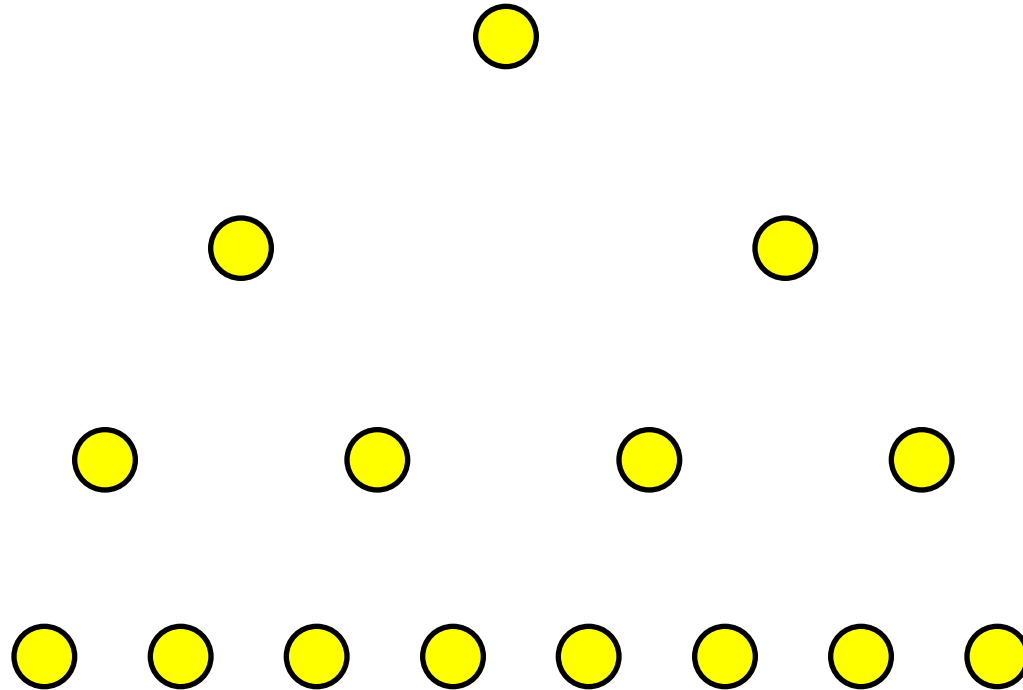
Constellation Model:  
Fully connected shape model



Implicit Shape Model:  
Star-Model w.r.t. Reference Point

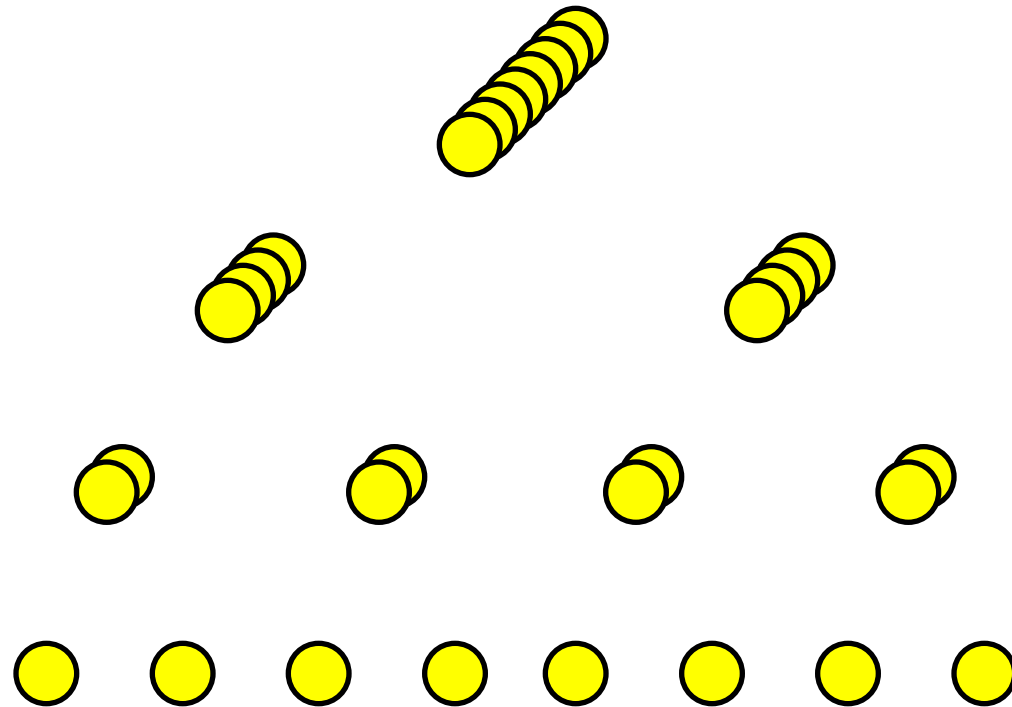


# Multiskalen- Repräsentation



- Bildpyramiden nicht ausdrucksstark genug

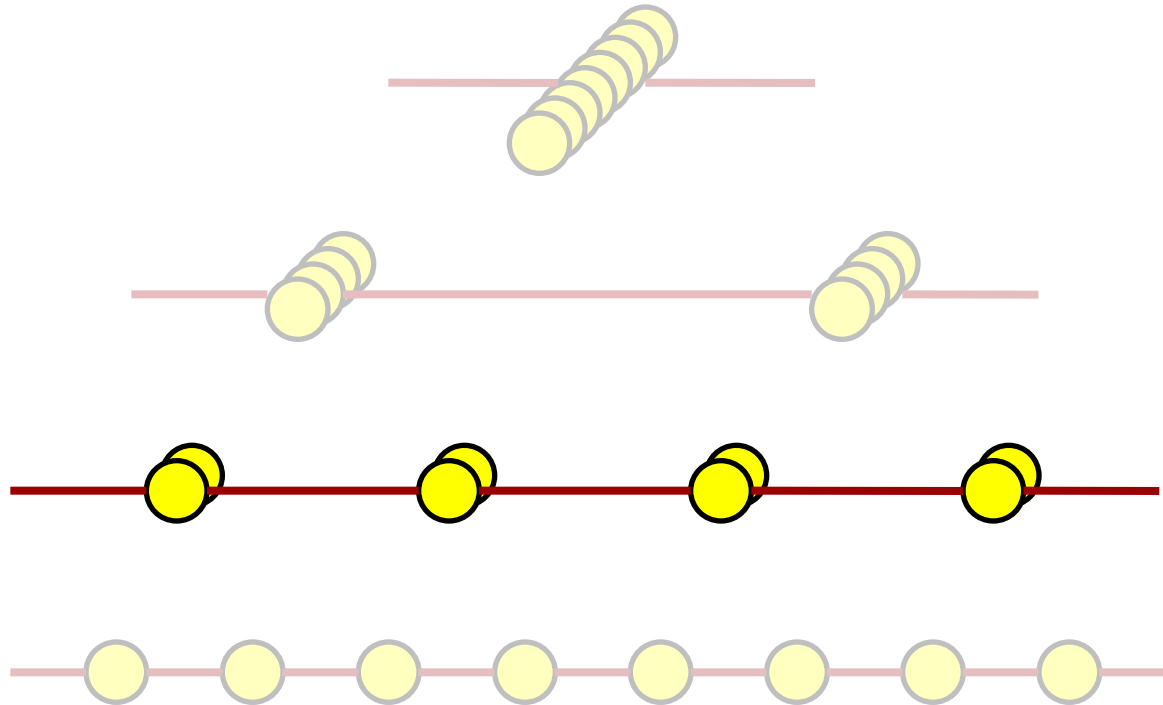
# Erhöhung der Merkmalsanzahl mit abnehmender Auflösung



- Reichhaltige Repräsentationen auch in höheren Ebenen

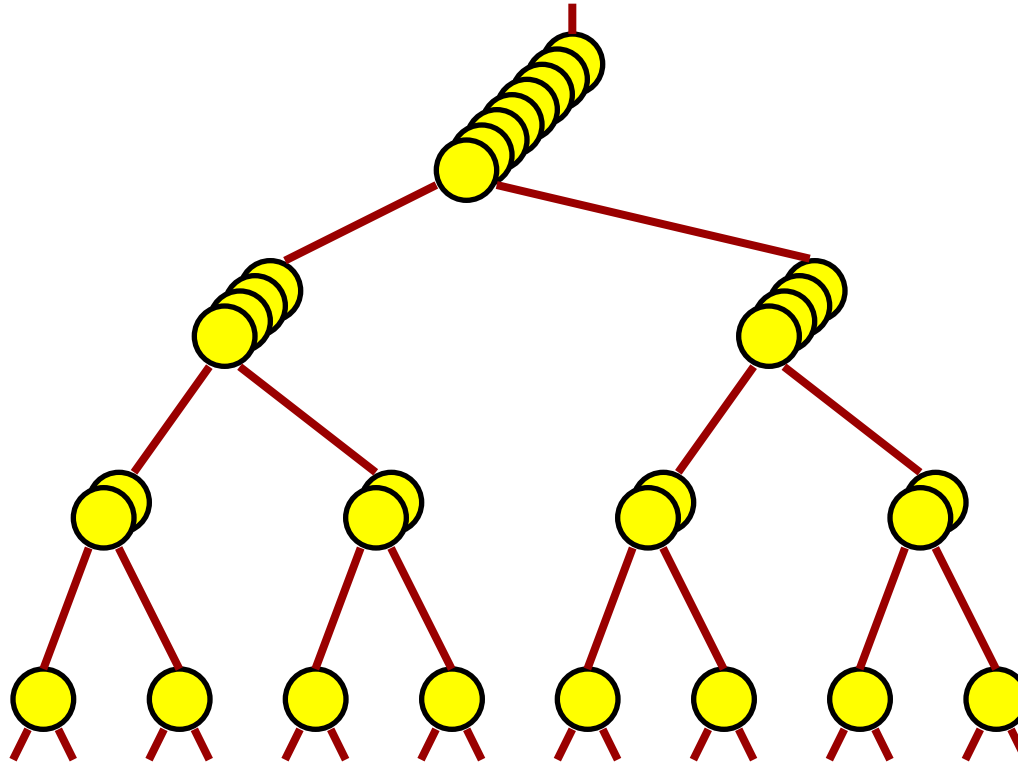


# Modellierung horizontaler Abhängigkeiten



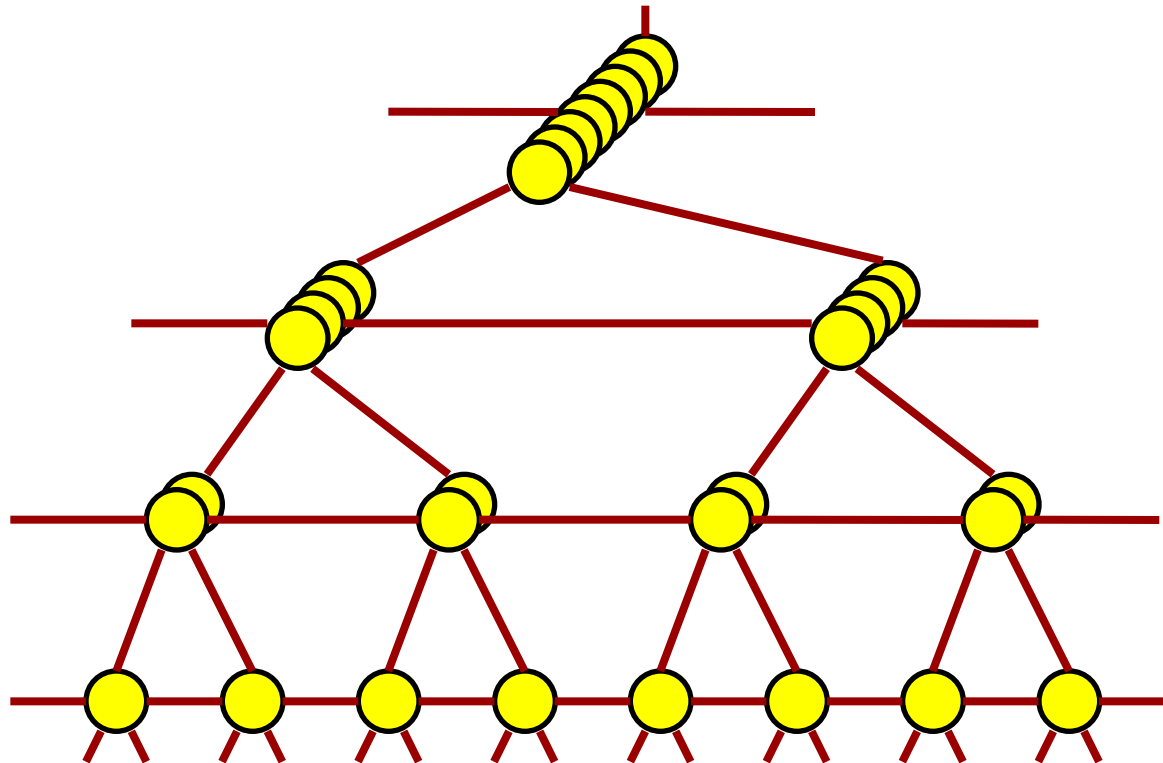
- 1D: HMM, Kalman-Filter, Partikelfilter
- 2D: Markov Random Fields
- Entscheidung für eine Beschreibungsebene ist problematisch
- Ignoriert vertikale Abhängigkeiten, flache Modelle skalieren nicht

# Modellierung Vertikaler Abhängigkeiten



- Strukturgraphen, etc.
- Ignoriert horizontale Abhängigkeiten

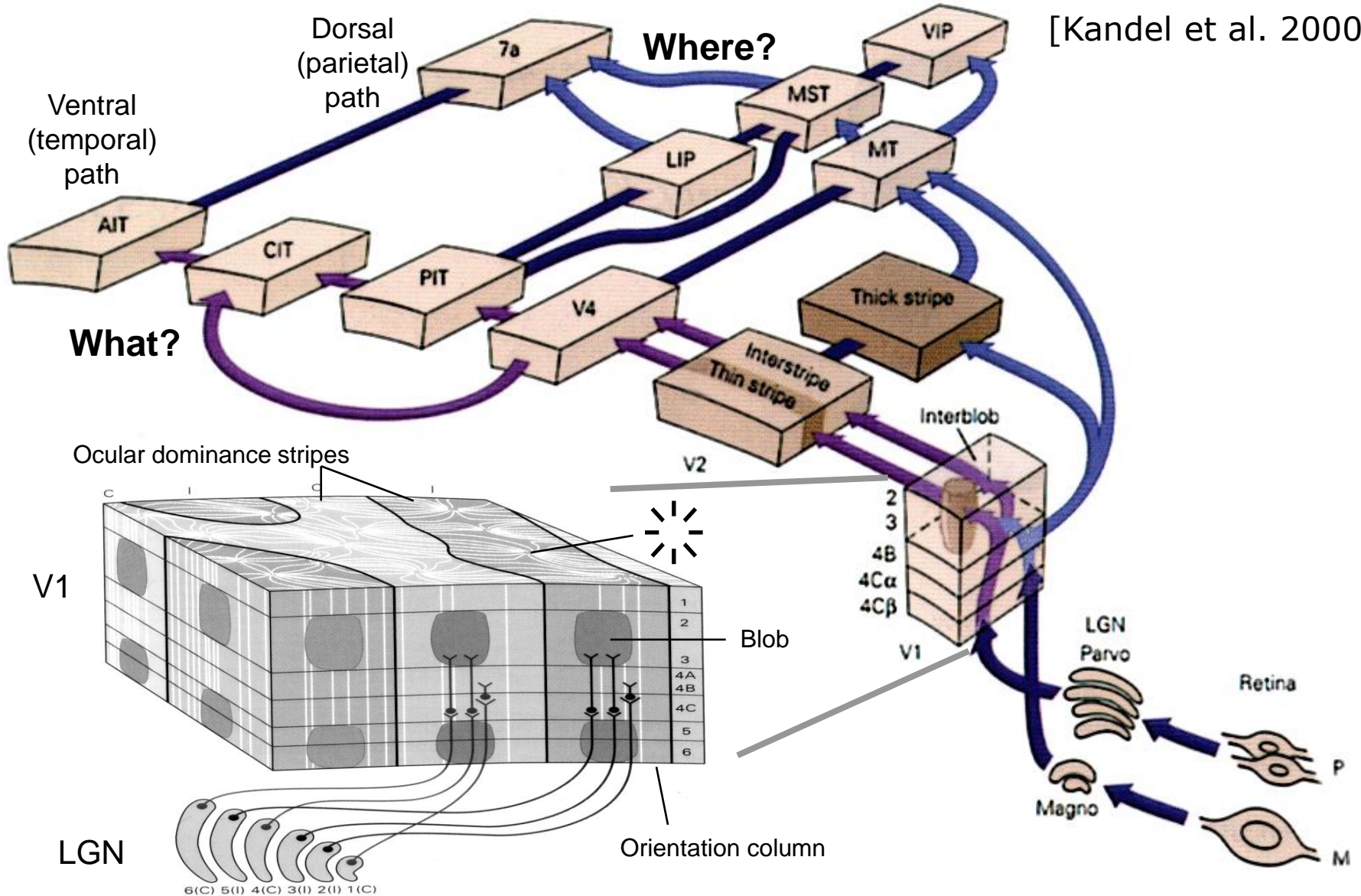
# Horizontale und Vertikale Abhängigkeiten



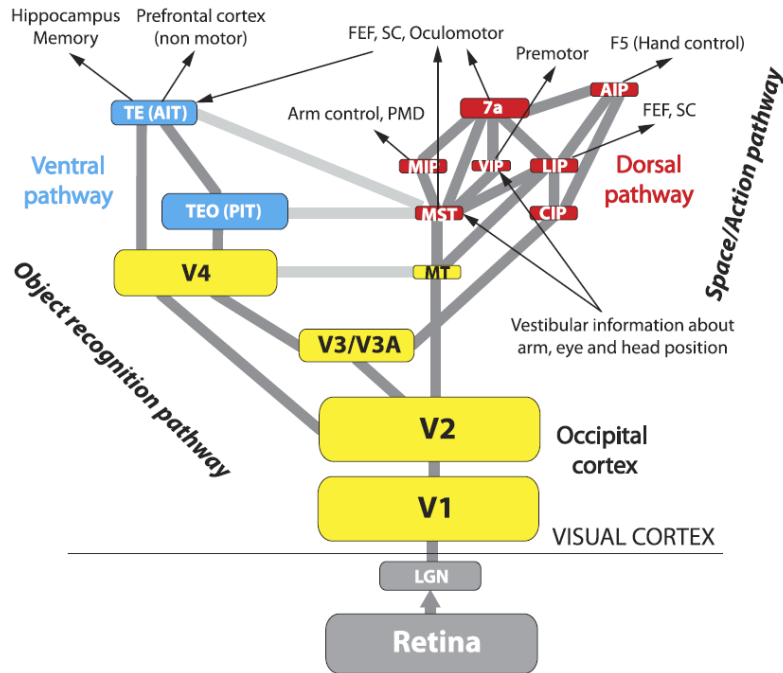
- Problem: Zyklen machen exakte Inferenz unmöglich
- Idee: Nutze approximative Inferenzverfahren

# Visuelles System des Menschen

[Kandel et al. 2000]



# Visuelle Verarbeitungshierarchie

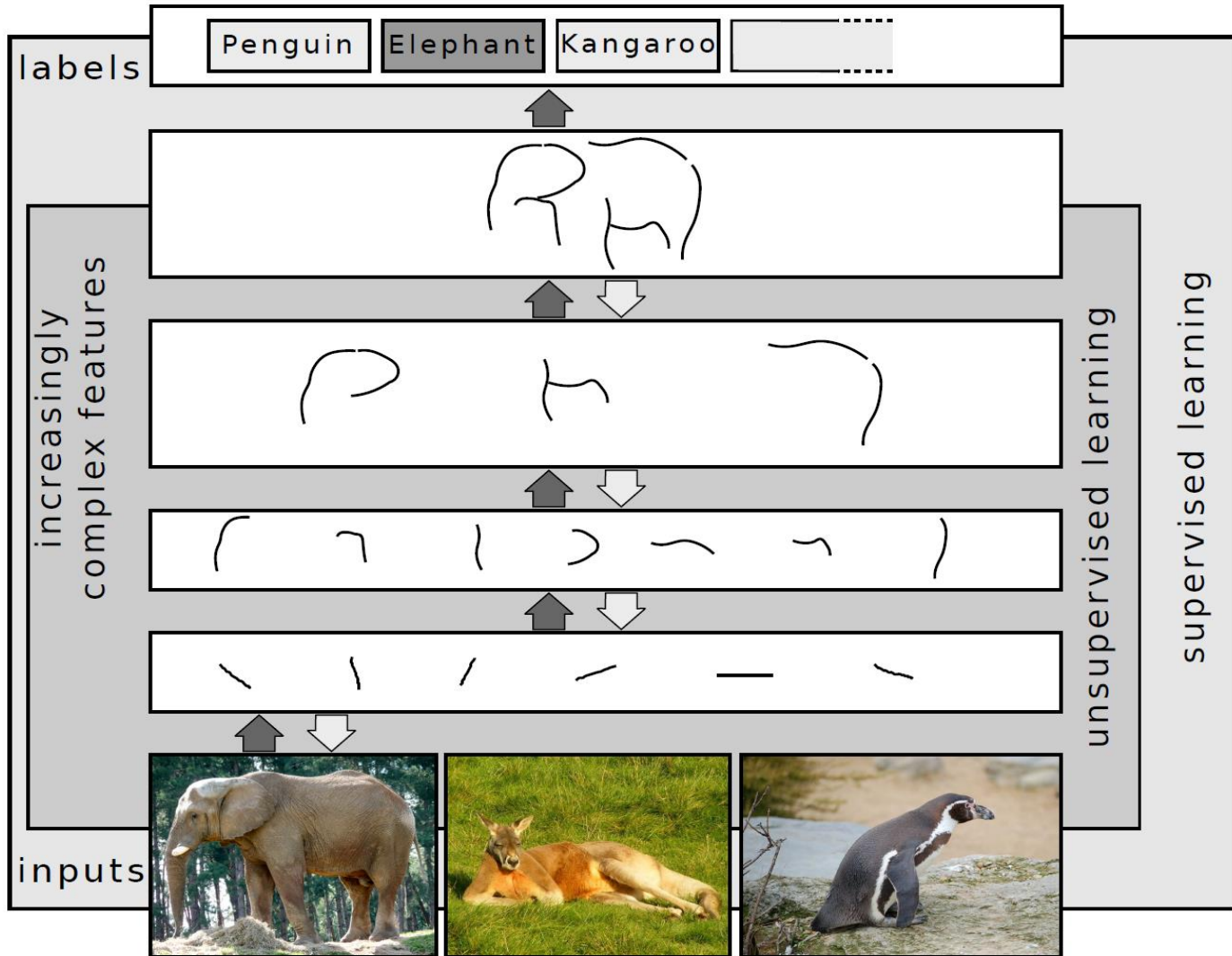


- Steigende Komplexität
- Steigende Invarianz
- Alle Verbindungen bidirektional
- Mehr Feedback als Feedforward
- Laterale Verbindungen wichtig

Area	TE (AIT)	AIP	7a	MIP	VIP	LIP
RF size						
Task						
	ventral			dorsal		
TEO (PIT)						CIP
V4						MST
V3/V3A						MT
V2						V3/V3A
V1						V2
LGN						V1
Retina (ganglion cells)						Retina (ganglion cells)
Retina (receptors)						Retina (receptors)
Area	RF size	Color	2D Shape	3D Shape	Motion	RF size
Area	RF size	Color	2D Shape	3D Shape	Motion	RF size

[Krüger et al., TPAMI 2013]

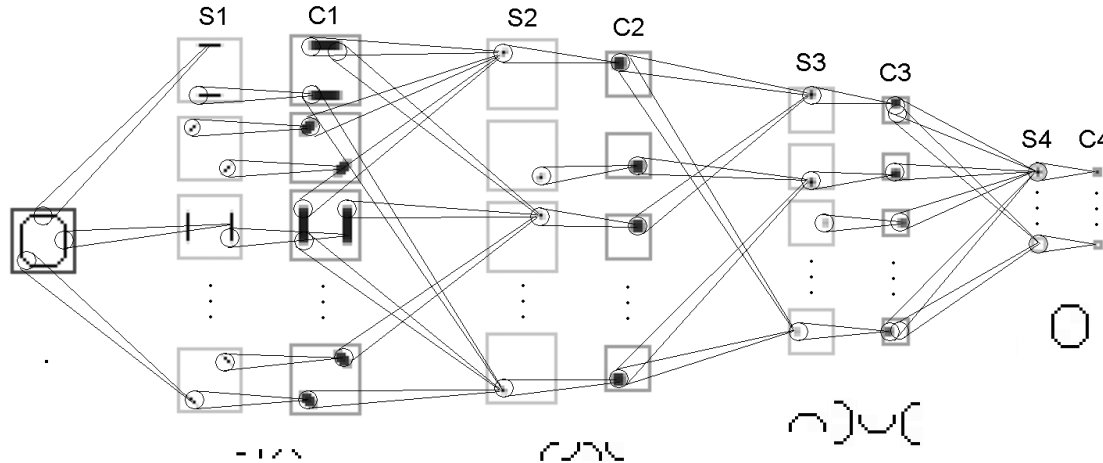
# Repräsentationsebenen



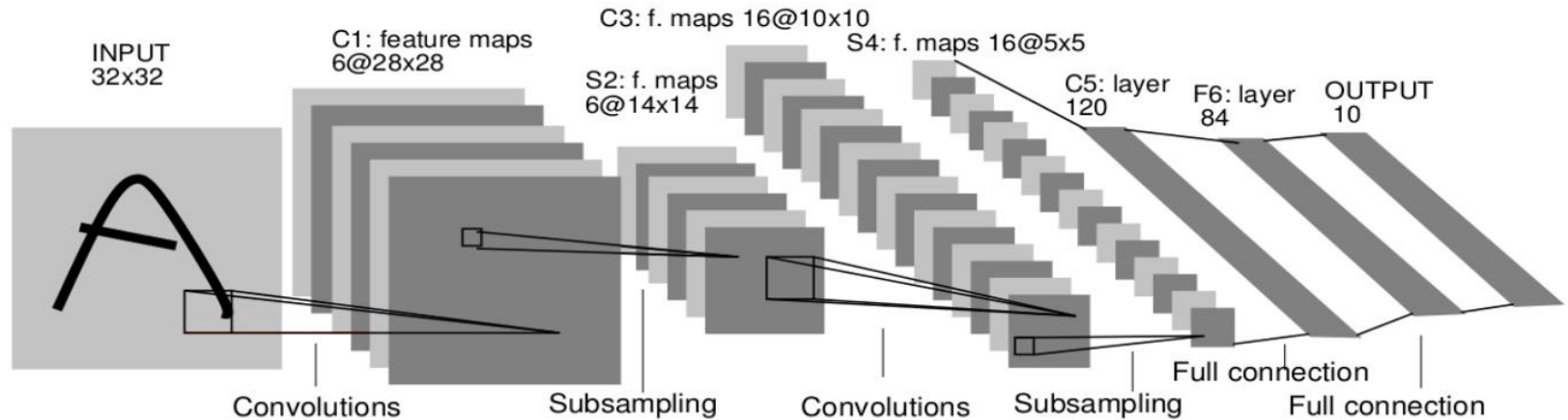
[Schulz and Behnke, KI 2012]

# Konvolutionale Neuronale Netze

## ■ Neocognitron: Fukushima 1980

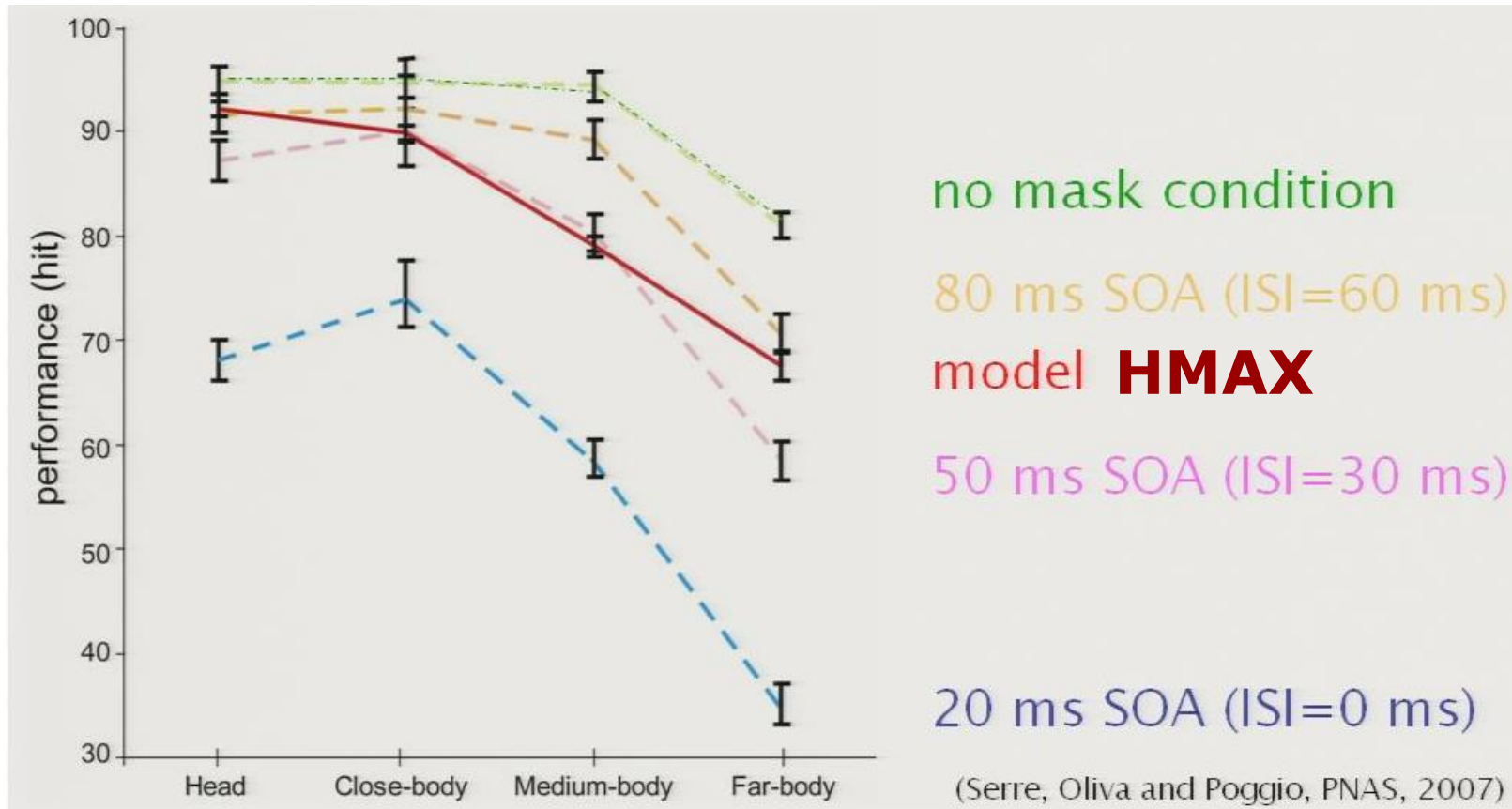


## ■ Überwachtes Training: LeCun 1989



# Feed-Forward-Modelle können Performanz nicht erklären

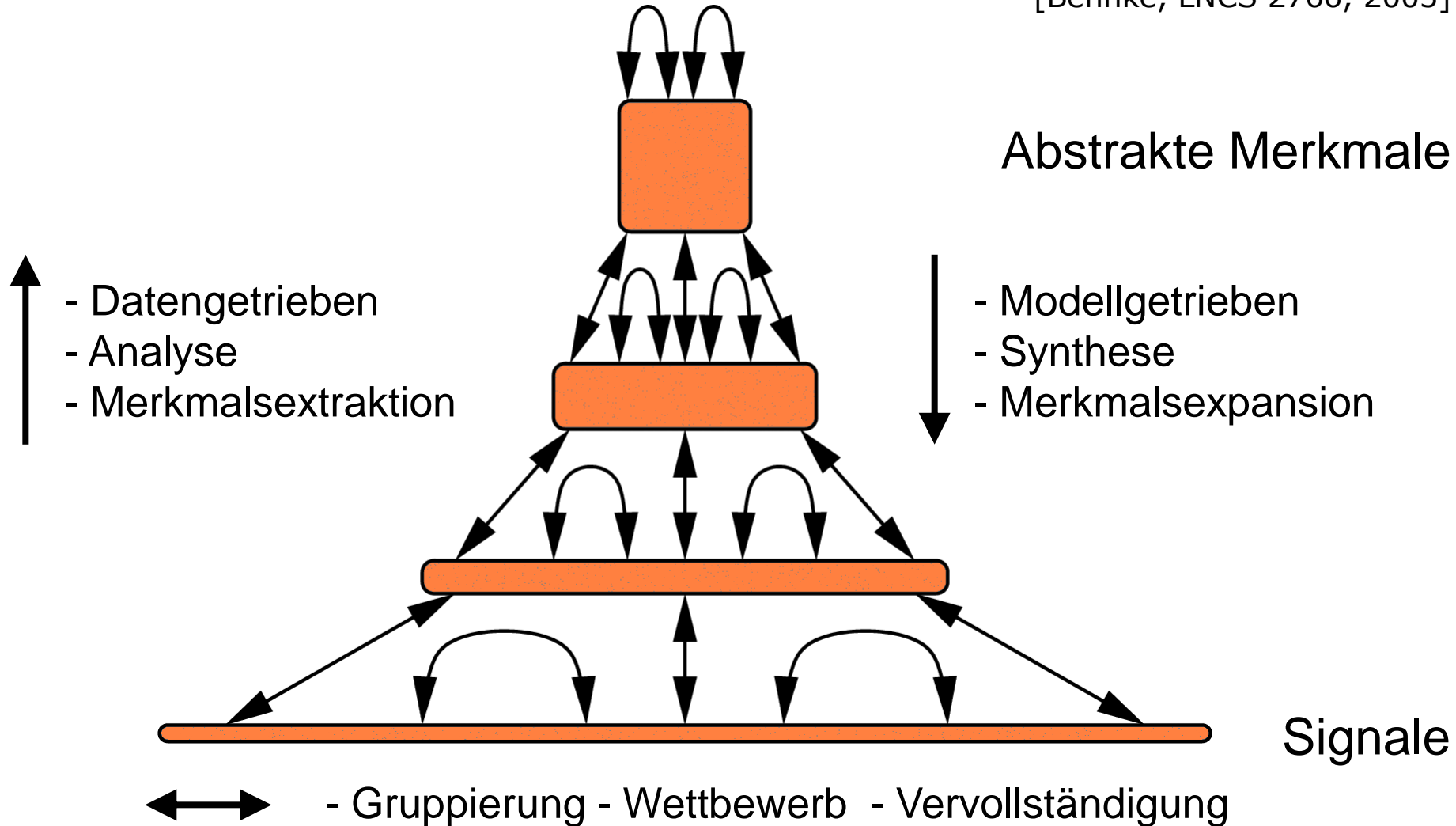
- Performanz steigt mit Beobachtungszeit





# Neurale Abstraktionspyramide

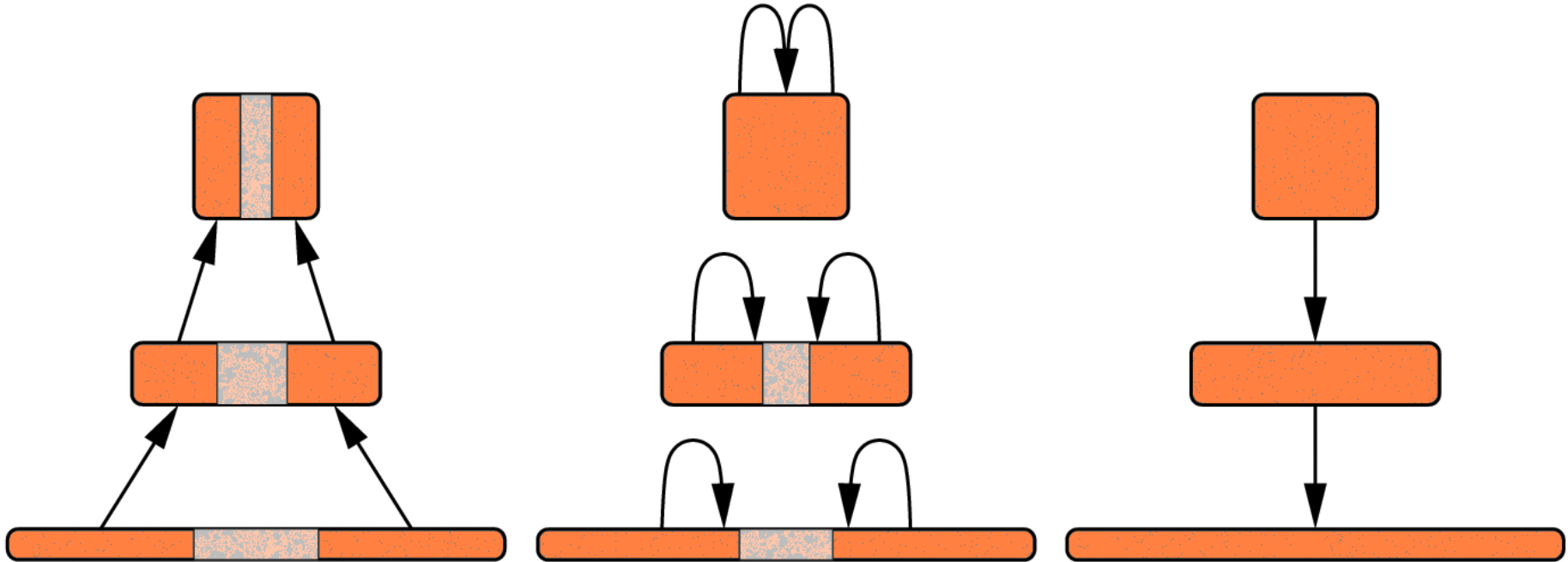
[Behnke, LNCS 2766, 2003]



# Iterative Interpretation

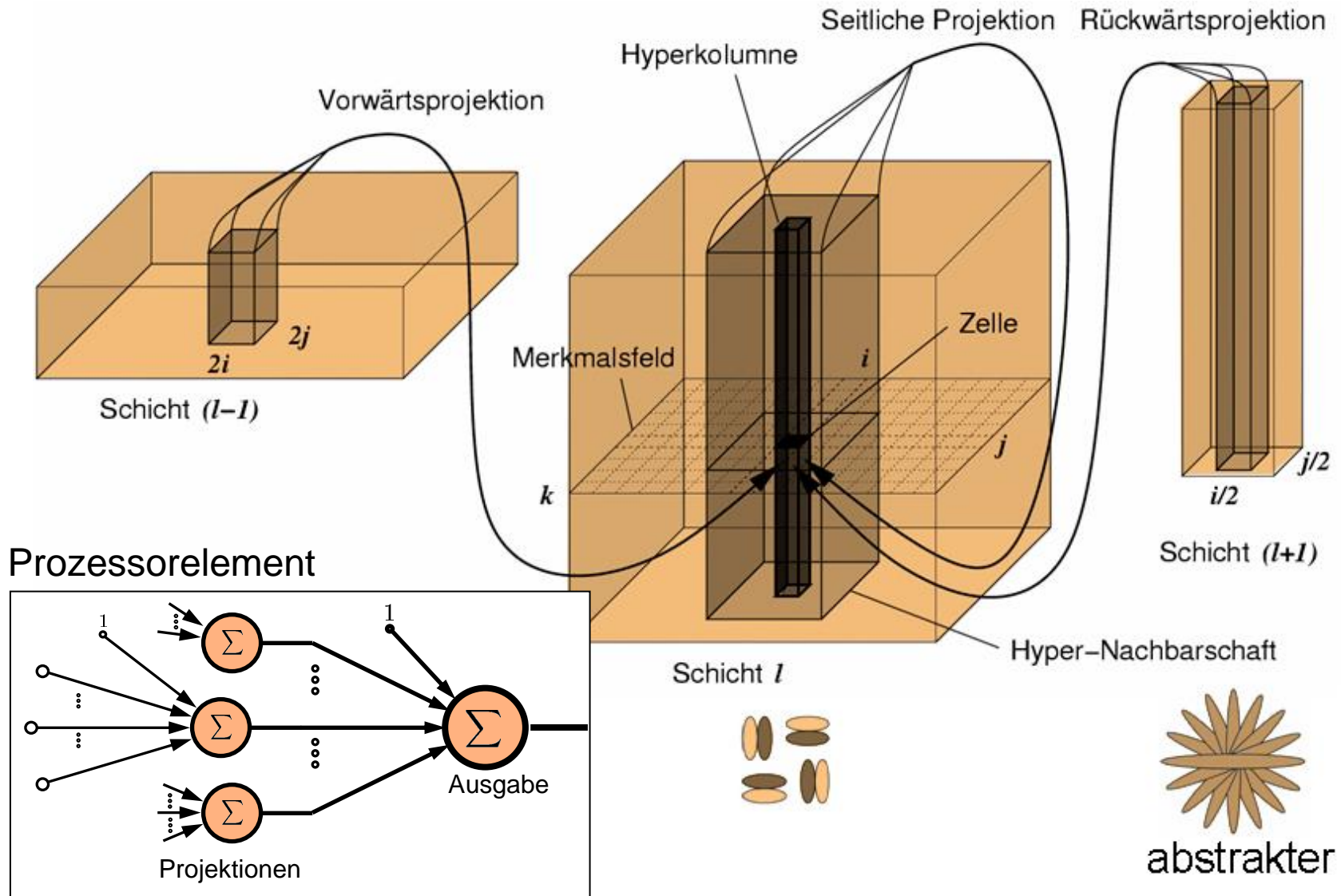
[Behnke, LNCS 2766, 2003]

- Interpretiere die einfachsten Stellen zuerst



- Nutze Teil-Interpretationen als Kontext zur Auflösung von Mehrdeutigkeiten

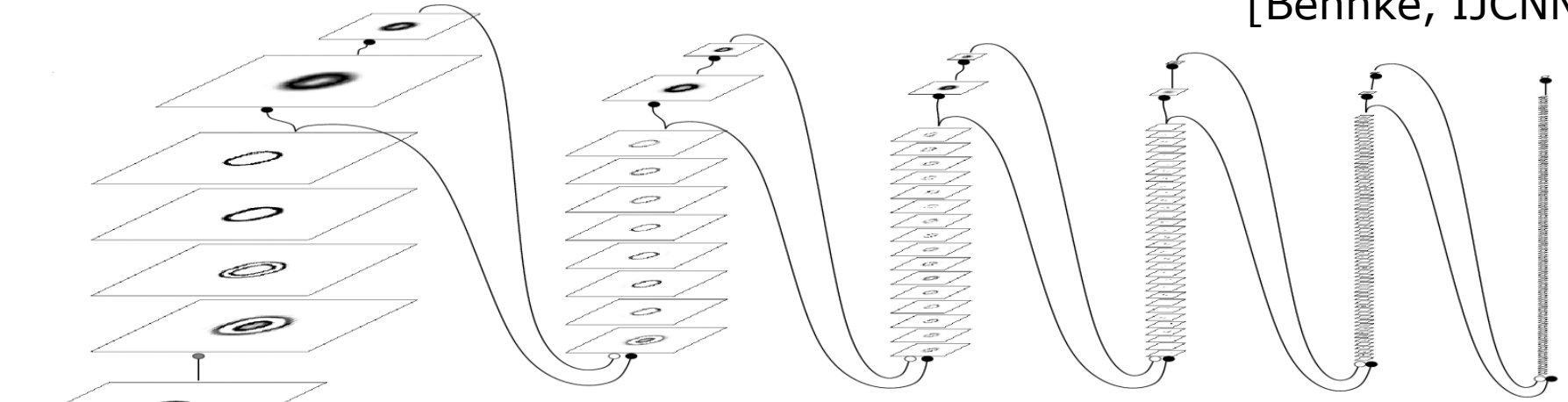
# Lokal Rekurrente Verbindungen



[Behnke, LNCS 2766, 2003]

# Lernen einer Merkmalshierarchie

[Behnke, IJCNN'99]



32x32 x 4

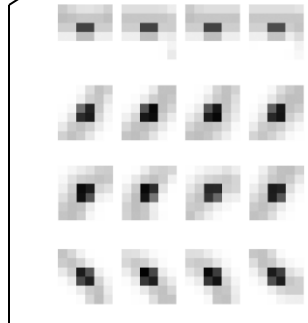
16x16 x 8

8x8 x 16

4x4 x 32

2x2 x 64

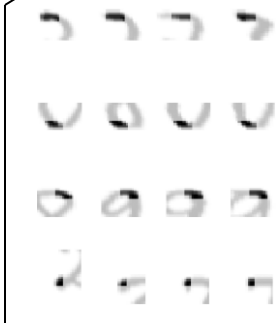
1x1 x 128



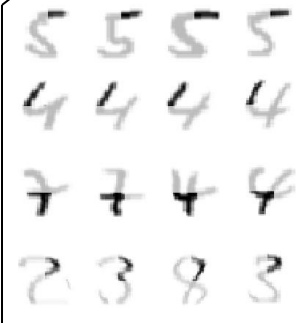
Kanten



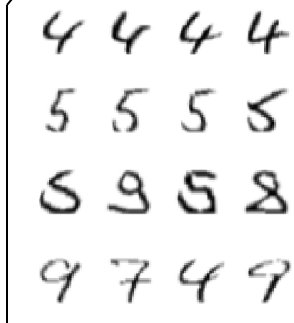
Linien



Kurven



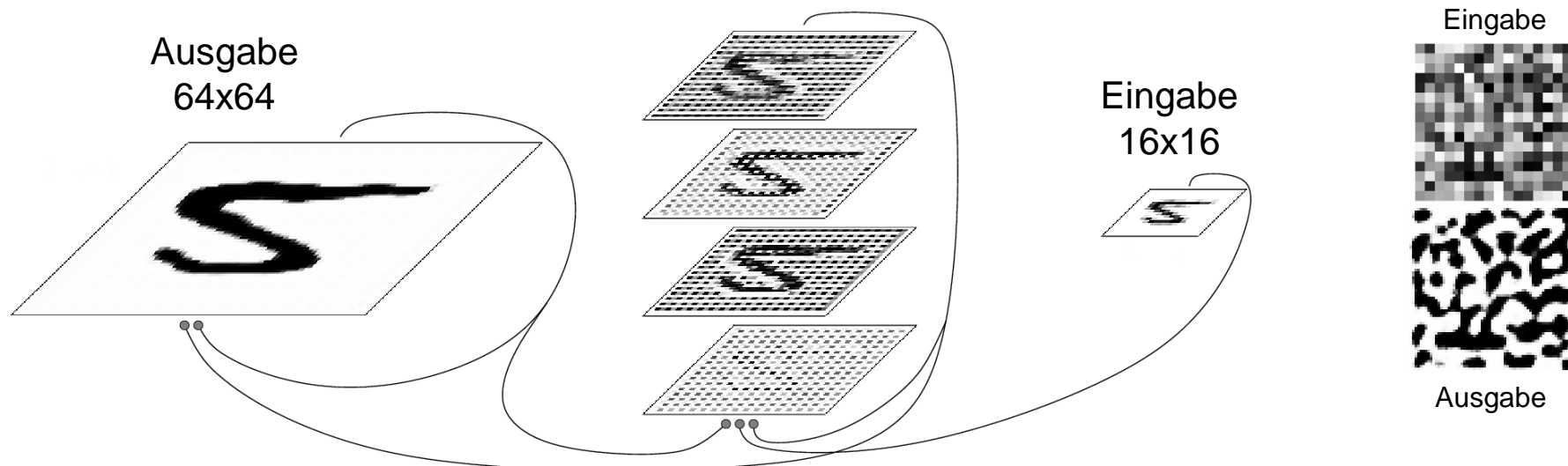
Teile



Ziffern

# Superresolution

[Behnke, IJCAI'01]

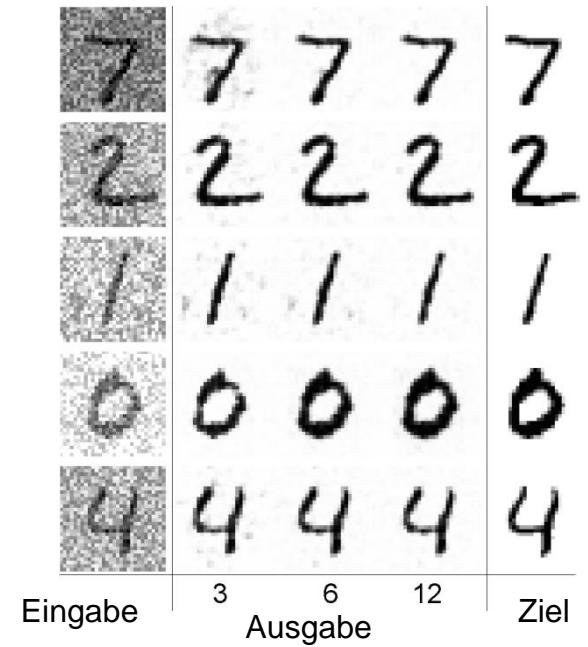
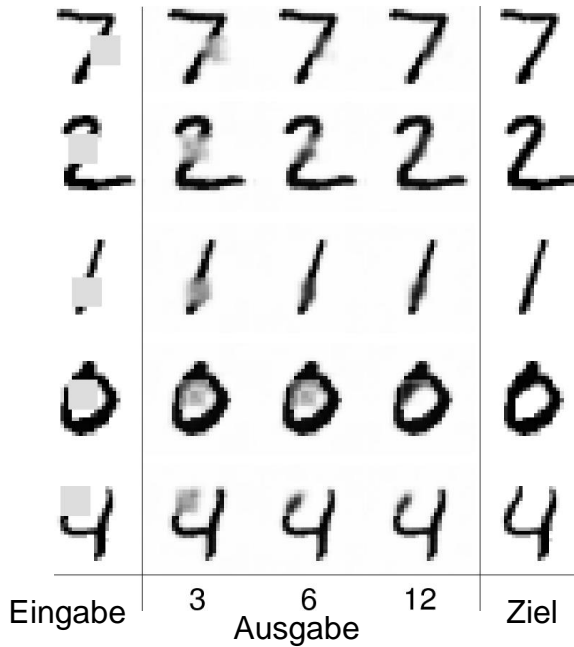
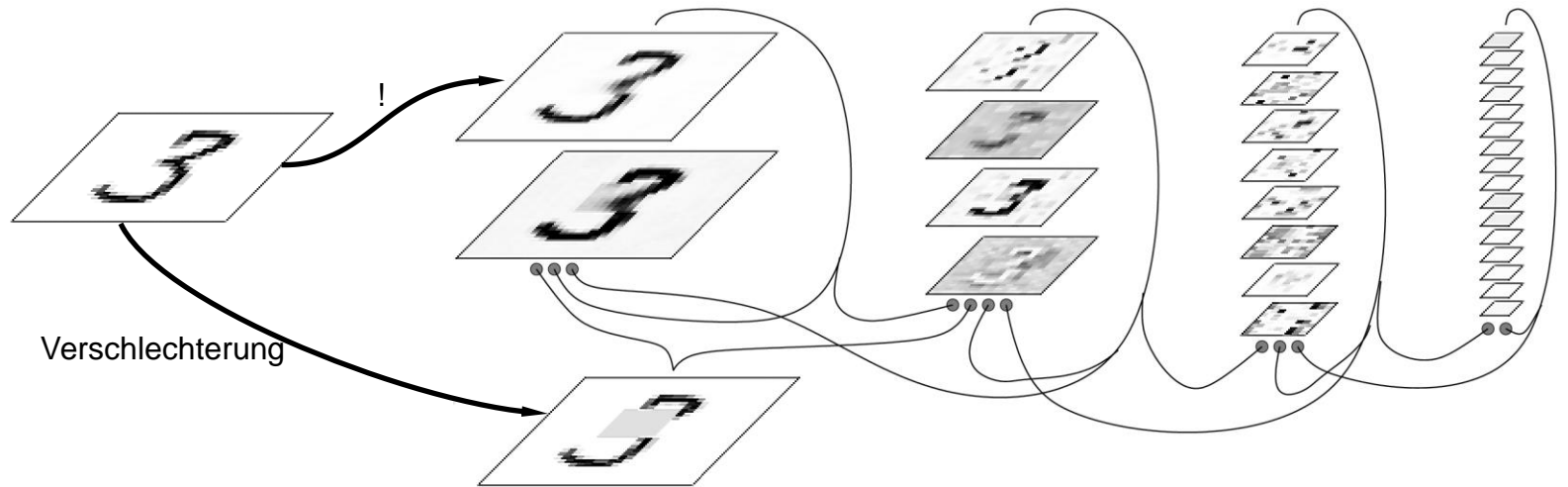


5	5	5	5	5	5
1	1	1	1	1	1
4	4	4	4	4	4
3	3	3	3	3	3
6	6	6	6	6	6
Eingabe	2	3	5	10	Ziel

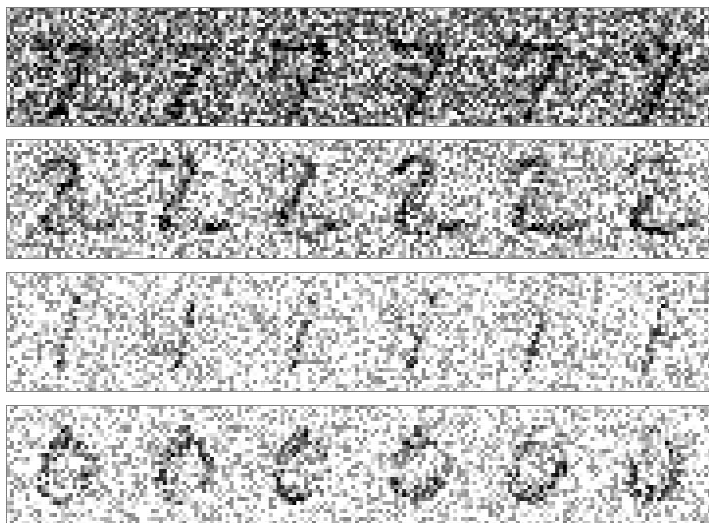
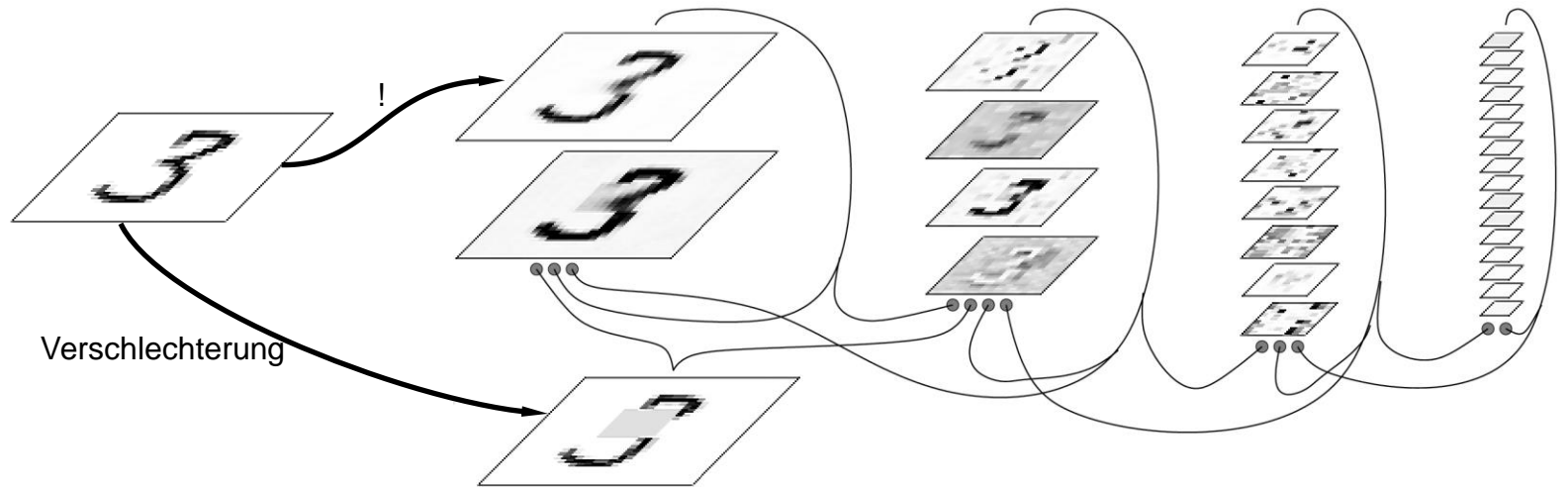
Ausgabe



# Ziffern-Rekonstruktion [Behnke, IJCAI'01]

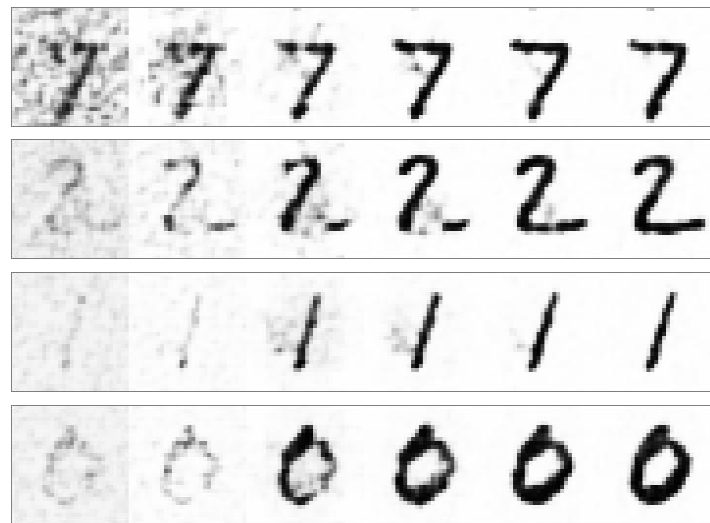


# Ziffern-Rekonstruktion [Behnke, IJCAI'01]



1 2 4 7 11 16

Eingabe



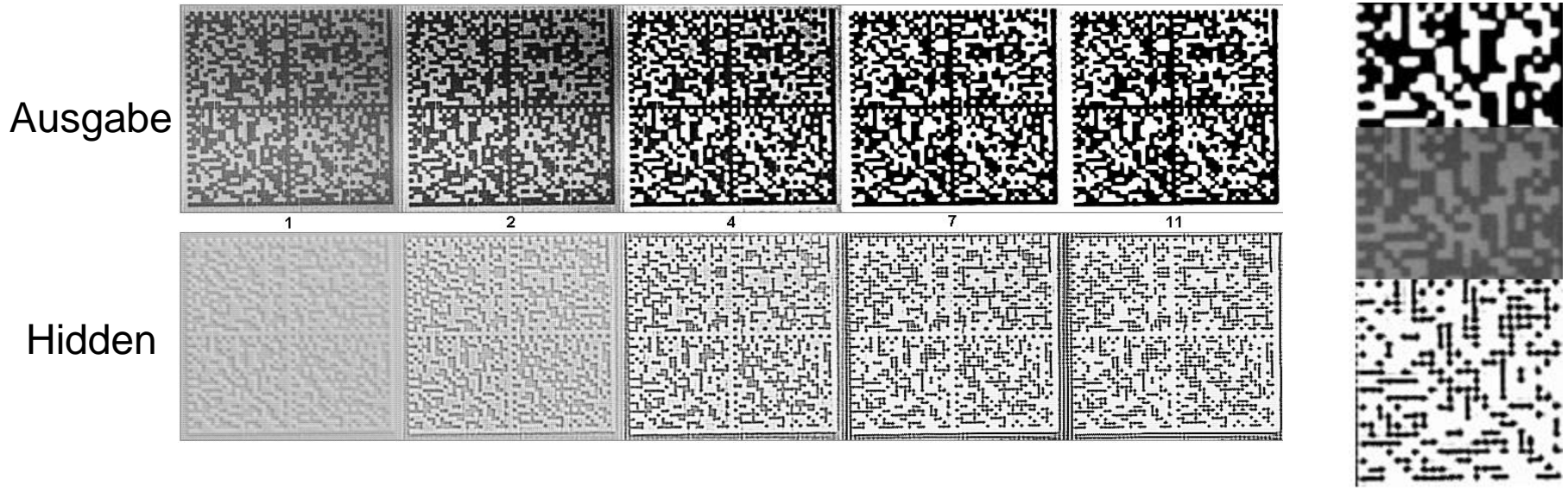
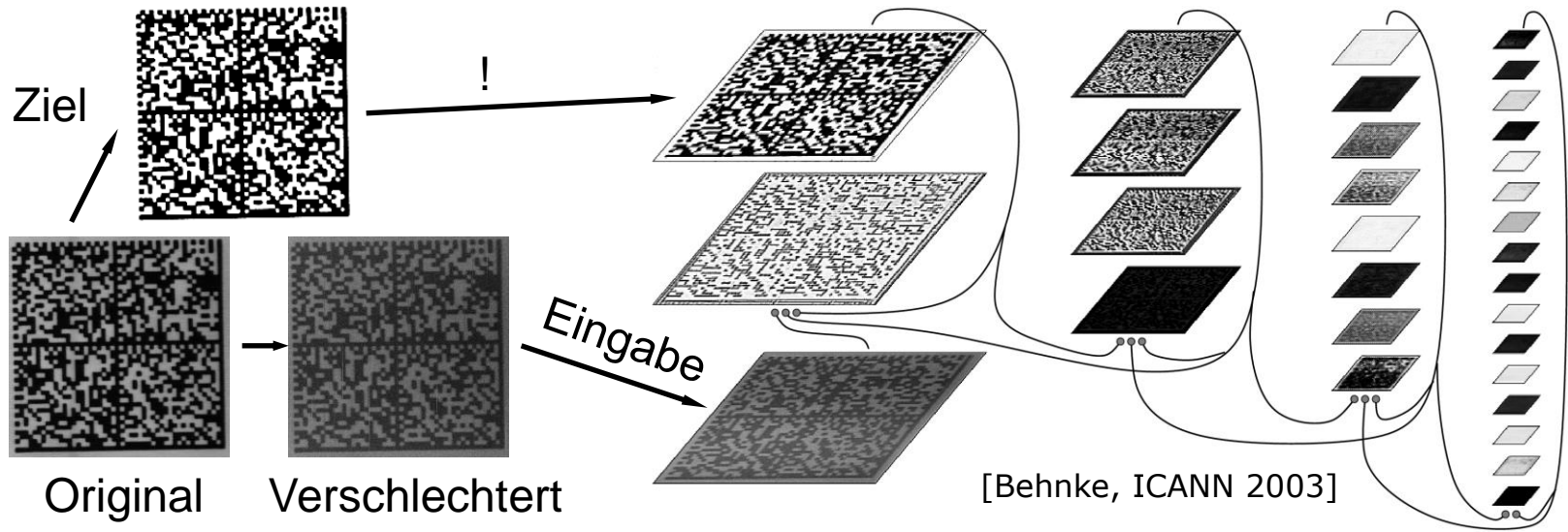
1 2 4 7 11 16

Ausgabe



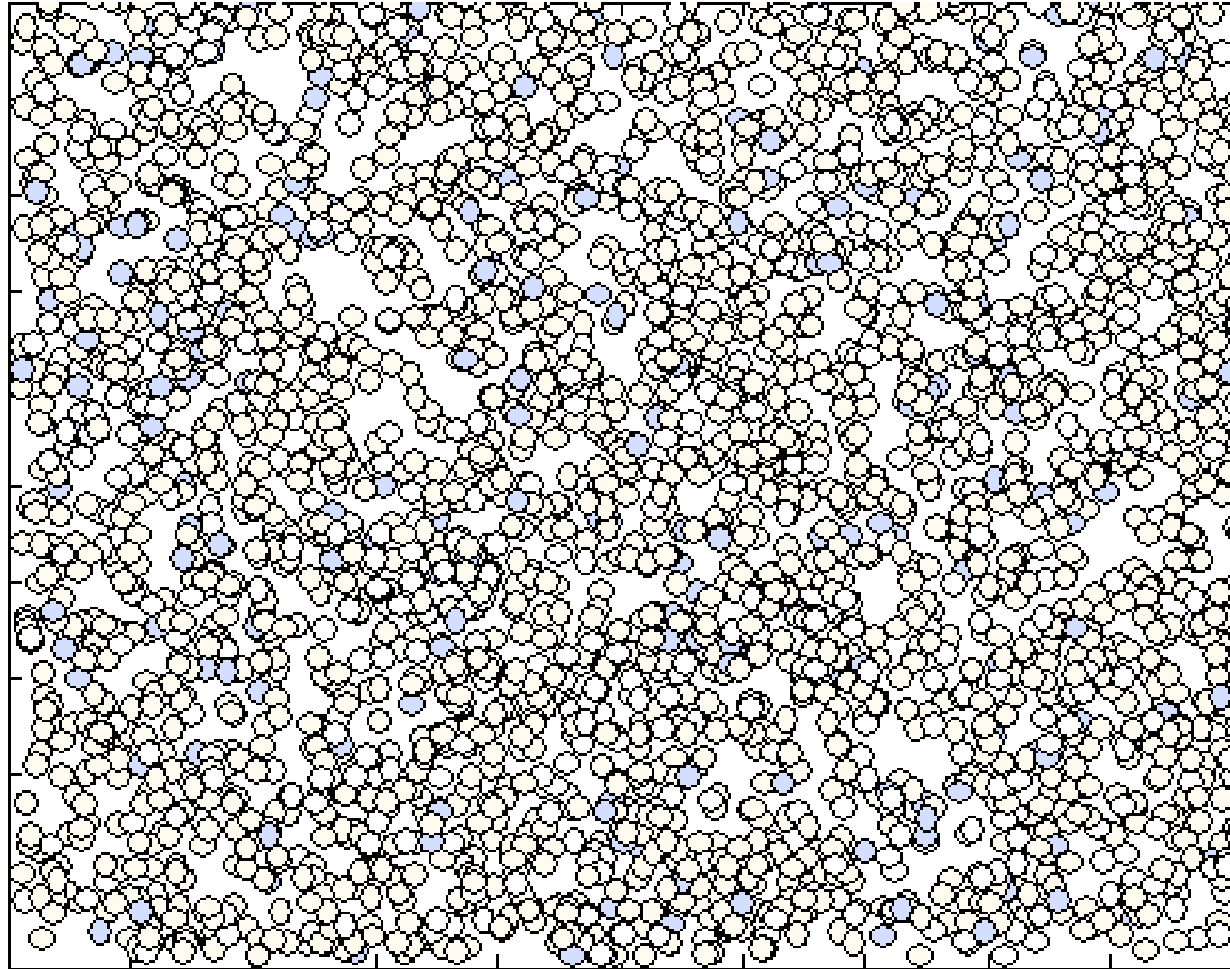
Ziel

# Binarisierung von Matrixcodes





# Kontinuierliche Attraktoren



- Lokale Erregung und globale Hemmung
- Stabile Aktivitätsblobs können verschoben werden

# Augen-Detektion

[Behnke, KES'03]

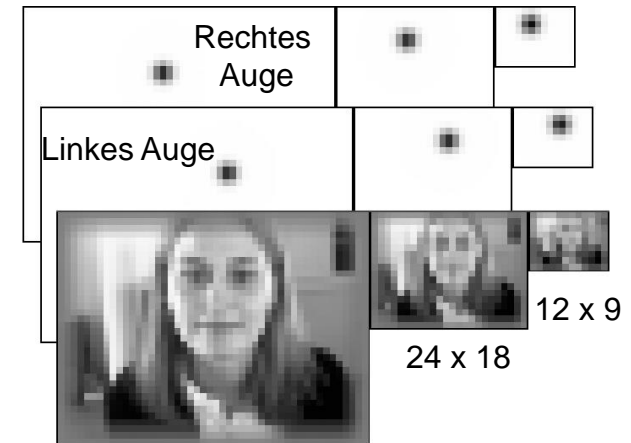
- BioID-Daten:
  - 1521 Bilder
  - 23 Personen



- Kodiere Augenpositionen mit Blobs



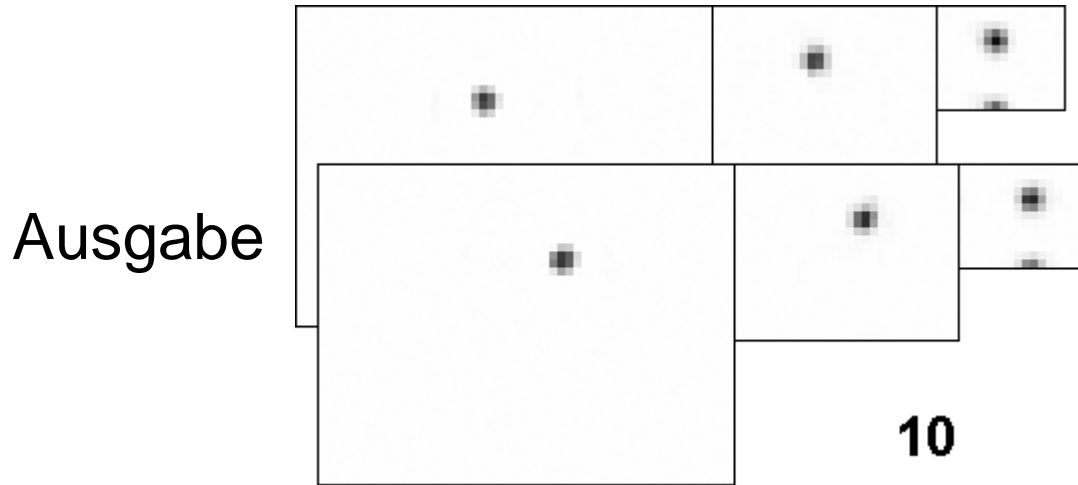
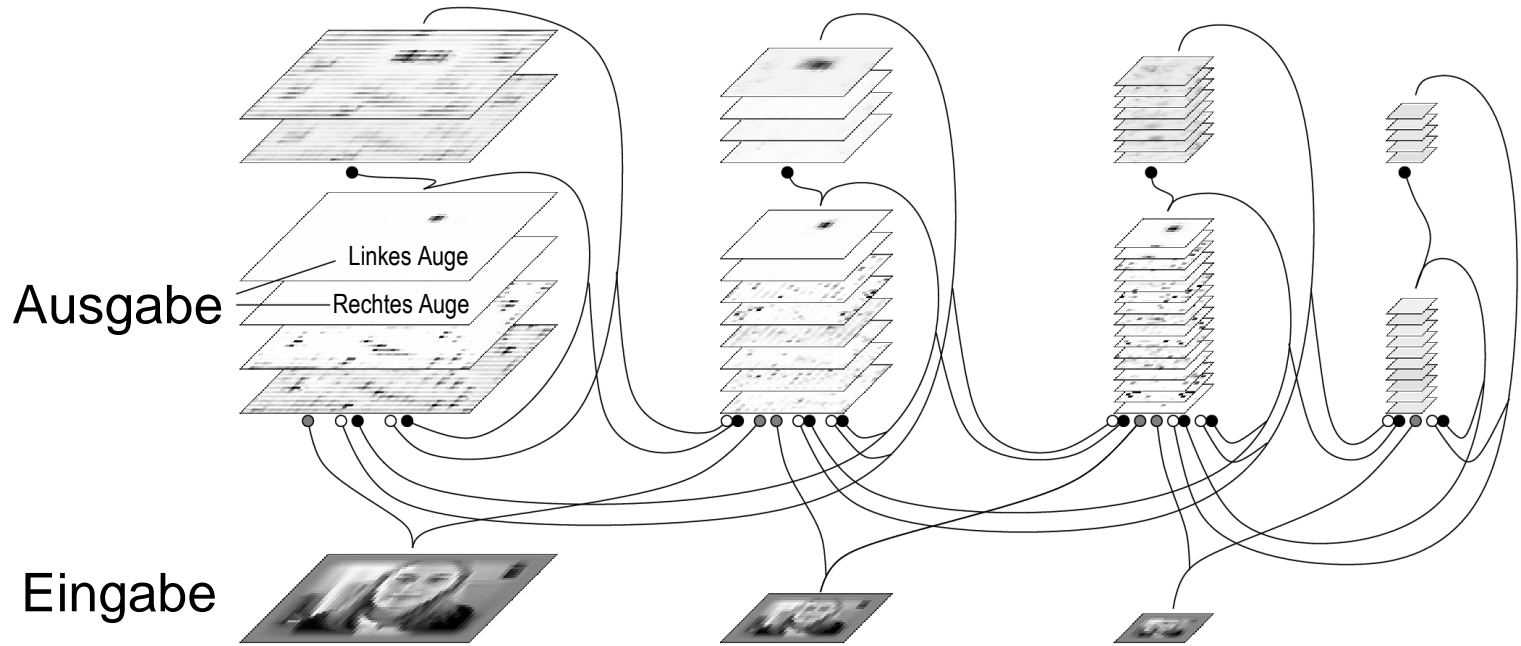
384 x 288



48 x 36

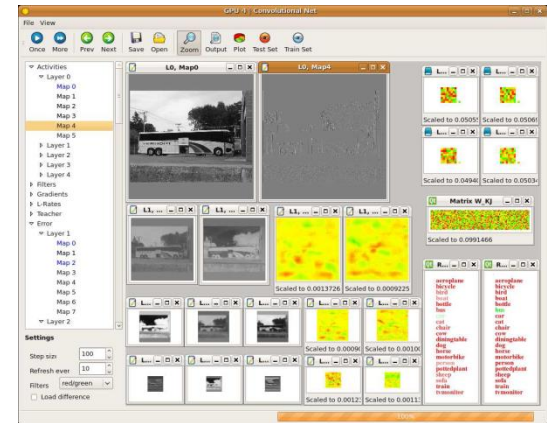
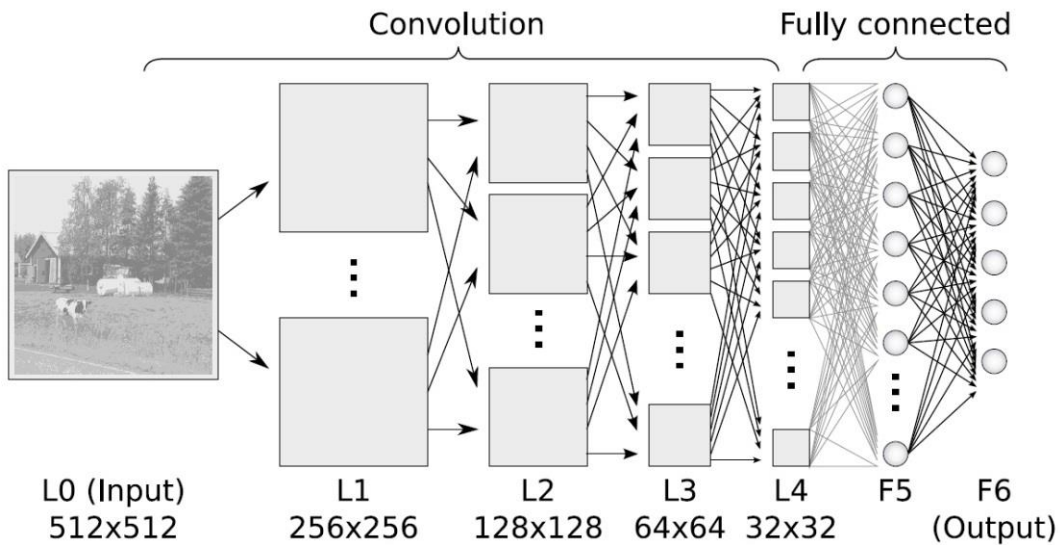
# Augen-Detektion

[Behnke, KES'03]



# GPU-Implementierung (CUDA)

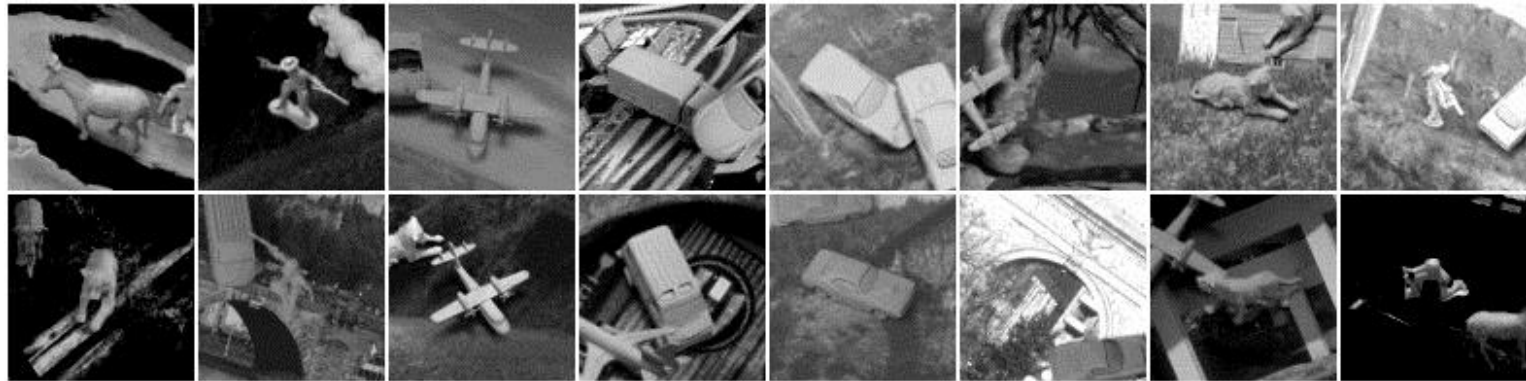
- Preiswerte Parallelrechner
- Programmierbar mit CUDA
- Konvolutional [Scherer & Behnke, 2009]



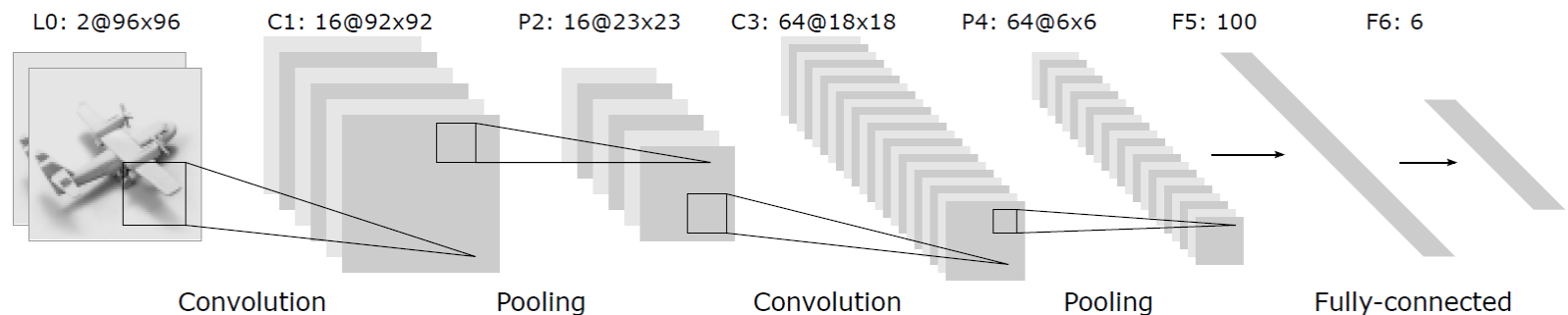
- Lokale Konnektivität [Uetz & Behnke, 2009]

# Objekterkennung: NORB

- 10 Kategorien, verrauscht



- **Backpropagation durch Max-Pooling**

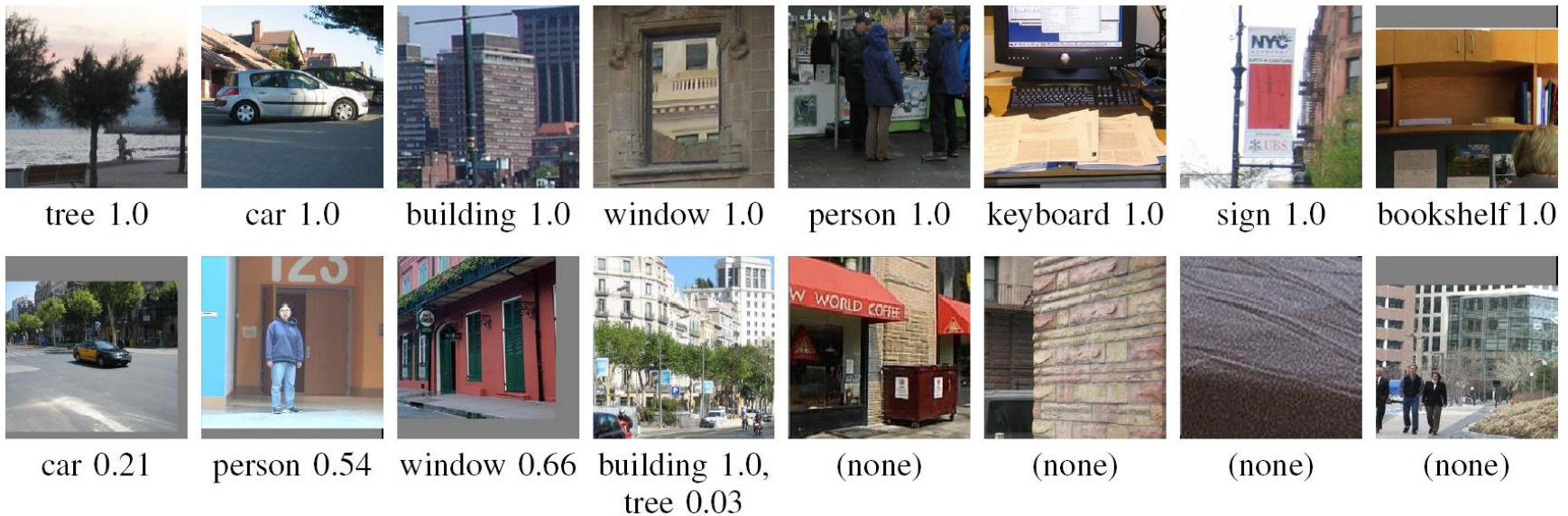


Testfehler: 5,6% (LeNet7: 7.8%)

[Scherer, Müller, Behnke, ICANN'10]

# Objekterkennung: LabelMe

- 50,000 Farbbilder (256x256)
- 12 Klassen + Clutter (50%)

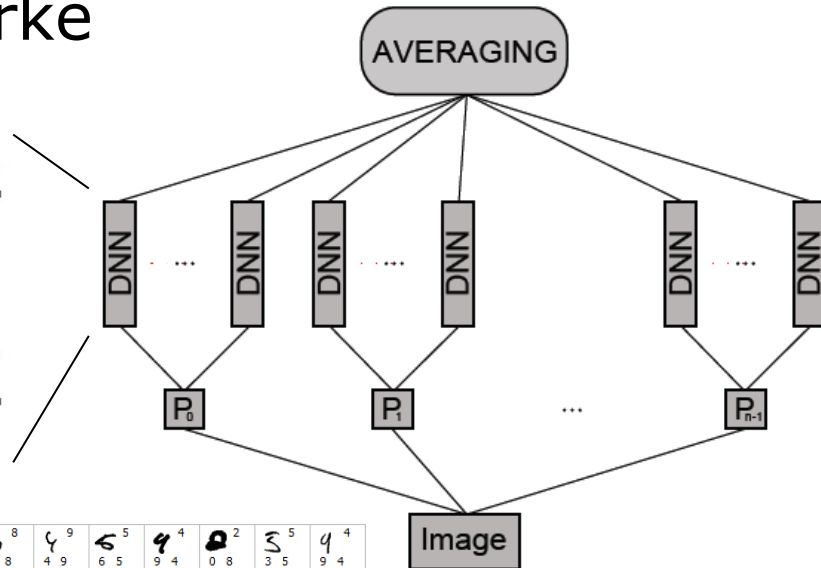
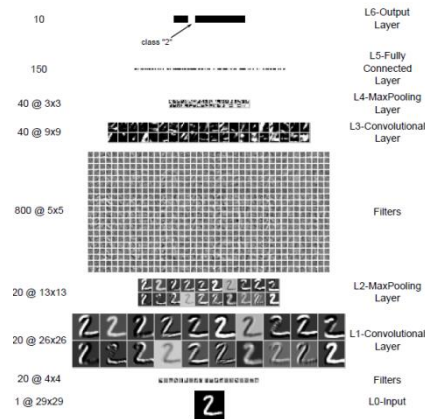
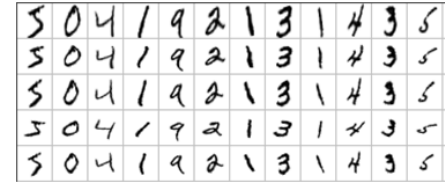


- Fehler TRN: 3.77%; TST: 16.27%
- Erkennungsgeschwindigkeit: 1.356 Bilder/s

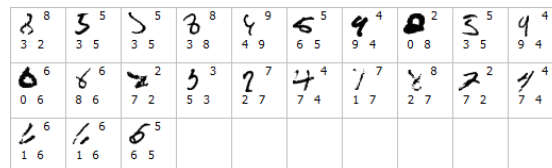
[Uetz, Behnke, ICIS2009]

# Multi-Column Deep Convolutional Networks

- Verschiedene Vorverarbeitungen
- Training mit Verzerrungen
- Mitteln tiefer Netzwerke



- MNIST: 0.23%
- NORB: 2.7%
- CIFAR10: 11.2%
- Verkehrszeichen: 0.54%



[Ciresan et al. CVPR 2012]

# ImageNet-Challenge

- 1.2 Millionen Bilder
- 1000 Kategorien
- Hierarchische Kategorie-Struktur (WordNet)



Golf cart (motor vehicle, self-propelled vehicle, wheeled vehicle, ... Egyptian cat (domestic cat, domestic animal, animal)

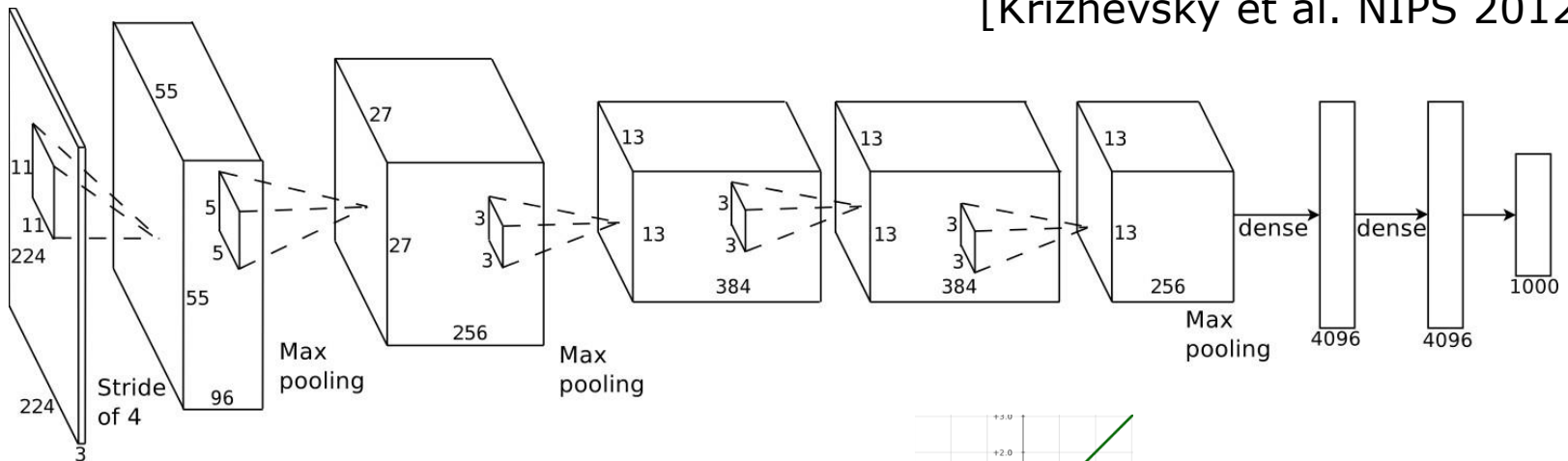
- Aufgabe: Erkenne Objekt im Bild
- Zusätzliche Detektionen werden kaum bestraft
- Hierarchische Fehlerberechnung

[Deng et al. 2009]

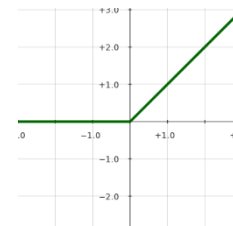


# AlexNet

[Krizhevsky et al. NIPS 2012]



- Rektifizierende Transferfunktion
- 650.000 Neuronen
- 60.000.000 Parameter
- 630.000.000 Verbindungen
- Trainiert mit Dropout und Daten-Augmentierung
- Test von 10 Teilbildern
- ILSVRC-2012: top-5 error 15.3%



96 gelernte Low-level-Filter

# Klassifikationsergebnisse



**mite**

**container ship**

**motor scooter**

**leopard**

	<p><b>mite</b></p> <p>black widow</p> <p>cockroach</p> <p>tick</p> <p>starfish</p>		<p><b>container ship</b></p> <p>lifeboat</p> <p>amphibian</p> <p>fireboat</p> <p>drilling platform</p>		<p><b>motor scooter</b></p> <p>go-kart</p> <p>moped</p> <p>bumper car</p> <p>golfcart</p>		<p><b>leopard</b></p> <p>jaguar</p> <p>cheetah</p> <p>snow leopard</p> <p>Egyptian cat</p>
--	--	--	--	--	---	--	--



**grille**

**mushroom**

**cherry**

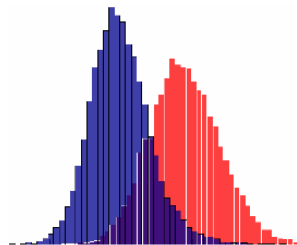
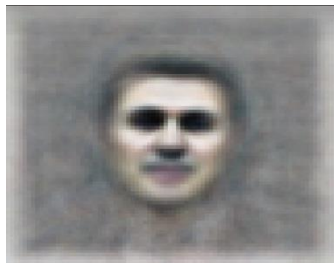
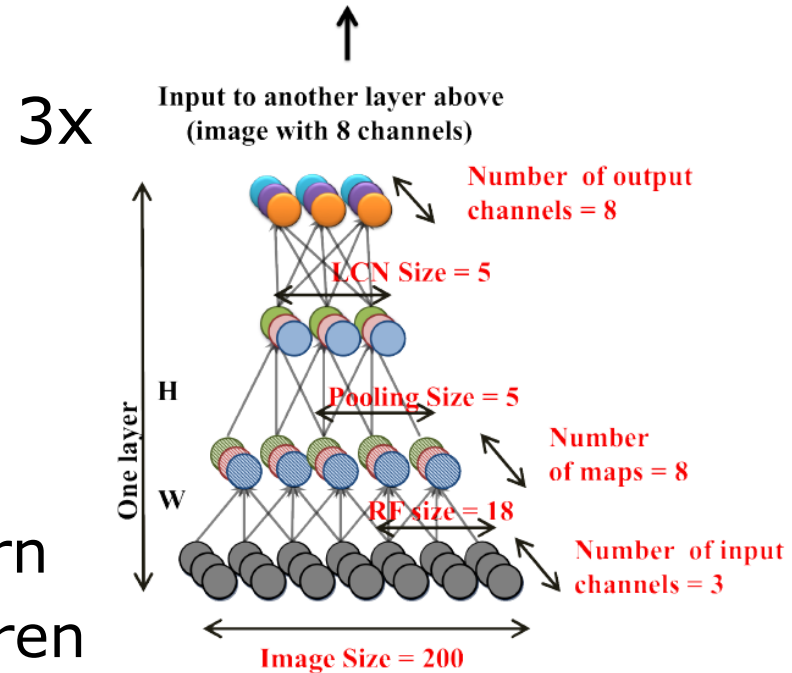
**Madagascar cat**

	<p><b>convertible</b></p> <p>grille</p> <p>pickup</p> <p>beach wagon</p> <p>fire engine</p>		<p><b>agaric</b></p> <p>mushroom</p> <p>jelly fungus</p> <p>gill fungus</p> <p>dead-man's-fingers</p>		<p><b>dalmatian</b></p> <p>grape</p> <p>elderberry</p> <p>ffordshire bullterrier</p> <p>currant</p>		<p><b>squirrel monkey</b></p> <p>spider monkey</p> <p>titi</p> <p>indri</p> <p>howler monkey</p>
--	---	--	---	--	---	--	--

[Krizhevsky et al.  
NIPS 2012]

# Unüberwachtes Lernen von Merkmalen

- 9 Schichten
- Lokale Verbindungsstruktur
- Spärlicher Autoencoder
- L2-Pooling
- Lokale Kontrastnormalisierung
- 1 Milliarde Verbindungen
- Trainiert auf 10 Millionen Bildern
- Unüberwacht gelernte Detektoren



- Überwachte ImageNet2011-Resultate (14M Bilder, 22K Kategorien): 15.8%  
[Le et al. 2012]

# Objektkategorisierung Besser als Menschen



GT: horse cart  
1: horse cart  
 2: minibus  
 3: oxcart  
 4: stretcher  
 5: half track



GT: birdhouse  
1: birdhouse  
 2: sliding door  
 3: window screen  
 4: mailbox  
 5: pot



GT: forklift  
1: forklift  
 2: garbage truck  
 3: tow truck  
 4: trailer truck  
 5: go-kart



GT: letter opener  
 1: drumstick  
 2: candle  
 3: wooden spoon  
 4: spatula  
 5: ladle



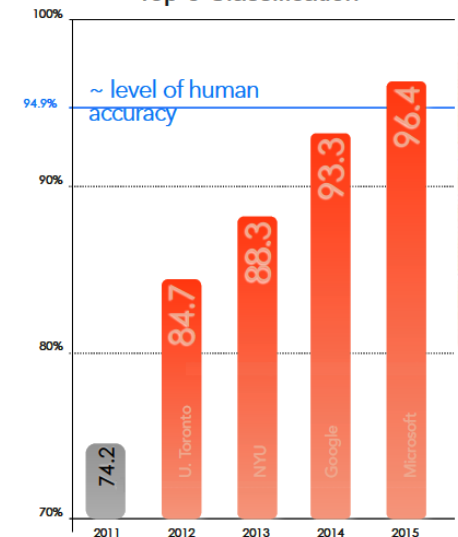
GT: coucal  
1: coucal  
 2: indigo bunting  
 3: lorikeet  
 4: walking stick  
 5: custard apple



GT: komondor  
1: komondor  
 2: patio  
 3: llama  
 4: mobile home  
 5: Old English sheepdog



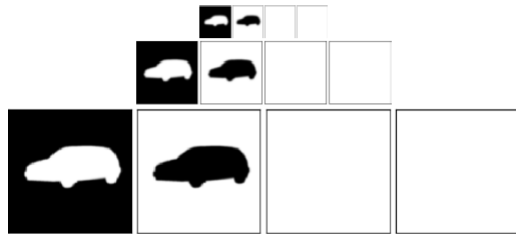
GT: yellow lady's slipper  
1: yellow lady's slipper  
 2: slug  
 3: hen-of-the-woods  
 4: stinkhorn  
 5: coral fungus



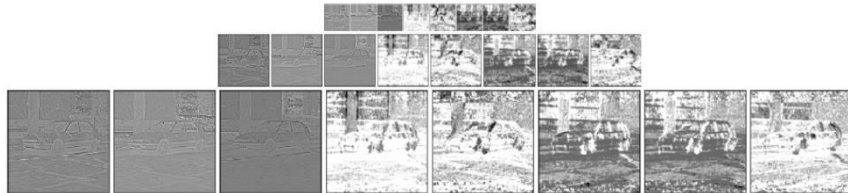
[He et al. 2015]

# Semantische Segmentierung

- Klassen-Annotation pro Bildpunkt

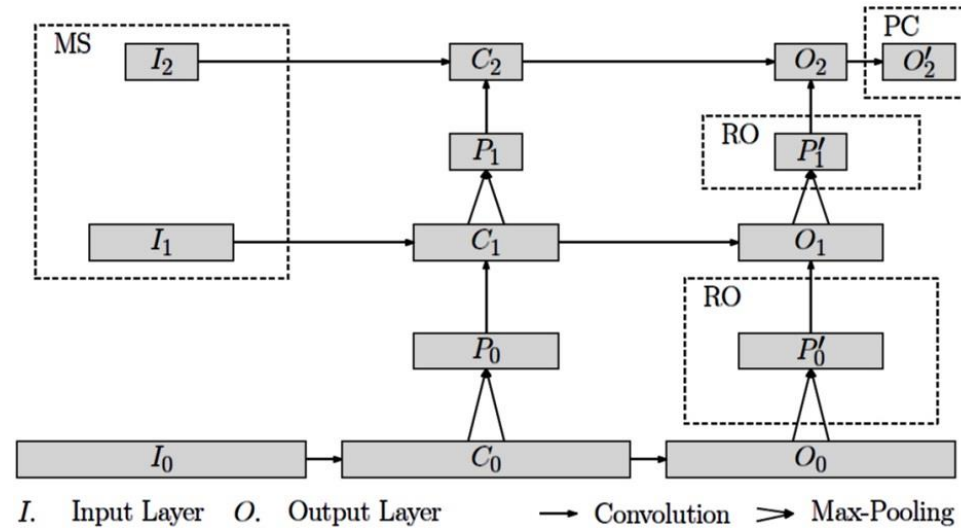


- Multiskalen-Eingabekanäle

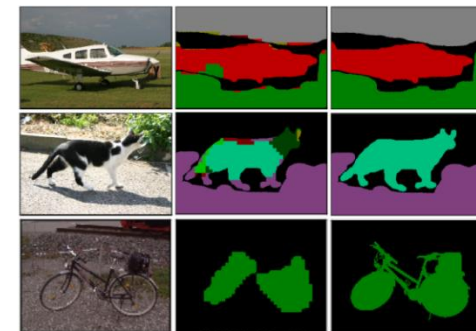


- Evaluiert auf MSRC-9/21 und INRIA Graz-02 Daten

[Schulz, Behnke 2012]



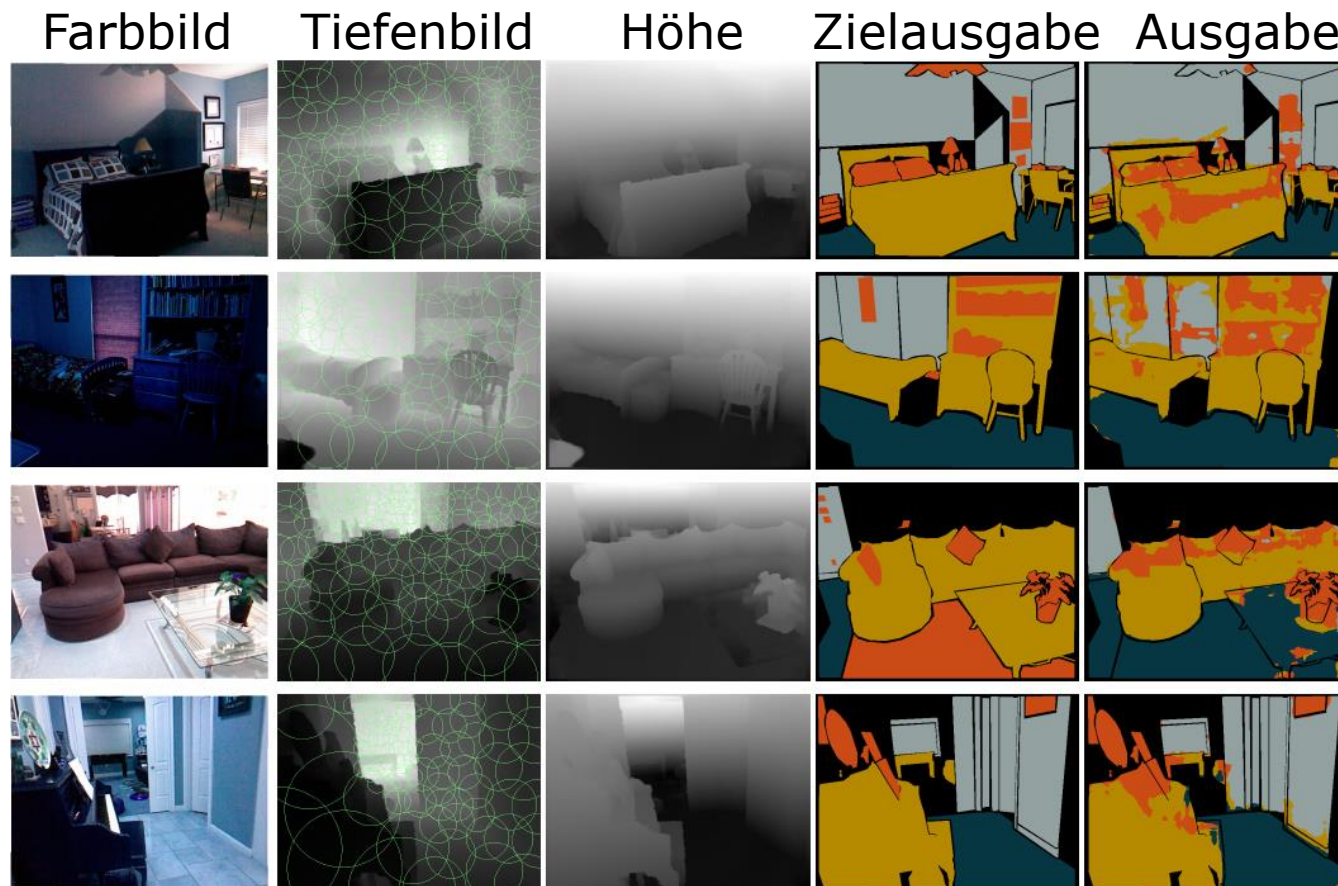
Eingabe    Ausgabe    Ziel



Eingabe    Ausgabe    Ziel

# Kategorisierung von Oberflächen

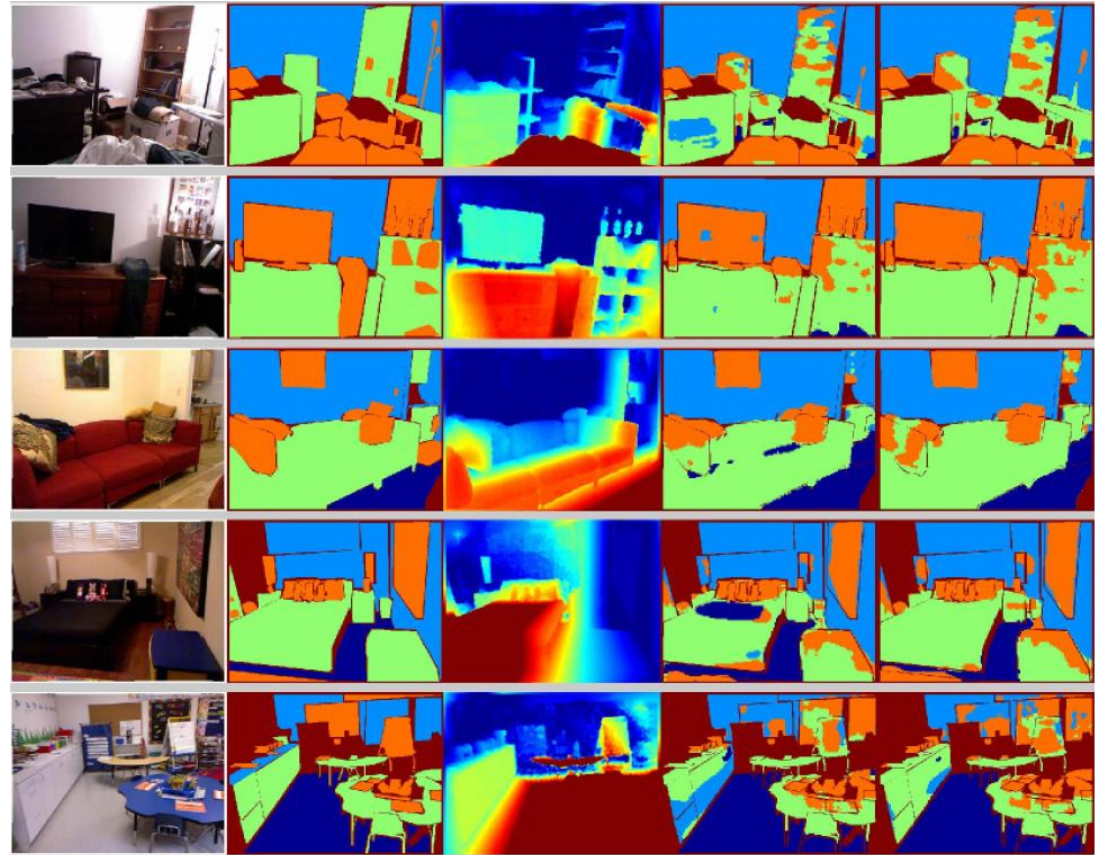
- RGB-D-Eingaben, Höhe über Boden geschätzt
- Skalierung der Eingabe mit der Tiefe



# Geometrische und Semantische Merkmale für RGB-D Objektklassensegmentierung

- Neues **geometrisches** Merkmal: Wandabstand
- **Semantische** Merkmale vortrainiert aus ImageNet
- Beide helfen signifikant

[Husain et al. RA-L 2016]



RGB

Truth

DistWall

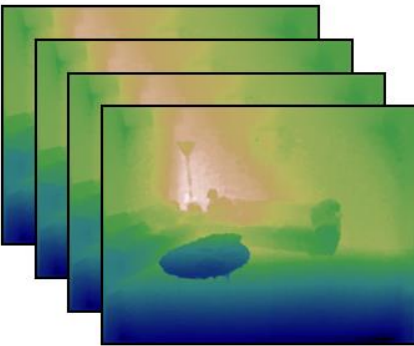
OutWO

OutWithDist

# Objektklassensegmentierung von RGB-D-Video

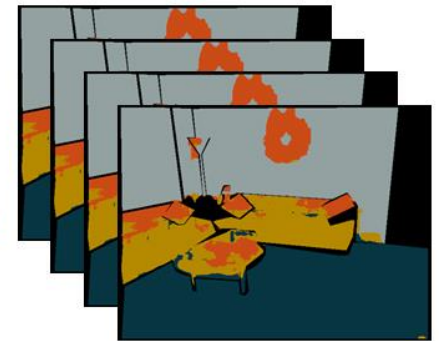
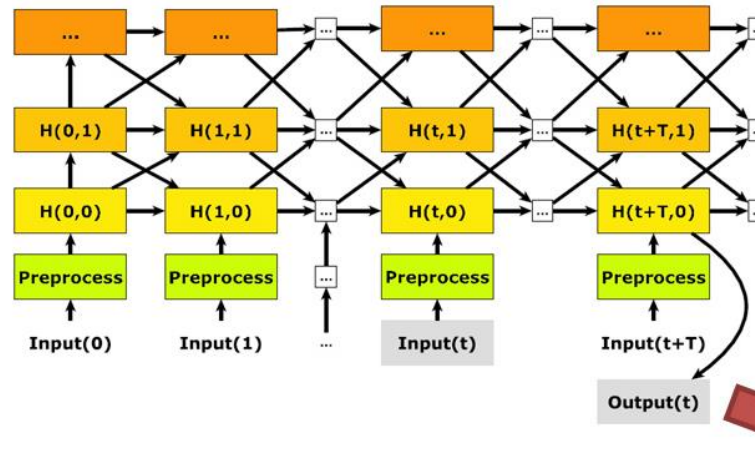
- Eingabe: RGB-D-Video (NYU Depth V2)

RGB



Tiefe

## Neuronale Abstraktionspyramide



Ausgabe

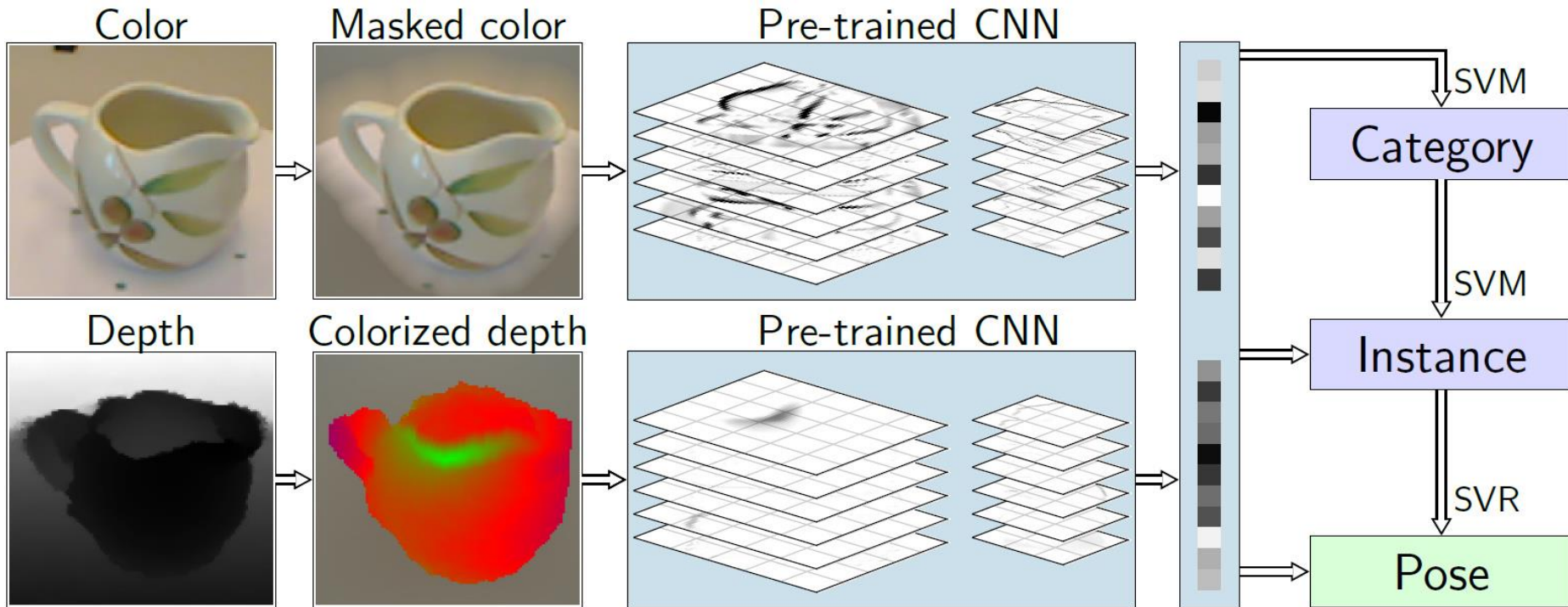
- Rekursive Berechnung ermöglicht effiziente zeitliche Integration

[Pavel, Schulz, Behnke, Neural Networks 2017]



# RGB-D Objekterkennung und Posenschätzung

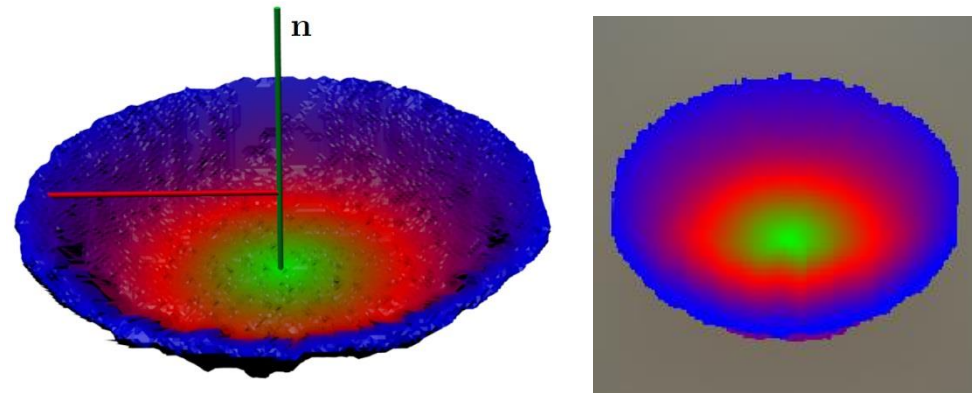
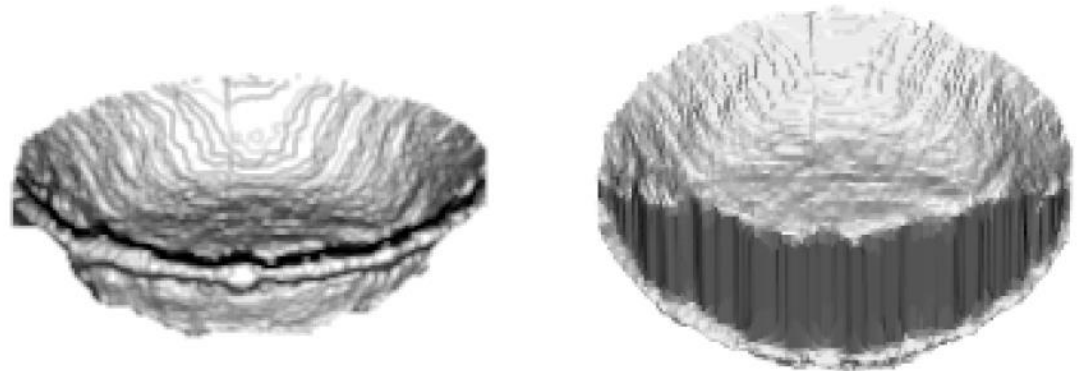
- Vortrainierte Merkmale von ImageNet



[Schwarz, Schulz, Behnke, ICRA2015]

# Kanonische Ansicht, Einfärbung

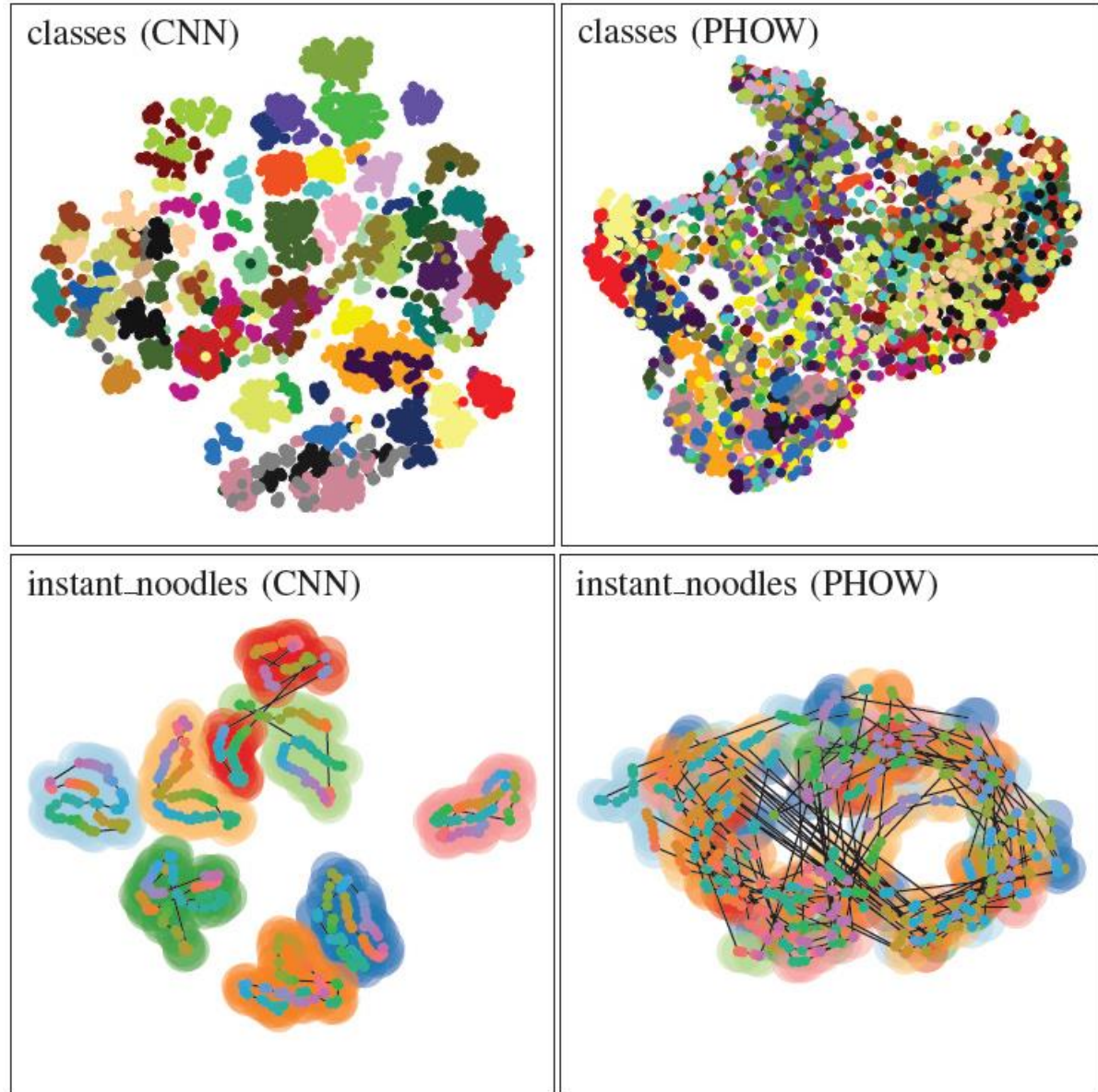
- Objektansichten aus verschiedenen Höhen
- Rendern einer kanonischen Ansicht
- Einfärbung anhand Distanz von Mittelachse



[Schwarz, Schulz, Behnke, ICRA2015]

# Merkmale Entwirren die Daten

- t-SNE Einbettung



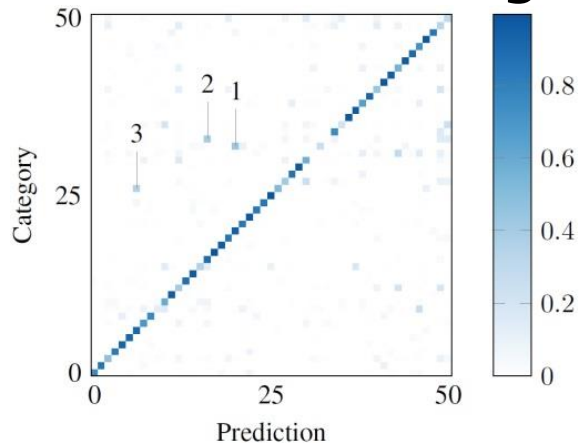
[Schwarz, Schulz,  
Behnke, ICRA2015]

# Erkennungsrate

## ■ Verbesserte Kategorie- und Instanzerkennung

Method	Category Accuracy (%)		Instance Accuracy (%)	
	RGB	RGB-D	RGB	RGB-D
Lai <i>et al.</i> [1]	74.3 ± 3.3	81.9 ± 2.8	59.3	73.9
Bo <i>et al.</i> [2]	82.4 ± 3.1	87.5 ± 2.9	<b>92.1</b>	92.8
PHOW[3]	80.2 ± 1.8	—	62.8	—
<b>Ours</b>	<b>83.1 ± 2.0</b>	88.3 ± 1.5	92.0	<b>94.1</b>
<b>Ours</b>	<b>83.1 ± 2.0</b>	<b>89.4 ± 1.3</b>	92.0	<b>94.1</b>

## ■ Verwechslungen



1: Kännchen / Kaffetasse



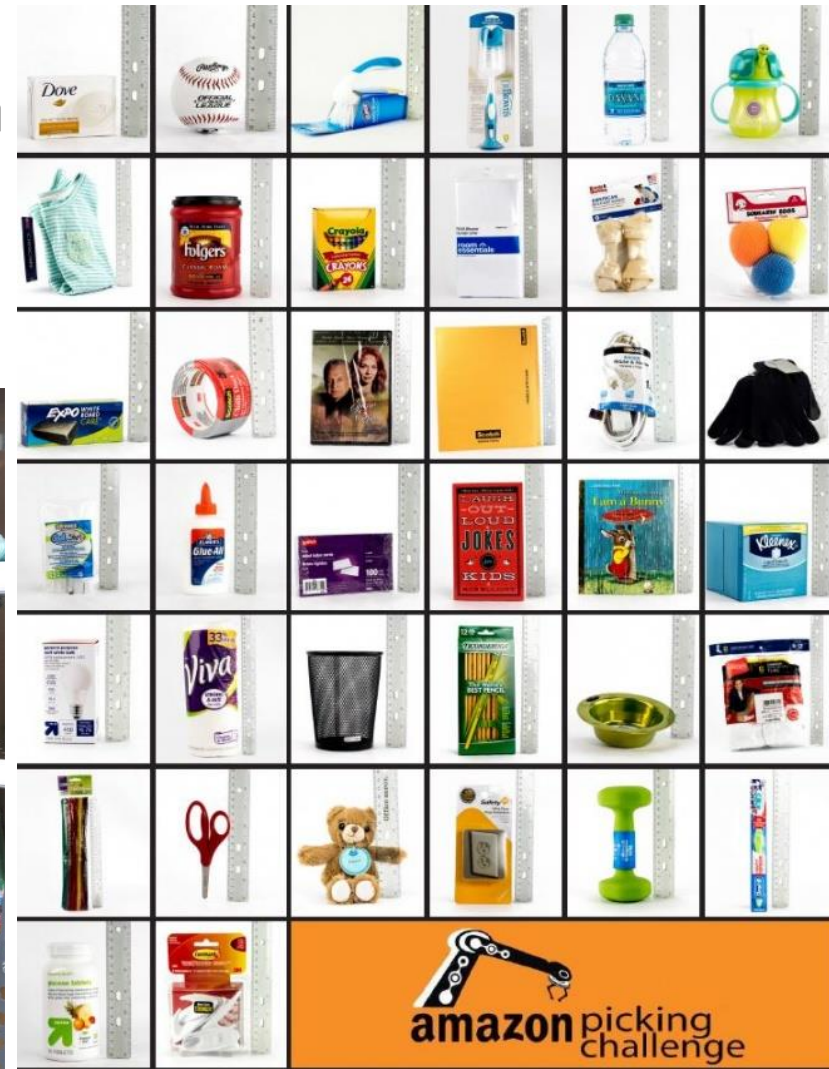
2: Pfirsich / Schwamm



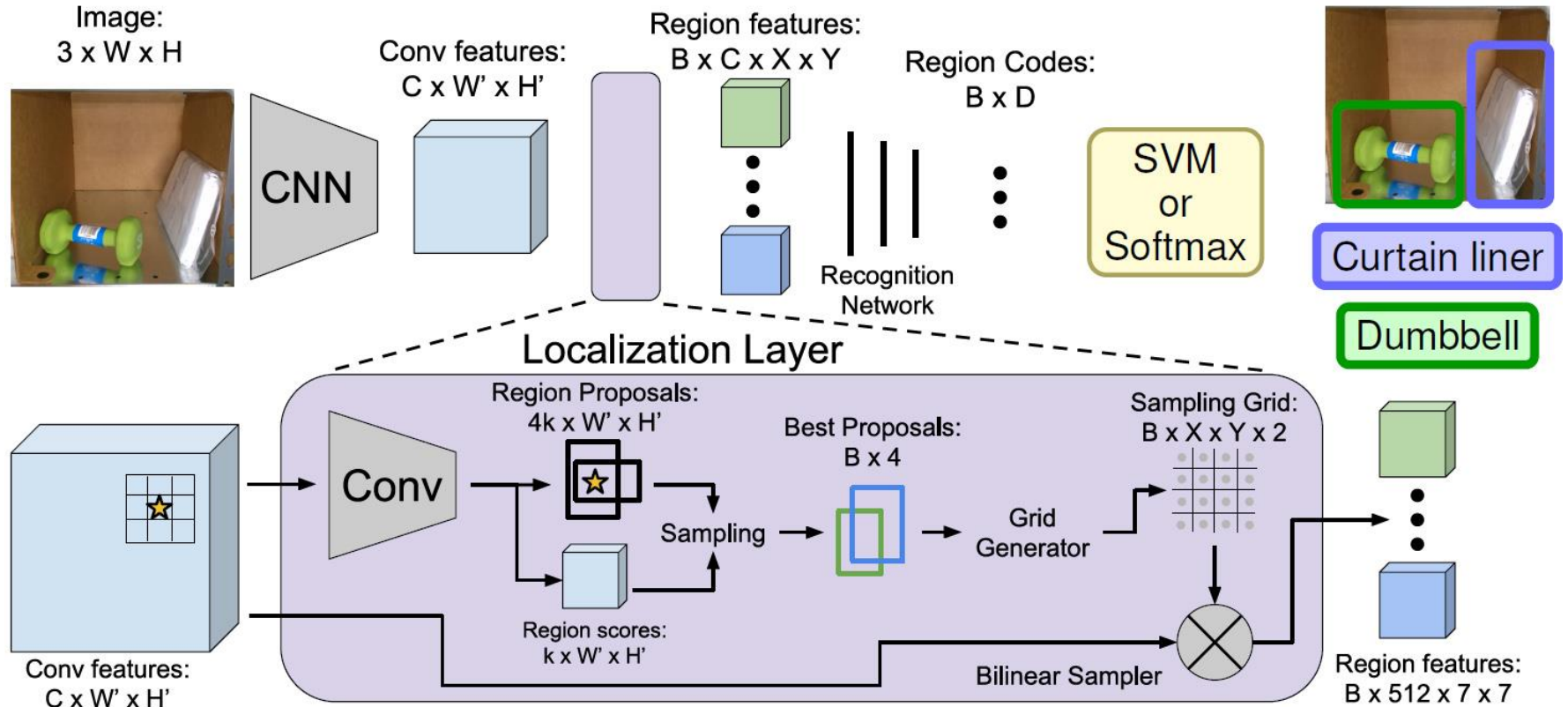
[Schwarz, Schulz, Behnke, ICRA2015]

# Amazon Picking Challenge 2016

- Große Objektvielfalt
- Problematische Eigenschaften
  - Transparent
  - Glänzend
  - Deformierbar
  - Schwer
- Einlagern
- Auslagern



# Objectdetektion



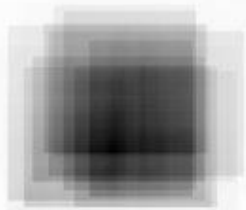
[Adapted from Johnson et al. CVPR 2016]

[Schwarz et al. ICRA 2017]

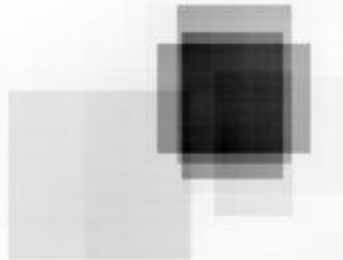
# Beispiel-Detektionen



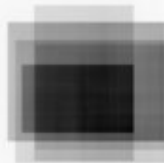
Gloves



Glue sticks



Sippy cup

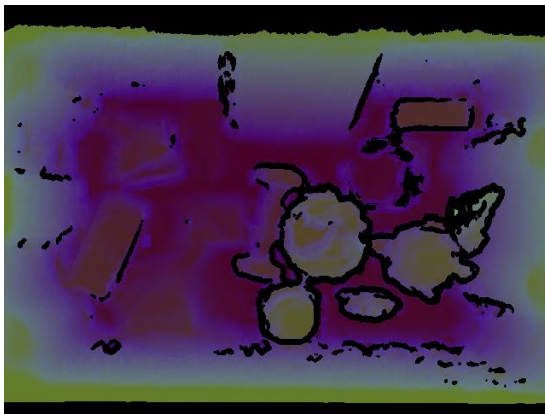


[Schwarz et al. ICRA 2017]

# Semantische Segmentierung

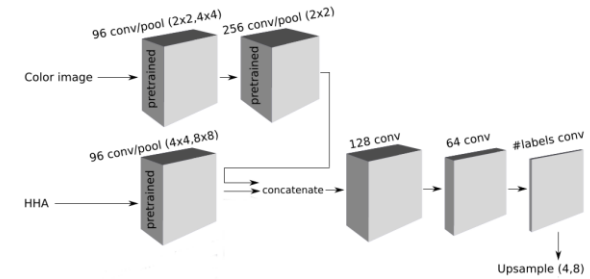
## ■ Tiefes Konvolutionales Netzwerk

RGB



HHA

Ausgabe



[Husain et al. RA-L 2016]

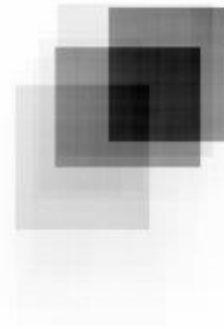


# Kombination von Detektion und Segmentierung

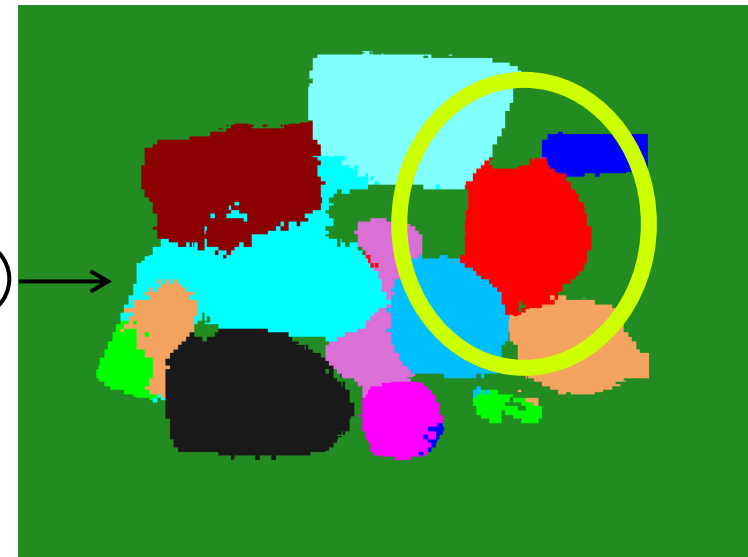
- Pixelweise Multiplikation



Detektion



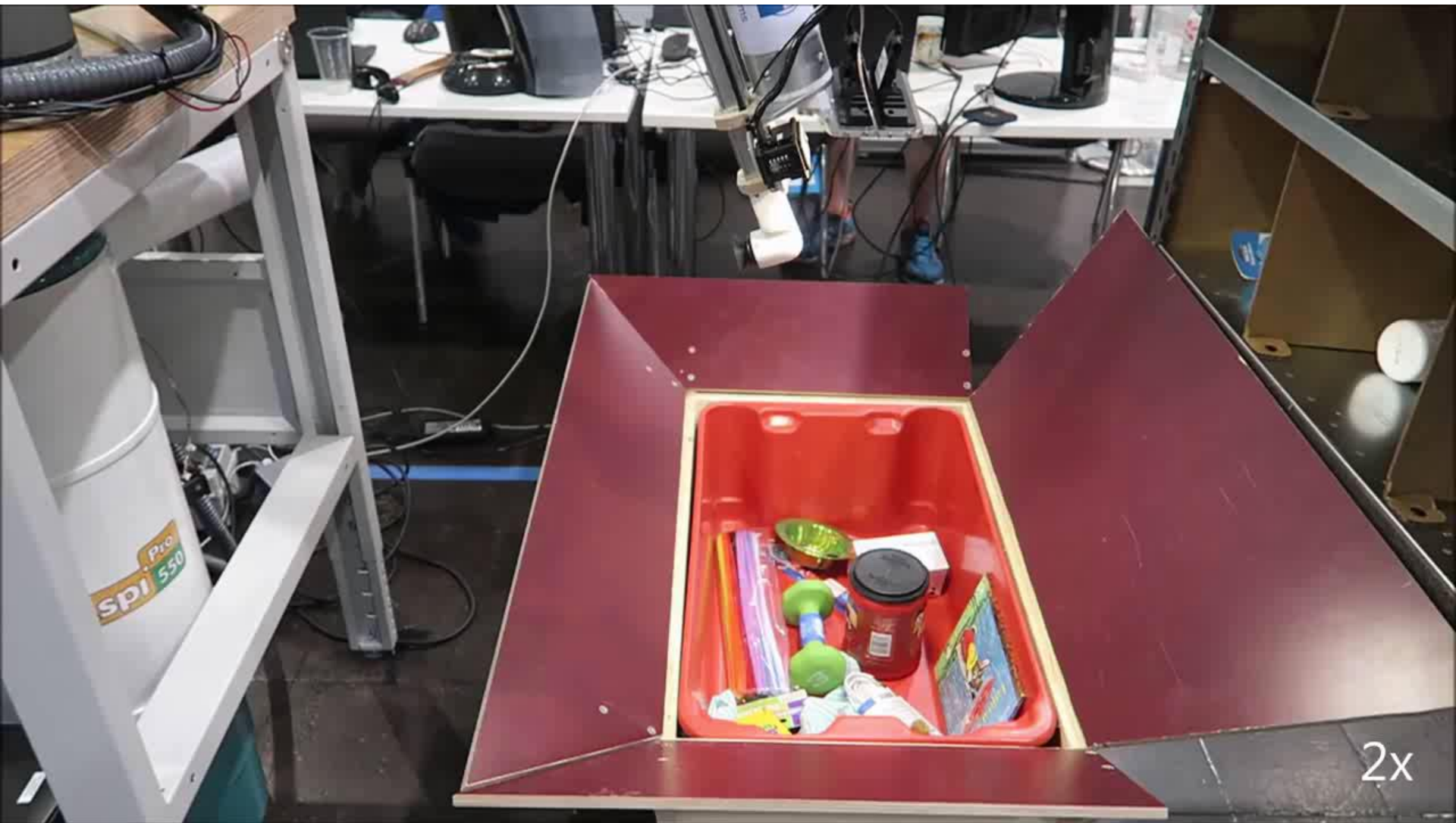
Ausgabe



Segmentierung



# Beispiel: Stowing



[Schwarz et al. ICRA 2017]

# Beispiel: Picking



[Schwarz et al. ICRA 2017]

# Detektion von Werkzeugen

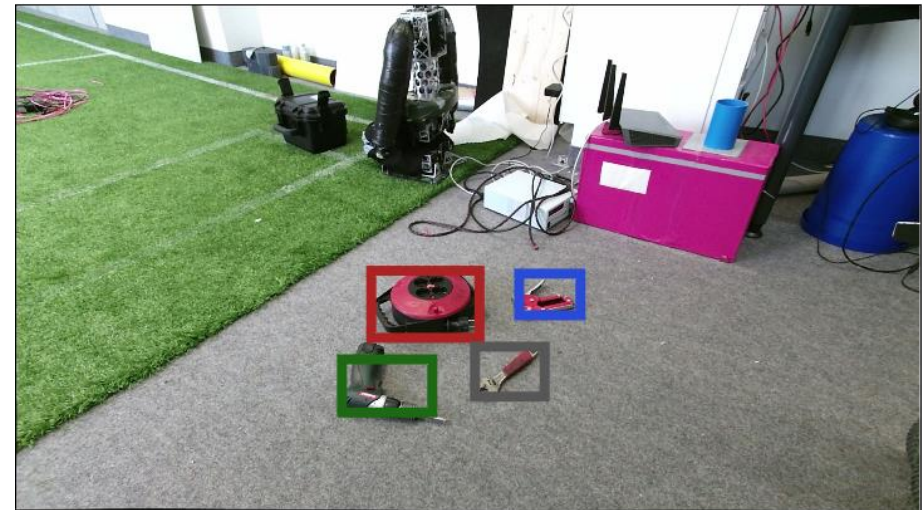
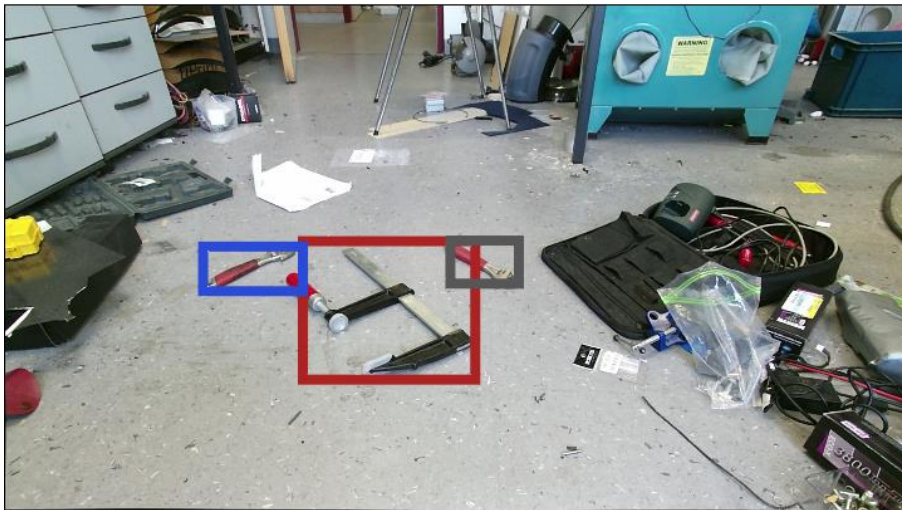


extension\_box stapler driller clamp [background]

Resolution	Clamp	Door handle	Driller	Extension	Stapler	Wrench	Mean
	AP / F1	AP / F1	AP / F1	AP / F1	AP / F1	AP / F1	AP / F1
720×507	0.881/0.783	0.522/0.554	0.986/0.875	1.000/0.938	0.960/0.814	0.656/0.661	0.834/0.771
1080×760	0.926/0.829	0.867/0.632	0.972/0.893	1.000/0.950	0.992/0.892	0.927/0.848	0.947/0.841
1470×1035	0.913/0.814	0.974/0.745	1.000/0.915	1.000/0.952	0.999/0.909	0.949/0.860	0.973/0.866

[Schwarz et al. IJRR 2017]

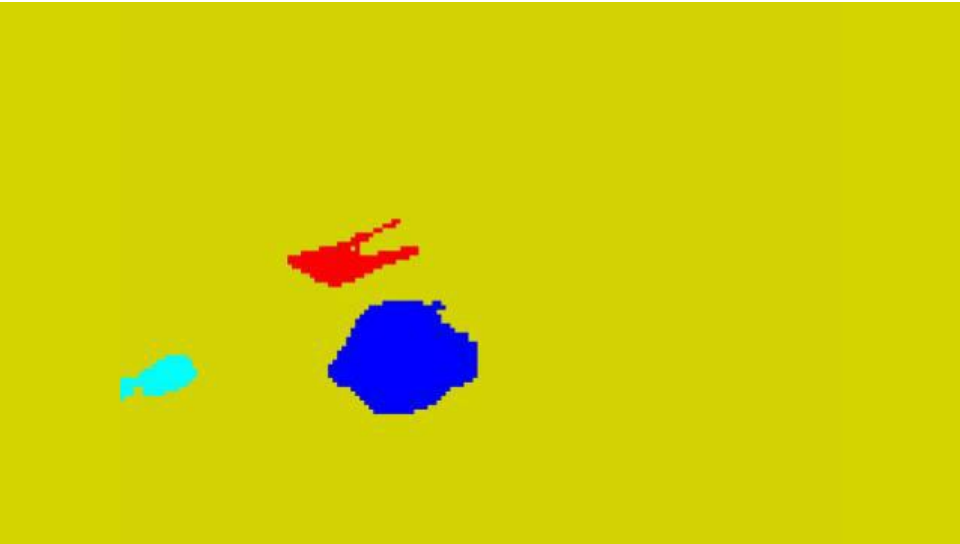
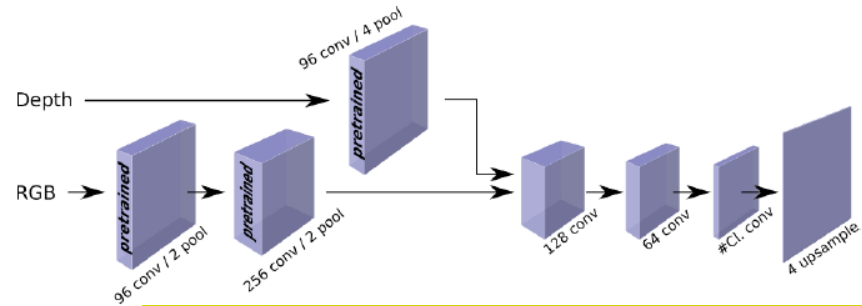
# Detektion von Werkzeugen



# Semantische Segmentierung

## ■ Tiefes CNN

[Husain et al. RA-L 2016]



Pixel Accuracy:

Clamp	Door handle	Driller	Extension	Stapler	Wrench	Background	Mean
0.727	0.751	0.769	0.889	0.775	0.734	0.992	0.805

# MBZIRC Challenge 2



2x

# Auswahl des Schraubenschlüssels

- DenseCap-Objektdetektion
- Trainingsmenge: 100 Stereo-Bildpaare



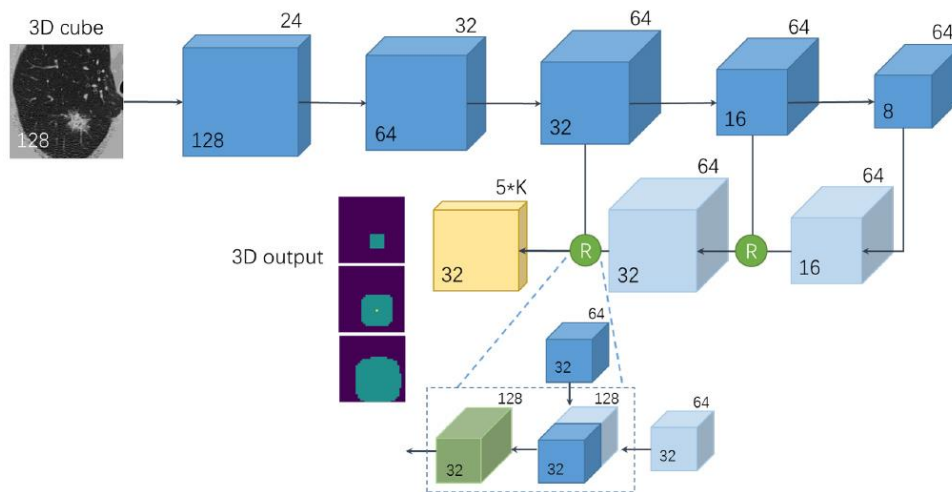


# Erkennung von Lungenkrebs

- Kaggle Data Science Bowl 2017, 1 Millionen US\$ Preisgeld
- Gegeben: CT-Scan der Lunge
- Gesucht: Wahrscheinlichkeit von Diagnose im nächsten Jahr
- Gewinner-Ansatz:

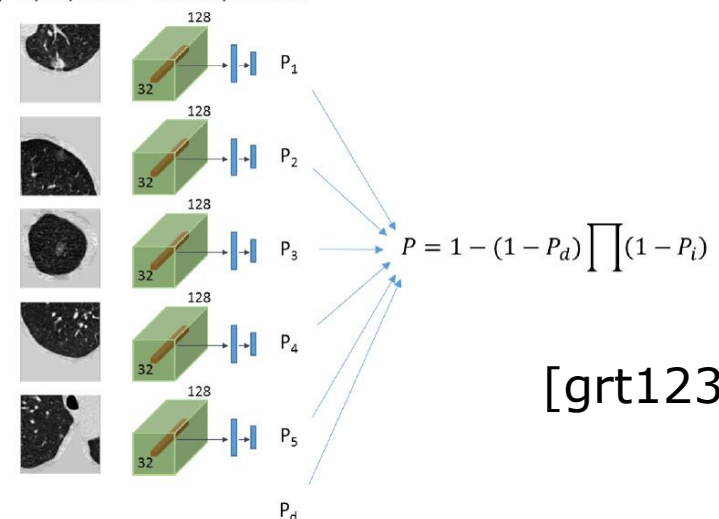


## 3D-Nodule-Detektion



## Wahrscheinlichkeitsberechnung

Top 5 proposals + dummy nodule



[grt123]

# Erkennung von Hautkrebs

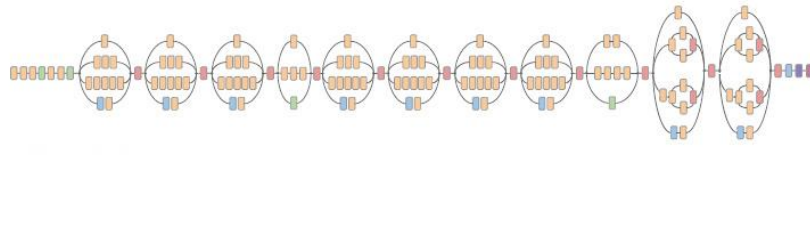
## ■ CNN-Training auf 129.450 Bildern

Skin lesion image

Deep convolutional neural network (Inception v3)

Training classes (757)

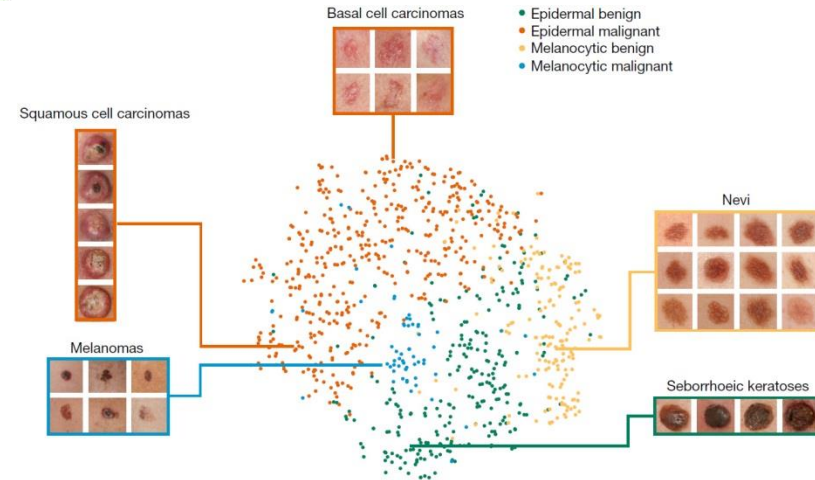
Inference classes (varies by task)



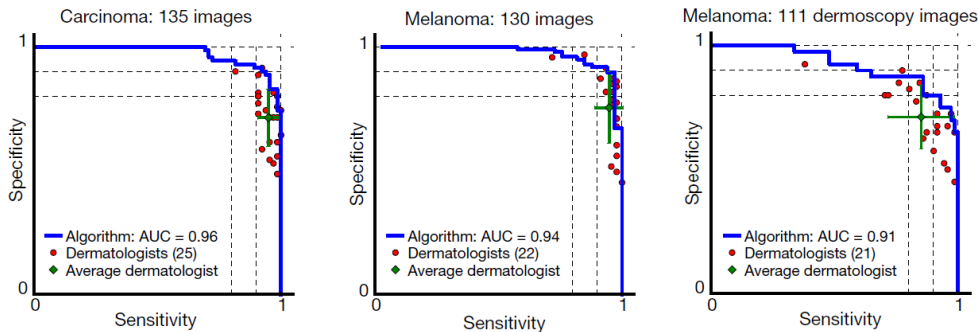
- Acral-lentiginous melanoma
- Amelanotic melanoma
- Lentigo melanoma
- ...
- Blue nevus
- Halo nevus
- Mongolian spot
- ...

- ⊕ ● 92% malignant melanocytic lesion
- ⊖ ● 8% benign melanocytic lesion

## ■ Gelernte Merkmale



## ■ Vergleich zu Dermatologen

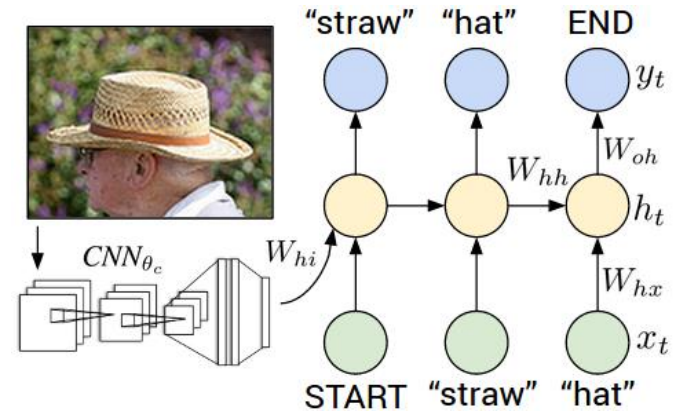


[Esteva et al. Nature 2017]

# Generieren von Bildbeschreibungen

- Rekurrentes neuronales Netz

[Karpathy, Fei-Fei 2015]



man in black shirt is playing guitar.

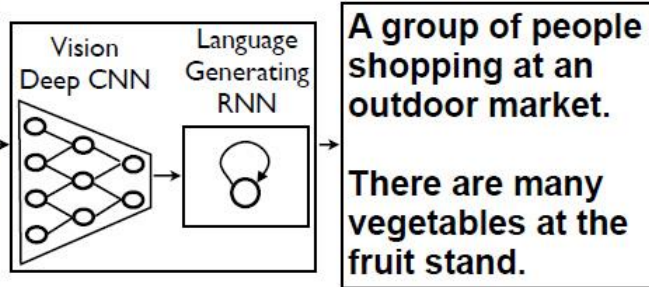


construction worker in orange safety vest is working on road.



two young girls are playing with lego toy.

# Generieren von Bildbeschreibungen



A group of young people playing a game of frisbee.



A herd of elephants walking across a dry grass field.



Two hockey players are fighting over the puck.



A close up of a cat laying on a couch.



A skateboarder does a trick on a ramp.



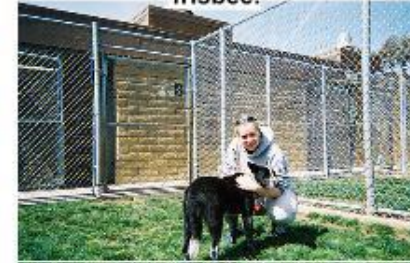
A little girl in a pink hat is blowing bubbles.



A red motorcycle parked on the side of the road.



A dog is jumping to catch a frisbee.



A refrigerator filled with lots of food and drinks.



A yellow school bus parked in a parking lot.



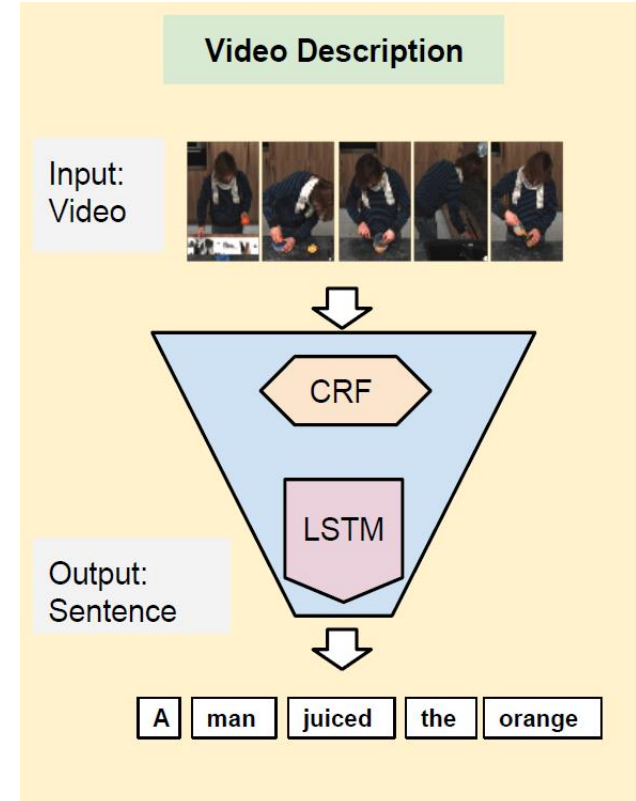
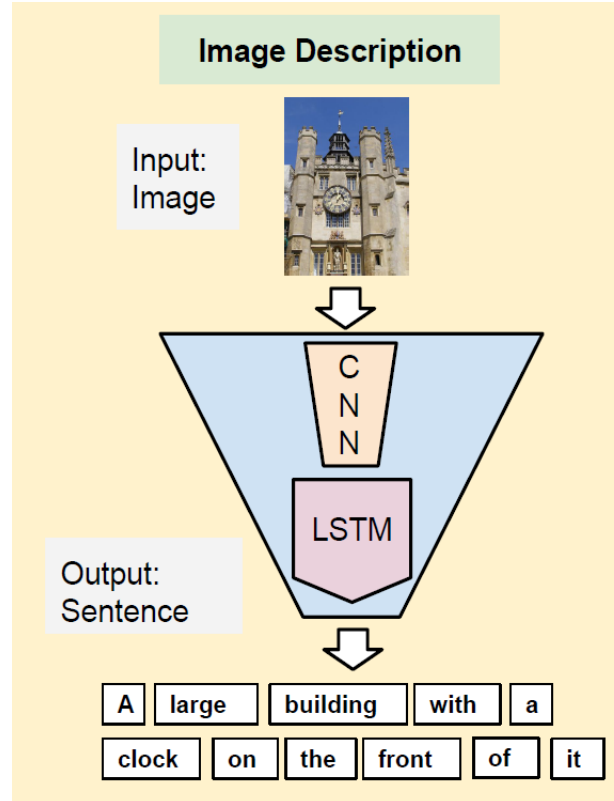
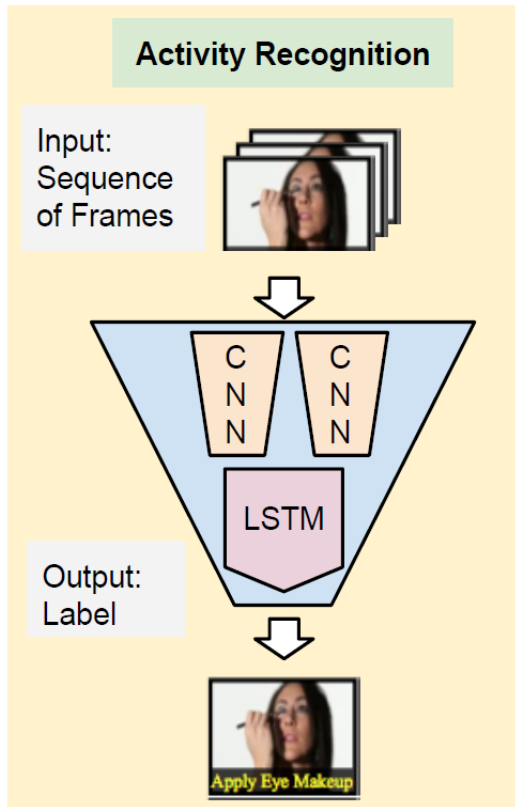
Describes without errors

Describes with minor errors

Somewhat related to the image

Unrelated to the image

# Aufgaben-spezifische Modellvarianten

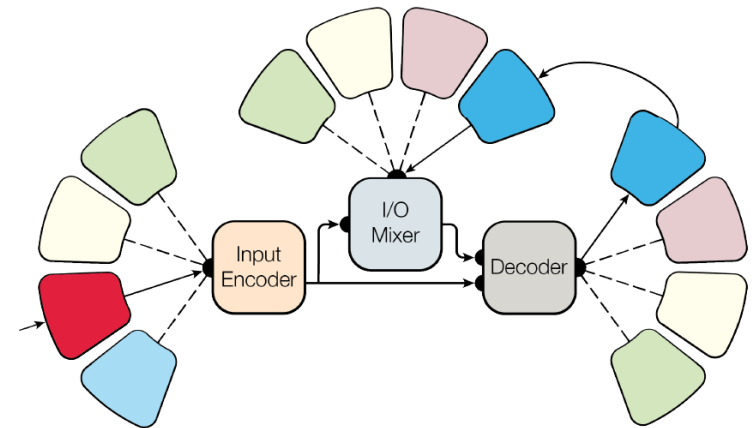


- Gemeinsames Lernen von Merkmalsextraktion und Zeitverhalten verbessert Performanz

[Donahue et al. CVPR 2015]

# Gleichzeitiges Training verschiedener Aufgaben

- Multimodell mit Aufmerksamkeit, Modulen, ...
- Gemeinsames Training verbessert Performanz



To English

“A man that is sitting in front of a suitcase”



To Category

Category 127 (Male Human)

“Last week, Kigali raised the possibility of military retaliation after shells...”

To French

“La semaine dernière, Kigali a soulevé la possibilité de représailles militaires après avoir débarqué des coquilles...”

“Can you give our readers some details on this?”

To German

“Können Sie unseren Lesern einige Details dazu geben?”

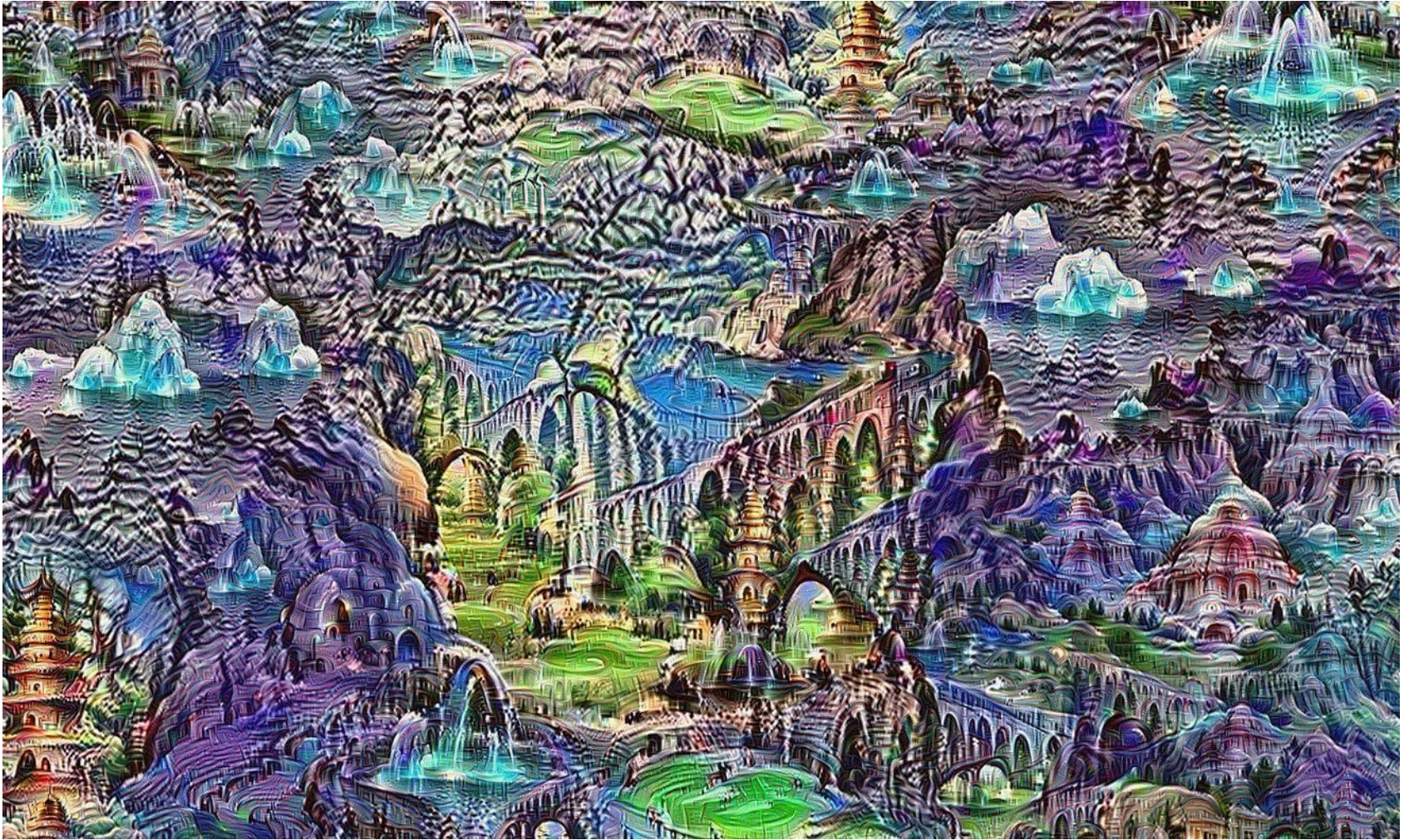
The above represents a triumph of either apathy or civility

To Parse

“S NP DT JJS /NP VP VBZ NP NP DT NN /NP PP IN NP NP NN /NP CC NP NN /NP /NP /PP /NP /NP . /S”

[Kaiser et al. 2017]

# Träumende Tiefe Netzwerke



[Mordvintsev et al 2015]

# Übertragung des Malstils

Original



Turner



van Gogh



Munch

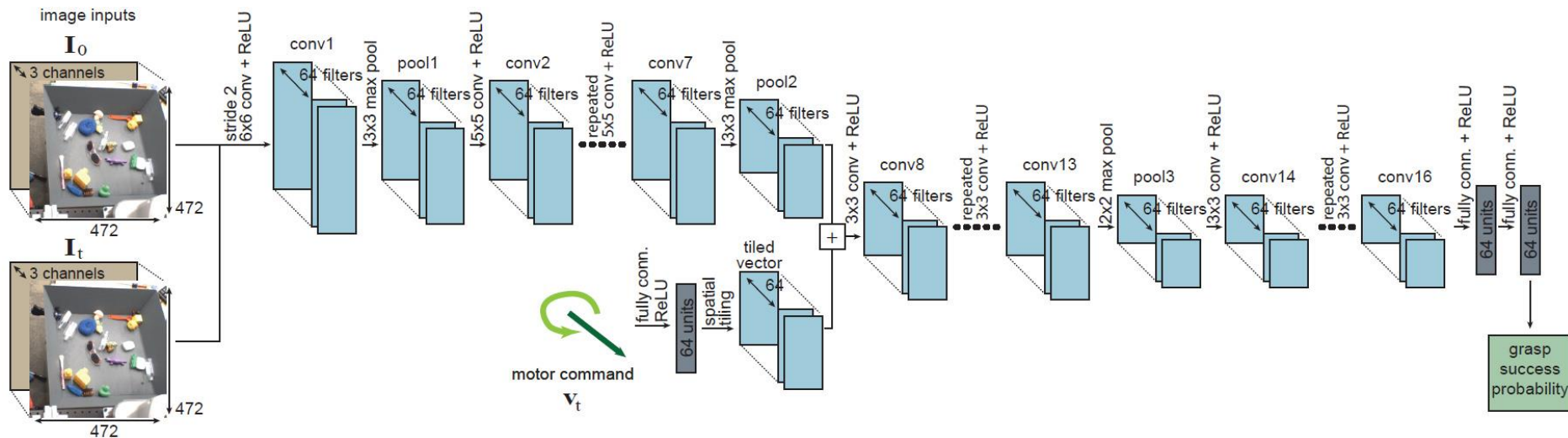


[Gatys et al. 2015]

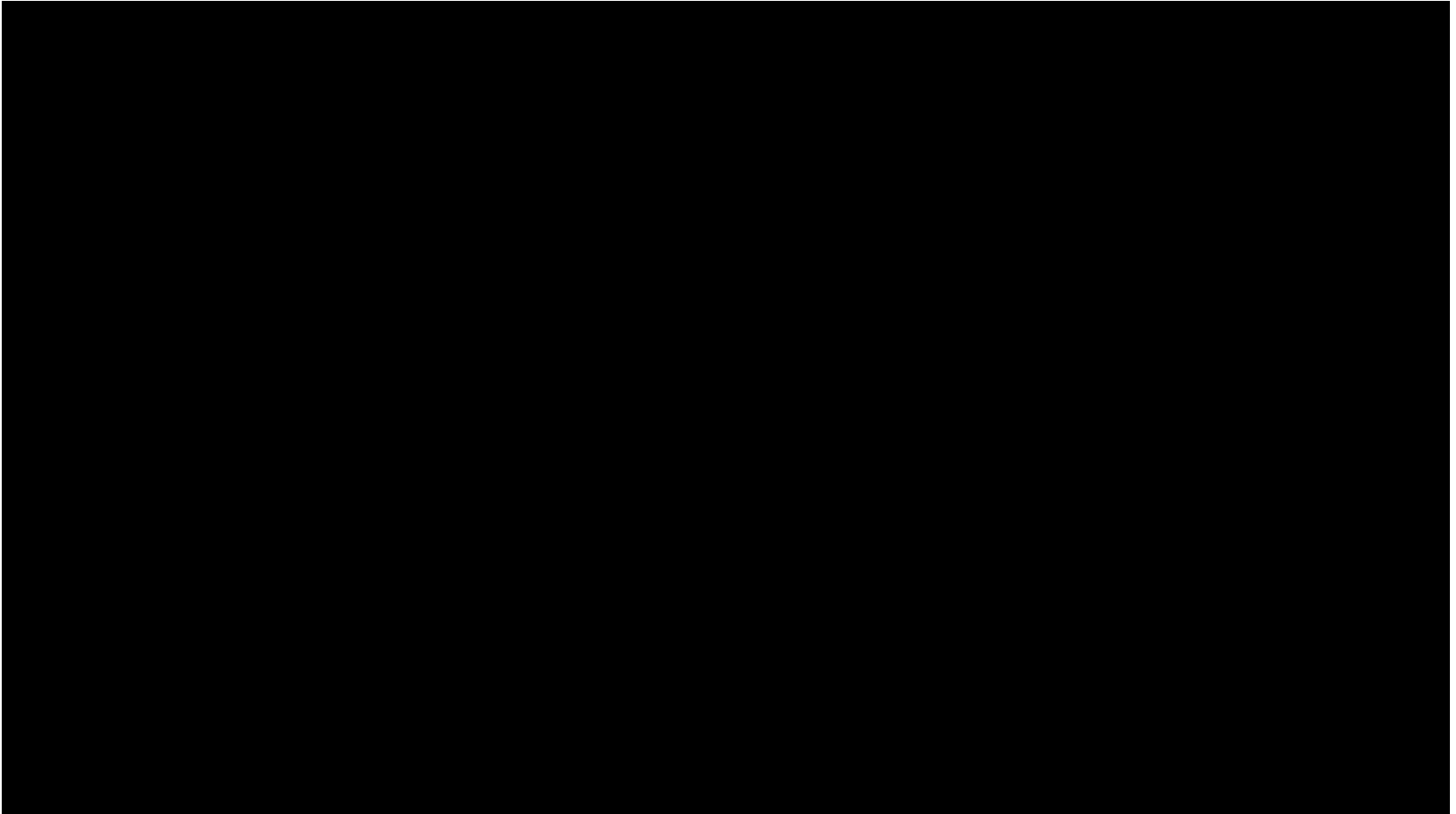


# Google: Lernen der Kamera-Greifer Koordination für Griff in die Kiste

- Zahlreiche Objekte ungeordnet in Kiste
- Unkalibrierte Farbkamera
- Konvolutionales Neuronales Netz zur Vorhersage des Greiferfolgs



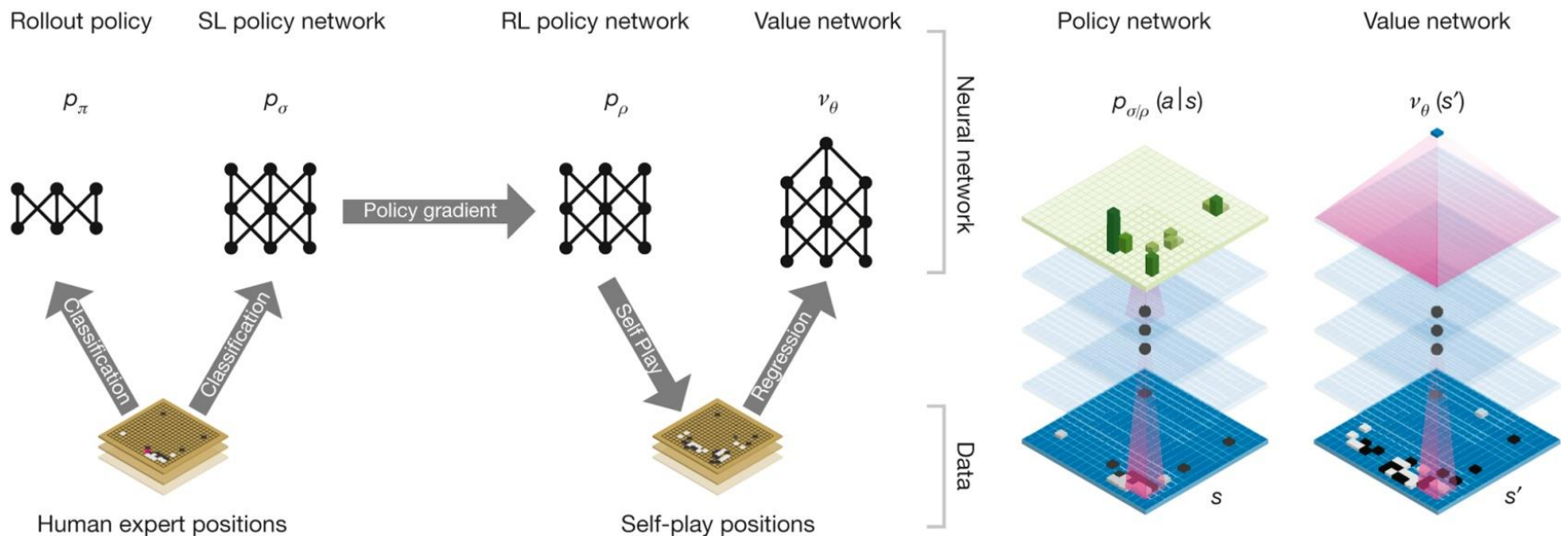
# NVidia: BB8 Selbstfahrendes Auto



- Deep learning - Netzwerk (DAVENET)
- Ende-zu-Ende-Training mit Beispielen von Menschen

# Google DeepMinds: AlphaGo

- Tiefe konvolutionale neuronale Netze, um Nutzen und Aktionen aus Spielsituation vorherzusagen
- Überwachtes Training mit Spielen von Menschen
- Verstärkendes Lernen durch Spiel gegen sich selbst
- Monte Carlo-Simulation mit Bewertung durch Netze



# Zusammenfassung

- Deep Learning hat visuelle Mustererkennung und Spracherkennung revolutioniert
- Aktuelle Forschung mit dem Ziel der Realisierung höherer kognitiver Fähigkeiten (General AI)
- Ende-zu-Ende Training von Gesamtsystemen zur Generierung von Aktionen
- Künstliche Intelligenz wird Auswirkungen auf alle Lebensbereiche haben
- Gesellschaftliche Herausforderungen wie Sicherheit, Haftungsfragen, Auswirkungen auf die Arbeitswelt, Privatsphäre und Datenschutz

**Vielen Dank für Ihre  
Aufmerksamkeit!**