

# Analyse dynamischer Szenen mit LSTM

Patrick Nagorski  
Master Informatik - Grundseminar SS17  
Betreuer: Prof. Dr. Andreas Meisel

02.05.2017

# Gliederung

- ▶ Motivation
- ▶ Grundlagen
- ▶ LSTM-Netze
- ▶ Verwandte Arbeiten
- ▶ Ausblick
- ▶ Konferenzen

# Motivation

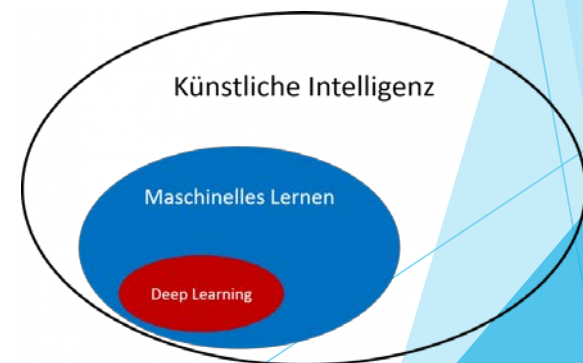
- ▶ Wie bin ich zu diesem Thema gekommen?
  - ▶ WP - Robot Vision
  - ▶ Gespräch mit Betreuer



Quelle: [01]

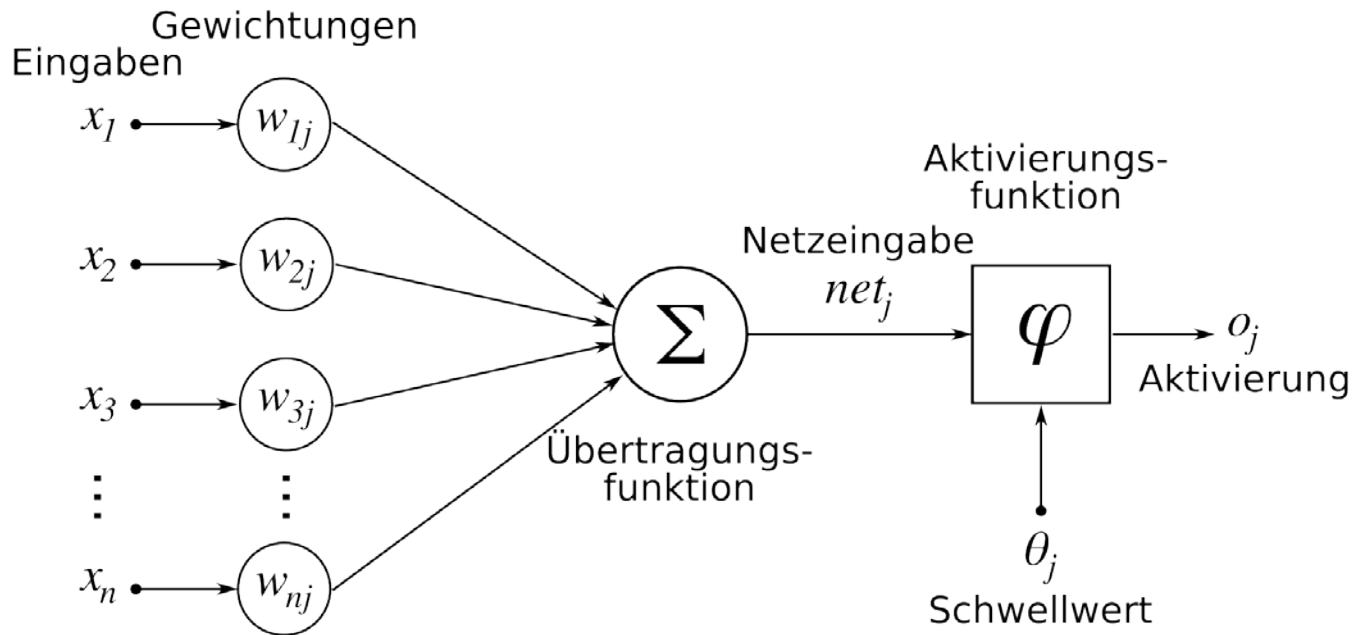
# Grundlagen - Maschinelles Lernen

- ▶ künstliche Generierung von Wissen aus Erfahrung
- ▶ Lernen: aus Bekanntem auf Unbekanntes schließen (neuronales Netz)
- ▶ künstliches System lernt aus (vielen) Beispielen
  - ▶ Ziel nach Beendigung der Lernphase -> Beispiele verallgemeinern



Quelle: [02]

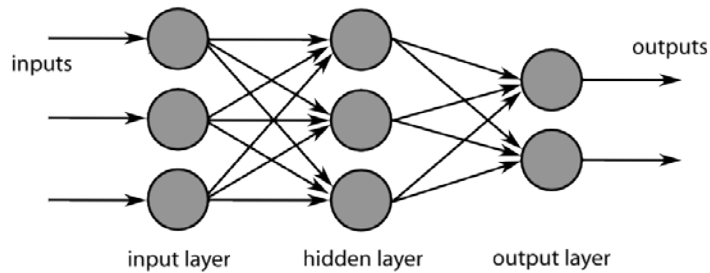
# Grundlagen - Das Neuron



Quelle: [03]

# Grundlagen - Arten neuronaler Netze

Feedforward-Netz:

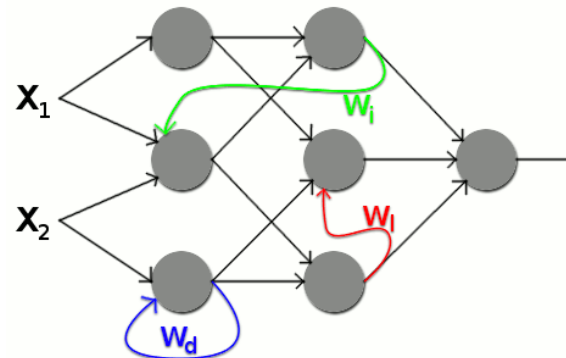


Quelle: [04]

Eigenschaften Feedforward-Netz:

- Kein Gedächtnis
- Ausgangssignale abhängig nur von Eingangssignalen

Rekurrentes neuronales Netz:



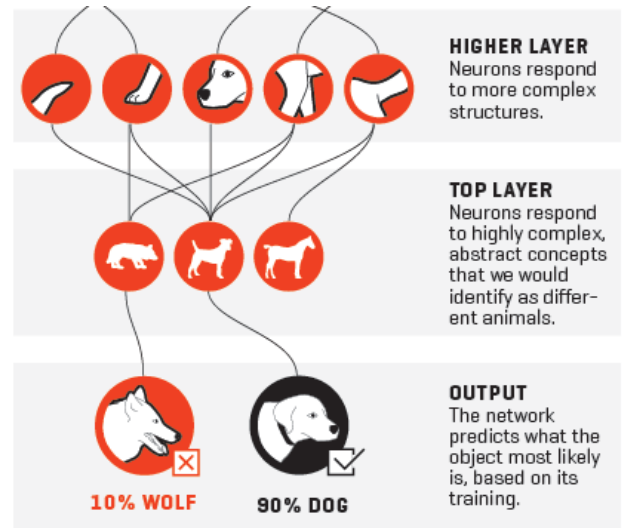
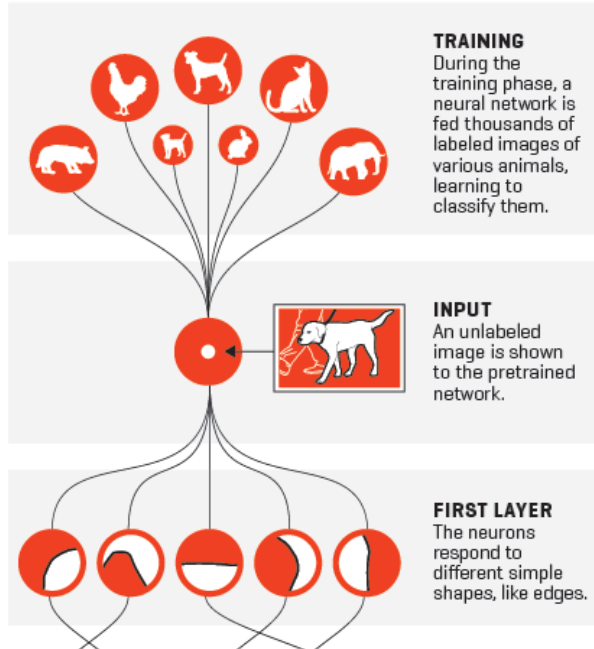
Quelle: [05]

Eigenschaften rekurrentes neuronales Netz:

- Rückkopplung
- Gedächtnis
- Ausgangssignale abhängig von Eingangssignalen und von der zeitlichen Vorgeschichte
- Nur kurze Zeiträume -> LSTM!

# Grundlagen - Beispiel eines neuronalen Netzes

## HOW NEURAL NETWORKS RECOGNIZE A DOG IN A PHOTO



Quelle: [06]

# Grundlagen - Arten des maschinellen Lernens

- ▶ Überwachtes Lernen (engl. supervised learning)
  - ▶ Algorithmus lernt eine Funktion aus gegebenen Paaren von Ein- und Ausgaben
  - ▶ während des Lernens „Lehrer“, der den korrekten Funktionswert zu einer Eingabe bereithält
- ▶ Ablauf:
  - ▶ Anlegen der Eingabe
  - ▶ Verarbeitung der Eingabe
  - ▶ Vergleich der Ausgabe mit dem erwünschten Wert (Fehler)
  - ▶ Verkleinern des Fehlers durch Modifikation der Gewichte
- ▶ Ergebnis nach Training: Unbekannte ähnliche Eingabe -> korrekte Ausgabe



# LSTM-Netze (Long short-term memory)

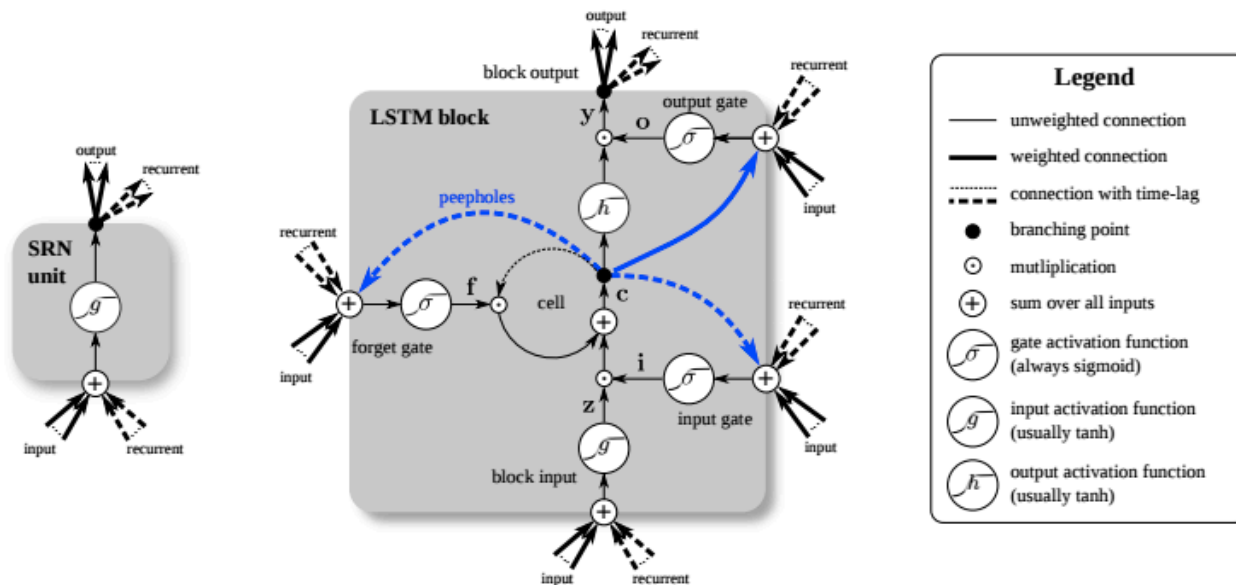


Figure 1. Detailed schematic of the Simple Recurrent Network (SRN) unit (left) and a Long Short-Term Memory block (right) as used in the hidden layers of a recurrent neural network.

Quelle: [07]

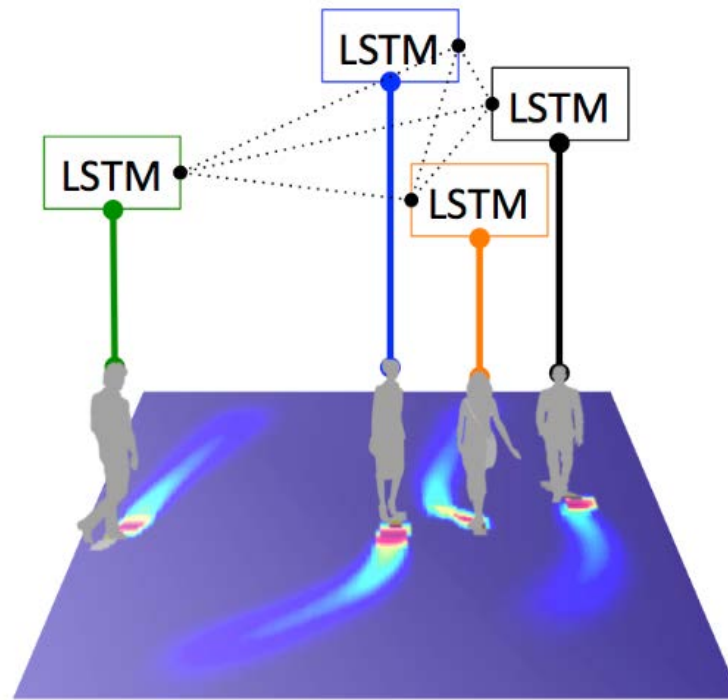
## Eigenschaften LSTM-Netz:

- Sowohl kurze, als auch lange Zeiträume zwischen Ursache und Wirkung
- Löschen nicht mehr benötigter Informationen

# Verwandte Arbeiten

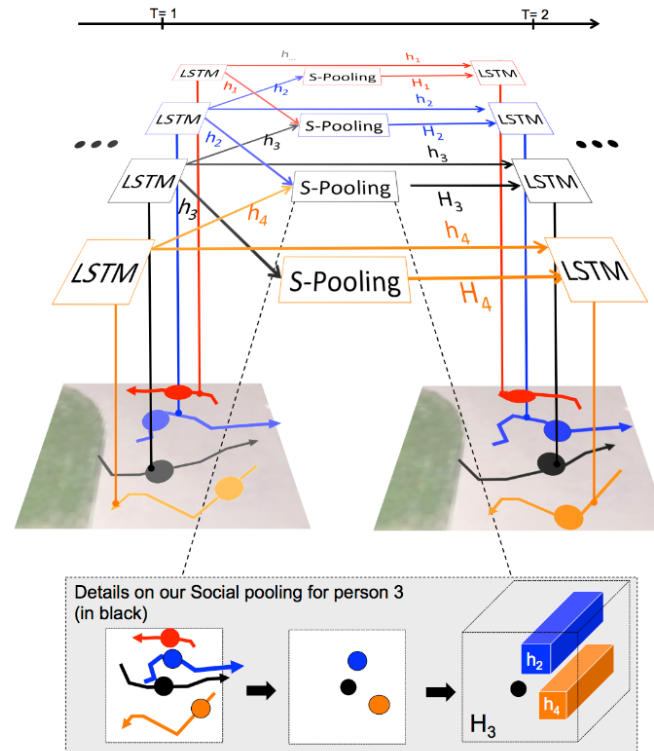
- ▶ Social LSTM: Human Trajectory Prediction in Crowded Spaces (2016)
- ▶ Real-Time Pedestrian Detection With Deep Network Cascades (2015)
- ▶ A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation (2016)

# Social LSTM: Human Trajectory Prediction in Crowded Spaces (2016)



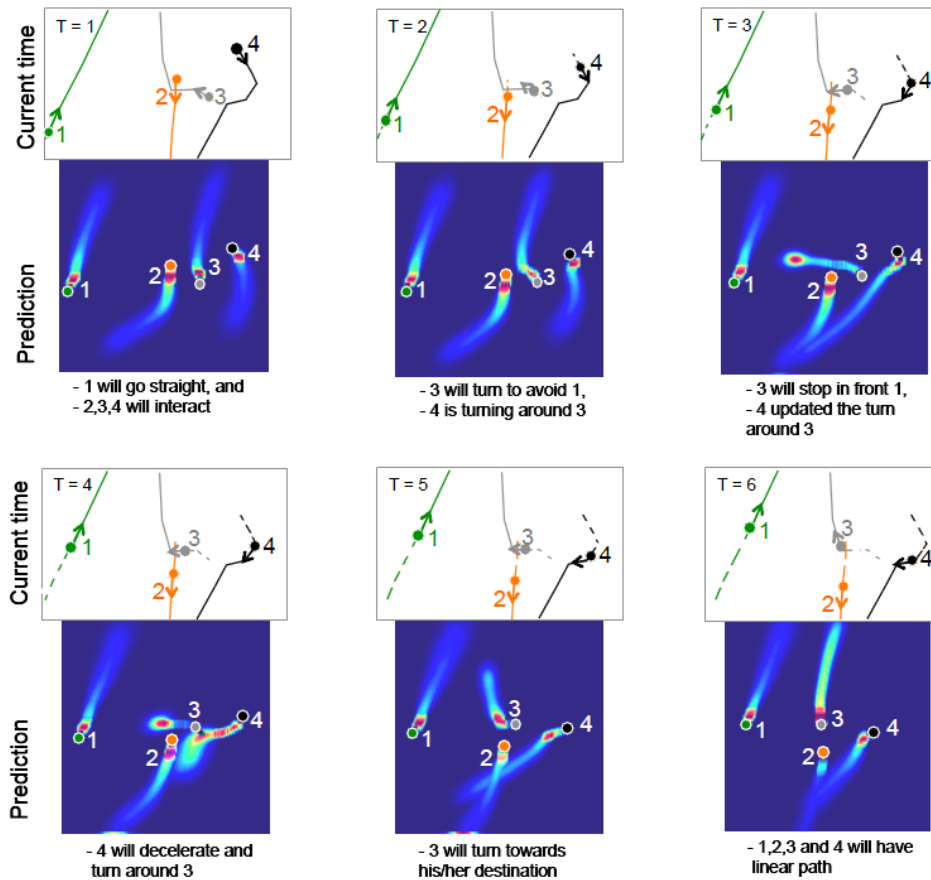
Quelle: [08]

# Social LSTM: Human Trajectory Prediction in Crowded Spaces (2016)



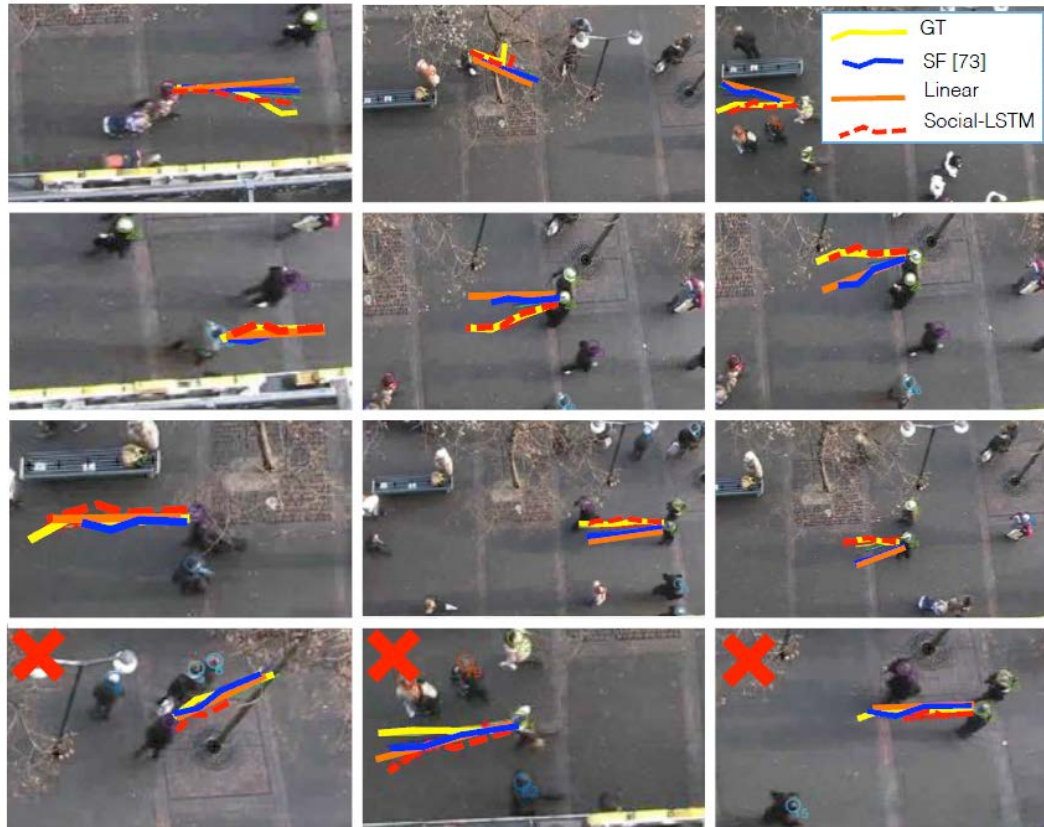
Quelle: [08]

# Social LSTM: Human Trajectory Prediction in Crowded Spaces (2016)



Quelle: [08]

# Social LSTM: Human Trajectory Prediction in Crowded Spaces (2016)



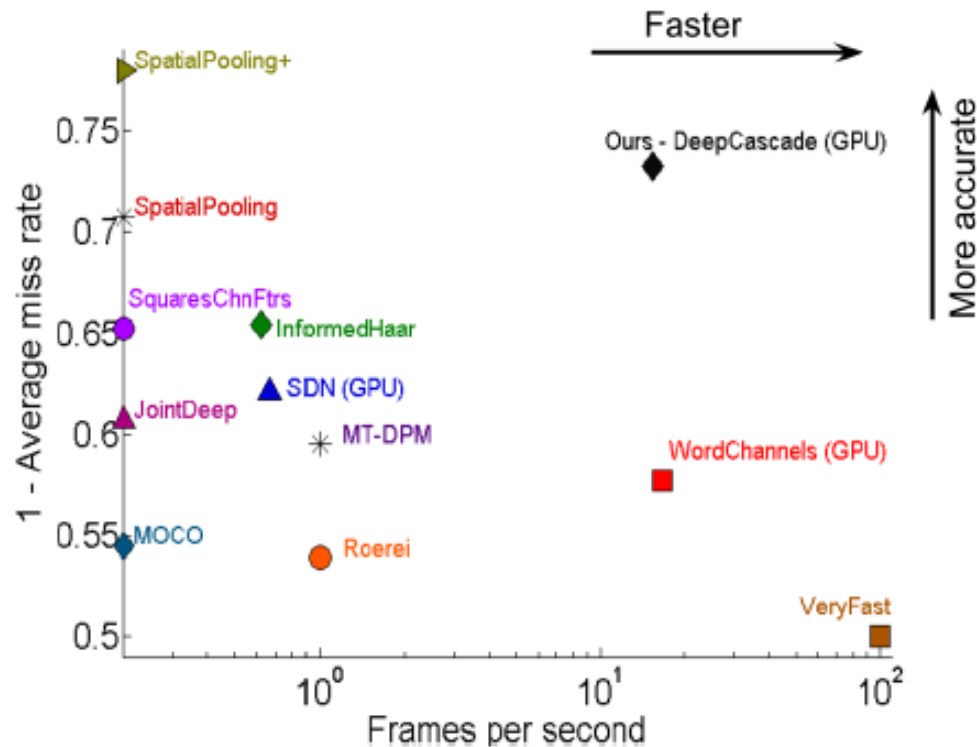
Quelle: [08]

# Social LSTM: Human Trajectory Prediction in Crowded Spaces (2016)

Metric	Methods	Lin	LTA	SF [73]	IGP* [60]	LSTM	our O-LSTM	our Social-LSTM
Avg. disp. error	ETH [49]	0.80	0.54	0.41	<b>0.20</b>	0.60	0.49	0.50
	HOTEL [49]	0.39	0.38	0.25	0.24	0.15	<b>0.09</b>	0.11
	ZARA 1 [39]	0.47	0.37	0.40	0.39	0.43	<b>0.22</b>	<b>0.22</b>
	ZARA 2 [39]	0.45	0.40	0.40	0.41	0.51	0.28	<b>0.25</b>
	UCY [39]	0.57	0.51	0.48	0.61	0.52	0.35	<b>0.27</b>
	Average	0.53	0.44	0.39	0.37	0.44	0.28	<b>0.27</b>
Avg. non-linear disp. error	ETH [49]	0.95	0.70	0.49	0.39	0.28	<b>0.24</b>	0.25
	HOTEL [49]	0.55	0.49	0.38	0.34	0.09	<b>0.06</b>	0.07
	ZARA 1 [39]	0.56	0.39	0.41	0.54	0.24	<b>0.13</b>	<b>0.13</b>
	ZARA 2 [39]	0.44	0.41	0.39	0.43	0.30	0.20	<b>0.16</b>
	UCY [39]	0.62	0.57	0.54	0.62	0.31	0.20	<b>0.16</b>
	Average	0.62	0.51	0.44	0.46	0.24	0.17	<b>0.15</b>
Final disp. error	ETH [49]	1.31	0.77	0.59	<b>0.43</b>	1.31	1.06	1.07
	HOTEL [49]	0.55	0.64	0.37	0.37	0.33	<b>0.20</b>	0.23
	ZARA 1 [39]	0.89	0.66	0.60	0.39	0.93	<b>0.46</b>	0.48
	ZARA 2 [39]	0.91	0.72	0.68	0.42	1.09	0.58	<b>0.50</b>
	UCY [39]	1.14	0.95	0.78	1.82	1.25	0.90	<b>0.77</b>
	Average	0.97	0.74	<b>0.60</b>	0.69	0.98	0.64	0.61

Quelle: [08]

# Real-Time Pedestrian Detection With Deep Network Cascades (2015)



Quelle: [09]



# Real-Time Pedestrian Detection With Deep Network Cascades (2015)

Dataset	Avg. miss rate	FPS
Roerei [5]	46.13	1
MOCO[7]	45.5	
WordChannels [9]	42.3	16
JointDeep[29]	39.3	
MT-DPM+Context [40]	37.64	
ACF+SDt [34] (w. motion)	37.3	
SDN [25]	37.8	0.67
LFOV [3]	35.8	3.6
InformedHaar [42]	34.6	0.63
Hosang et al. [22]	32.4	
SpatialPooling [31]	29.0	0.13
Hosang et al. (w. extra data) [22]	23.3	
SpatialPooling+/Katamari (w. motion): [6, 31]	22.0	
DeepCascade	31.11	15
DeepCascadeID on indep. data	30.17	15
DeepCascadeED w. extra data	26.21	15

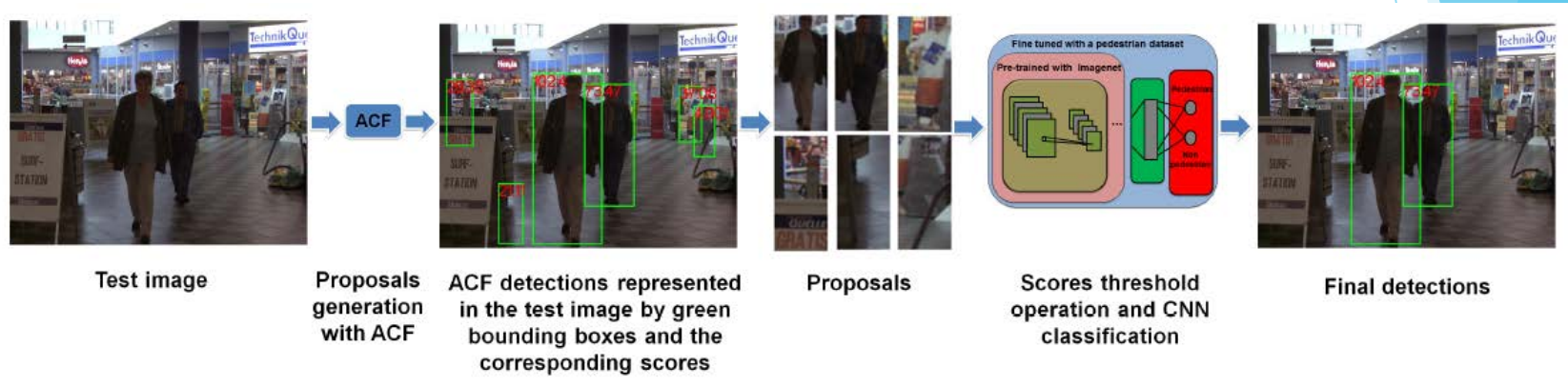
Quelle: [09]

# A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation (2016)



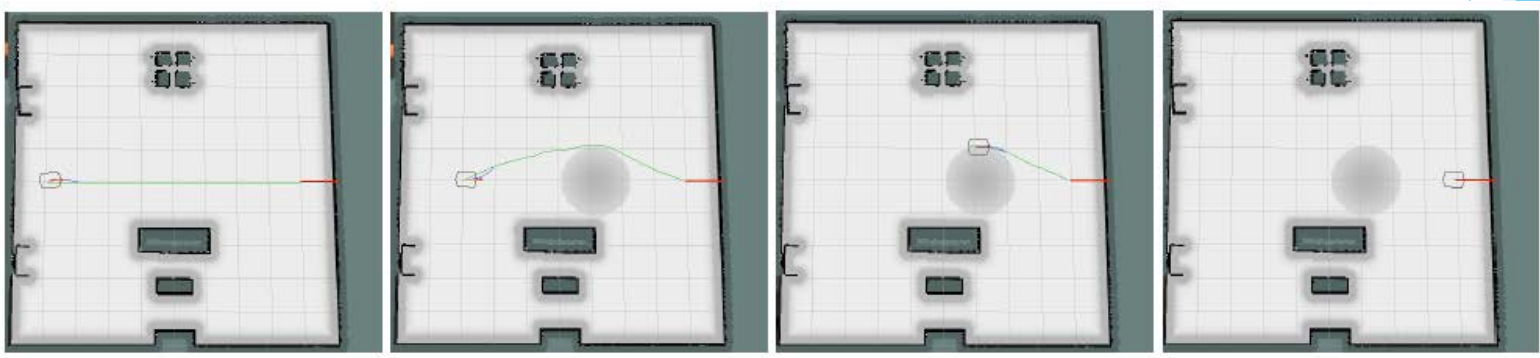
Quelle: [10]

# A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation (2016)



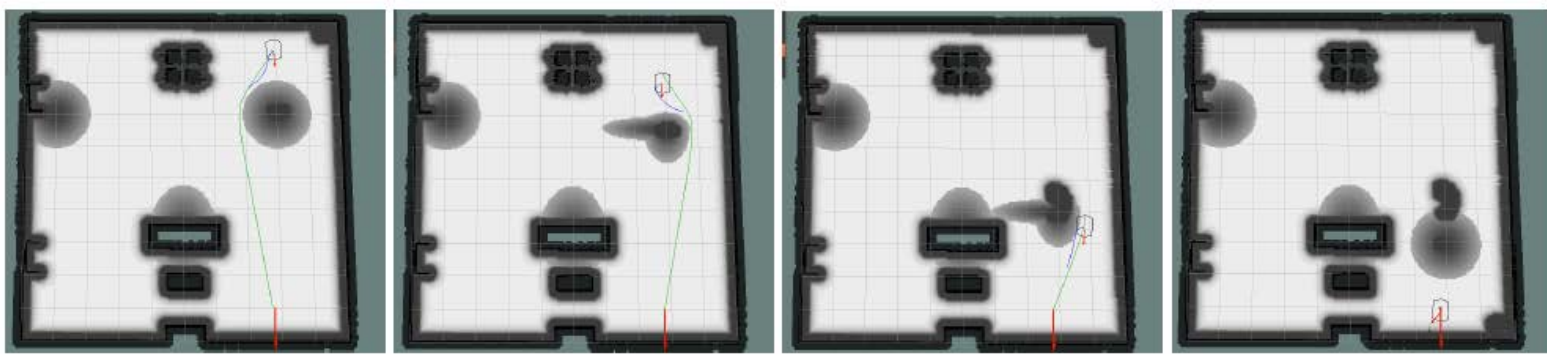
Quelle: [10]

# A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation (2016)



Quelle: [10]

# A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation (2016)



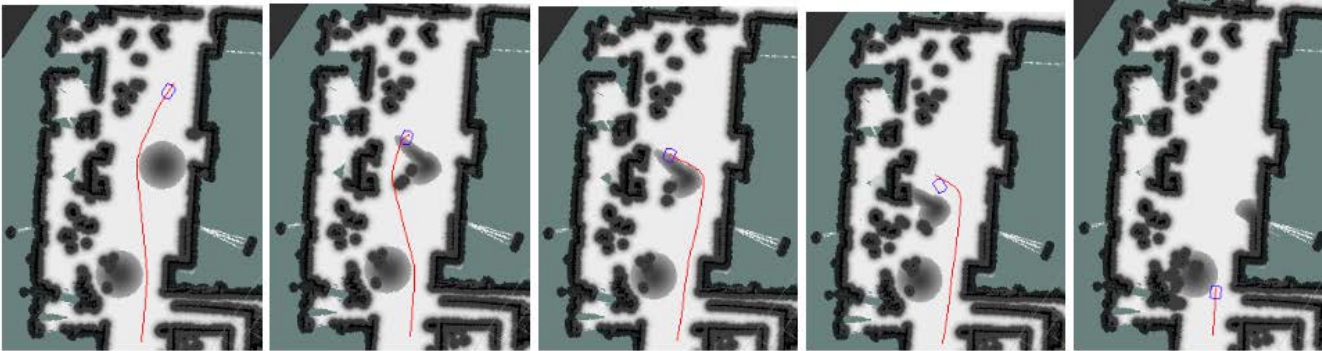
Quelle: [10]

# A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation (2016)



Quelle: [10]

# A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation (2016)



Quelle: [10]

# Probleme aus verwandten Arbeiten

- ▶ Trefferquote verbessern
  - ▶ Zusätzliche Sensorinformationen
  - ▶ Multi LSTM-Netze -> Abstimmung
- ▶ Genauigkeit vs. Performance (Echtzeit)
- ▶ Kamera im Auto
  - ▶ Hardware eingeschränkt
  - ▶ Geschwindigkeit des Autos
- ▶ Extrahieren zusätzlicher Informationen

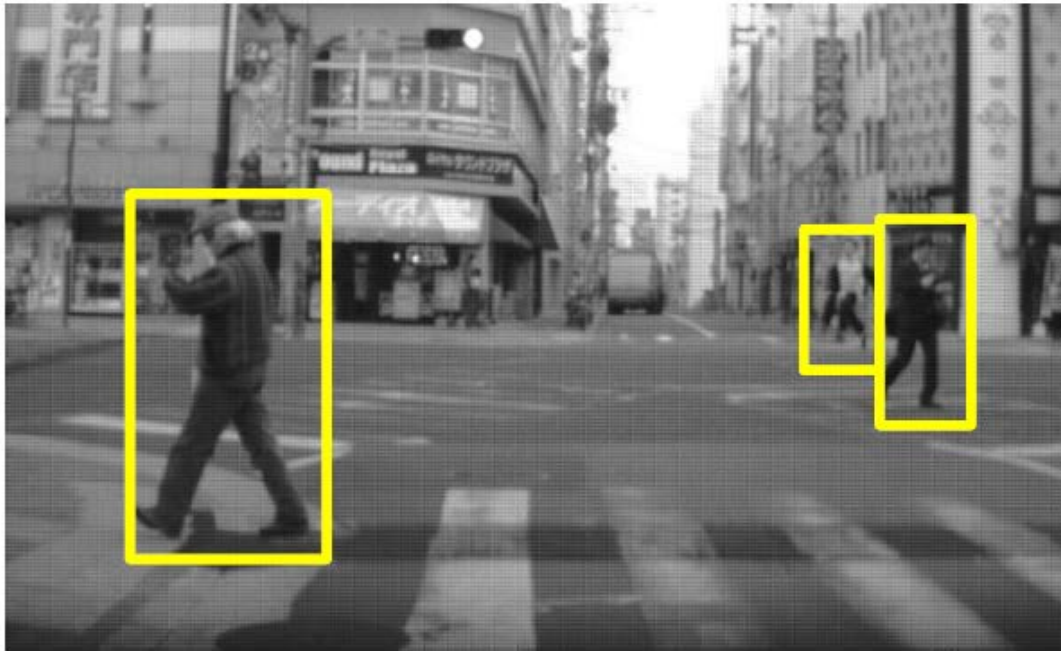


# Probleme aus verwandten Arbeiten



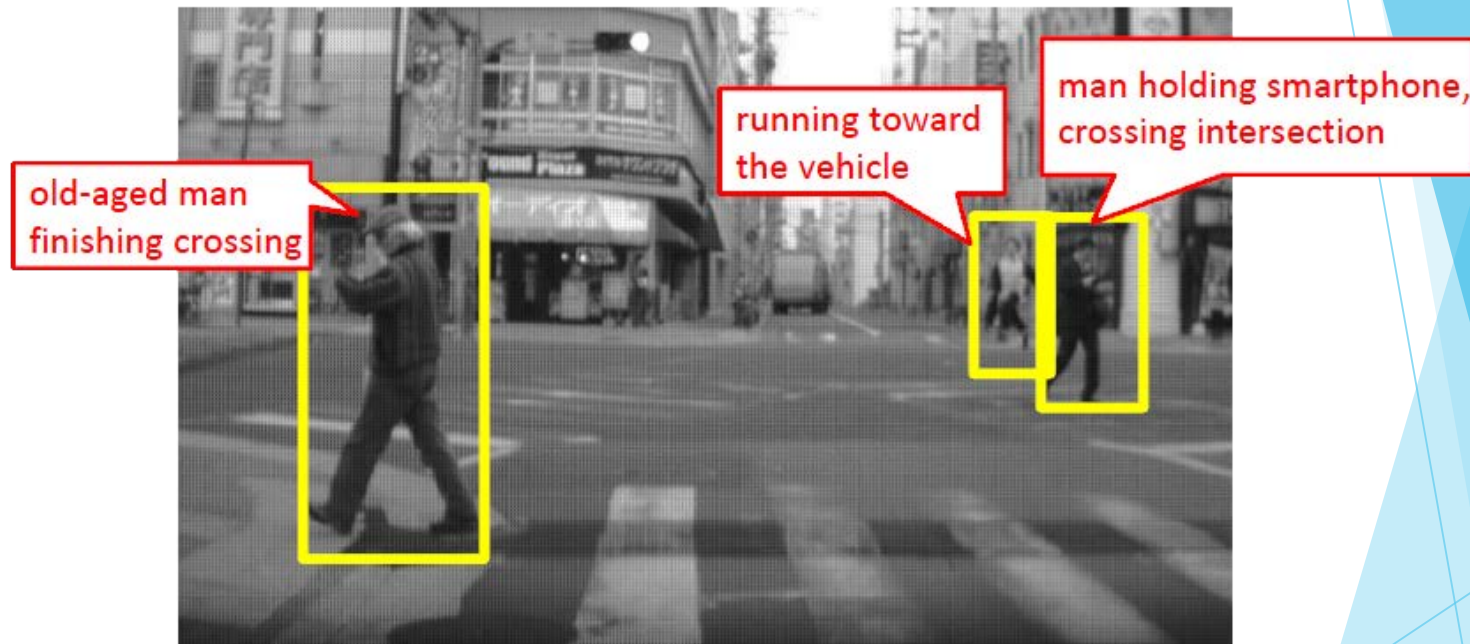
Quelle: [11]

# Probleme aus verwandten Arbeiten



Quelle: [11]

# Probleme aus verwandten Arbeiten



Quelle: [11]

# Ausblick

- ▶ Tieferes Verständnis von LSTM-Netzen erlangen
- ▶ Ausschuchen und Kennenlernen der Werkzeuge
  - ▶ Tensorflow (Google)
    - ▶ <https://www.tensorflow.org/>
  - ▶ CNTK - Microsoft Cognitive Toolkit (Microsoft)
    - ▶ <https://www.microsoft.com/en-us/research/product/cognitive-toolkit/>
  - ▶ Caffe
    - ▶ <http://caffe.berkeleyvision.org/>
  - ▶ Keras
    - ▶ <https://keras.io/>
- ▶ Trainingsumgebung kennenlernen
  - ▶ Datensammlung - 250.000 Bilder von Fußgängern
    - ▶ Caltech Pedestrian Detection Benchmark
    - ▶ [http://www.vision.caltech.edu/Image\\_Datasets/CaltechPedestrians/](http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/)

# Ausblick

- ▶ Generierung eigener dynamischer Systeme (Matlab)
  - ▶ Kein Problem bezüglich Trainingsdaten
- ▶ Fußgängererkennung in Sequenzen -> LSTM
  - ▶ Eigentliches Problem

# Konferenzen

- ▶ The AI Summit - 9.-10. Mai 2017 - London
  - ▶ <https://theaisummit.com/london/>
- ▶ Deep learning for object detection and neural network deployment - Alison Lowndes (NVIDIA) - 23. Mai 2017 - London
  - ▶ <https://conferences.oreilly.com/strata/strata-eu/public/schedule/detail/57754>
- ▶ The AI Conference - 2. Juni 2017 - San Francisco
  - ▶ <https://aiconference.ticketleap.com/helloworld/details>
- ▶ AI and Machine Learning World 2017 - 13.-15. Juni 2017 - London
  - ▶ <https://tmt.knect365.com/ai-machine-learning-world/>

# Quellen - Literatur

- ▶ A. Alahi et al., Social LSTM: Human Trajectory Prediction in Crowded Spaces, 2016
- ▶ A. Angelova et al., Real-Time Pedestrian Detection With Deep Network Cascades, 2015
- ▶ D. Ribeiro, A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation, 2016
- ▶ I. Sato & H. Niihara, Beyond Pedestrian Detection: Deep Neural Networks Level-Up Automotive Safety, 2014
- ▶ V. Gomez, Object Detection for Autonomous Driving Using Deep Learning, 2015
- ▶ T. Chen, Going Deeper with Convolutional Neural Network for Intelligent Transportation, 2015
- ▶ A. Meisel, Vorlesungsfolien - Robot Vision, 2015
- ▶ A Beginner's Guide to Recurrent Networks and LSTMs - <https://deeplearning4j.org/lstm.html>

# Quellen - Bilder

- [01] <https://www.iis.fraunhofer.de/de/ff/lok/proj/kotag.html>
- [02] <https://www.intergator.de/2016/11/01/deep-learning-der-generalschluessel-fuer-dokumentanalysen-teil-2/>
- [03] [https://de.wikipedia.org/wiki/K%C3%BCnstliches\\_Neuron](https://de.wikipedia.org/wiki/K%C3%BCnstliches_Neuron)
- [04] [https://de.wikipedia.org/wiki/Deep\\_Learning](https://de.wikipedia.org/wiki/Deep_Learning)
- [05] [https://de.wikipedia.org/wiki/Rekurrentes\\_neuronales\\_Netz](https://de.wikipedia.org/wiki/Rekurrentes_neuronales_Netz)
- [06] <http://fortune.com/ai-artificial-intelligence-deep-machine-learning/>
- [07] <https://deeplearning4j.org/lstm.html>
- [08] A. Alahi et al., Social LSTM: Human Trajectory Prediction in Crowded Spaces, 2016
- [09] A. Angelova et al., Real-Time Pedestrian Detection With Deep Network Cascades, 2015
- [10] D. Ribeiro, A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation, 2016
- [11] I. Sato & H. Niihara, Beyond Pedestrian Detection: Deep Neural Networks Level-Up Automotive Safety, 2014



Danke für Ihre Aufmerksamkeit!  
Gibt es noch Fragen?