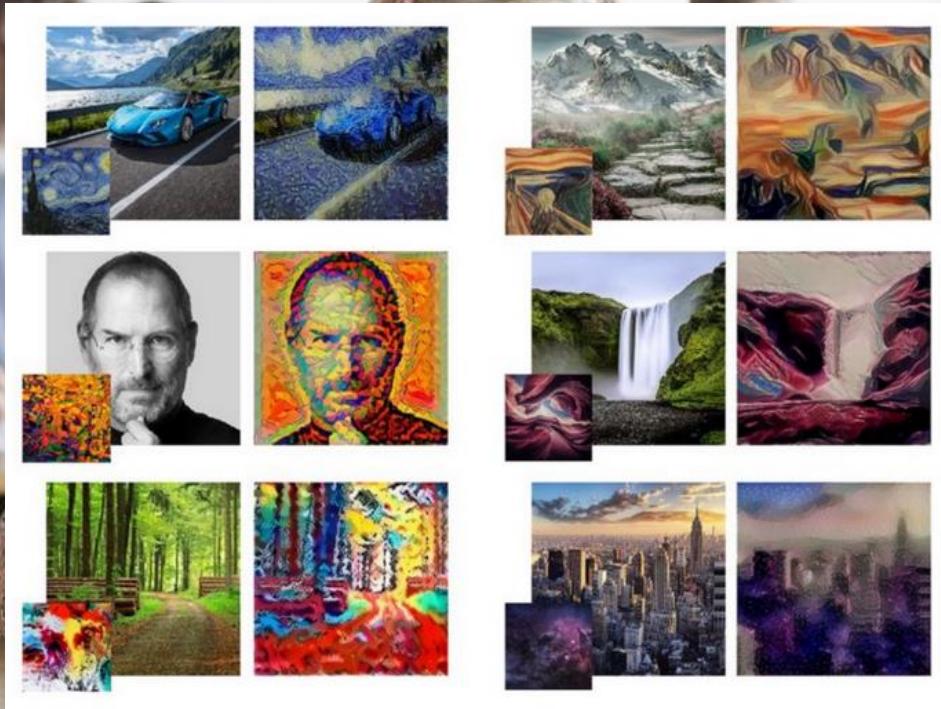


Which machine learning technique is used in this case (and in Deepfakes)?

Schwierigkeitsgrad	Art des Wissens Abfragewissen (Vorlesung)	Anwendungswissen (Literatur)
Einfach	Green	Yellow
Mittel	Yellow	Red
Schwierig	Red	Red



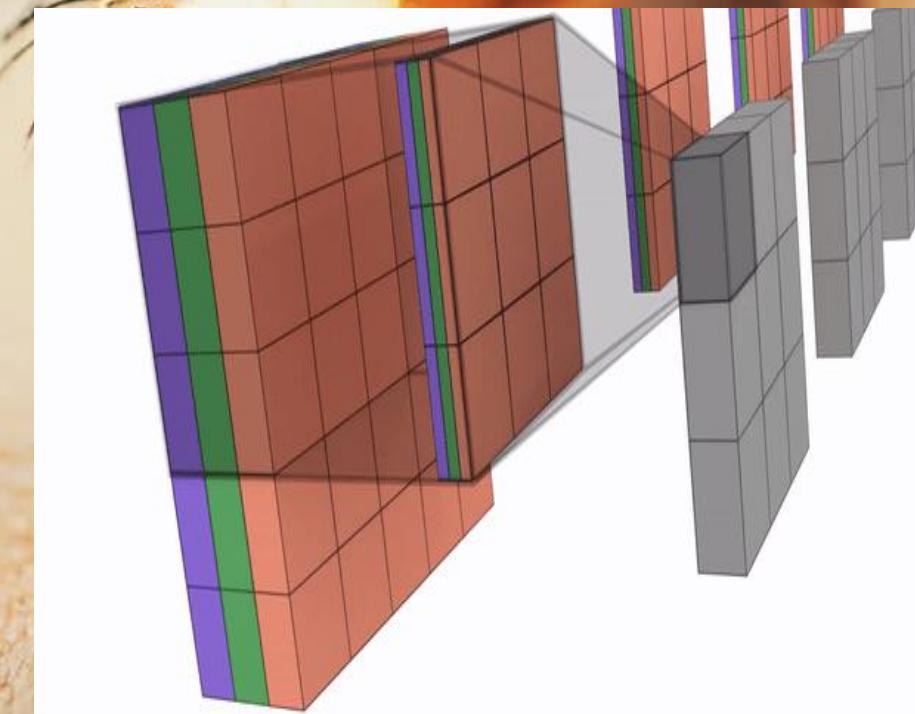
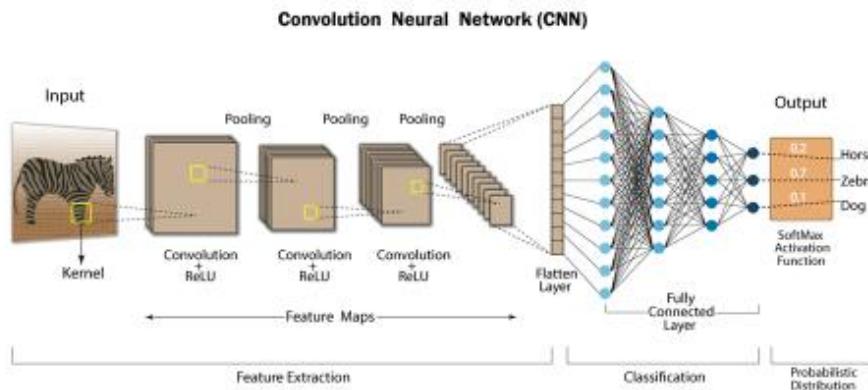
- a) Convolutional Neural Networks
- b) Transfer Learning
- c) Generative Adversarial Networks
- d) Reinforcement Learning
- e) Transformer

8 Machine Learning I

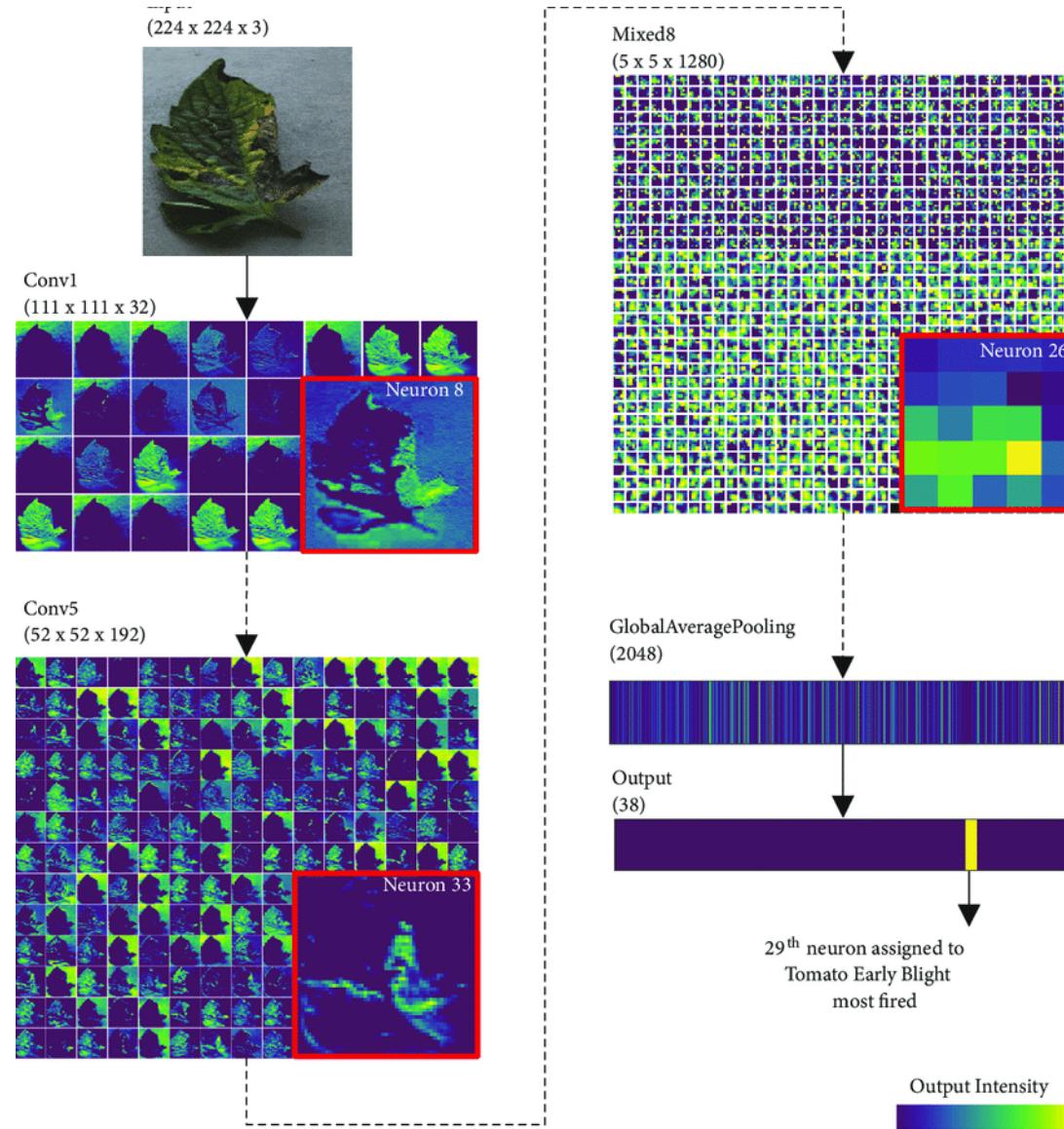
- Neural Networks & Deep Learning

(4) Basics of Convolutional Neural Networks (CNN)

CNN: Grundstruktur am Beispiel



CNN: Eine Beispielanwendung



8 Machine Learning I

- Neural Networks & Deep Learning

Content:

1. Motivation
2. Basics of
Neural Networks (NN)
3. TensorFlow Playground
4. Basics of Convolutional
Neural Networks (CNN)
5. Deep Learning (DL) by
Deepmind: AlphaGo, Zero...
6. Deep NN in
Tesla Autonomous Driving
7. Summary

0
1
2
3
4
5
6
7
8
9

8 Machine Learning I

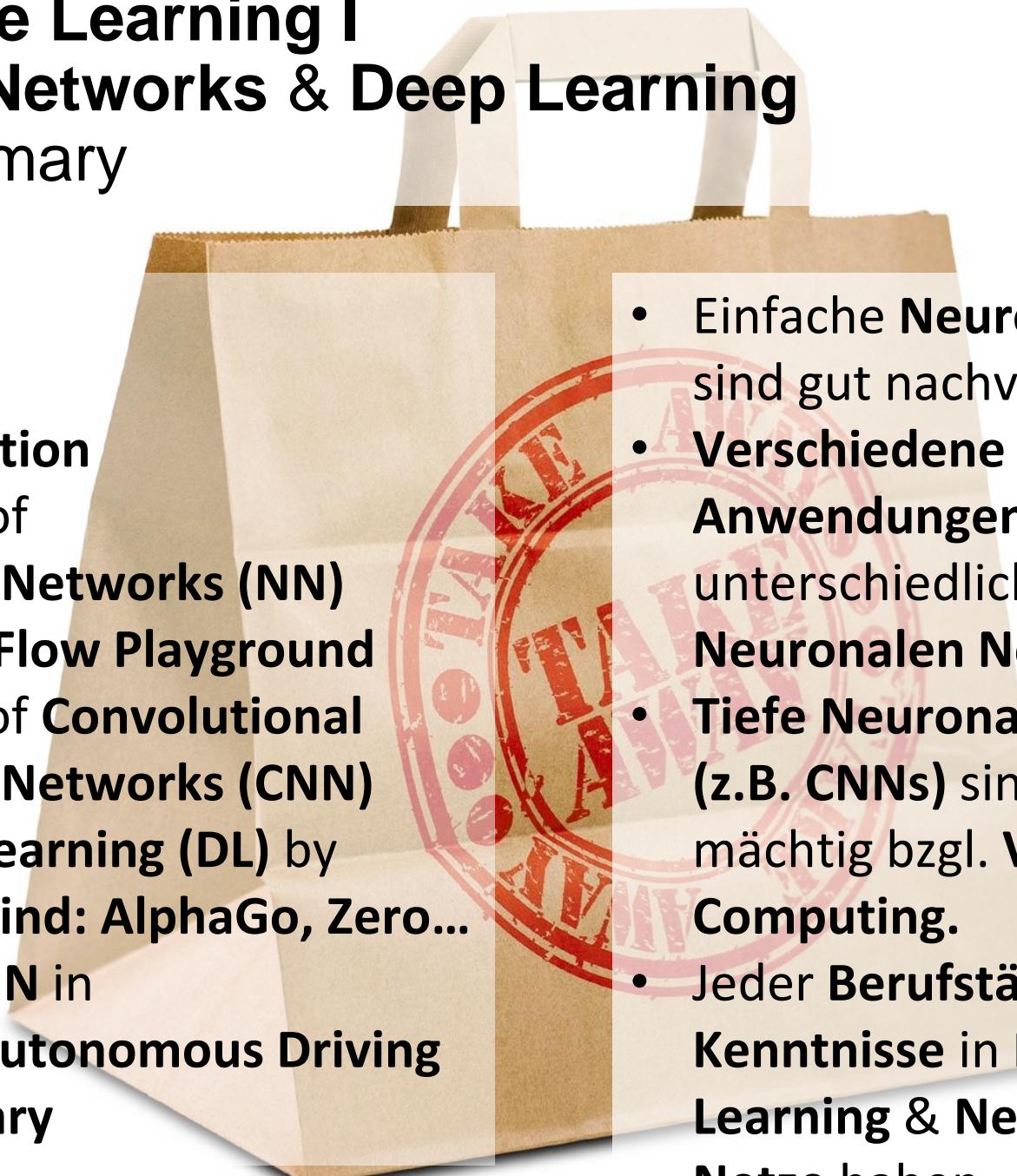
- Neural Networks & Deep Learning

(7) Summary

Content:

1. Motivation
2. Basics of Neural Networks (NN)
3. TensorFlow Playground
4. Basics of Convolutional Neural Networks (CNN)
5. Deep Learning (DL) by Deepmind: AlphaGo, Zero...
6. Deep NN in Tesla Autonomous Driving
7. Summary

- Einfache **Neuronale Netze** sind gut nachvollziehbar.
- Verschiedene Anwendungen benötigen unterschiedliche **Neuronale Netze**.
- Tiefe **Neuronale Netze** (z.B. CNNs) sind sehr mächtig bzgl. **Visual Computing**.
- Jeder **Berufstätige** sollte Kenntnisse in **Machine Learning & Neuronale Netze** haben.



MIT Introduction to Deep Learning 6.S191

The screenshot shows a YouTube video player. The main content area displays a diagram of a deep learning architecture. It starts with an 'INPUT' image of a car key. This is followed by a sequence of operations: 'CONVOLUTION + RELU', 'POOLING', 'CONVOLUTION + RELU', and another 'POOLING'. These steps are grouped under the heading 'FEATURE LEARNING'. The output of these features is then processed through a 'FLATTEN' layer, followed by a 'FULLY CONNECTED' layer, and finally a 'SOFTMAX' layer to produce classification results for 'CAR', 'TRUCK', 'VAN', and 'BICYCLE'. Below the diagram, applications of this architecture are listed: 'Detection', 'Semantic segmentation', and 'End-to-end robotic control'. To the right of the diagram, there is a video frame showing a man (Alexander Amini) speaking at a podium. The video player interface includes a progress bar (29:47 / 37:20), a 'Applications' link, and various interaction buttons like 'ABONNIEREN' (Subscribe) and 'CHATWIEDERGABE ANZEIGEN' (Show Chat Replay). The MIT Deep Learning logo is visible in the top right corner of the video frame.

An Architecture for Many Applications

INPUT CONVOLUTION + RELU POOLING CONVOLUTION + RELU POOLING

FEATURE LEARNING

Detection
Semantic segmentation
End-to-end robotic control

CLASSIFICATION

CAR TRUCK VAN
BICYCLE

MIT Deep Learning

IntroToDeepLearning.com

Massachusetts Institute of Technology

6.S191 Introduction to Deep Learning
introtodeeplearning.com @MITDeepLearning

1/28/20

29:47 / 37:20 • Applications >

Alexander Amini
86.200 Abonnenten

ABONNIEREN

CHATWIEDERGABE ANZEIGEN

Text im Video suchen

3128 24 TEILEN SPEICHERN ...

MIT Introduction to Deep Learning 6.S191: Lecture 3

109

Michael Amberg

Todays Content:

- 1. Motivation**
- 2. Basics of Neural Networks (NN)**
- 3. TensorFlow Playground**
- 4. Basics of Convolutional Neural Networks (CNN)**
- 5. Deep Learning (DL) by Deepmind: AlphaGo, Zero...**
- 6. Deep NN in Tesla Autonomous Driving**
- 7. Summary**





Deep learning

From Wikipedia, the free encyclopedia

Deep learning (also known as **deep structured learning**) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised.^{[1][2][3]}

Deep-learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, machine vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.^{[4][5][6]}

Artificial neural networks (ANNs) were inspired by information processing and distributed communication nodes in **biological systems**. ANNs have various differences from biological brains. Specifically, neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analog.^{[7][8][9]}

The adjective "deep" in deep learning comes from the use of multiple layers in the network. Early work showed that a linear **perceptron** cannot be a universal classifier, and then that a network with a nonpolynomial activation function with one hidden layer of unbounded width can on the other hand so be. Deep learning is a modern variation which is concerned with an unbounded number of layers of bounded size, which permits practical application and optimized implementation, while retaining theoretical universality under mild conditions. In deep learning the layers are also permitted to be heterogeneous and to deviate widely from biologically informed **connectionist** models, for the sake of efficiency, trainability and understandability, whence the "structured" part.

Part of a series on
Machine learning
and
data mining

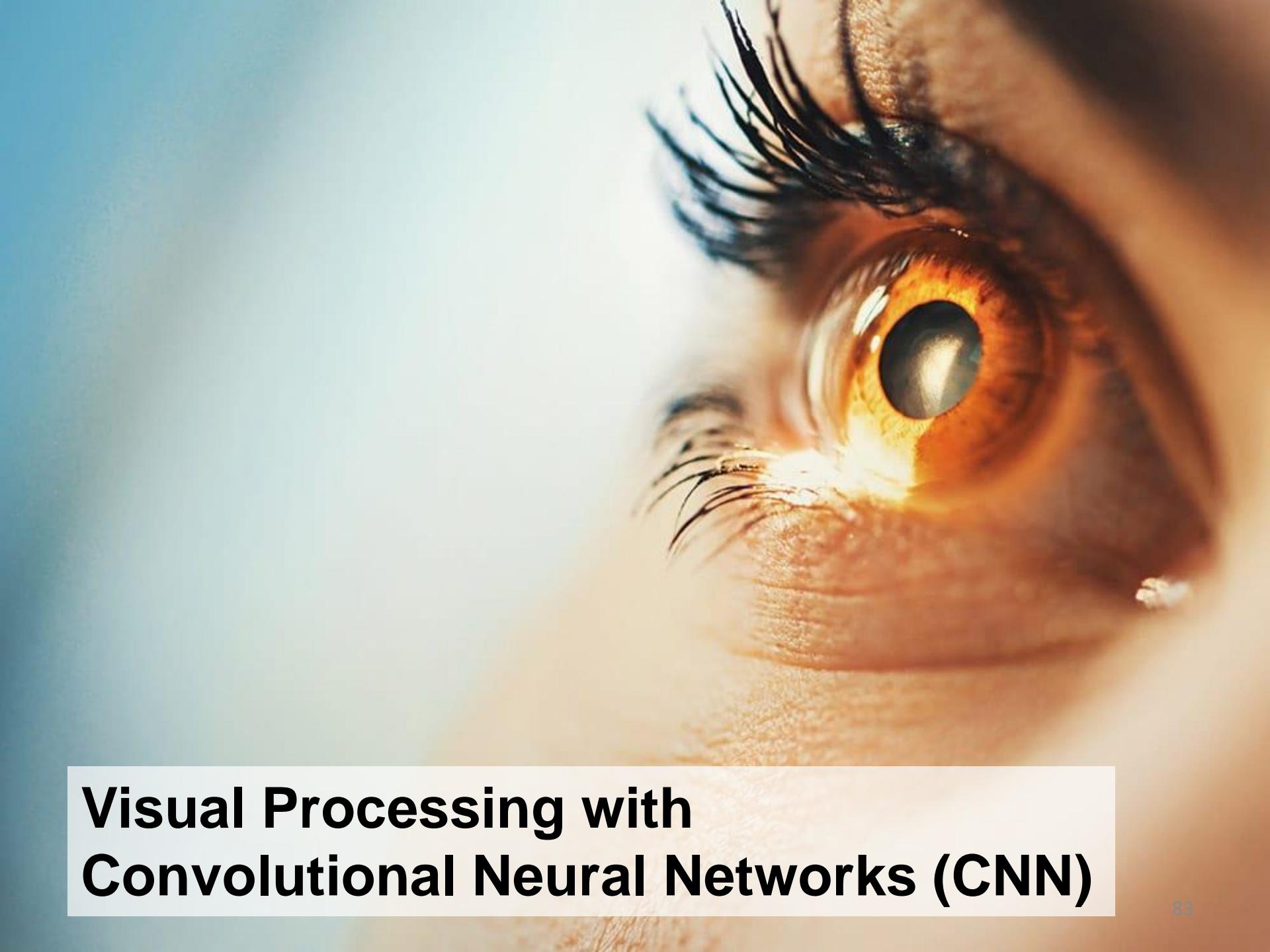
Problems	[show]
Supervised learning	[show]
(classification • regression)	
Clustering	[show]
Dimensionality reduction	[show]
Structured prediction	[show]
Anomaly detection	[show]
Artificial neural network	[show]
Reinforcement learning	[show]
Theory	[show]
Machine-learning venues	[show]
Glossary of artificial intelligence	[show]
Related articles	[show]

V · T · E

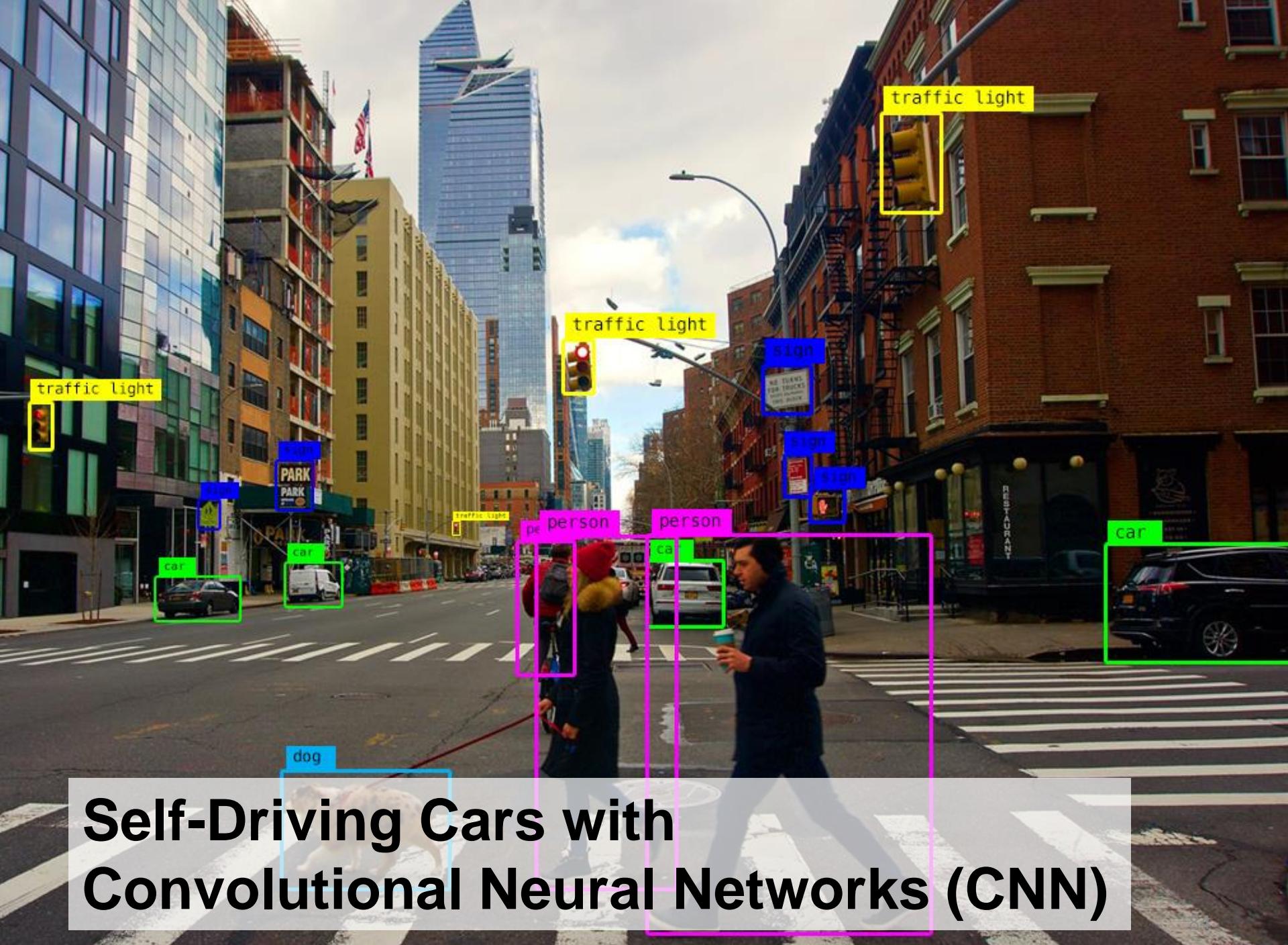
Part of a series on
Artificial intelligence

Major goals	[show]
--------------------	------------------------

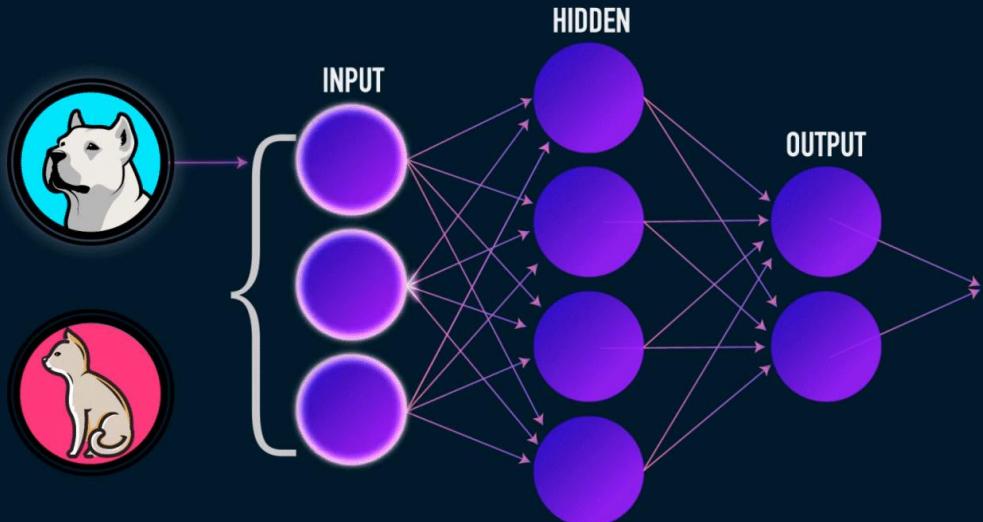
V · T · E



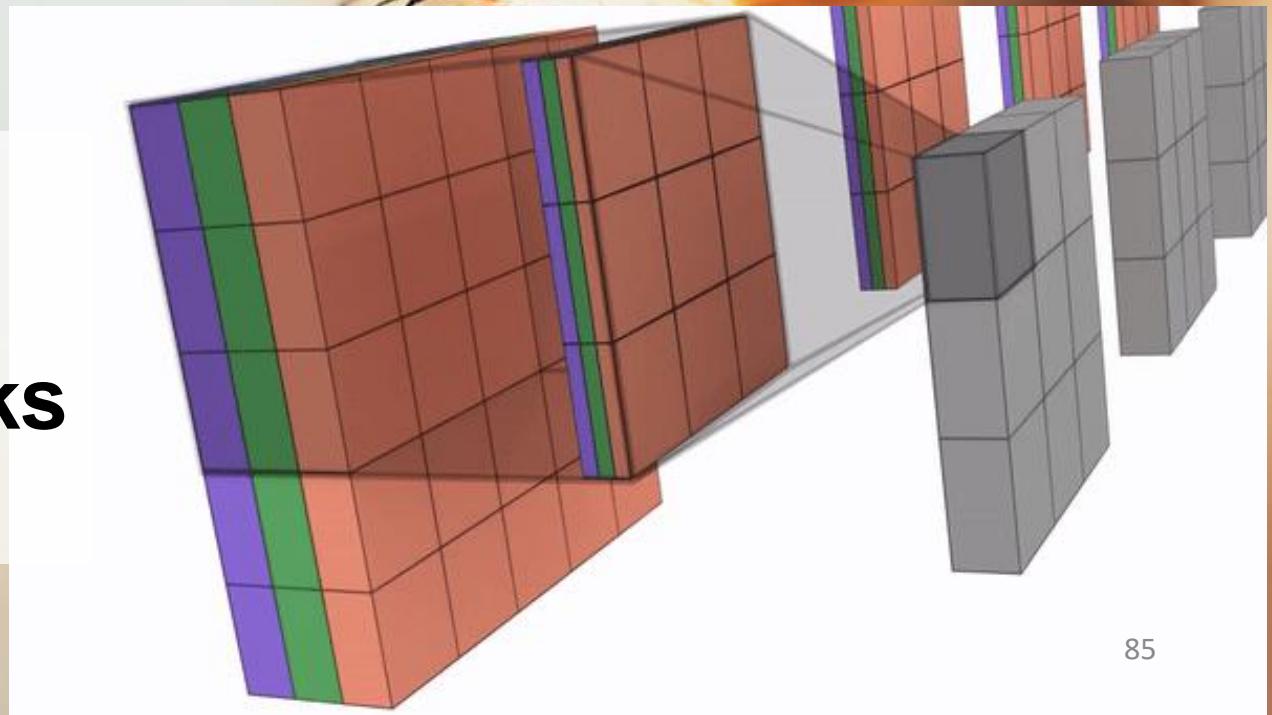
Visual Processing with Convolutional Neural Networks (CNN)



Self-Driving Cars with Convolutional Neural Networks (CNN)

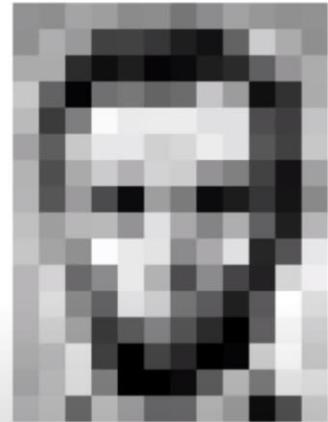


Basics of Convolutional Neural Networks (CNN)



CNN: Grundlegende Problemstellung (1/3)

Was Computer sehen: Ein Bild ist eine Folge von Bildpunkten (Pixel).



What the computer sees															
157	153	174	168	160	162	129	151	172	161	155	156	155	182	163	74
156	182	163	74	75	62	93	17	110	210	180	154	180	180	50	14
180	180	50	14	34	6	10	33	48	106	159	181	206	109	5	124
206	109	5	124	131	111	120	204	166	15	56	180	194	68	137	251
194	68	137	251	237	239	239	228	227	87	71	201	172	106	207	233
172	106	207	233	233	214	220	239	228	98	74	206	188	88	179	209
188	88	179	209	185	215	211	158	139	75	20	169	189	97	165	84
189	97	165	84	10	168	134	11	31	62	22	148	199	168	191	193
199	168	191	193	158	227	178	143	182	106	36	190	205	174	195	252
205	174	155	252	236	231	149	178	228	43	95	234	190	216	116	149
190	216	116	149	236	187	84	150	79	38	218	241	190	224	147	108
190	224	147	108	227	210	127	102	35	101	255	224	190	214	173	66
190	214	173	66	103	143	95	50	2	109	249	215	187	196	235	75
187	196	235	75	1	81	47	0	6	217	255	211	183	202	237	145
183	202	237	145	0	0	12	108	200	138	243	236	195	206	123	207
195	206	123	207	177	121	123	200	175	13	96	218				

An image is just a matrix of numbers [0,255]!
i.e., 1080x1080x3 for an RGB image

Let's identify key features in each image category



Nose,
Eyes,
Mouth



Wheels,
License Plate,
Headlights



Door,
Windows,
Steps

CNN: Grundlegende Problemstellung (2/3)

Eine Objekterkennung in Bildern ist nicht einfach,
da das Umfeld und weitere Rahmenbedingungen dies erschweren.

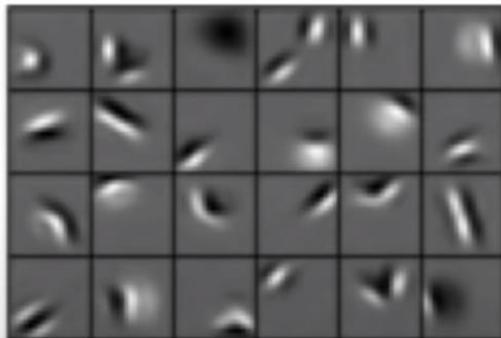


CNN: Grundlegende Problemstellung (3/3)

Kann ein **Computer** die **relevanten Aspekte (Features)** zur **Objekterkennung** aus **Daten** **selbstständig erlernen**, ohne sich von den Rahmenbedingungen „**ablenken**“ zu lassen?

Can we learn a **hierarchy of features** directly from the data instead of hand engineering?

Low level features



Edges, dark spots

Mid level features



Eyes, ears, nose

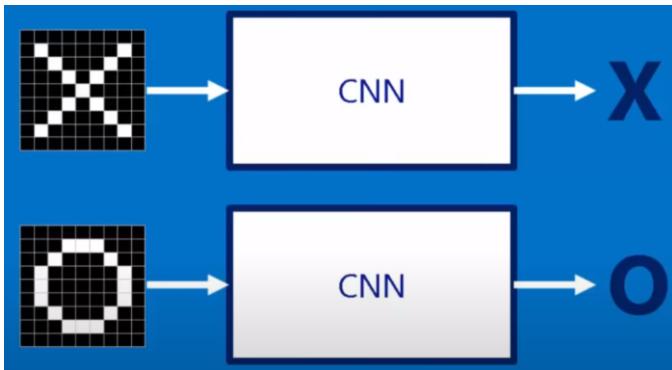
High level features



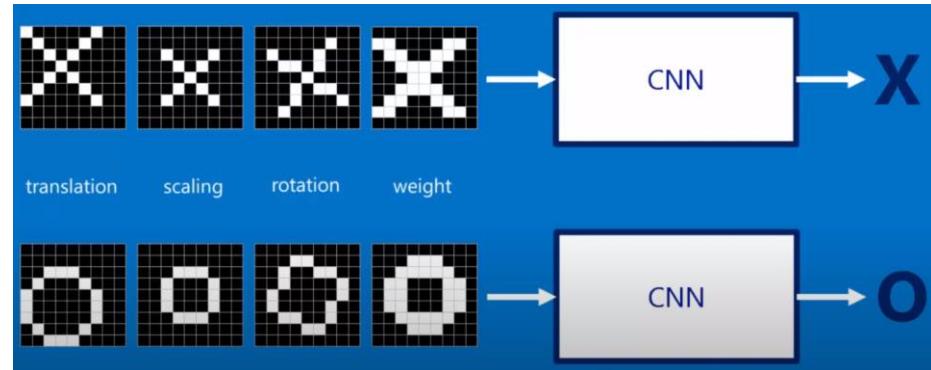
Facial structure

CNN: Kernidee am Beispiel Objekterkennung (1/5)

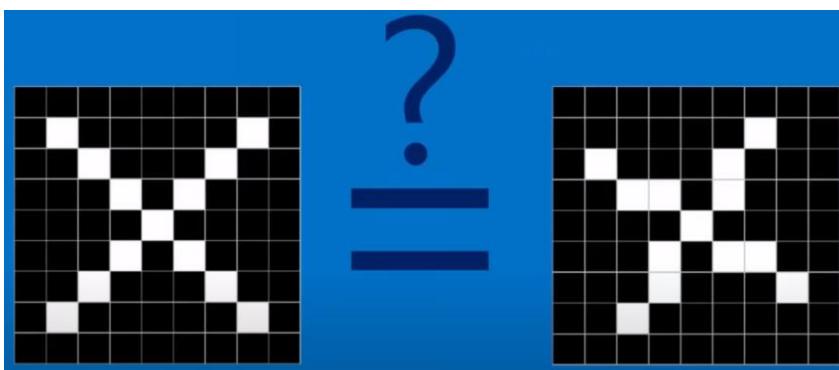
Problemstellung



Herausforderung



Problem



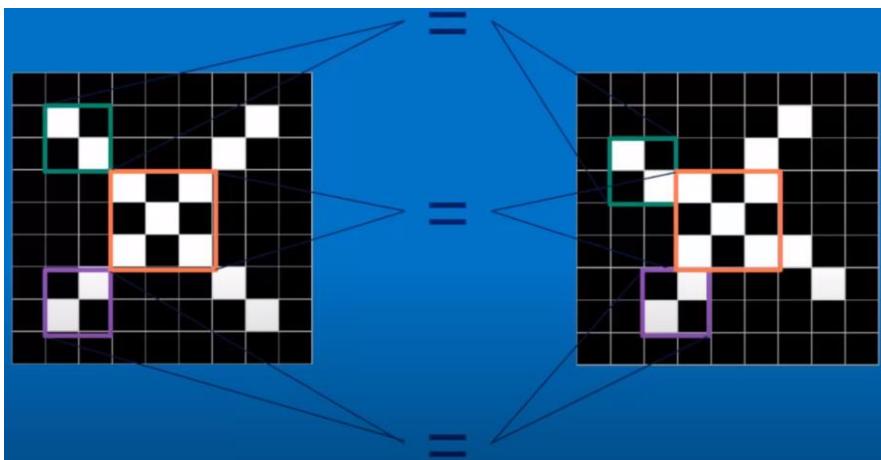
Problem (Computerdarstellung)

This diagram provides a numerical representation of the convolution process. The input image is a 4x4 matrix of -1s and 1s. The kernel is a 3x3 matrix with a central value of 1. The output is a 2x2 matrix where the central element is highlighted in blue, representing the result of the convolution step.

-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	-1	-1	1	-1	-1	-1
-1	-1	-1	1	-1	-1	1	-1	-1
-1	1	-1	-1	-1	-1	1	-1	-1
-1	-1	-1	-1	-1	-1	-1	1	-1

CNN: Kernidee am Beispiel Objekterkennung (2/5)

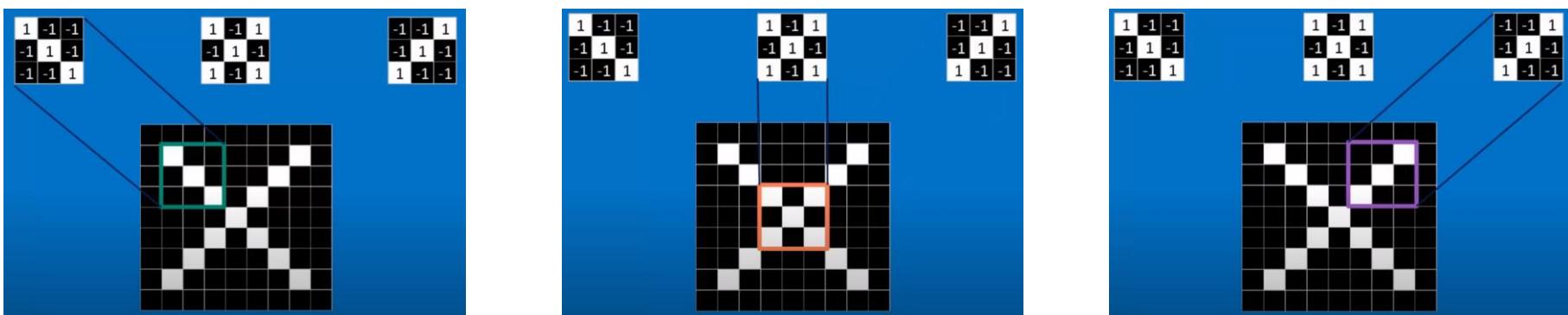
Idee: Features (Filter) helfen



Relevante Features (Filter) identifizieren

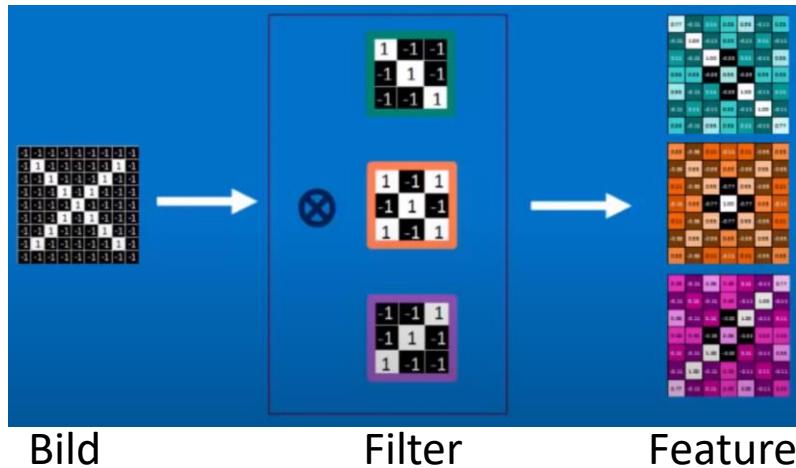
$\begin{matrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{matrix}$	$\begin{matrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{matrix}$	$\begin{matrix} -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{matrix}$
---	---	---

Features (Filter) anwenden: Suche nach Features im Bild



CNN: Kernidee am Beispiel Objekterkennung (3/5)

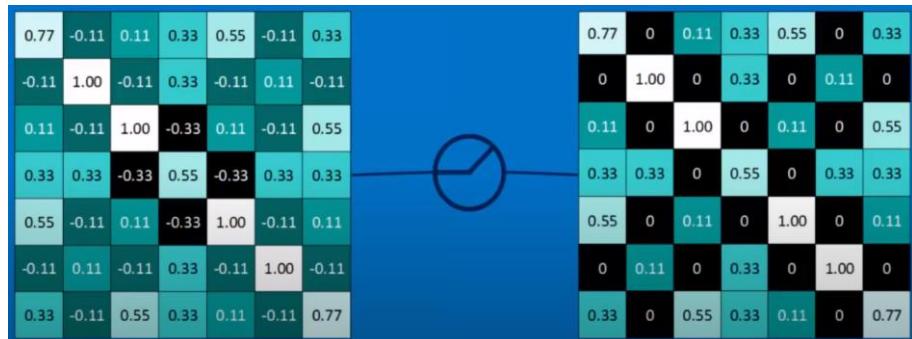
Stapel von mehreren Features (Filtern)



Pooling (Komplexität reduzieren & ungenauer werden)



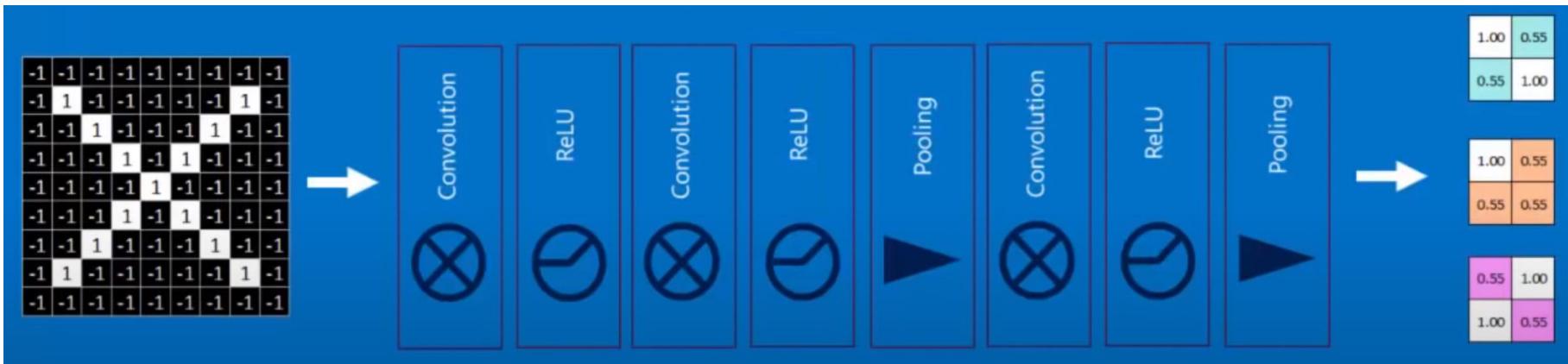
ReLU (Unwichtiges entfernen)



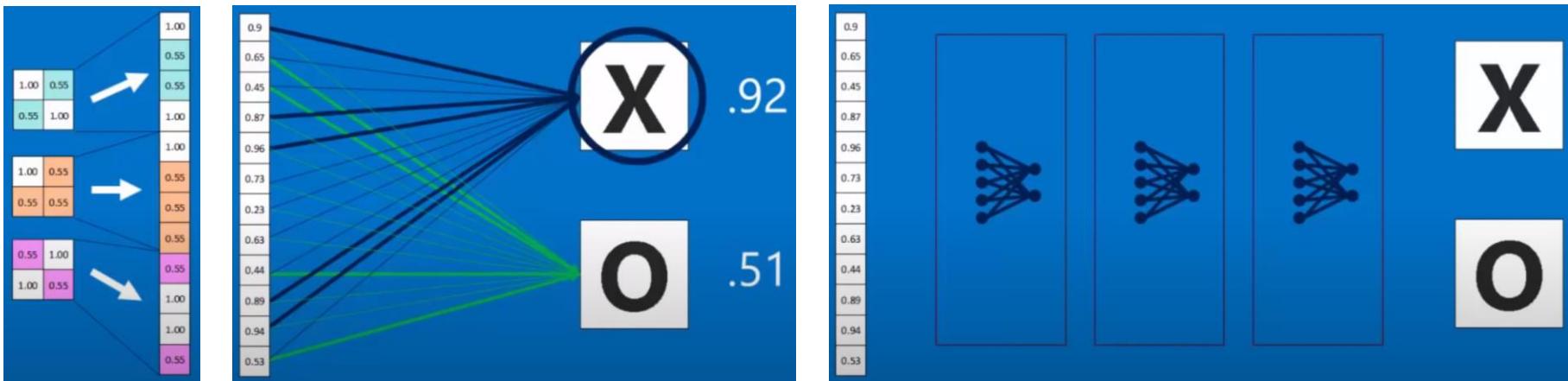
CNN: Kernidee am Beispiel Objekterkennung (4/5)

Diese Operationen können mehrfach wiederholt werden:

Features suchen (Conv./Falten), Ergebnis bereinigen (ReLU) und komprimieren (Pooling).



Ergebnis (die finalen Features) ebnen (engl. flatten) und bewerten (evtl. mehrstufig)



How Convolutional Neural Networks work, 2016

www.youtube.com/watch?v=FmpDlaiMleA

CNN: Kernidee am Beispiel Objekterkennung (5/5)

Backpropagation (Trainieren)

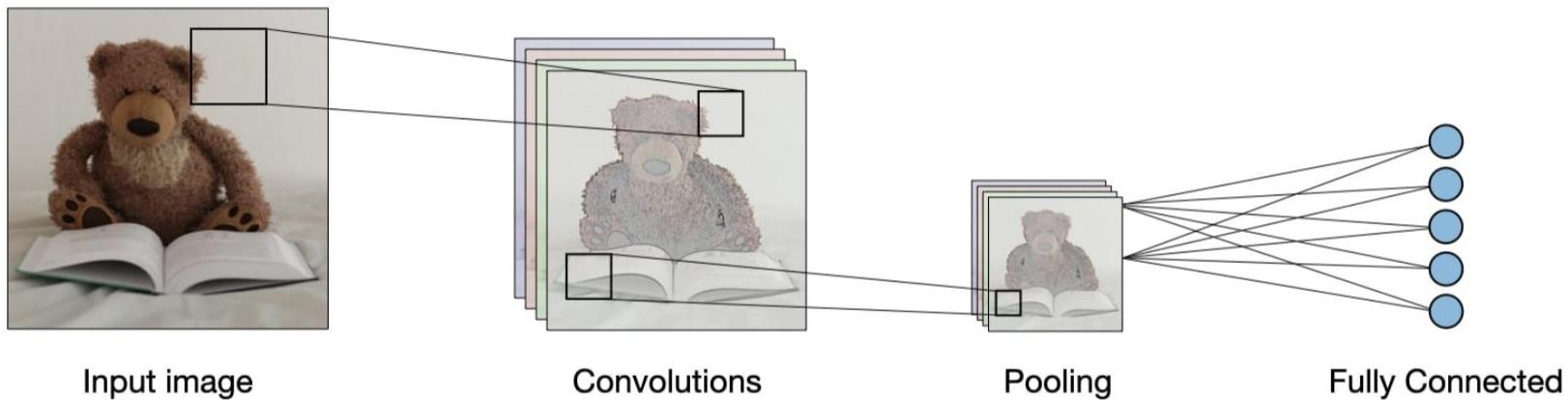
Backprop

	Right answer	Actual answer	Error
X	1	0.92	0.08
O	0	0.51	0.49



Zentrale Bausteine klassischer CNN am Beispiel

- **Architecture of a traditional CNN** — Convolutional neural networks, also known as CNNs, are a specific type of neural networks that are generally composed of the following layers:



The convolution layer and the pooling layer can be fine-tuned with respect to hyperparameters that are described in the next sections.

Google, 2017: Attention is All You Need

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar
Google Research
nikip@google.com

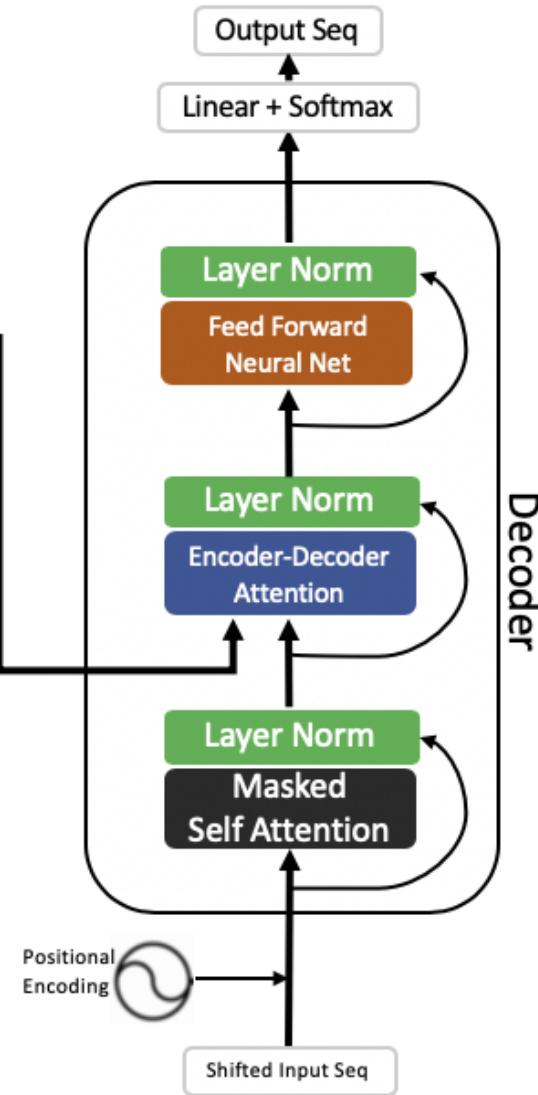
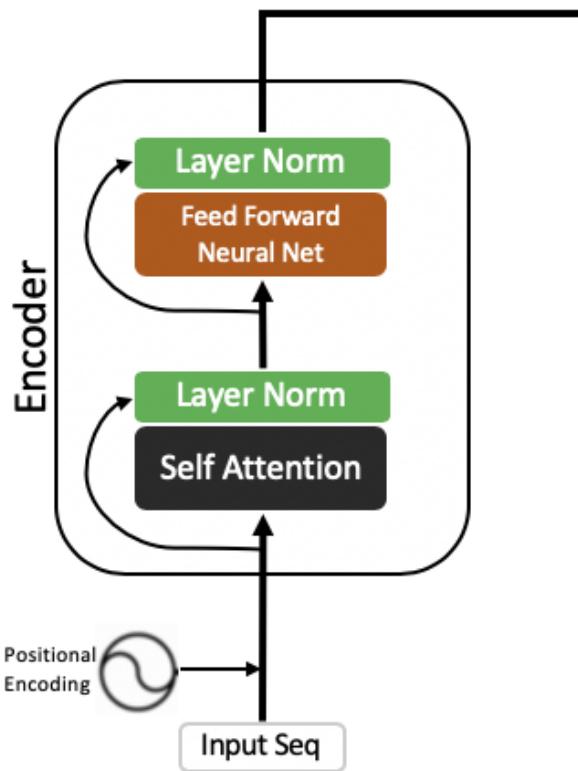
Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

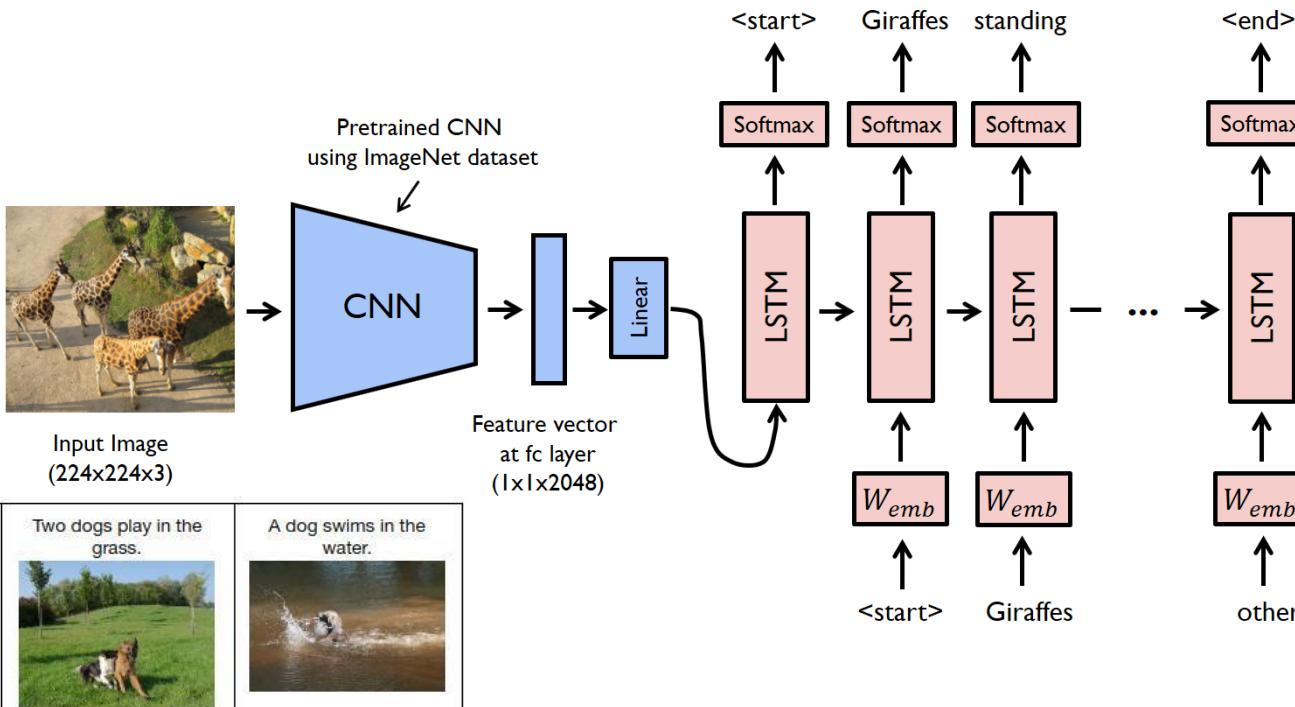
Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on recurrent convolutional neural networks that include an encoder and decoder. Recurrent models also connect the encoder and decoder via a shared recurrent mechanism. We propose a new simple network architecture based solely on attention mechanisms, dispensing with recurrence entirely. Experiments on two machine translation tasks show that our model is superior in quality while being more parallelizable and requiring less time to train. Our model achieves 28.4 BLEU on the English-to-German translation task, improving over the existing best ensembles, by over 2 BLEU. On the WMT 2014 English-to-French task, our model establishes a new single-model state-of-the-art BLEU score after training for 3.5 days on eight GPUs, a small fraction of the best models from the literature. We show that the Transformer can also succeed on other tasks by applying it successfully to English constituent ordering and limited training data.



Kombinierter Einsatz von RNN und LSTM



RNN und LSTM können auch mit **Feed-Forward-Netze** (z.B. **Convolutional Neural Networks**) kombiniert werden. Sie bringen damit „Memory“ hinein.

Anwendungen sind z.B.:
Textuelle Beschreibung von Bildern (**Image Captioning**) oder Automatische Generierung von Untertiteln (**Video Captioning**).

Google, 2017: Attention is All You Need

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Llion Jones*
Google Research
llion@google.com

Noa Golany*
Google Brain
noamg@ai.goo

Aidan老人
University of
Edinburgh
aidan@inf.ed.ac.uk

Clémentine

Niki Dyer*

Torko Elmiussin*

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

Google, 2017: Attention is All You Need

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkorei
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukasz.kaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

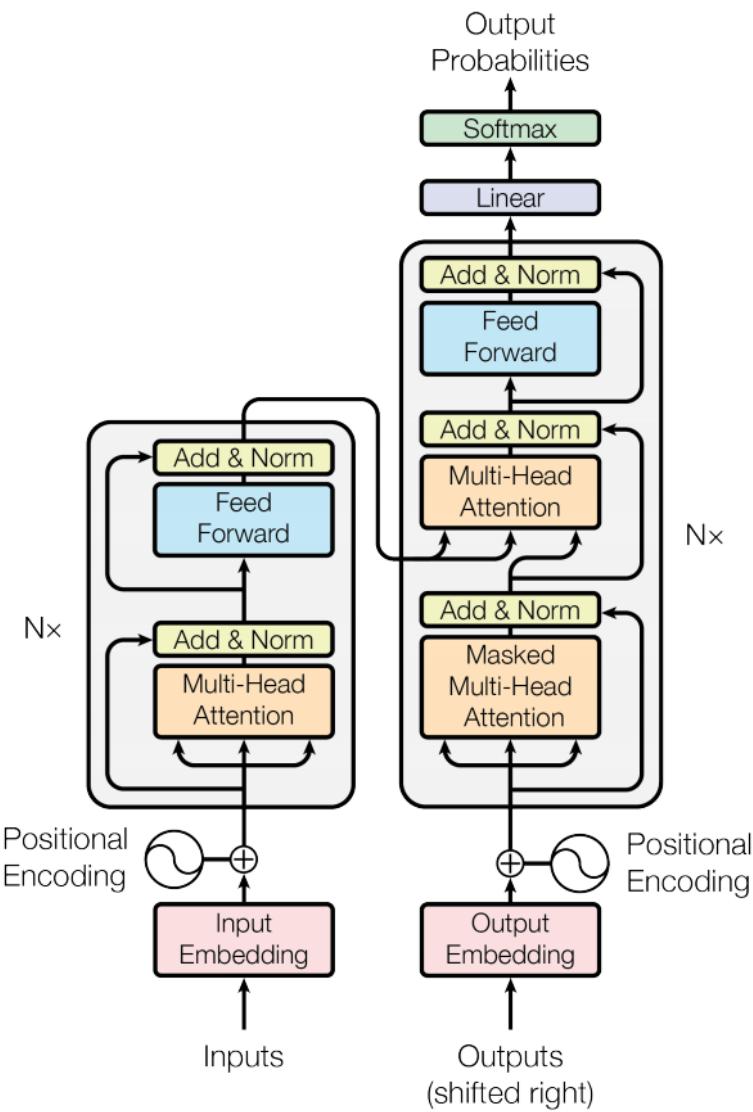


Figure 1: The Transformer - model architecture.