

Vergleich der Deep-Learning-Ansätze “YOLO” und “R-CNN” für den Bereich der bildbasierten Objekterkennung

Dr. Michael Thrun

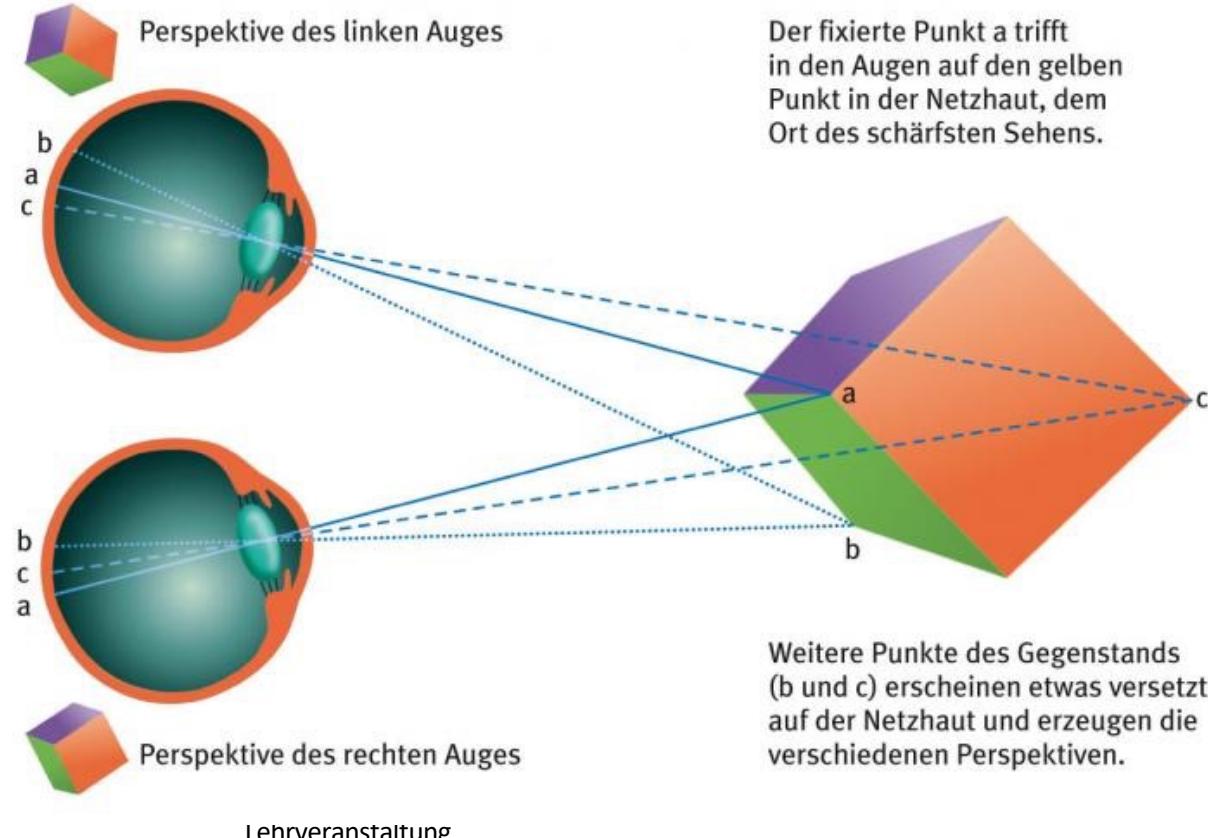
<https://github.com/Mthrun/Deeplearning4ObjectDetection>

Lernziele

1. Datenbionische Konzepte der Objekterkennung im Maschinellen Lernen
2. Vorstellung „R-CNN“ und „YOLO“
3. Vor- und Nachteile dieser Algorithmen

Natürliche Objekterkennung bei Tieren

- Raubtiere segmentieren das Bild und haben fünf Möglichkeiten in Ihren visuellen Systemen, um getarnte Beutetiere zu erkennen [Regan, 2000]
 - Luminanz
 - Bewegung
 - Farbe
 - Textur
 - Binokulare Disparität



Natürliche Objekterkennung bei Tieren

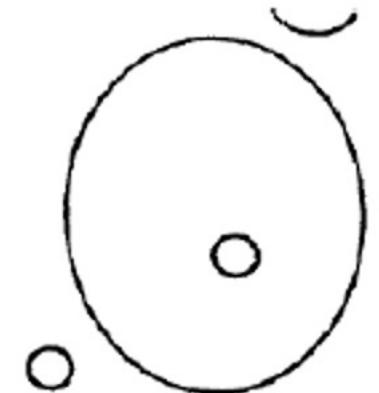
- Raubtiere segmentieren das Bild und haben fünf Möglichkeiten in Ihren visuellen Systemen, um getarnte Beutetiere zu erkennen [Regan, 2000]
 - Luminanz
 - Bewegung
 - Farbe
 - Textur
 - Binokulare Disparität

Theorien der Objekterkennung

- In der Erkennung konkurrieren zwei verschiedene Theorien
 - In den kantenbasierenden Theorien sind weitgehend nur die Konturen von Objekten relevant
 - Hier hat Farbe keinen Einfluss.
 - In oberflächenbasierten Theorien hat die Oberfläche eine gleichwertige Bedeutung zum Umriss
 - Dadurch ist Farbe von Bedeutung.

Beispiele kantenbasierender Theorie der Objekterkennung

- a) Objekterkennung die Möglichkeit einer Erkennung der Form und der räumlichen Relation zwischen Umrissen benötigt
- b) Ein Objekt wird vor der Summe seiner Komponenten erkannt
- c) Doppeldeutigkeiten sind möglich
- d) Umrandungen müssen nicht geschlossen sein



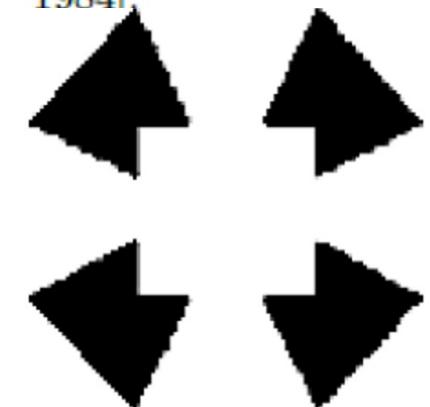
(a) Gesicht nur bei korrekter Anordnung der einzelnen Elemente erkennbar [Ullman, 1984].



(b) Die Zahl wird zuerst erkannt [Ullman, 1984].



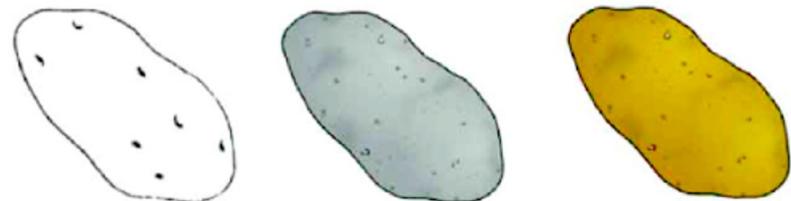
(c) Saxophonspieler oder Gesicht?



(d) Quadrat aus 4 uneindeutigen Komponenten.

Beispiele Oberflächenbasierte Theorie der Objekterkennung

f) Objektoberflächeninformationen wie Farbe und Textur sind selbst bei Zeichnungen für einfache Objekterkennung wesentlich



(f) Die gezeichnete Kartofel ist nur durch die Farbe erkennbar [Rossion u. a., 2004].

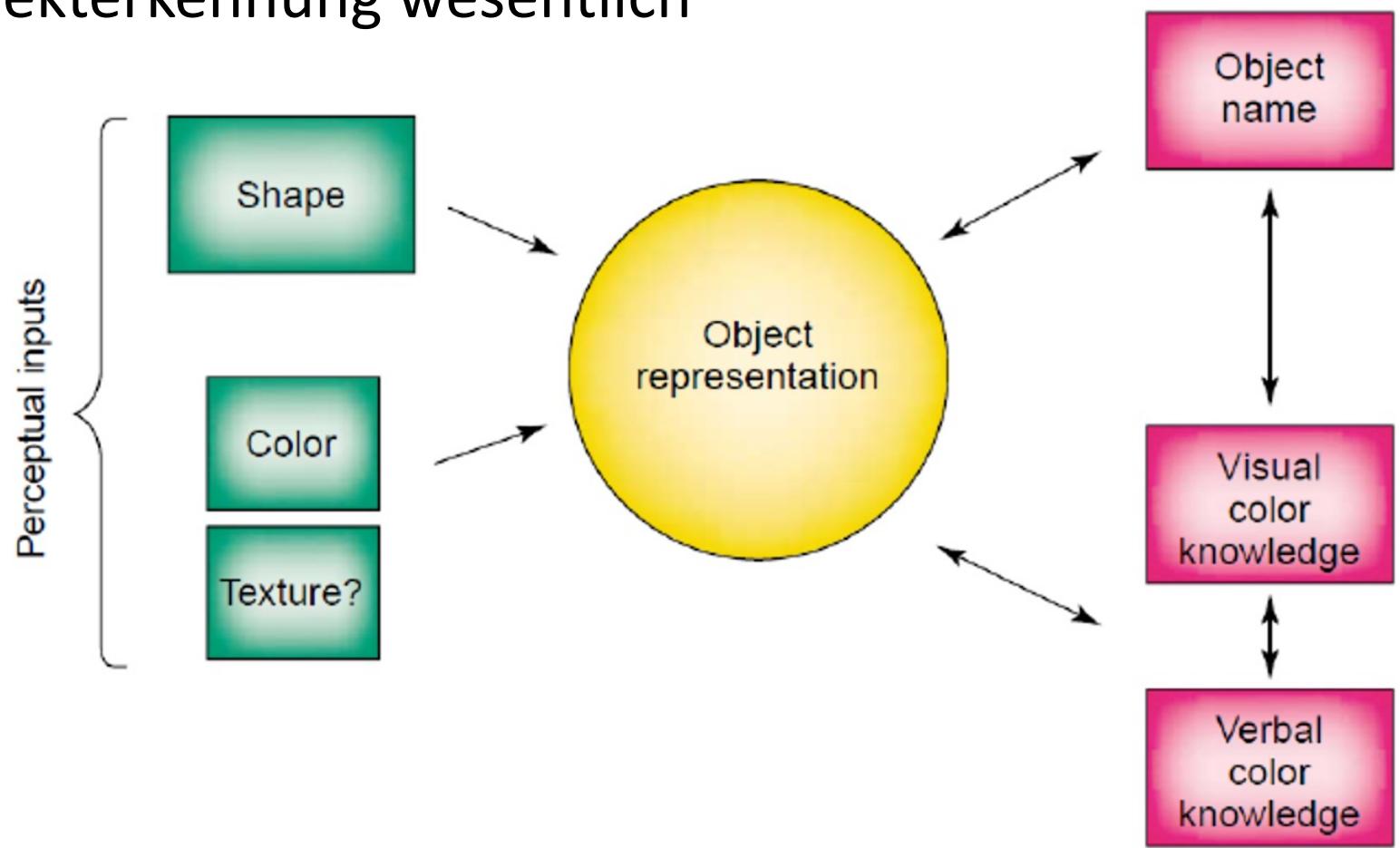
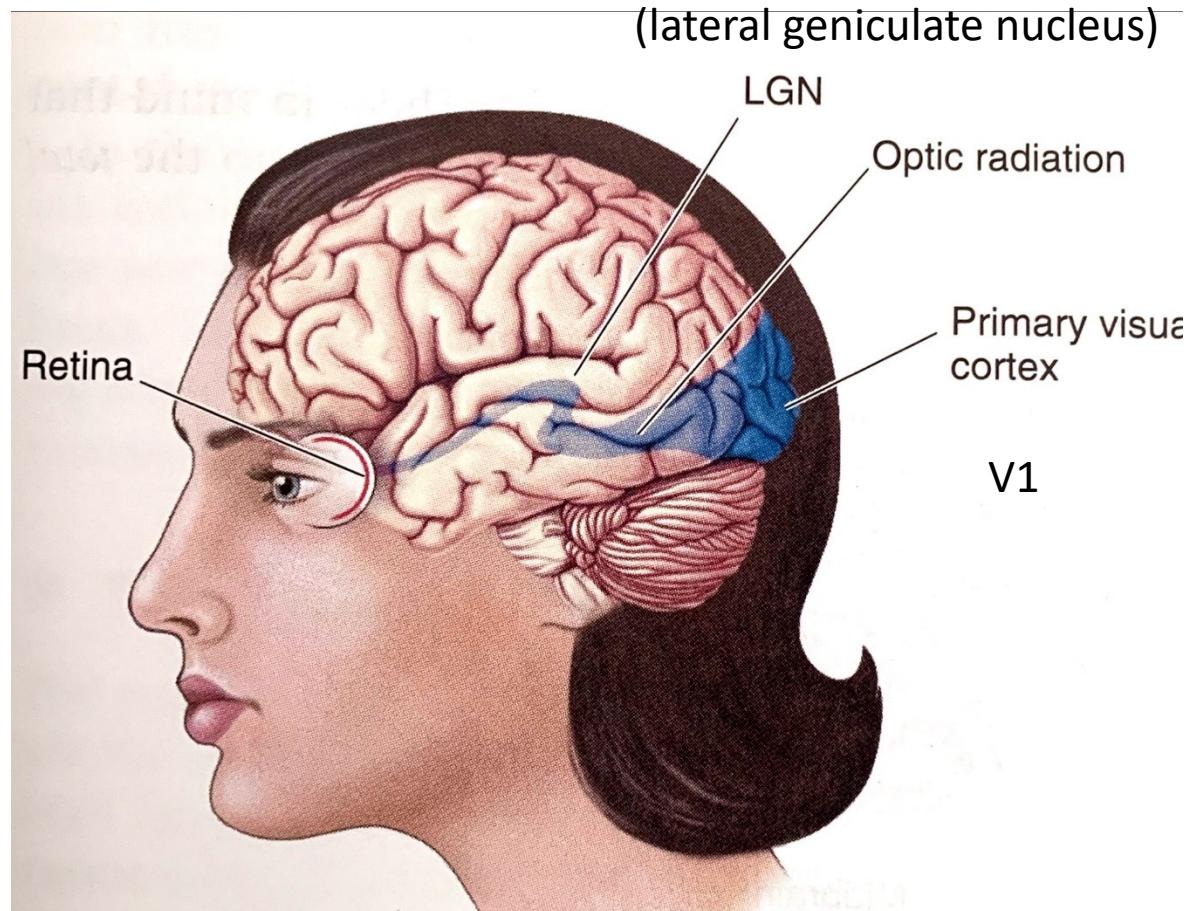
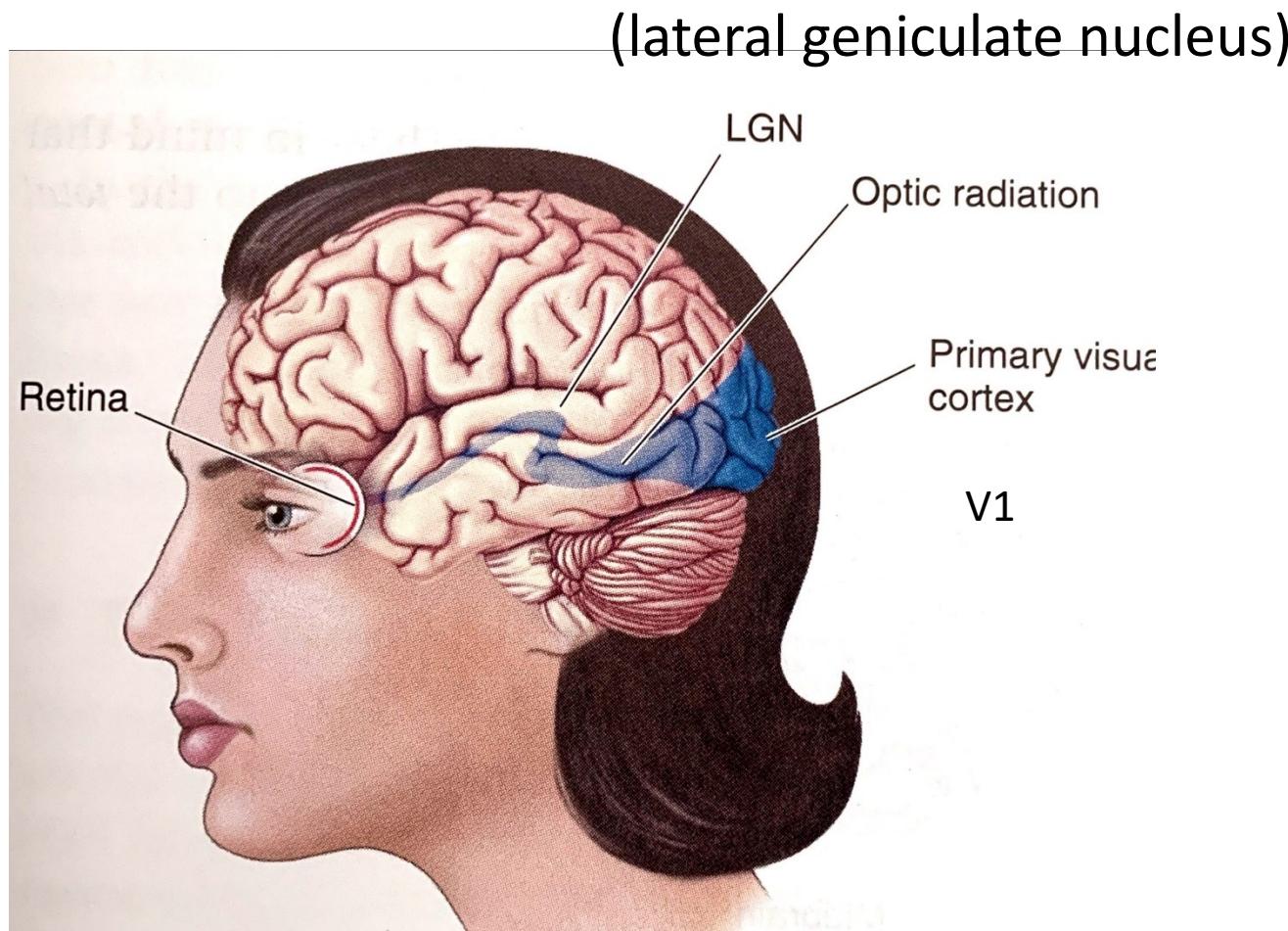


Abb. 1.2-2: Kontur- und Oberflächenmodell der Objekterkennung aus [Tanaka u. a., 2001].

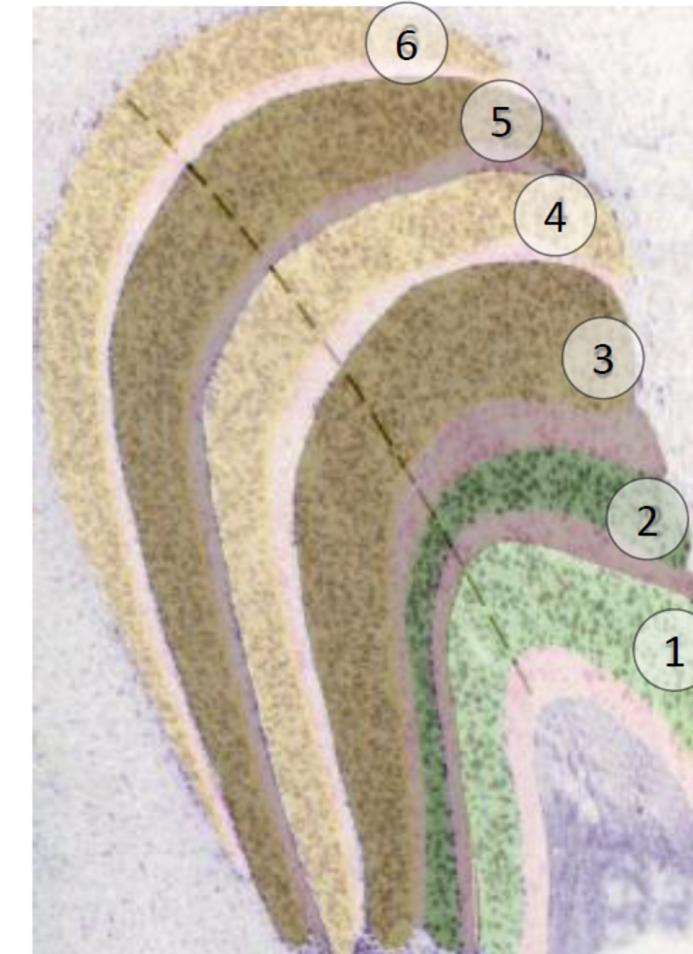
Objekterkennung & Wahrnehmung beim Menschen: Auge bis V1



Objekterkennung & Wahrnehmung beim Menschen: Auge bis LGN



lateral geniculate nucleus (LGN)



Objekterkennung & Wahrnehmung beim Menschen: Auge bis LGN

- In allen Schichten des CGL findet sich eine **retinotope** Abbildung der Umwelt.
 - Benachbarte Positionen im visuellen Feld, befinden sich in benachbarten Bereichen des CGL
- Schichten 1+2 codieren Bewegung Tiefe
 - Teil des "magnzellulärer Pfad"
- Schichten 3-6 codieren Farbe, Details
 - Teil des "parvozellulärer Pades"
- Zwischenschichten - Steuerung von Pupillenweite & reflektorischen Augenbewegungen
 - "Koniozellulärer Pfad"



Objekterkennung & Wahrnehmung beim Menschen: Auge bis LGN

- In allen Schichten des CGL findet sich eine **retinotope** Abbildung der Umwelt.
 - Benachbarte Positionen im visuellen Feld, befinden sich in benachbarten Bereichen des CGL
 - Schichten 1+2 codieren Bewegung Tiefe
 - Teil des “magnzellulärer Pfad”
 - Schichten 3-6 codieren Farbe, Details
 - Teil des “parvozellulärer Pfade”
- ⇒CGL herrscht eine strikte Trennung von Informationen der verschiedenen Pfade
- ⇒Information verbleibt im Bezug zu Bildkoordinaten



Objekterkennung ML versus Natürlich

- Seit den Arbeiten von Biederman (1987) geht man davon aus, dass die menschliche Objekterkennung eine Hierarchie von Teilen berücksichtigt
 - > Viele Algorithmen modellieren Objekte als Konstellationen (Weber, Welling, & Perona, 2000) oder Kompositionen (Ommer & Buhmann, 2010) von generischen Teilen
- Frage nach der Beziehung zwischen Objekt-Teilen und Objekten wurde häufig parallel im Maschinellem Lernen (ML) und der menschlichen Wahrnehmung untersucht.

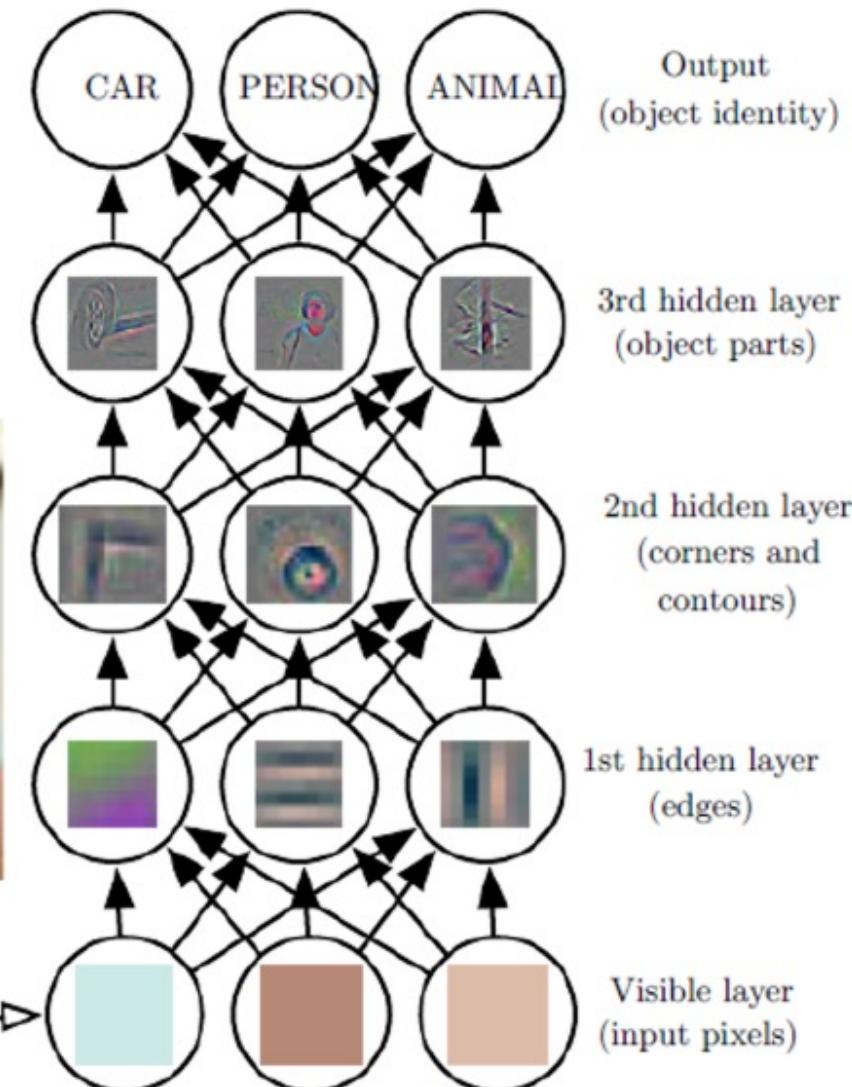
Transfer: Objekterkennung im Deep Learning

- Input daten sind Pixelwerte [r,b,g] werden der 1. Schicht des neuronalen Netzes präsentiert
- Bild wird durch verschiedene Zwischenschichten segmentiert
-> Automatische Feature Extraction
- > Deep learning = „*mehrschichtiges Lernen*“

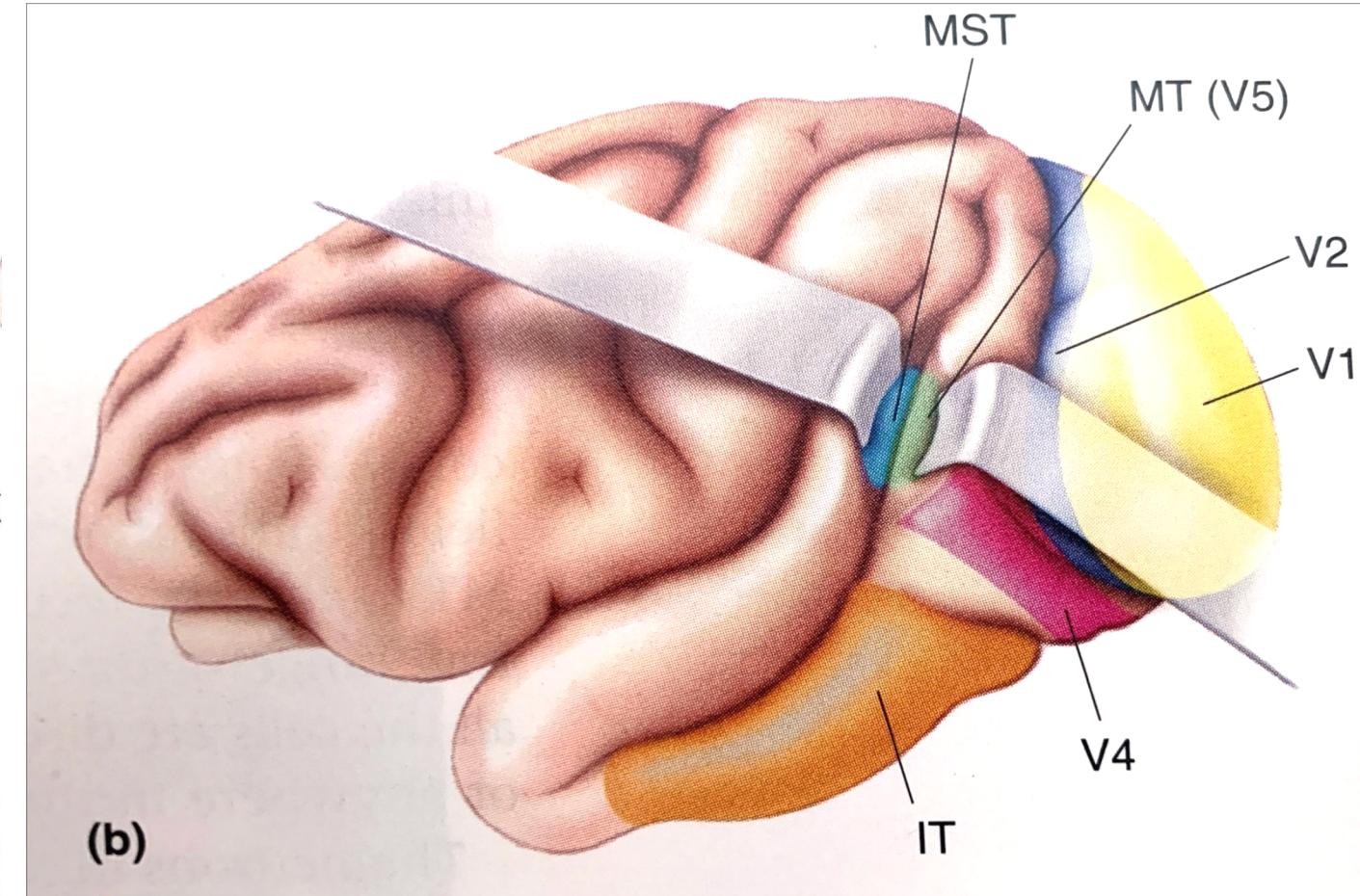
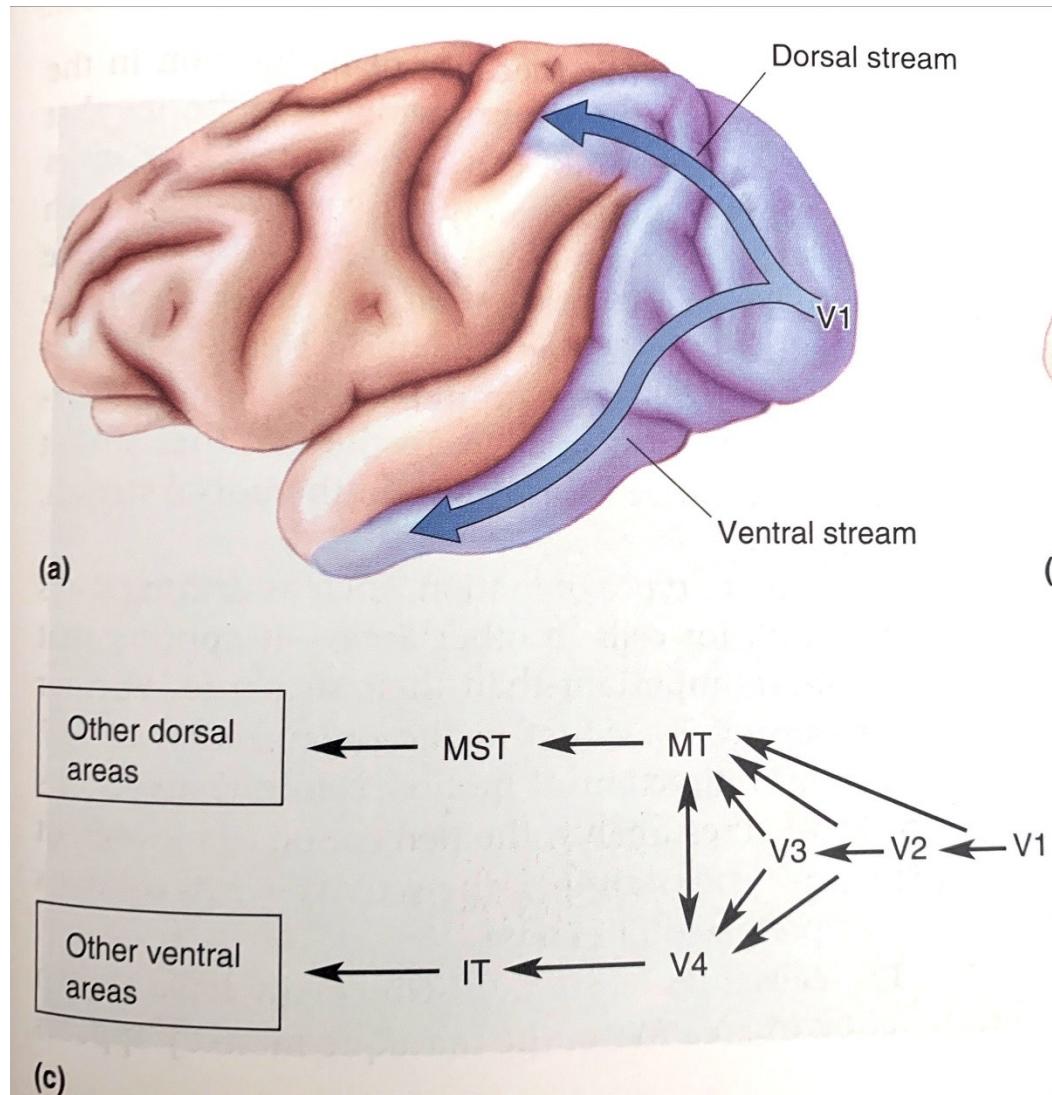
[Goodfellow et al., 2015]



Lehrveranstaltung

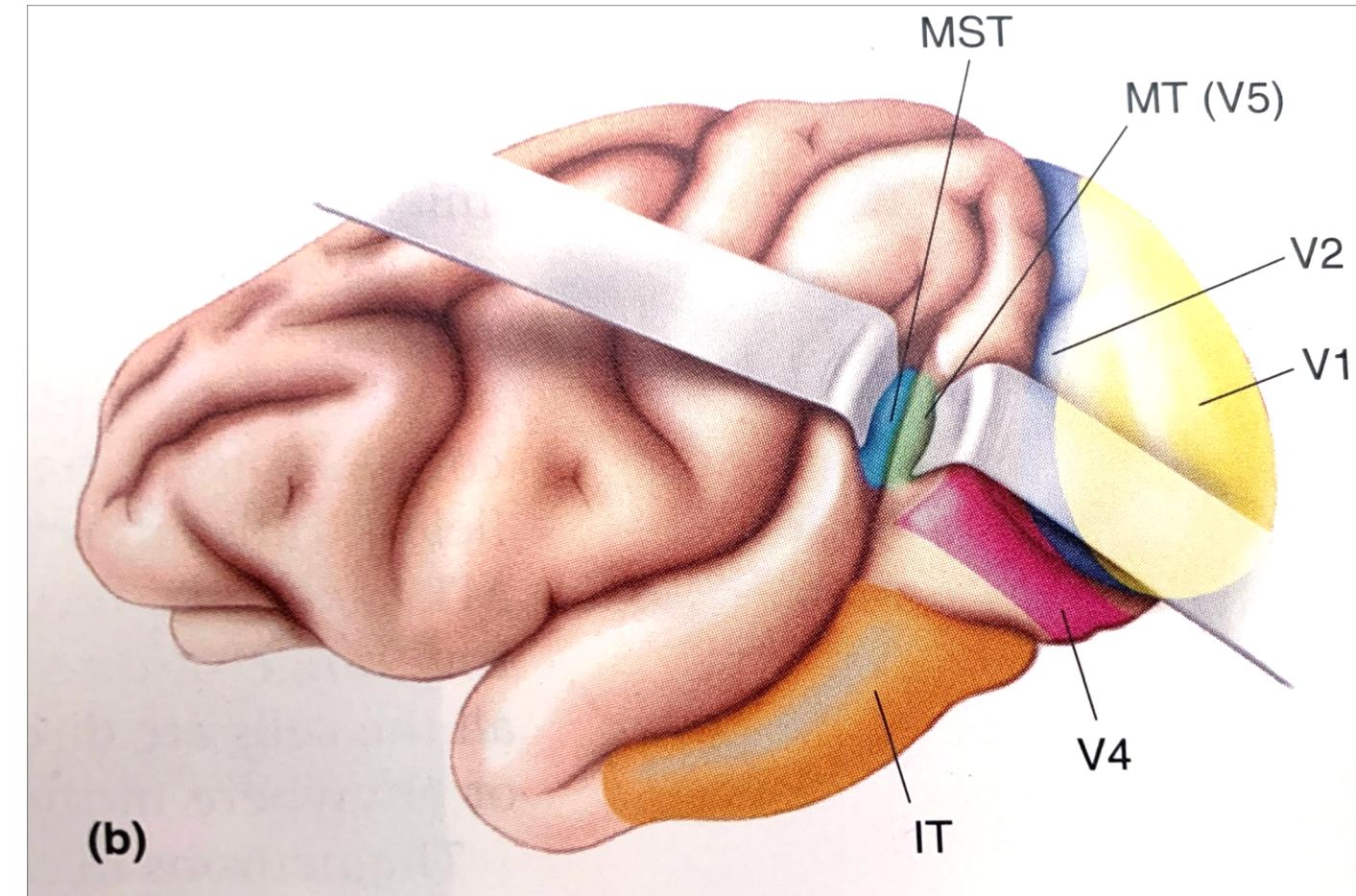
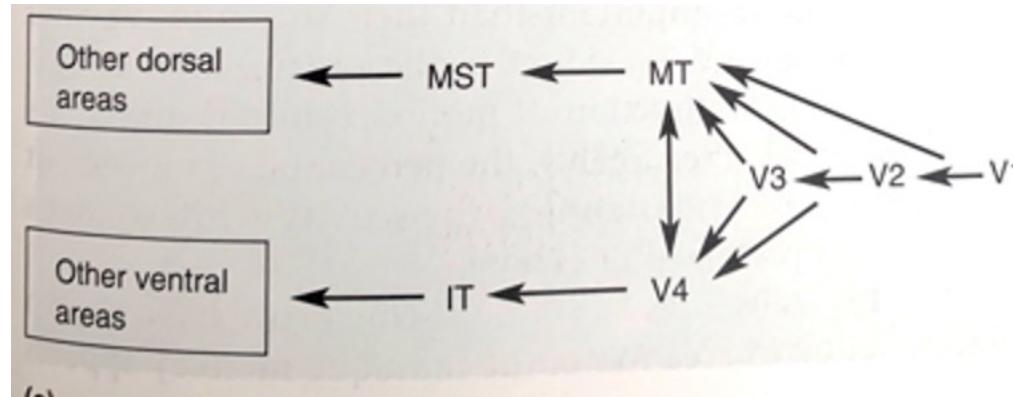


Objekterkennung beim Menschen: In V1 Aufteilung in verschiedene Aufgaben



Objekterkennung beim Menschen: In V1 Aufteilung in verschiedene Aufgaben

- Mittlere Schichten der Bildverarbeitung im Gehirn werden (V1 / V2) gelegentlich als **Faltung mit einer Gauss Funktion („Gabor“)** angesehen
- Objektbewegungen: MT (V5)
- Objektform: V2
- Objektform & Farbe: V4
- Komplexe Formen: IT



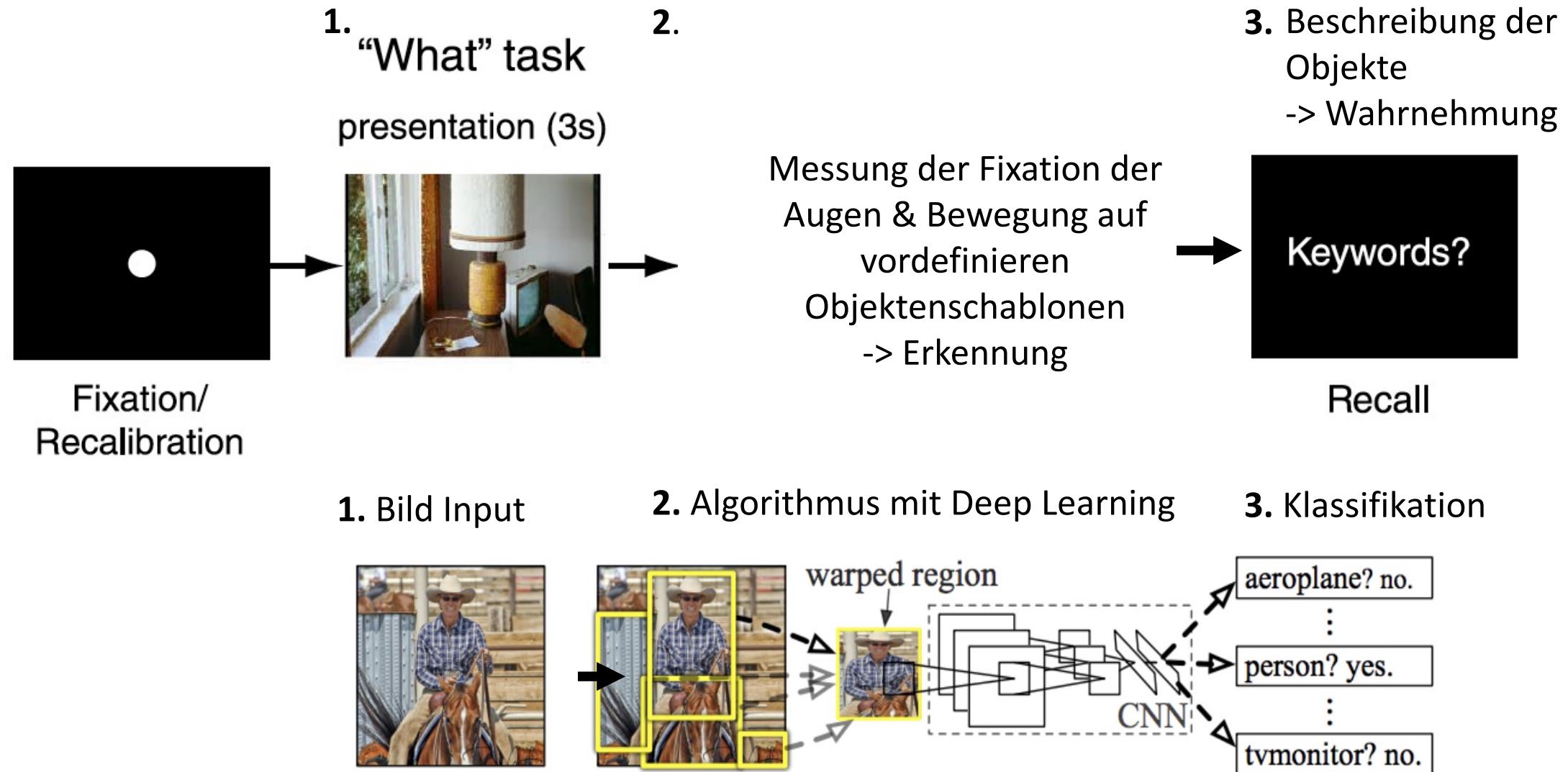
Objekterkennung LGN -> V1

- Neuronen besitzen Rezeptives Feld - Neuronen reagieren auf bestimmte Bereiche des gezeigten Bildes
- In V1 finden sich sog. „simple“ und „complex“ cells
 - Simple cells reagieren auf geradlinige Kanten und haben spezifische Orientierungen im Bezug zu ihrem rezeptivem Feld
 - Mehrere simple cells der gleichen Orientierung projizieren Information auf complex cells
- Im ML scheinen diese Merkmale auf mittlerer Ebene der Bildverarbeitung (V1) ideal für die Objekterkennung zu sein (Ullman, Vidal-Naquet, & Sali, 2002)
-> Die höheren Hirnareale (nach V2) werden bisher nicht nachgeahmt

Fazit Objekterkennung

- Sehen von Objekten wird entweder über Erkennung oder bewusste Wahrnehmung behandelt [Davidoff, 1991].
- Objekterkennung ist ein Problem der Klassifikation [Davidoff, 1991]
- Klassifikation eines Objektes im Bild ist die Aufgabe ein Label dem Objekt aufgrund des Inputs zuzuordnen
 - Es werden sog. Trainingsdaten gelernt mit vordefinierten Objekt Kategorieren
 - Generalisierung wird anhand von Testdaten überprüft
 - > Üblicherweise wird nur das Label von dem hervorstechendsten Objekt vergeben und gelernt [Li, Ma, Sajid et al., 2019]
 - > Üblicherweise steht dieses Objekt zentriert im Bild

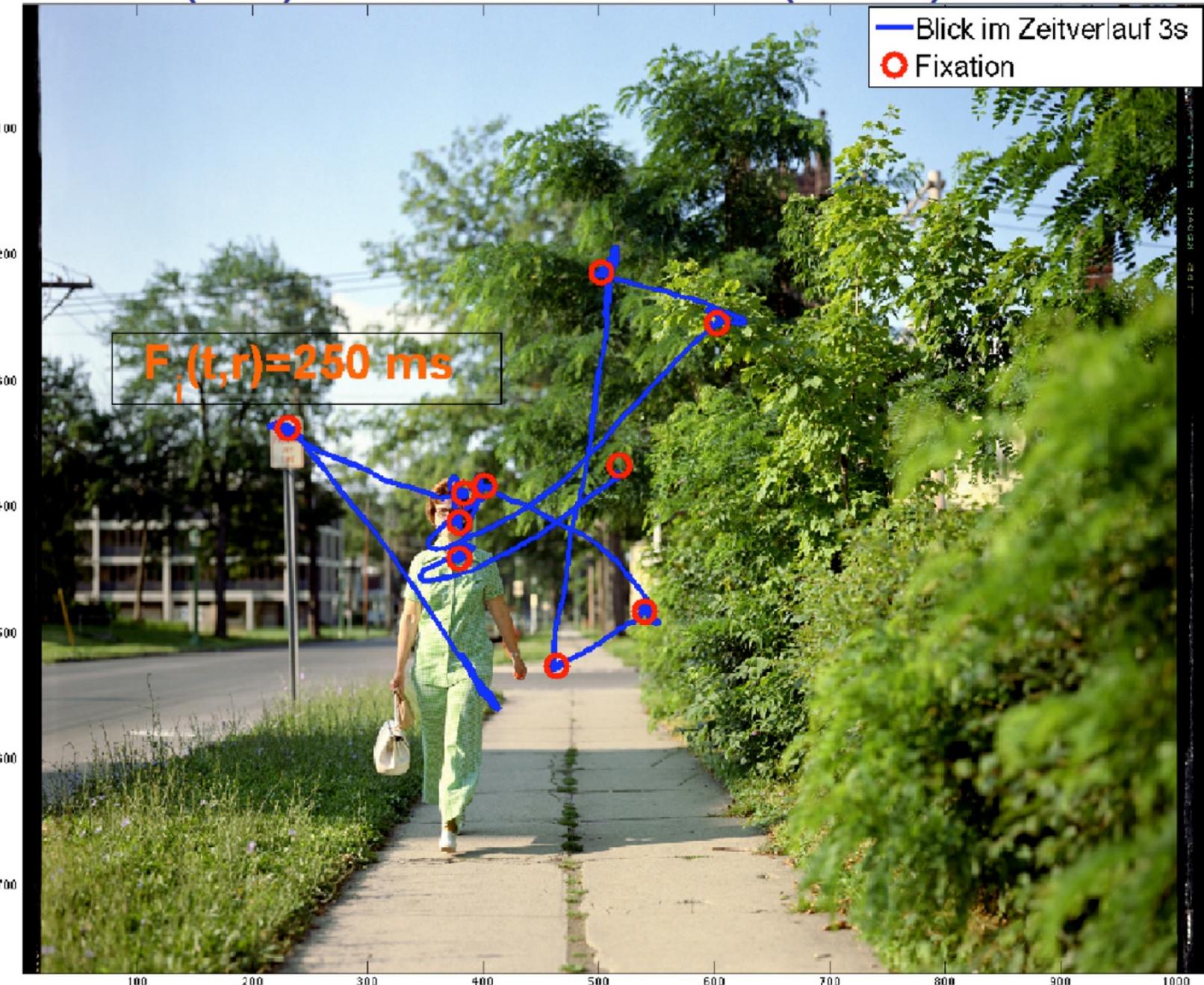
Praxis: Menschliche Objekterkennung vs Algorithmus



Beispiel Fixationen(rot) im Blickverlauf (blau)

1 Datenpunkt:
Summe der
Zeitdauern aller
Fixationen(rot) einer
Versuchsperson auf
einem Objekt einer
Farbbedingung eines
Bildes

Rechts:
Beispieldatensatz
einer
Versuchsperson



Beispiel Objekte

Objektschablonen für Fixationszuordnung:

Fixation auf einem Objekt bedeutet innerhalb der Objektschablone

5: house(3)

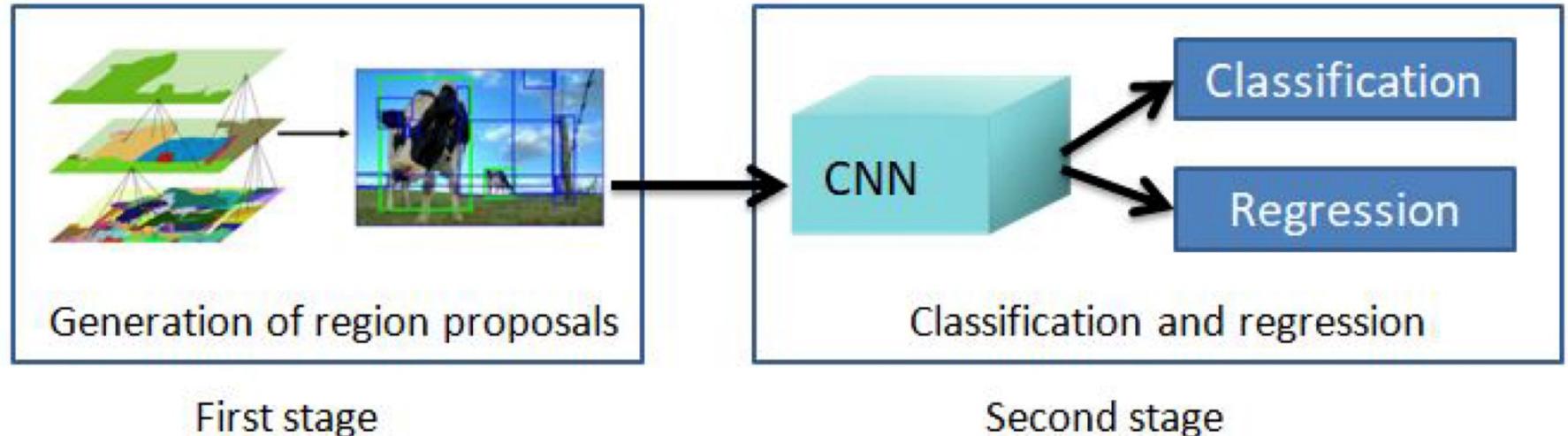


10: stair[1]

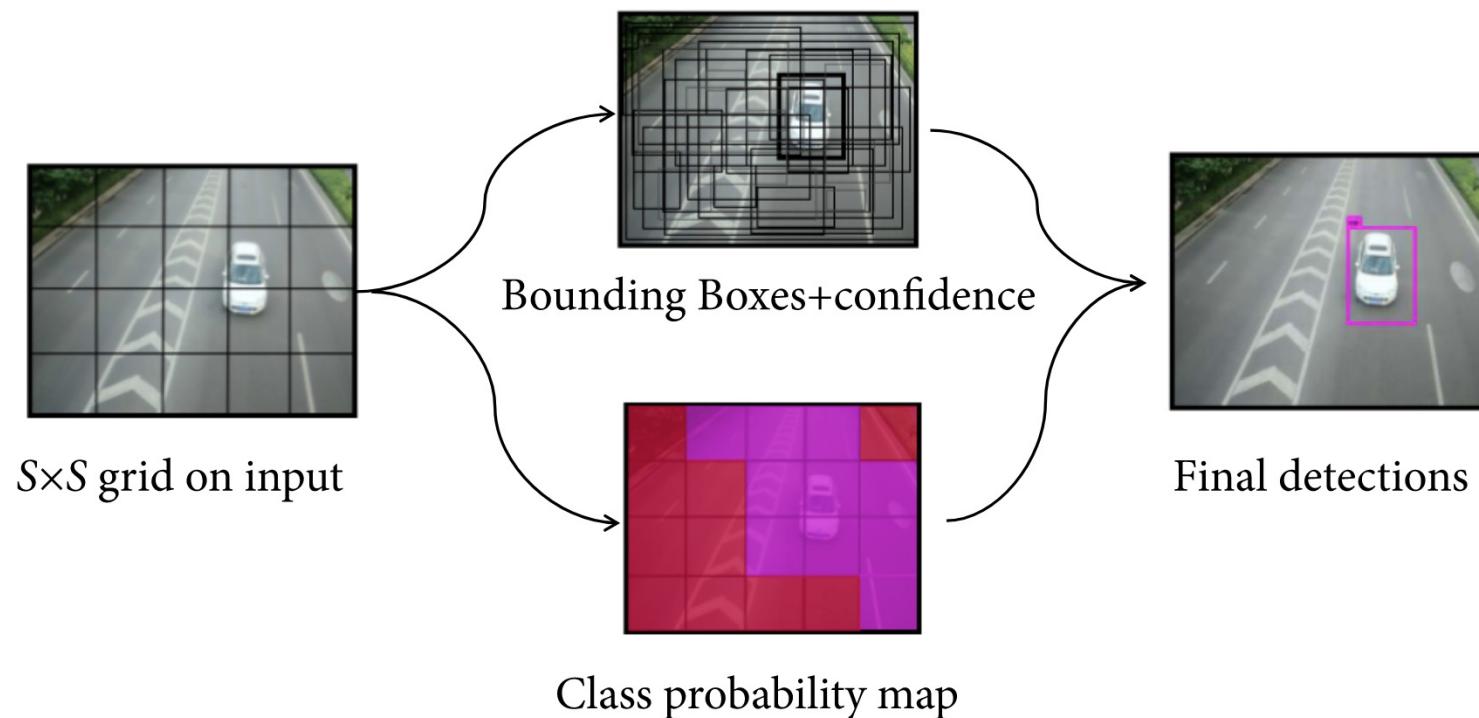


Praxis Algorithmen – Teile Bild so dass nur 1 Objekt pro Teilbild

- Two stage models



- One stage model



Beispielanwendung

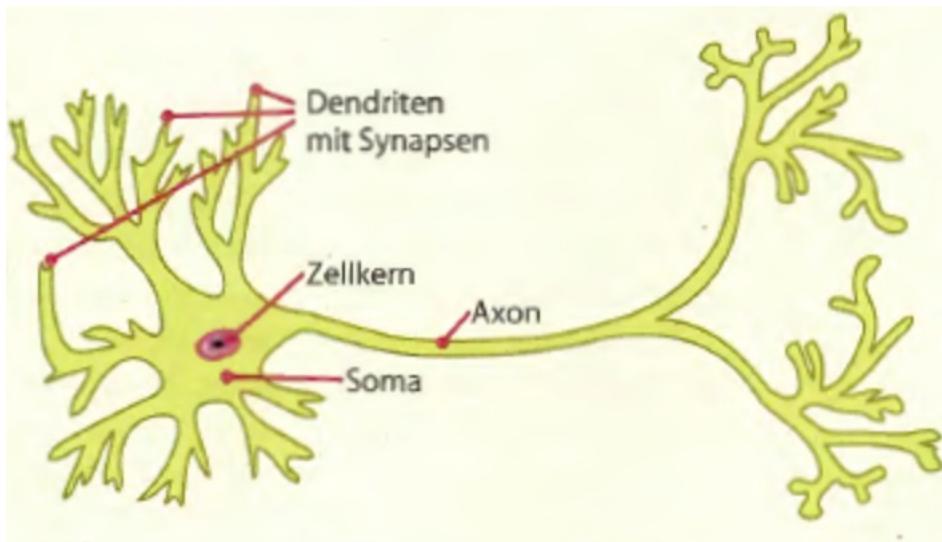
- Rmarkdown

Taxonomie der Typen von künstlichen Neuronalen Netzen [Ultsch, 1991]

Eigenschaft	Unterklassen	Wichtige Ausprägung
Neuronenmodell	Begrenzter Aktivierungsbereich	McP-Unit Logistische Unit Boltzmann-Unit Rectified Linear Units
	Unbegrenzer Aktivierungsbereich	Kohonen-Unit
Updatestrategie	Deterministisch	Synchron Parallel seriell
	Nicht-Deterministisch	
Struktur des Netzwerks (Topologie)	Vollständig verbunden	symmetrisch
	Nicht vollständig verbunden	Gerichtet geschichtet
Berechnungsmodus	Spreading activation	Feed-forward Bestmatch
	Relaxation	Simulated annealing
Lernalgorithmus	unüberwacht	Hebb'sche Regel Selbstorganisation
	überwacht	Delta-Regel Backpropagation Boltzmann Lernen

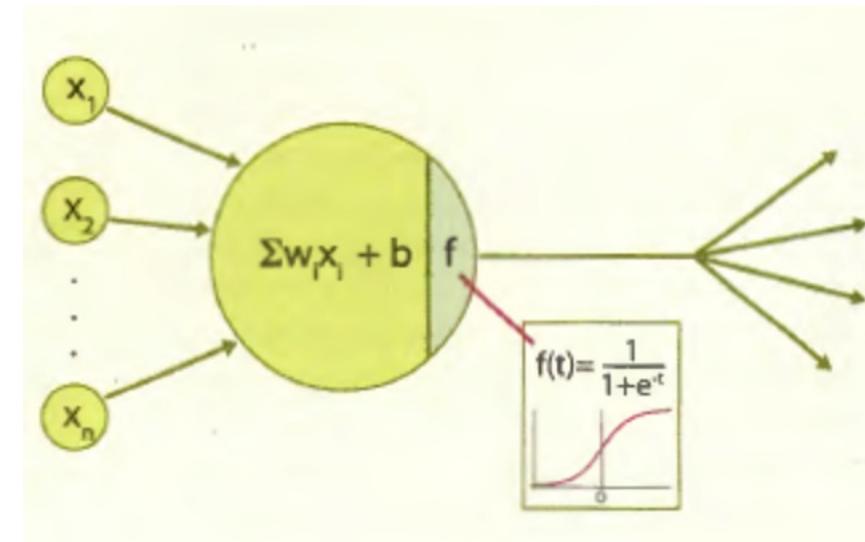
Nachahmung des Neuronenmodells

- Das Neuron nimmt Eingangssignale über seine Dendriten auf, die
 - das elektrische Signal an den Zellkörper weiterleiten
 - eine Berechnung durchführen
 - Und ein Aktionspotenzial auslösen

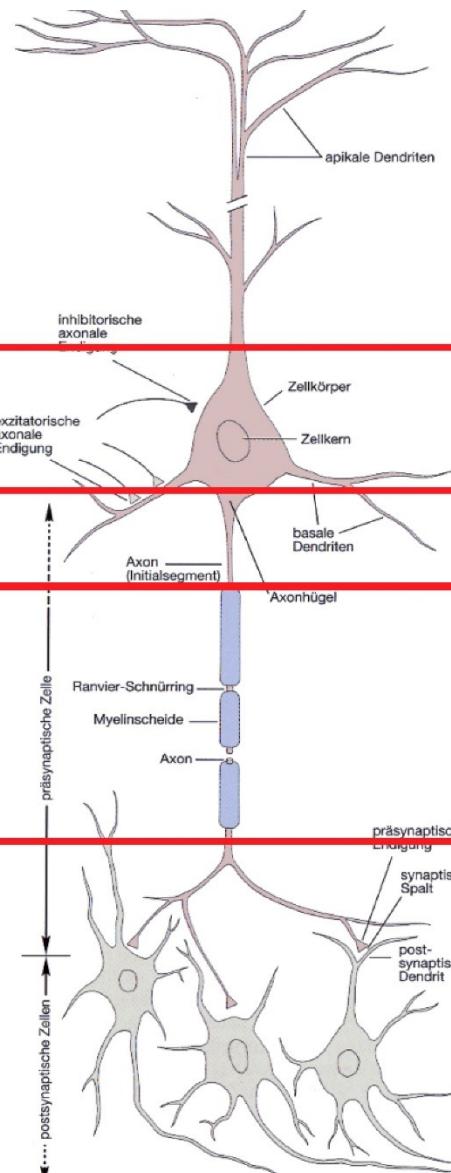


Beispiel der
„Datenbionik“:

Übertragung von
Ergebnissen
biologischer Methoden
und Systeme aus der
Natur in die Technik
zur Datenanalyse



Neuron in Reality is Highly Complex...



Input (analog) and processing

processing (non-linear integration)

“A/D conversion”

transmission (digital)
robust information transfer

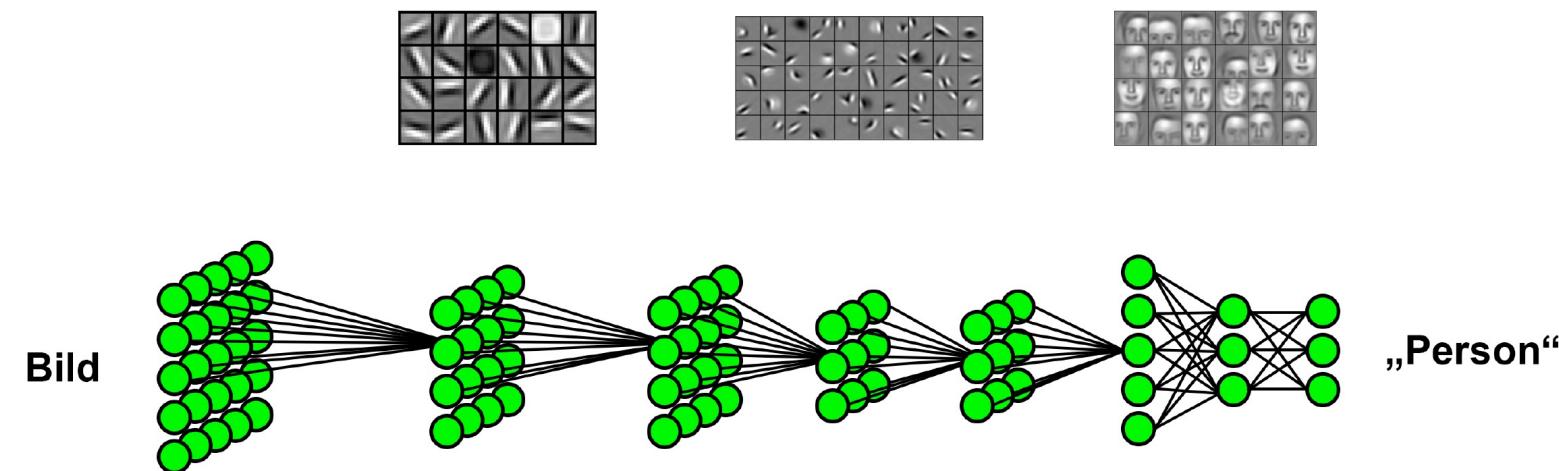
“D/A conversion”

- Dendrites: Integration of incoming signals from upstream neurons (convergence of information)
- Soma: Cell organelles, metabolic apparatus, cell nutrition, synthesis of transmitter substances, formation and degradation of vesicle membrane
- Axon: Transmission of the excitation of the cell (information) to downstream neurons (formation of the AP on the axon hill). Conducts information at up to 100 m/s
- Synapse: the axon splits into finer branches (collaterals) and thus reaches a large number of postsynaptic cells. A large number of synapses thus enable divergence of the transmitted signals
-

Taken from Neurophysics Lectures, Prof. Dr. Einhäuser-Treyer

Convolutional neural networks (CNN) [LeCun, 1989] I

- Spezielle Art von neuronalen Netzen zur Verarbeitung von Daten, die eine bekannte, gitterartige Topologie aufweisen
 - Time Series: 1 dimensional grid
 - Pixel picture: 2 D grid
- > Ausnutzung der Eigenschaft „retinotop“, d.h. jedes Neuron in jedem Layer hat rezeptives Feld
- > rezeptives Feld: festen Bereich von Eingabeneuronen, mit denen das Neuron verbunden ist, zentriert um die Position des Neurons



Convolutional neural networks (CNN) [LeCun, 1989] II

- Anstatt üblicher Matrix-Multiplikation die Faltung zweier Funktionen mit Operator $*$ verwendet
- Berechnung der Ausgabe des Neurons in einem Convolutional Layer entspricht einer Faltung (Convolution) der Eingabewerte (Werte der Eingangs-Neuronen) mit einer Maske, die sich aus den Kantengewichten ergibt
- Faltung (Convolution):
 - $f(t) = (x * w)(t) = \int x(a)w(t - a)da$
 - D.h. Multiplikation der Funktion $x(t)$ an jeder Stelle
 - **Mit einer Gewichtsfunktion $w(a)$ (Kernel),**
 - a nennt man Lag.
 - $f(t)$ wird auch als Feature (map) bezeichnet
 - für diskrete Funktionen:
 - $f(t) = \sum_{a=-\infty}^{\infty} x(a)w(t - a)$

2D Beispiel:

1	2	3
4	5	6
7	8	9

 \ast

^m _n	-1	0	1
-1	-1	-2	-1
0	0	0	0
1	1	2	1

x

w

$=$

-13	-20	-17
-18	-24	-18
13	20	17

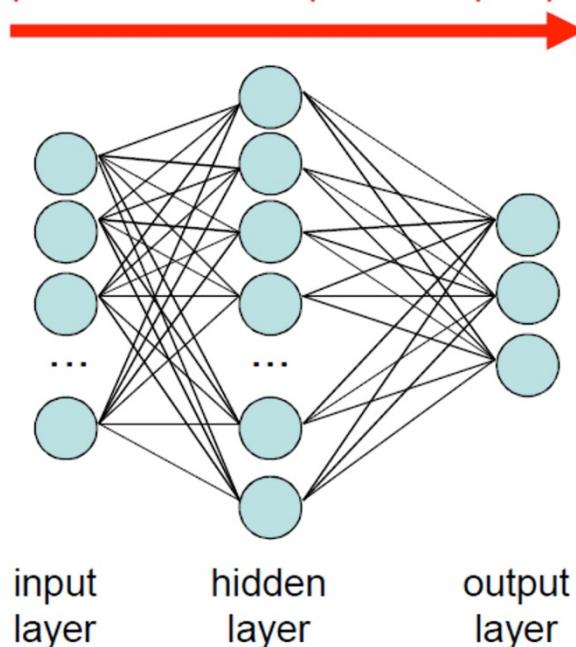
f

$\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} x_{(m-i)(n-j)} y_{(1+i)(1+j)}$

http://www.songho.ca/dsp/convolution/convolution2d_example.html

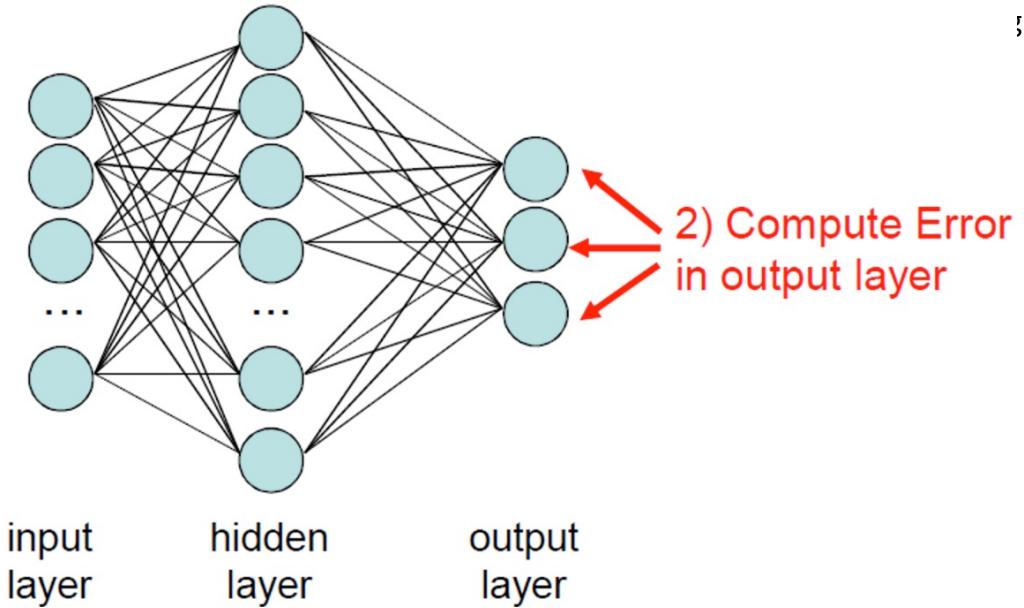
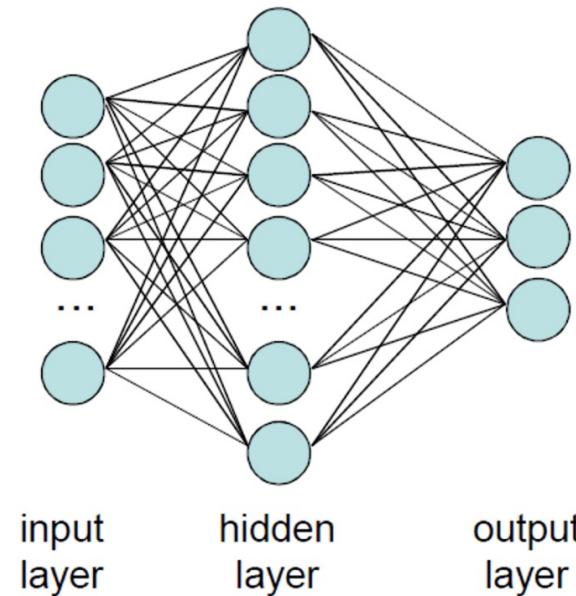
0) Initialization of the weights between the units with random values and Presentation of the input and desired output vector.

1) Compute current output for input pattern



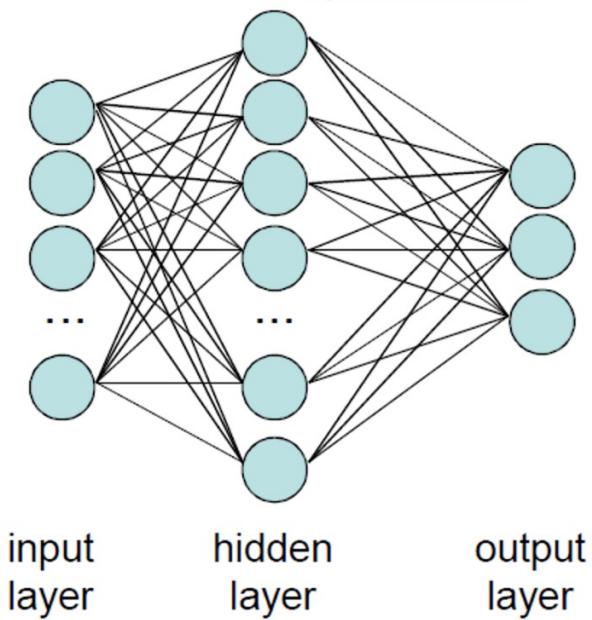
4) Now you have all the derivatives => update weights

ii) Update weights with gradient descent



3) Propagate back error

Compute Deltas of weights via „Delta Rule“



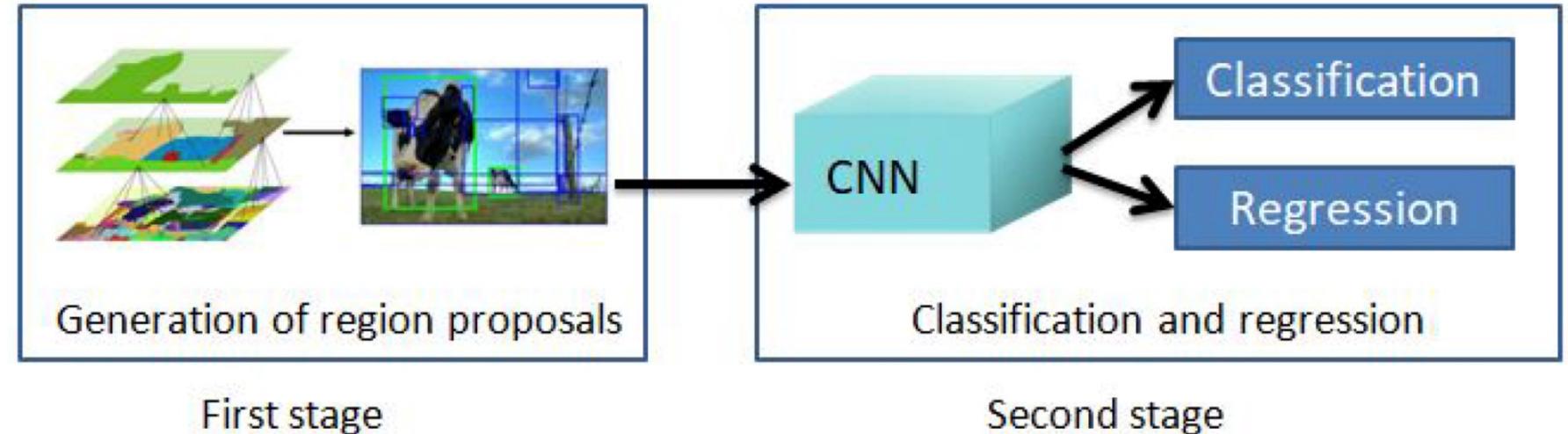
Backpropagation Prinzip

Convolutional Neural Networks (CNN)

Eigenschaft	Unterklassem	Wichtige Ausprägung
Neuronenmodell	Begrenzter Aktivierungsbereich	McP-Unit Logistische Unit Boltzmann-Unit Rectified Linear Units
	Unbegrenzer Aktivierungsbereich	Kohonen-Unit
Updatestrategie	Deterministisch	Synchron Parallel seriell
	Nicht-Deterministisch	
Struktur des Netzwerks (Topologie)	Vollständig verbunden	symmetrisch
	Nicht vollständig verbunden	Gerichtet geschichtet
Berechnungsmodus	Spreading activation	Feed-forward Bestmatch
	Relaxation	Simulated annealing
Lernalgorithmus	unüberwacht	Hebb'sche Regel Selbstorganisation
	überwacht	Delta-Regel Backpropagation Boltzmann Lernen
THRUN		

Erkennungstypen in der Praxis – Teile Bild so dass nur 1 Objekt pro Teilbild

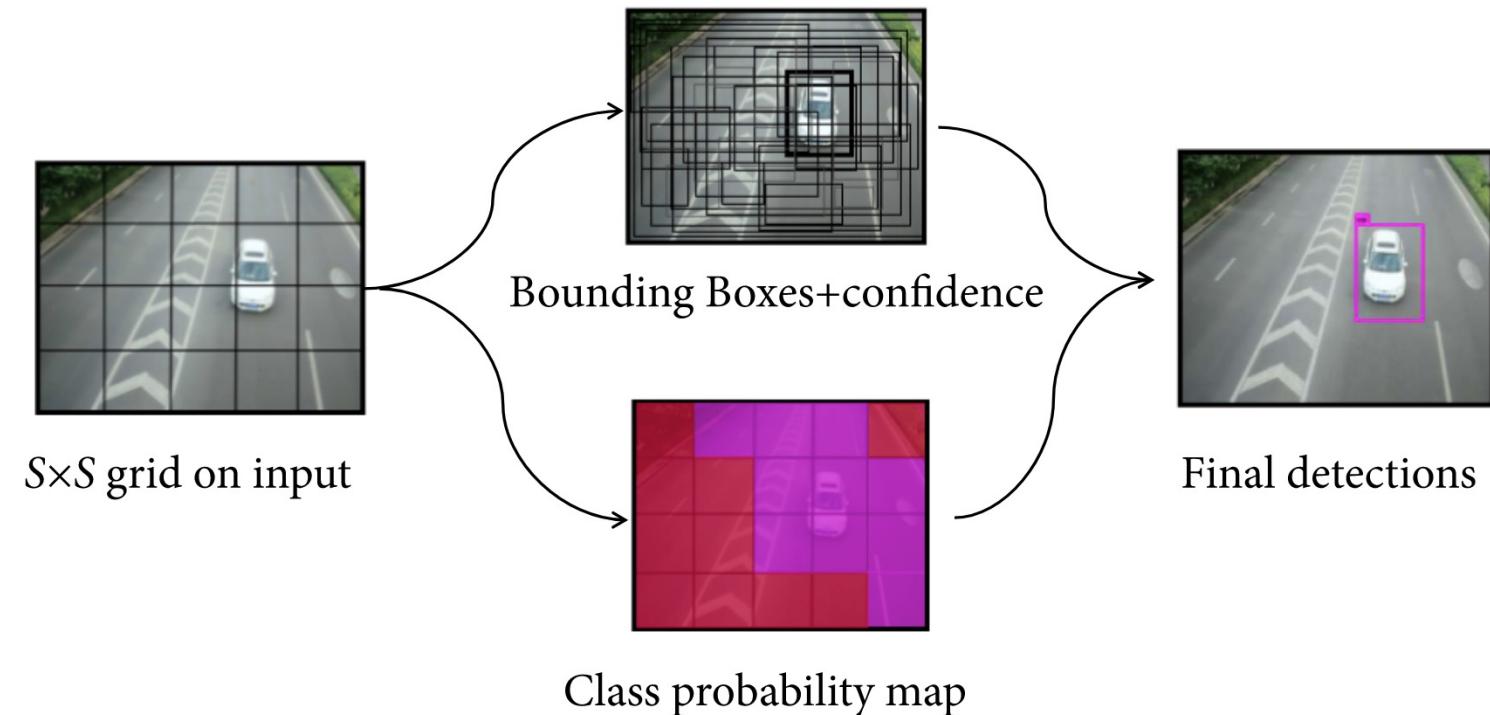
- Two stage models



First stage

Second stage

- One stage model



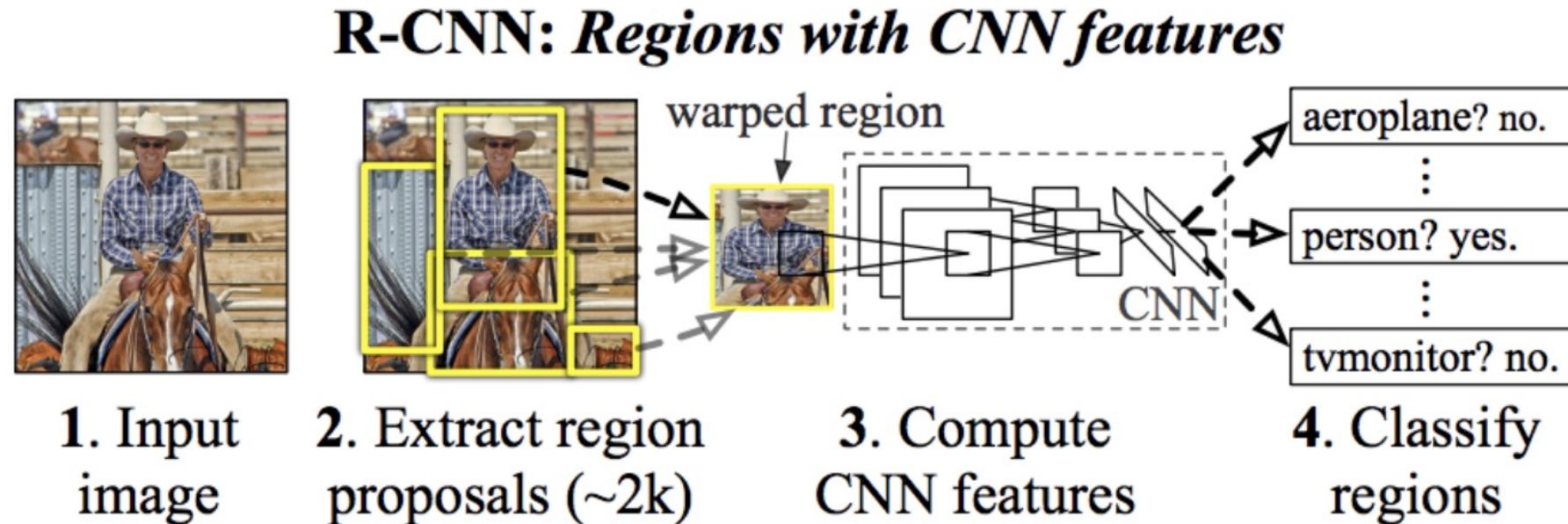
S \times S grid on input

Class probability map

Final detections

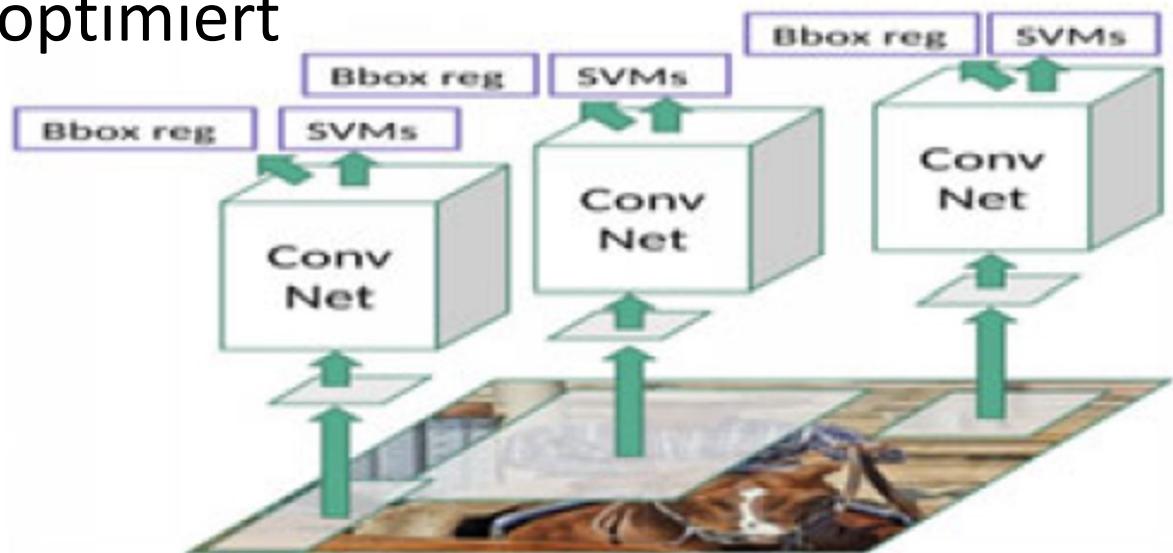
Region Based Convolutional Neural Networks (RCNN) - a two stage model

- Stufe 1: Selective Search zur Ermittlung von Regionen von Interesse ("Proposal") in Analogie zu „Fixationen“
- Stufe 2: Ein CNN pro Vorschlag zur Feature Extraction der Region



Details of Region-based CNN

1. Lokalisierungen von Objektpositionen durch Regionsvorschläge mit “selected search” [Uijlings et al., 2013]
 - 2000 Regionen werden unabhängig von Objektkategorie und Größe vorgeschlagen
2. Feature Extraction 4096 dim pro Region mit Cafee (CNN) [Girshick et al, 2014]
 - Pro Region ein CNN mit 5 convolutional und 2 fully connected layers
 - Damit wird ein linearer SVM pro Objektklasse trainiert [Girshick et al, 2015]
 - Bounding Boxes werden im Regression optimiert
 - R-CNN [Girshick et al 2014, 2015]
 - Extension Faster R-CNN
 - Extension Mask R-CNN



Fazit Objekterkennung I

- Typischerweise kann man zwei unterschiedliche Ansätze verfolgen
 - Representative Learning
 - „Nachahmung“ der Menschlichen Objekterkennung [Davidoff, 1991] mit deep learning [Goodfellow et al., 2015]
 - Classic Machine Learning
 - Auf Manueller Feature Extraction basierende Ansätze [Nixon&Aguado, 2012]

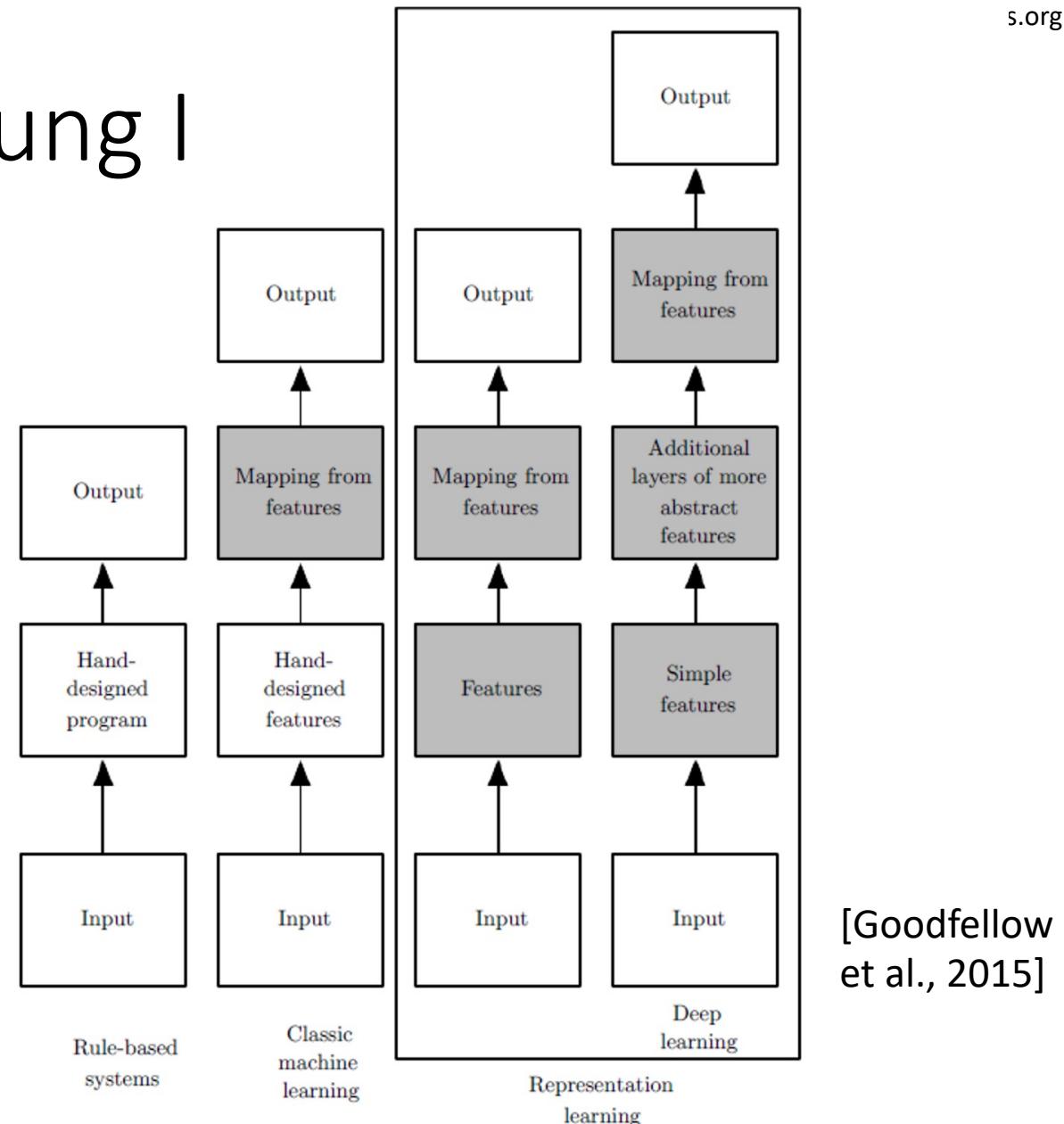


Figure 1.5: Flowcharts showing how the different parts of an AI system relate to each other within different AI disciplines. Shaded boxes indicate components that are able to learn from data.

Fazit Objekterkennung I

- Typischerweise kann man zwei unterschiedliche Ansätze verfolgen
 - Representative Learning
 - „Nachahmung“ der Menschlichen Objekterkennung [Davidoff, 1991] mit deep learning [Goodfellow et al., 2015]
 - Classic Machine Learning
 - Auf Manueller Feature Extraction basierende Ansätze [Nixon&Aguado, 2012]

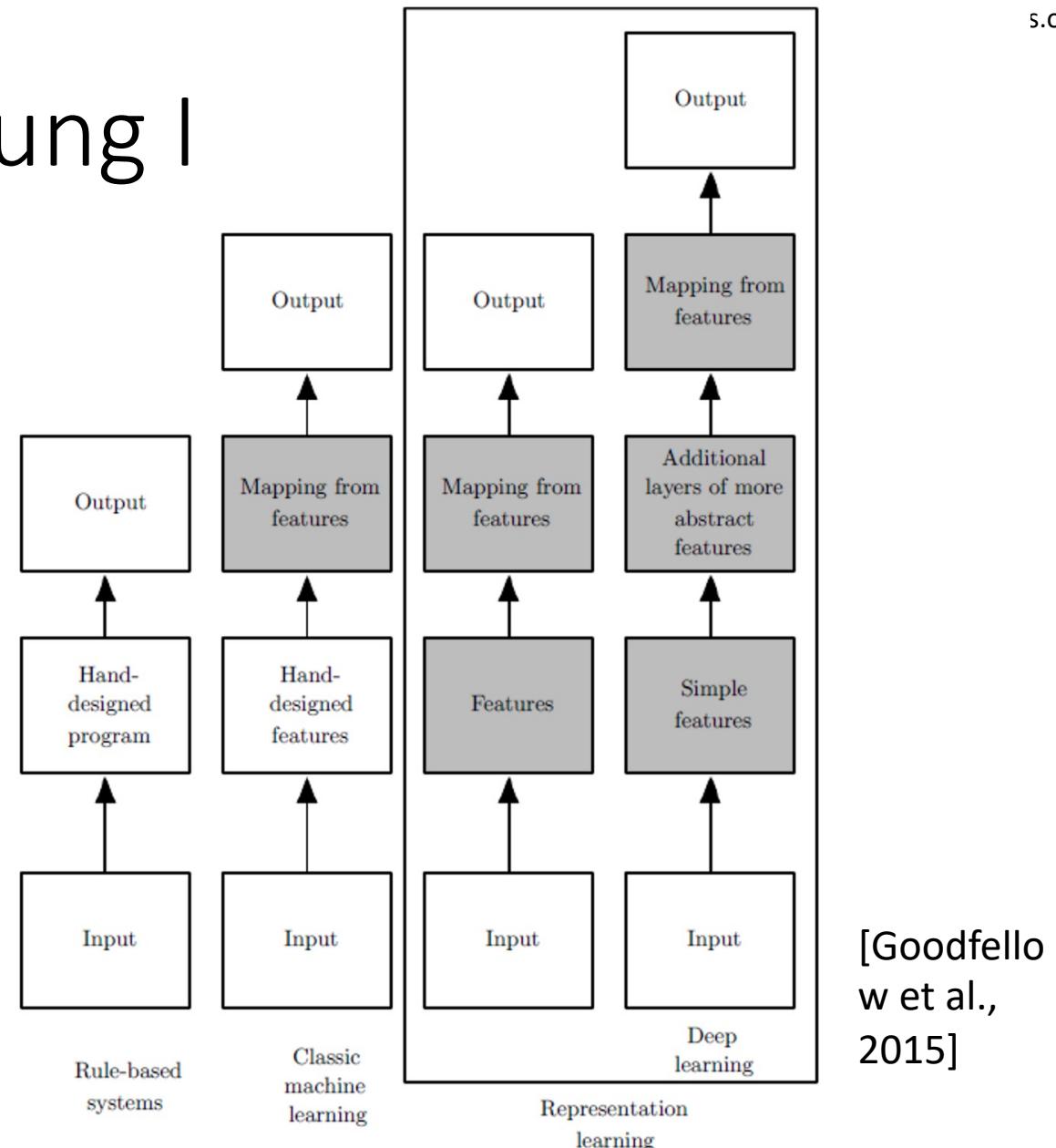
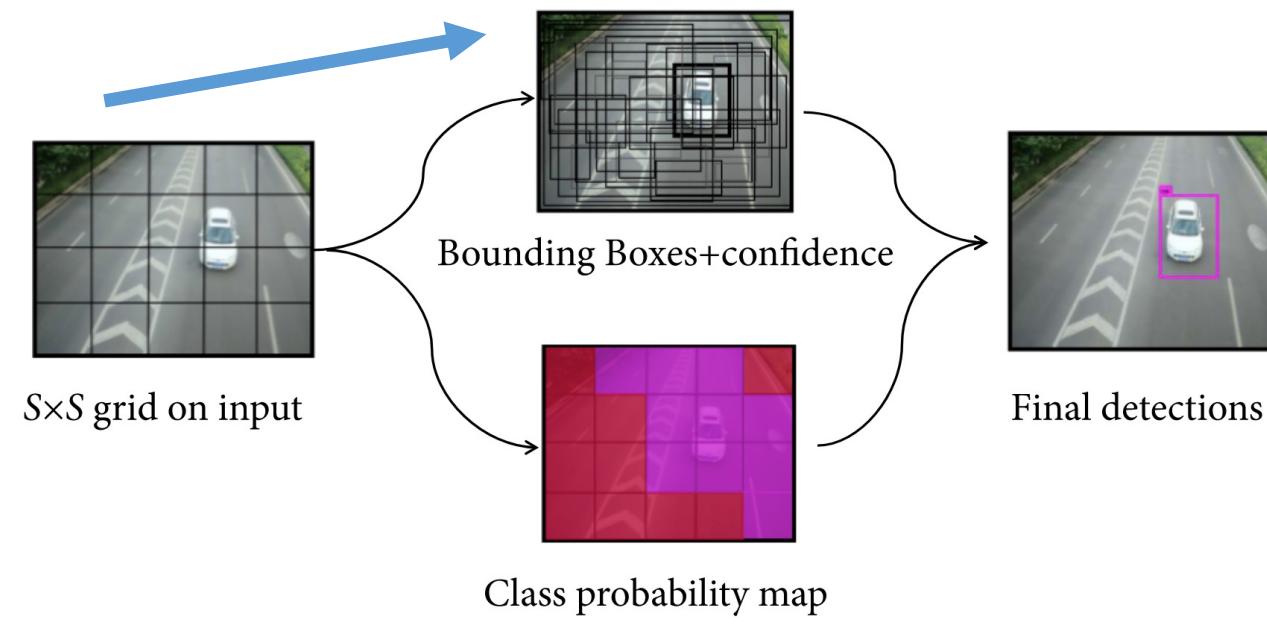


Figure 1.5: Flowcharts showing how the different parts of an AI system relate to each other within different AI disciplines. Shaded boxes indicate components that are able to learn from data.

YOLO [Redmon et al., 2015]

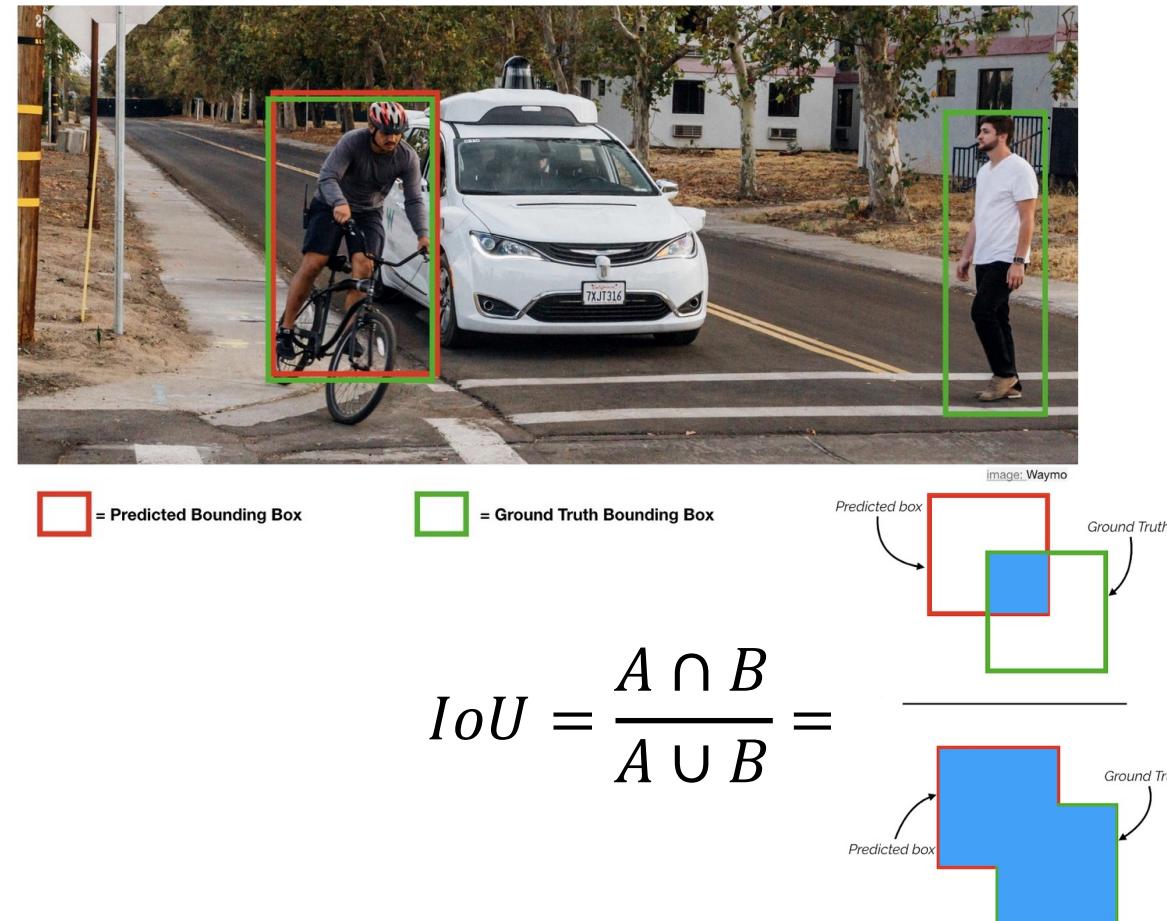
- YOLO sucht nach Teilen des Bildes, die mit hoher Wahrscheinlichkeit ein Objekt enthalten
 - Feature Extraction eines CNN des ganzen Bildes wird für Bestimmung aller Bounding Boxes gleichzeitig verwendet
- YOLO nimmt zunächst ein Bild auf und unterteilt es dann in $S \times S$ -Gitter.
 - Für jedes einzelne Raster S werden M Bounding Boxes gezeichnet und Konfidenzniveaus berechnet

$$K = \text{Pr}(Object) * IoU$$



Qualitätsmaße IoU

- “Intersection over Union (IoU)” bestimmt die Überdeckung der Fläche des **erkannten Segments A** mit dem korrekten Segment der **Ground-Truth Annotation B**

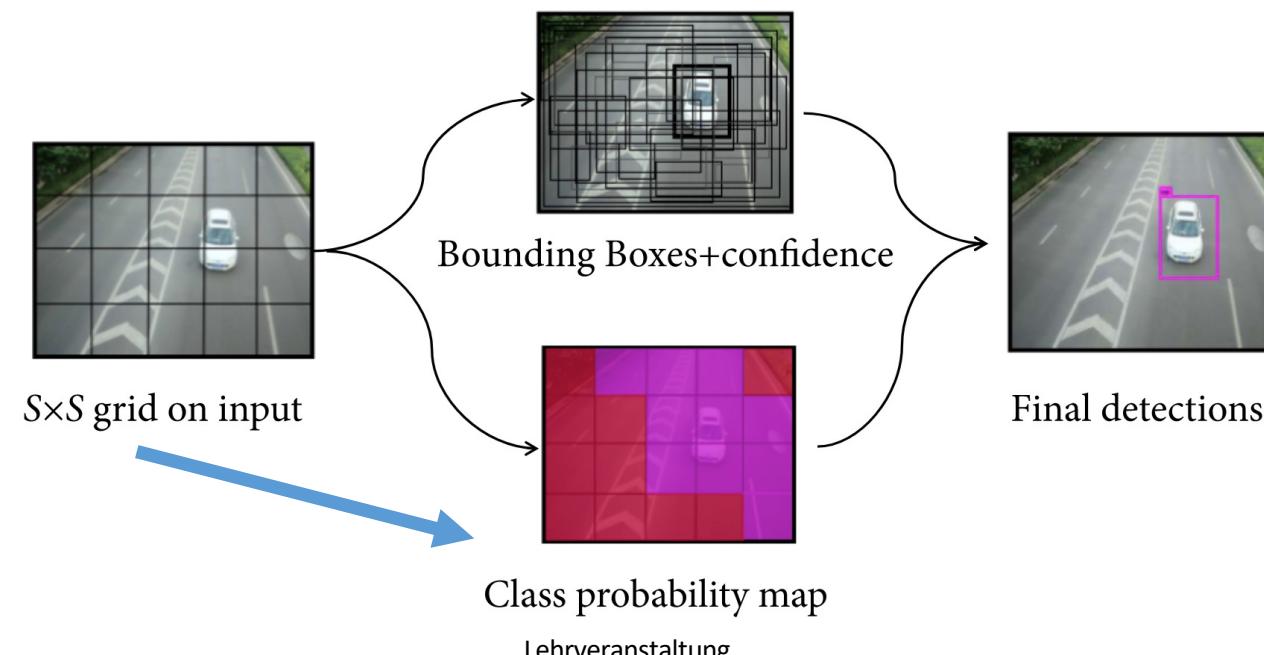


YOLO [Redmon et al., 2015]

- Für jeden Begrenzungsrahmen werden bedingte Klassenwahrscheinlichkeit berechnet

$$c = \Pr(\text{Object class} | \text{Object})$$

- Ein Begrenzungsrahmen (pink) wird ausgewählt, wenn seine Klassenwahrscheinlichkeit größer ist als
 - Klassenwahrscheinlichkeit der anderen Begrenzungsrahmen
 - Schwellenwert für die Klassenwahrscheinlichkeit;
- Andere Begrenzungsrahmen werden verworfen (rot)

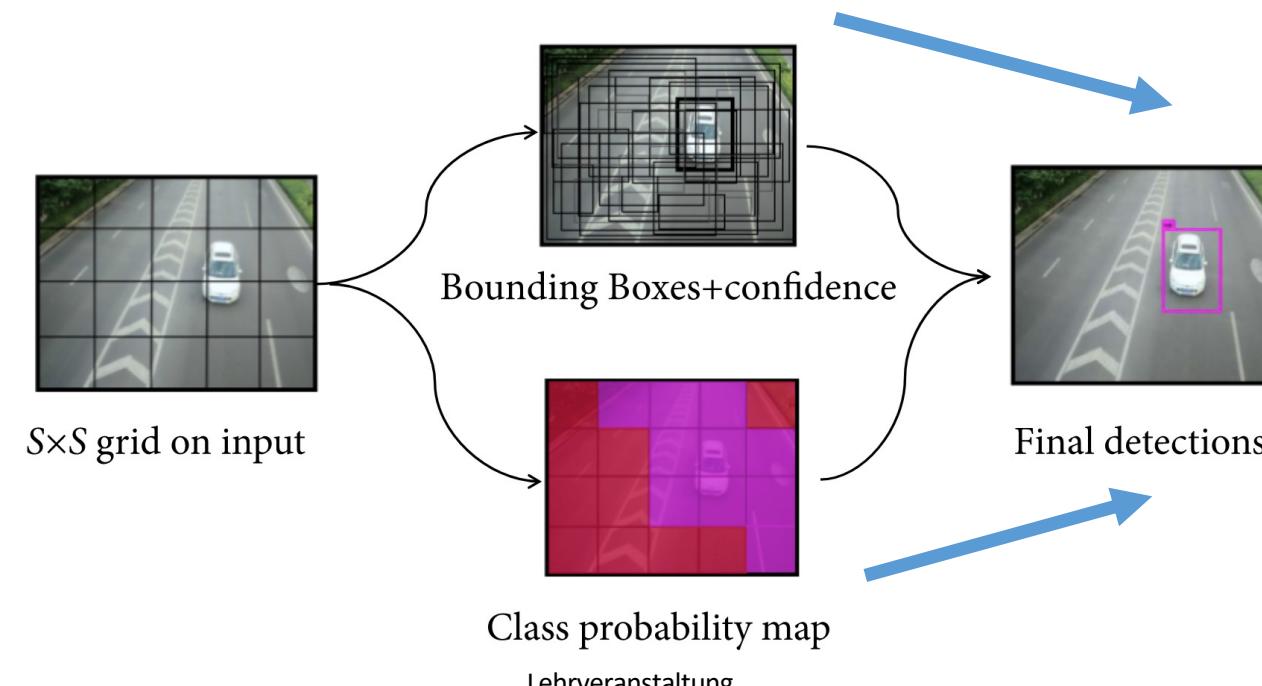


YOLO [Redmon et al., 2015]

$$K = \text{Pr}(Object) * IoU$$

$$c = \text{Pr}(Objectclass|Object)$$

- Der Score $K*c$ wird pro Bounding Box bestimmt und kodiert sowohl die Wahrscheinlichkeit dass sich die Objektklasse in Bounding der Box auftaucht wie auch wie gut die Bounding Box sich mit der Objektfläche überschneidet
- Die ausgewählte Bounding Box entspricht dem Objekt in einem Bild mit höchstem $K*c$



YOLO – CNN Details

- 24 CNN layer+2 fully connected layers

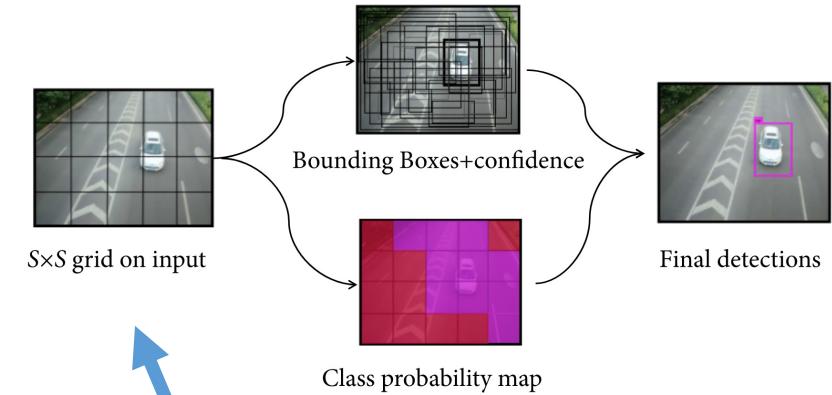
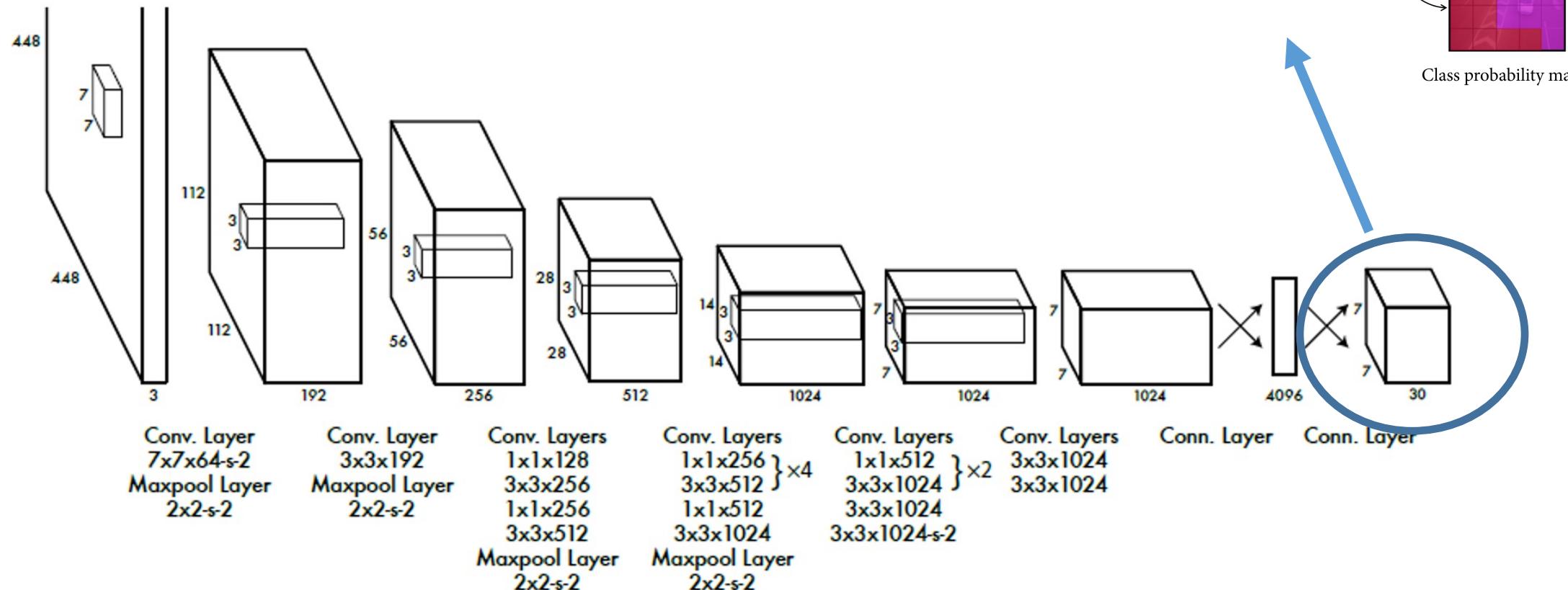


Figure 3: The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1×1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224×224 input image) and then double the resolution for detection.

Typische Datensätze für Training und Test

- PASCAL VOC Challenge (diverse Jahre) z.B. [Everingham et al., 2019]
 - 2012: 20 Objektklassen, 11.530 Bilder
- MS coco [Lin et al., 2014]
 - 330k Bilder
- SUN challenge
- VisDrone-DET2018 [Zhu et al., 2018]
 - 8599 Bilder von Dronen mit kleinen und dichten Objekten
- ImageNet- Large Scale Visual Recognition Challenge (ILSVRC) [Russakovsky et al., 2015]

⇒ Die Trainingsdaten definieren und schränken damit die Objektklassen ein, welche erkannt werden können!

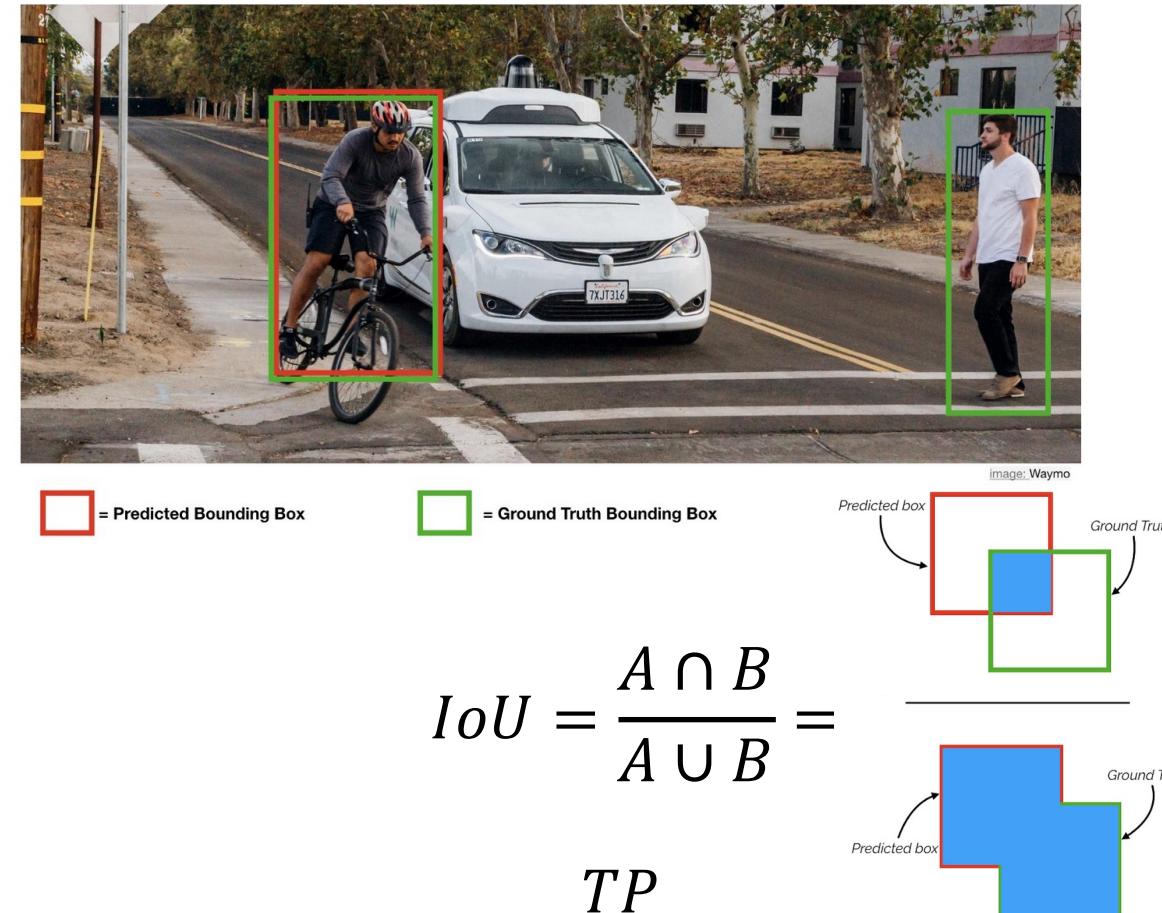
⇒ Typische Datensätze wie MS coco liefern auch eigene Qualitätsmaße (QMs) mit speziellen Bias

⇒ QM Bias kann für eigene Anwendungen von Nachteil sein [Thrun, 2021, Nature Scientific Reports]

Qualitätsmaße I

- *Klassifikationsfehler*
- “Intersection over Union (IoU)” bestimmt die Überdeckung der Fläche des **erkannten Segments A** mit dem korrekten Segment der **Ground-Truth Annotation B**
- Statt Flächeninhalt auszurechnen kann Berechnung bei Rasterbildern auch mithilfe der Bewertung der Pixel bestimmt werden
 - TP: True positives
 - FP: False positives
 - FN: False negatives

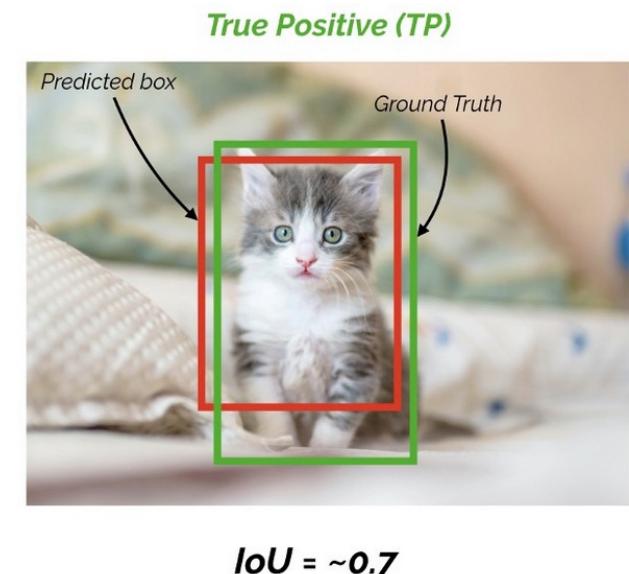
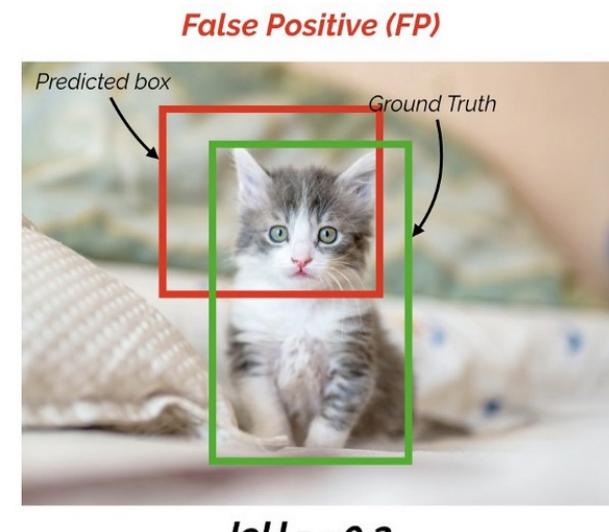
$$E = \frac{\# \text{ fehlerhaft klassifizierte Objekte}}{\# \text{ Objekte insgesamt}}$$



Qualitätsmaße II: Pixelbewertung durch P&R

- Zur Evaluation von Objekterkennung berechnen wir Precison & Recall (P&R) anhand des IoU-Wertes für einen bestimmten IoU-Schwellenwert.
 - Liegt der IoU-Schwellenwert beispielsweise bei 0,5 und der IoU-Wert für eine Vorhersage bei 0,7
 - Dann klassifizieren wir die Vorhersage als True Positive (TF)
 - Liegt der IoU-Schwellenwert dagegen bei 0,3
 - Wird die Vorhersage als Falsch positiv (FP) eingestuft.

=> Abhängigkeit von P&R durch IoU-Schwellenwert



Qualitätsmaße III: mAP

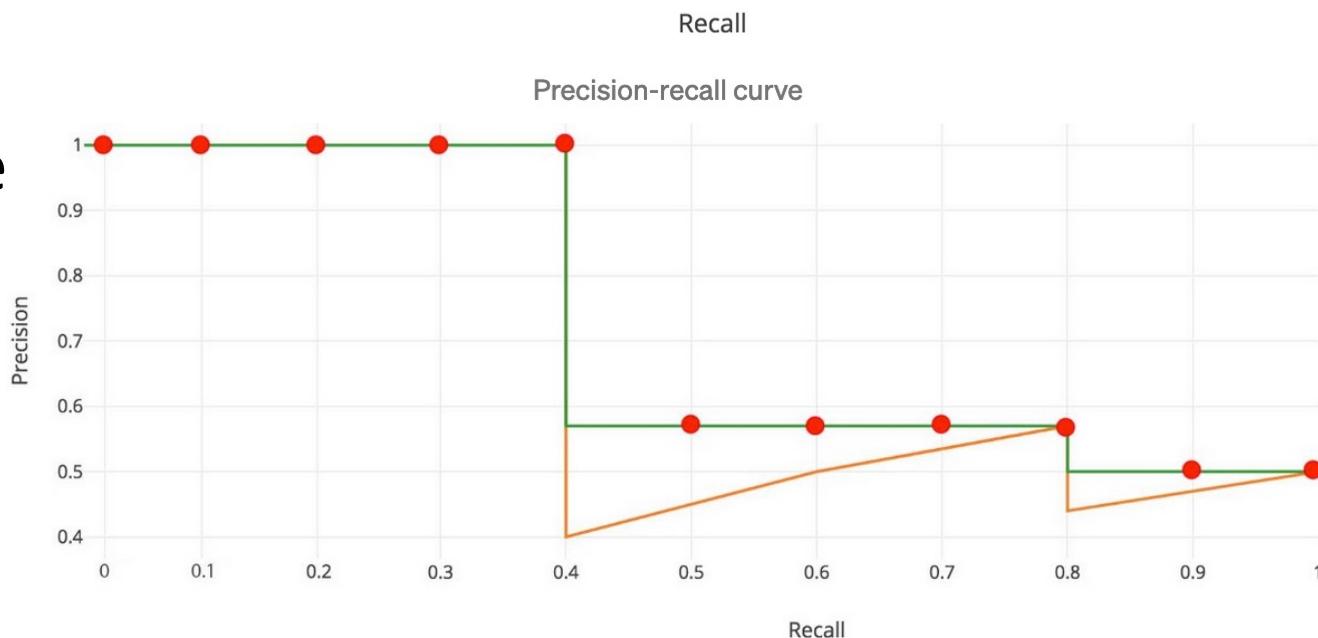
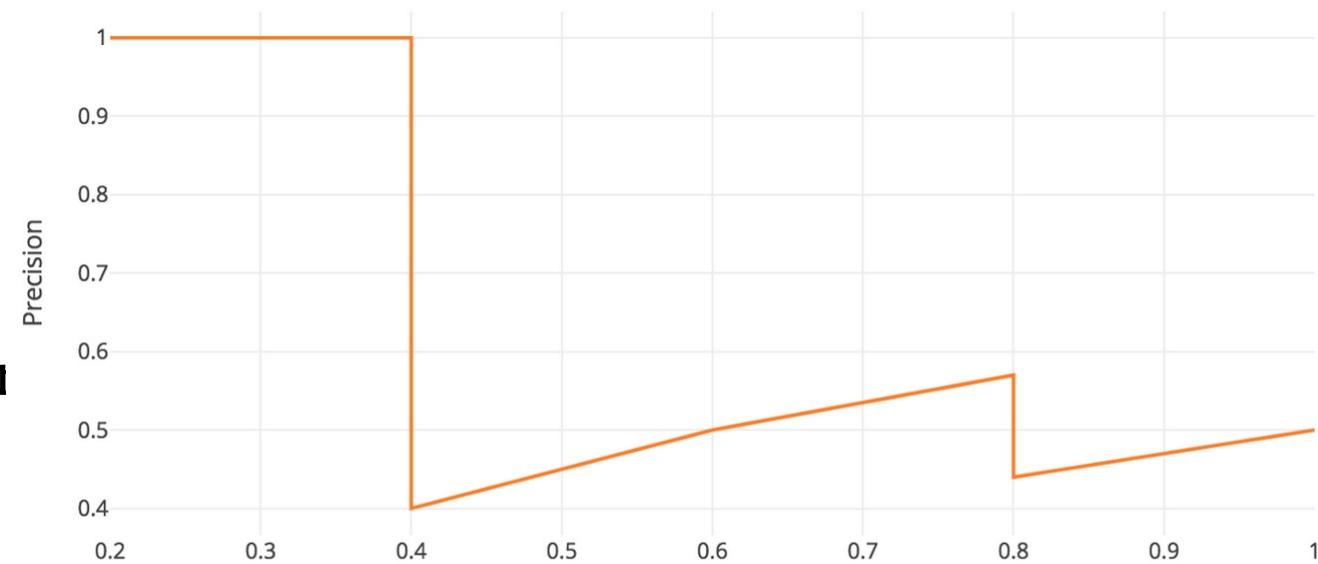
- mAP ist der MW des Maximums der Precisions an verschiedenen Recall werten [Bharati/Pramanik, 2009]
 - Benötigt vom Datensatz abhängige Definition der Precision Recall Kurve
 - Berechnet sich über Precision und Recall jedes Detektions Ereignisses
 - Precision ist Anteil korrekt erkannter Objekte und allen erkannten Objekten
 - Recall ist der korrekte erkannte Anteil und allen Objekten (der Annotation)

$$P(\text{Objekt}) = \frac{TP}{TP+FP}$$

$$R(\text{Objekt}) = \frac{TP}{TP+FN}$$

Precision Recall Kurve

- Average Precision (AP) is the Fläche unter der Precision Recall Kurve
 - Kurve basiert auf einem Ranking der Objekte geordnete nach Recall
 - Kurve wird manchmal interpoliert
- $AP = \int_0^1 p(r)dr$
- mAP score ist der MW über alle alle Klassen oder alle IoU Schwellenwerte in einem komplexen und variierenden Berechnungsverfahren
 - Formel ändert sich je nach Datensatz



Vergleich YOLO vs R-CNN

- Viele Abwandlungen dieser beiden Ansätze
 - Üblicherweise Trade-off der Accuracy vs Geschwindigkeit [Bharati/Pramanik, 2009; Li et al, 2019]
 - Evaluierung üblicherweise abhängig von diversen Experimententscheidungen [Sultana et al., 2019]
 - Feature extractors (Basis ANN), Bildauflösung, IoU threshold, bounding box encoding, training dataset, parameter configurations,
- R-CNN
 - 1 Bilder pro 2 Sekunde (0.5 FPS)
 - Hohe Genauigkeit (70% mAP, Pascal VOC 2007)
- YOLO
 - 45 Bilder pro Sekunde
 - Genauigkeit geringer i.A. (63.4% mAP, Pascal VOC 2007)
 - Findet aber manchmal mehr Objekte wie R-CNN [Sumit et al. 2019]

[Sumit et al. 2019]

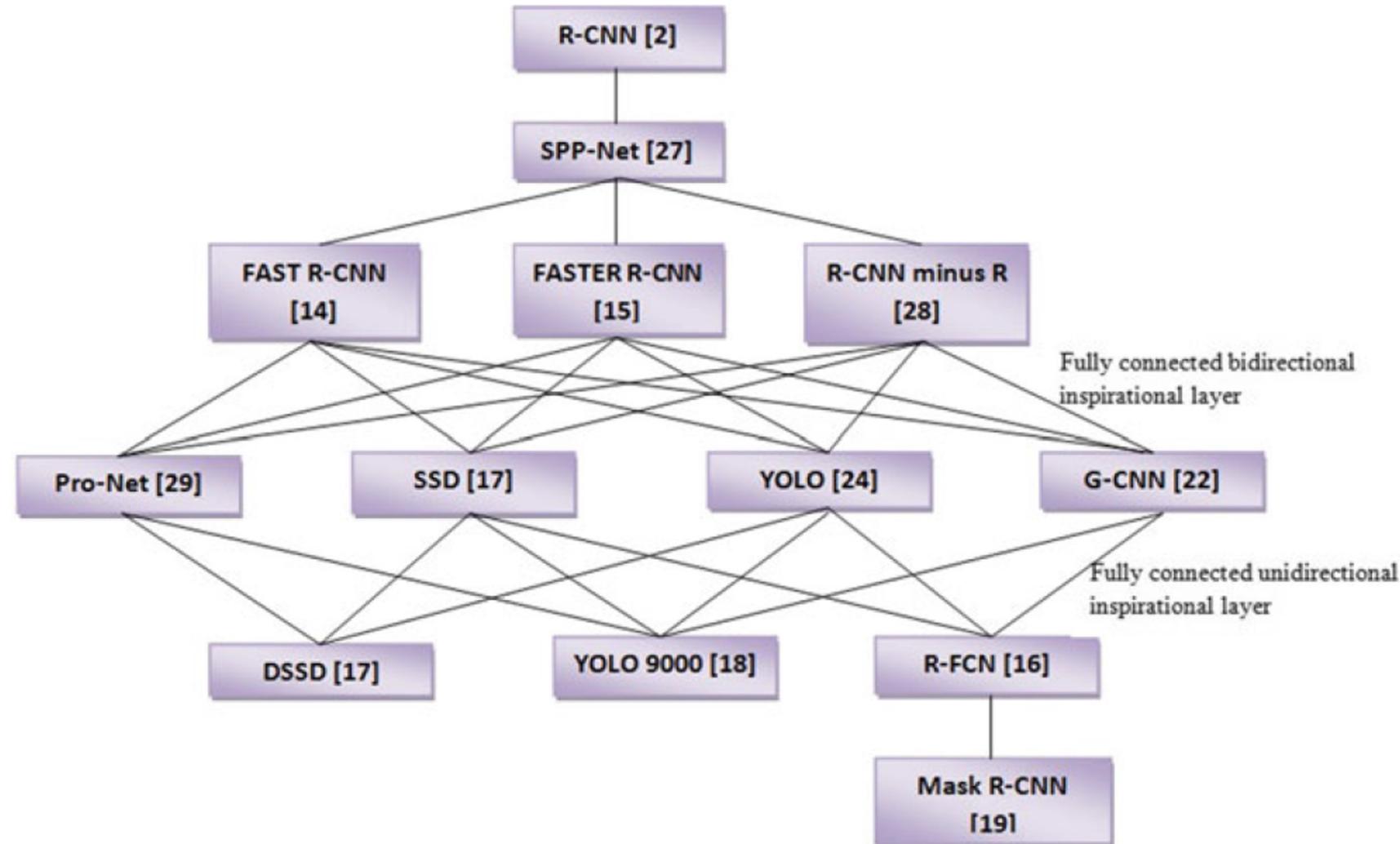
- Oben: Mask R-CNN erkennt Menschen nicht
- Unten Yolo erkennt alle Menschen



- Zerlegung in Regionen hat bei Mask-R-CNN nicht gut funktioniert

Ausblick & Schlussfolgerungen

- Viele Unterkategorien
- Trade-off Accuracy <-> Speed
- Wahl des Datensatzes und QMs ist kritisch für Performanz
- Two Stage models -> Bessere Genauigkeit
- One Stage Model -> schnellere Berechenbarkeit



Zusammenfassung I: Datenbionik könnte Lücken schließen und Methoden verbessern

- Datenbionik lehrt die Übertragung von Ergebnissen biologischer Methoden und Systeme aus der Natur in die Technik zur Datenanalyse
- Datenbionische Konzepte für Rechnergestützte Objekterkennung (RO) sind
 - Retinotopie und Schichtaufbau des Neuronalen Netzes
 - Einfache Simulation eines Neurones und nur Simulation mittleren Hirnstrukturen durch Deep learning
 - Oberflächenbasierte Objekterkennung

=> Viele Lücken in der jetzigen Technik könnten in interdisziplinärer Forschung geschlossen werden

Zusammenfassung II: Objekterkennung durch Deep Learning (ODL) eher für spezialisierte Anwendungen nützlich

- ODL Algorithmen teilen Bild in Regionen und klassifizieren 1 Objekt pro Region in der Hoffnung dass
 - Nur 1 Objekt in Region existiert
 - Objektklasse durch Trainingsdaten bekannt ist
- GIGA (Garbage in Garbage out): Prinzip bei automatischer Feature Extraction
 - Rmarkdown Beispiel zeigt den Fallstrick der ODL
 - Simple Änderung der Eingabedaten führt zu riesigen Fehlern weil Netz nicht versteht was es tut
=> Es werden keine höheren Areal des Gehirnes simuliert
=> Deep learning ist nur künstliches Können aber nicht künstliche Intelligenz
- Benchmarking von RO-Systemen wie R-CNN und YOLO ist biased
 - Diverse Entscheidungen des Experimentators fließen ein
 - Wahl des Qualitätsmaßes
 - Wahl des Datensatzes
 - Objektgröße & -Anzahl, Klassen, Objektverdeckung,...