

Analyse komplexer Szenen mit Hilfe von Convolutional Neural Networks

Anwendungen 1

Vitalij Stepanov

HAW-Hamburg

24 November 2011

Inhalt

- Motivation
- Alternativen
- Problemstellung
- Anforderungen
- Lösungsansätze
- Zielsetzung
- Risiken

Motivation

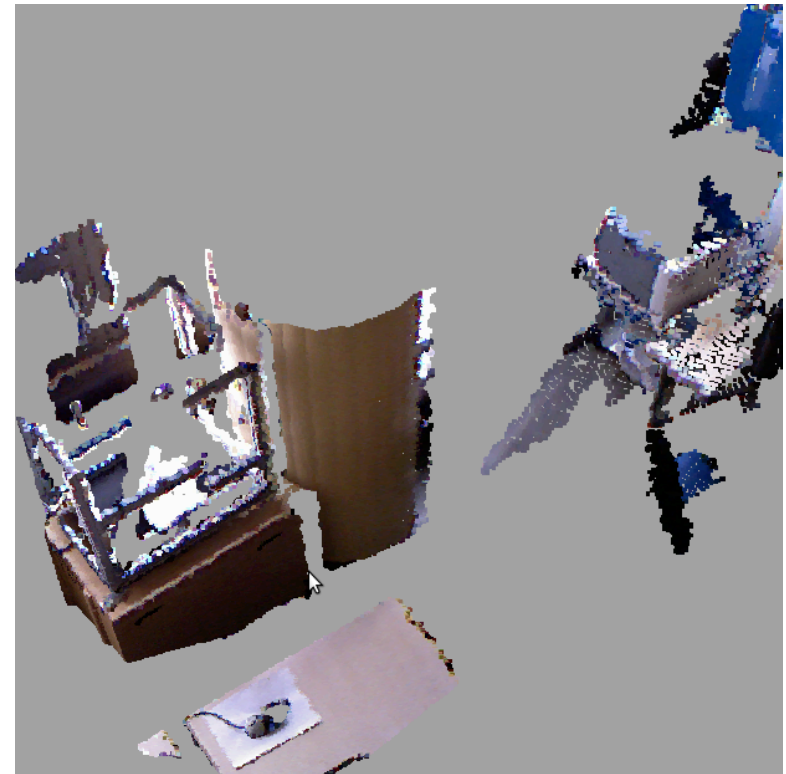
- Robot Vision
 - Lokalisieren von Objekten
 - Personen
 - Erkennen von Hindernissen
- Fahrerassistenz
 - Verkehrszeichen Erkennen
 - Erkennen von Fußgänger
 - Eye Tracking

Alternativen

- Laserscanner



- Kinect Aufnahme



Problemstellung

- Vielfalt
- Perspektive (Rotation)
- Belichtungsverhältnisse
- Verrauscht
- Verzerrung (affine Transformation)



Anforderungen

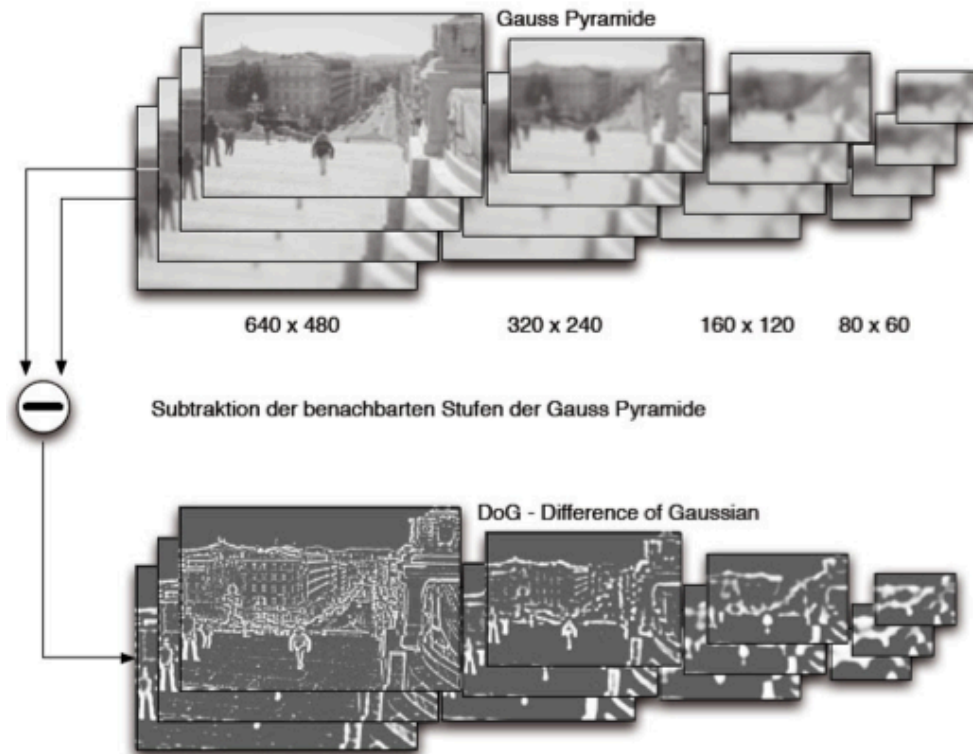
- hoher Grad an Generalisierung
- hohe Anzahl zu erkennende Objekte
- Effizienz
- Zuverlässigkeit

Lösungsansätze

- SIFT
- Neuronale Netze
- Convolutional Neural Networks

SIFT

- Scale-Invariant Feature Transform
- Unterteilung in Merkmalspunkte (extrema)
- Filterung (Stabilitätsanalyse)
- Bestimmung der Hauptorientierungen
- 128-dimensionaler Merkmalsvektor

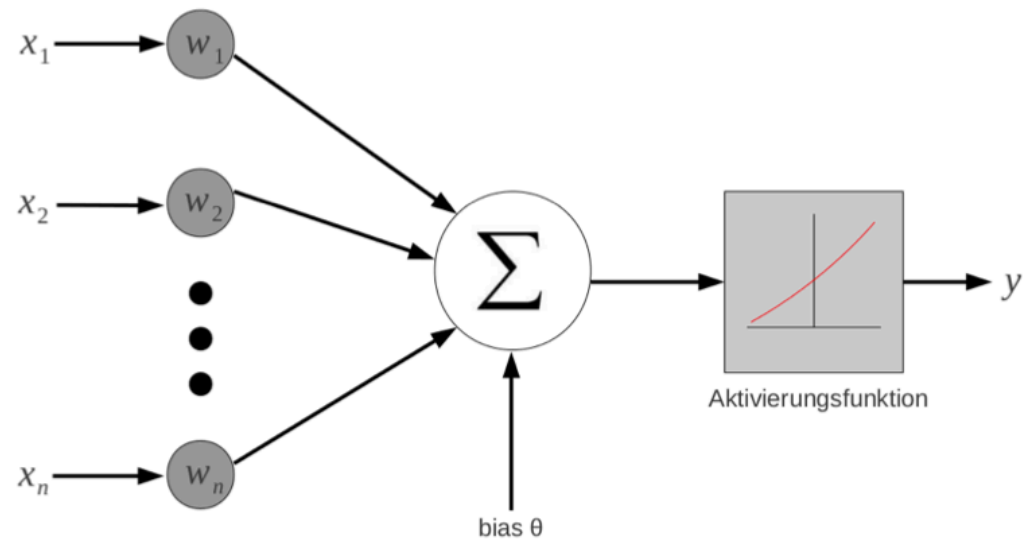


SIFT

- + unempfindlich gegen Koordinatentranslation
 - + robust gegen Belichtungsvariation
 - + gegen Bildrauschen,
 - + geringeren geometrischen Deformation
-
- erkennt nur die bekannten Objekte
 - keine Generalisierung

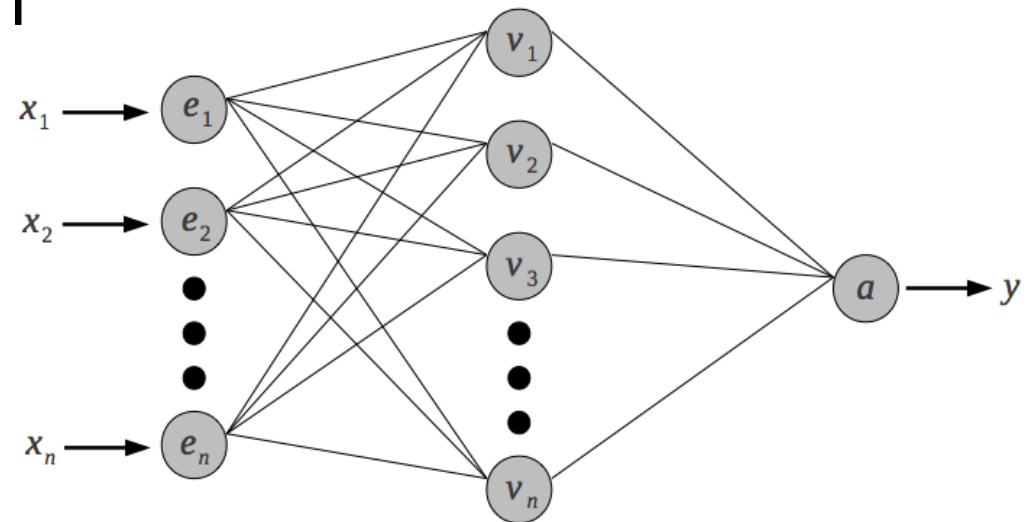
Neuronale Netze

- Perzeptron Neuron
 - mehrere Eingänge
 - Gewichtet
 - Summen aus gewichteten Eingängen
 - Schwellwert (bias)
 - Aktivierungsfunktion
 - Schrittfunktion
 - Sigmoid



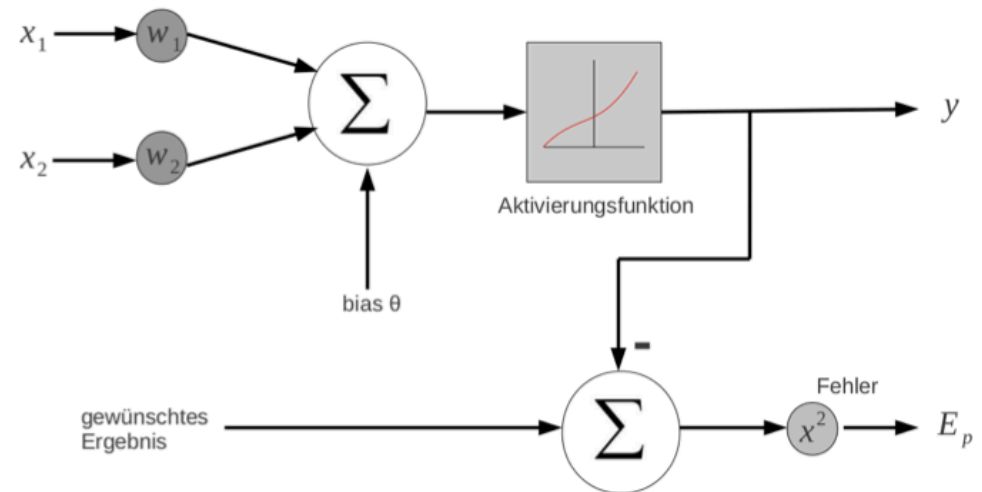
Neuronale Netze

- Neuronen werden zum Netz verbunden
- Eingangsschicht
- versteckte Schichten
- Vollvermascht
- Ausgangsschicht



Neuronale Netze

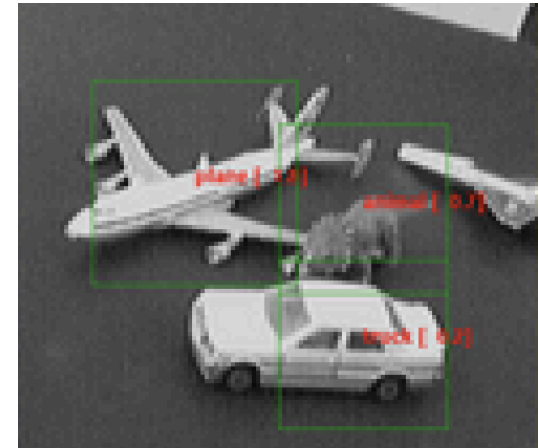
- Trainieren von NN
 - Einstellen von Gewichten
 - vom Ausgang
 - Differenzvektor bilden
 - Gewichte anpassen
- Lernalgorithmen
 - Backpropagation (Gradientenabstieg)
 - Quickprop
 - Boostalgorithmus (bekannt/unbekannt)



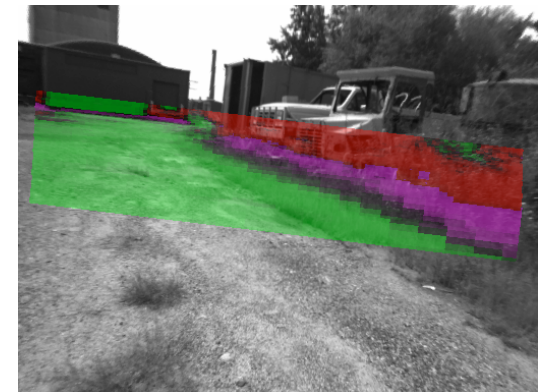
[1]

Convolutional Neural Networks

- Yann Le Cun, New York University
 - Schrifterkennung (LeNet 5) [4]
 - Generische Objekterkennung (LeNet 7) [5]
 - Off-Road Driving [6]
- Eindhoven University of Technology
 - Speed Sign Recognition [2]
- Tobias Hassenklöver, HAW-Hamburg
 - Gesichtserkennung [1]



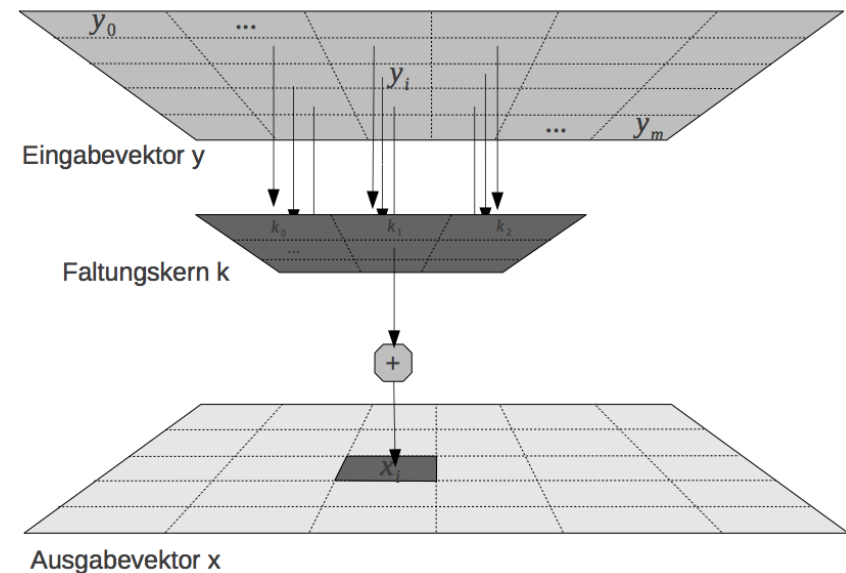
[5]



[6]

Convolutional Neural Networks

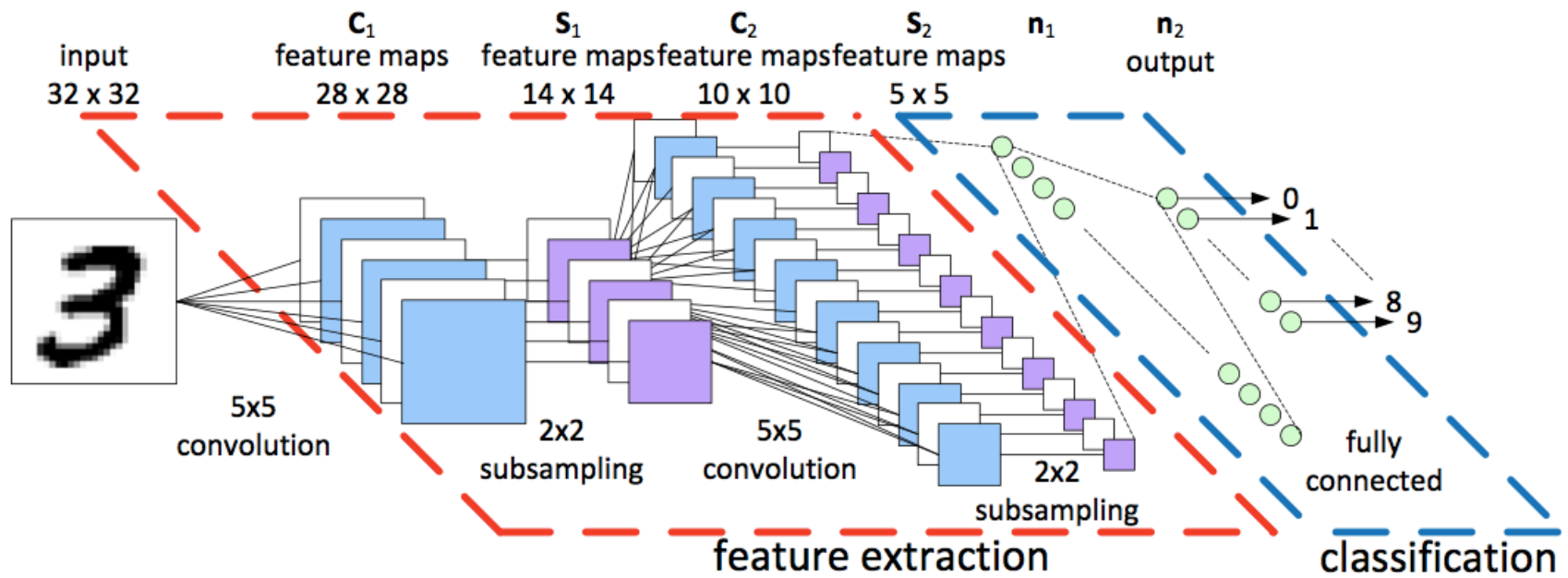
- Convolution-Schicht
 - Faltungsmaske
 - flexible Größe
 - auch komplex möglich
- Subsampling-Schicht
 - mindert die Bildgröße
 - erhöht Abstraktion
- Perceptron-Schicht
 - führt die Klassifizierung durch



[1]

Convolutional Neural Networks

- parallele Faltungen
- feature maps



Convolutional Neural Networks

- Einfach skalierbar auf größere Bilder
- Aufteilung des Bildes in Fenster
- LeNet 7
 - Auflösung von 96x96
 - 4.66 Mio. Multiplikationsakkumulationen (MAK)
 - Das Bild mit 240x240 mit 169 Teilbilder
 - erfordert 47.5 Mio. MAK
 - non-CNN Klassifizierer erfordern 788 Mio. MAK
 - Fehler von 6,2%
- Speed Sign Recognition
 - Auflösung 1280x720 35,7 fps.
 - Erkennungsrate 99,81%

Convolutional Neural Networks

- + einfach skalierbar
 - + hochgradig variant
 - + zuverlässig
 - + effizient
-
- Empirische Erreichung der Ergebnissen
 - Optimale Lösung kann nicht garantiert werden

Ziele

- Untersuchung des Lernverhaltens
- Vorbereitungsmethoden
- Trainingsdaten
- Testdaten
- Training
- Implementierung

Risiken

- hohe Ressourcenanforderungen
- explodierende Menge der Trainingsdaten
- Overlearning
- Nicht geeignete Trainingsdaten
- Unzuverlässigkeit
- hoher Fehler

Danke für die Aufmerksamkeit.

Fragen???

Literatur

[1] Tobias Hassenklöver. Klassifikation hochvarianter Muster mit Faltungsnetzwerken.

[2] Peemen, Maurice*, Mesman, Bart, Corporaal, Henk. Speed Sign Recognition by Convolutional Neural Networks.

<http://parse.ele.tue.nl/research/projects>

[3] Carsten Fries, 2011. Kamerabasierte Identifizierung und Lokalisierung von Gegenständen für flexible Roboter

[4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner 1998. Gradient-based learning applied to document recognition.

<http://yann.lecun.com/exdb/lenet/>

Literatur

[5] LeCun, Huang, and Bottou, 2004. Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting

<http://www.cs.nyu.edu/~yann/research/norb/index.html>

[6] Sermanet et al. 2009. A Multi-Range Architecture for Collision-Free Off-Road Robot Navigation

<http://www.cs.nyu.edu/~yann/research/lagr/index.html>

[7] Stefan Duffner 2007. Face image analysis with convolutional neural networks.