
DLM Semesterprojekt: Bestimmung der 3D-Boundingbox eines Roboters

GruppeNR: 1, Namen: Tim Eisenacher (1870755) und Paul Manea()
EMail: 1870755@stud.hs-mannheim.de

Abstract

This Latex and Style file are modified versions from NeurIPS 2018.[2]

1 Einleitung

Objekterkennung ist einer der vielversprechendsten und am stärksten im Forschungsfokus stehenden Bereiche von *Computer Vision* und *Machine Learning* [5]. Besonders im Bereich der medizinischen und industriellen Innovationen konnten dadurch bereits jetzt schon große Fortschritte erzielt werden [7, 1]. So können beispielhaft Operationsroboter stark von einer auf *Deep-Learning*-Algorithmen basierenden Erkennung und Lokalisation der Operationsinstrumente profitieren [6]. Auch in der Tumordiagnostik kann so die Erkennungsgenauigkeit gegenüber einer menschlichen Klassifikation bei gleichzeitig signifikant niedrigeren Kosten verbessert werden [3].

Die Anforderungen an die Bildverarbeitung sind dabei in diesen Bereichen besonders groß. Zum einen sorgen immer hochauflösendere Bilder für enorme Datenmengen und damit Hardwareanforderungen. Zum anderen ist die Anzahl an möglichen Klassen und damit Featurekombinationen bei der Objekterkennung enorm. Aufgrund ihrer hohen Trainingsperformanz und der Fähigkeit herausragend effizient Features in Bildern zu erkennen, eignen sich *Convolutional-Neural-Networks* (CNN) besonders gut als Lösungsansatz der beschriebenen Probleme [5].

In der vorliegenden Arbeit wird ein *Deep-Learning*-Modell zur Bestimmung der *3D-Boundingbox* eines Roboters implementiert und evaluiert. Dafür erfolgt anfangs die Abgrenzung und grobe Erläuterung des *Deep-Learnings* (DL) im Kontext des *Machine-Learnings* (ML). Der Fokus liegt dabei auf dem verwendeten DL-Konzept der CNN's und damit verbundener Probleme und Herausforderungen. Anschließend erfolgt eine Abhandlung der für die Implementierung verwendeten Netzstruktur und Metriken. Die Implementierung wird zunächst zur Lösung eines zweidimensionalen Problems erstellt und dann anschließend auf drei Dimensionen erweitert. Die Entwicklung erfolgt in Python 3.7 unter Verwendung des Tensorflow-Frameworks in der Spider und Eclipse IDE. Die Evaluation der Implementierung geschieht anhand eines Vergleiches der geschätzten *Boundingbox* mit der gelabelten *Boundingbox* der Testdaten. Als Datengrundlage für Training und Evaluation dient ein synthetisch generierter und vorgelabelter Datensatz mit RGB-Bildern.

2 Grundlagen und Stand der Technik

Salvaris beschreibt *Machine Learning* als einen Zweig der Computerwissenschaften, bei dem Computern beigebracht wird anhand von Trainingsdaten Entscheidungen zu treffen. Typische Anwendungsgebiete des ML sind Klassifikation, Regression und Clustering. DL ist ein Teilgebiet des ML bei dem besonders komplexe Neuronale Netze mit vielen Schichten und Neuronen Verwendung finden. Ein weitere wesentliche Abgrenzung stellt die Merkmalsextraktion dar. Also die Extraktion der Eigenschaften eines Objekts, die ausschlaggebend für etwaige Klassenzugehörigkeiten sind. Diese entscheidenden Merkmale müssen dem DL-Modell nicht vorgegeben werden, sondern werden von dem Algorithmus selbst gefunden. Dieser Umstand stellt mit die größten Herausforderungen aber auch Chancen beim DL dar [?, S.32-47]. Im Folgenden werden nur die für die vorliegende Arbeit besonders wichtigen und speziell angepassten Methoden und Aspekte des DL erläutert. Für weiterführende grundlegende Informationen zum Beispiel zu Neuronen, Schichttypen, Aktivierungsfunktionen und zum Gradientenabstiegsverfahren wird auf [4] verwiesen.

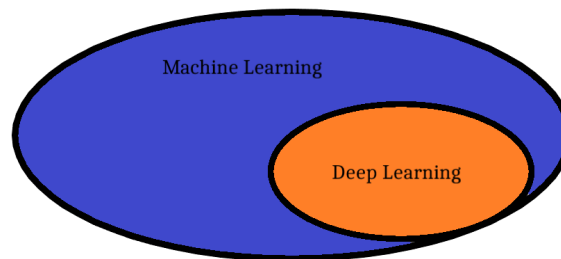


Figure 1: Abgrenzung Deep Learning zu Machine Learning.

2.1 Convolutional Neural Networks

2.2 Loss-Funktionen und Metriken

3 Methoden

4 Ergebnisse

4.1 Headings: second level

4.1.1 Headings: third level

5 Fazit und Ausblick

Vielleicht ganz kurz

Paragraphs

6 Citations, figures, tables, references

6.1 Citations

References

- [1] Abstract and :. Deep learning for medical image processing: Overview, challenges and future muhammad imran razzak, saeeda naz and ahmad zaib. 2016.
- [2] Md Atiqur, Rahman, Yang, and Wang. Optimizing intersection-over-union in deep neural networks for image segmentation. 2015.
- [3] Angel Cruz-Roa, Hannah Gilmore, Ajay Basavanthally, Michael Feldman, Shridar Ganesan, Natalie N.C. Shih, John Tomaszewski, Fabio A. González, and Anant Madabhushi. Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Scientific Reports*, 7(1), apr 2017.
- [4] Aaron Courville Ian Goodfellow, Yoshua Bengio. *Deep Learning. Das umfassende Handbuch: Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze (mitp Professional)*. mitp Professionals, 2018.
- [5] Wanli Ouyang, Xiaogang Wang, Xingyu Zeng, Shi Qiu, Ping Luo, Yonglong Tian, Hongsheng Li, Shuo Yang, Zhe Wang, Chen-Change Loy, and Xiaoou Tang. Deepid-net: Deformable deep convolutional neural networks for object detection. 2014.
- [6] Robot-Assisted Surgery, Using Deep, and Learning. Automatic instrument segmentation in. 2018.
- [7] Jonathan Tremblay, Thang To, and Balakumar Sundaralingam. Deep object pose estimation for semantic robotic grasping of household objects. 2018.

6.2 Footnotes

Here are some samples you can use. Please delete.

6.3 Figures

6.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 1.



Figure 2: Sample figure caption.

Table 1: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	~ 100
Axon	Output terminal	~ 10
Soma	Cell body	up to 10^6

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules*. We strongly suggest the use of the booktabs package, which allows for typesetting high-quality, professional tables:

<https://www.ctan.org/pkg/booktabs>

This package was used to typeset Table 1.