Projektarbeit: Semantische Segmentierung von Unterwasserbilddaten in Echtzeit mit PID-Net

Prüfungsleistung für das Modul „Deep Learning für Computer Vision“

im Wintersemester 2022/23

des Master-Studiengangs

Informatik

an der Technischen Hochschule Mittelhessen

23. Januar 2023

Bearbeitet von:

Ben Riegel Matrikel-Nr.

Fabian Schwickert 5455436

Gutachter der Studienakademie Prof. Dr.-Ing. Seyed Eghbal Ghobadi

Inhaltsverzeichnis

[Inhaltsverzeichnis II](#_Toc124498884)

[Abbildungsverzeichnis III](#_Toc124498885)

[Abkürzungsverzeichnis IV](#_Toc124498886)

[1 Problemstellung, Ziel und Aufbau 1](#_Toc124498887)

[2 Semantische Segmentierung von Bilddaten 1](#_Toc124498888)

[3 Untersuchungsobjekt: PID-Net 1](#_Toc124498889)

[4 Untersuchungsumgebung: SUIM-Datensatz 1](#_Toc124498890)

[5 Untersuchungsergebnisse 1](#_Toc124498891)

[Literaturverzeichnis V](#_Toc124498892)

Abbildungsverzeichnis

**Es konnten keine Einträge für ein Abbildungsverzeichnis gefunden werden.**

Abkürzungsverzeichnis

A a

# Problemstellung, Ziel und Aufbau

Computer spielen in der heutigen Gesellschaft eine große Rolle. Im privaten, ge­sell­schaft­li­chen und betrieblichen Umfeld werden Computer eingesetzt, um Aufgaben zu au­to­ma­ti­sie­ren und dem Menschen Arbeit abzunehmen. In den letzten Jahren ist besonders die „Künst­­liche Intelligenz“ (KI) in den Vordergrund gerückt. Denn mit KI-Software sind Com­pu­ter dazu in der Lage, Aufgaben wie z. B. das Erkennen von Objekten in Bil­dern bes­ser als der Mensch zu erledigen. Ohne KI-Software sind Computer nur schwerlich da­zu fä­hig.[[1]](#footnote-2)

Anfang 2023 kann auch KI-Software noch nicht alle Probleme perfekt lösen, die z. B. beim au­­­to­no­­men Fahren anfallen. Für diese Probleme wird noch leistungsstärkere KI-Software be­­­nö­­­­tigt. Besonders für die Anwendung im autonomen Fahren ist es zudem ent­schei­dend, dass die KI-Software zuverlässig die richtigen Ergebnisse in Echtzeit liefert. Denn wenn die KI-Soft­­­ware Gefahren nicht schnell genug erkennt, kann es zu Unfällen kommen.[[2]](#footnote-3)

Ty­pi­scher­wei­se ist KI-Software, die in Echtzeit agieren soll, weniger komplex als KI-Soft­ware, die kei­ner solchen Zeitbeschränkung unterliegt. Denn je komplexer die KI-Software ist, desto mehr Berechnungen müssen für die Verarbeitung von Daten durchgeführt wer­den und desto länger benötigt die KI-Software für den Verarbeitungsprozess. Dadurch ist Echt­zeit-KI-Software al­­­­lerdings meist weniger zuverlässig und ungenauer als andere KI-Soft­ware. Denn je komplexer die KI-Software ist, desto mehr variable Parameter besitzt die KI-Software und desto genauer sind ihre Berechnungen. Für An­wen­dungs­­felder wie das autonome Fahren muss ein Kompromiss zwischen Zeiteffizienz und Ef­­fektivität ge­fun­den werden.[[3]](#footnote-4)

Ein besonders für das autonome Fahren relevanter Aufgabentyp ist die „Segmentierung“ von Bildern bzw. Videos. Bei einer Segmentierungsaufgabe sollen die in den Bilddaten ent­­haltenen Objekte erkannt und klassifiziert werden. Beim autonomen Fahren wird da­durch z. B. der Gegenverkehr, die Straße und auch Fußgänger erkannt.[[4]](#footnote-5)

Um bei derartig komplexen Aufgaben einen Kompromiss zwischen Zeiteffizienz und Ef­­fek­­­tivität zu finden, wurde in den letzten Jah­ren verschiedene neuartige KI-Software ent­wi­­­ckelt. Eines dieser KI-Softwaresysteme ist „PID-Net“. PID-Net ist ein „Künstliches Neu­ro­­nales Netz“ (KNN), das Ende 2022 veröffentlicht wurde und Anfang 2023 zu den zeit­ef­fi­­zientesten und effektivsten KNN-Modellen für die Semantische Segmentierung von Bild­da­­ten in Echtzeit zählt. Dabei ist der besagte Kompromiss zwischen Zeiteffizienz und Ef­­fek­­­tivität mit PID-Net besonders gut gelungen.[[5]](#footnote-6)

Das Ziel der vorliegenden Projektarbeit ist es, die Forschungsergebnisse zu dem Anfang 2023 noch jungen PID-Net-Modell zu bestätigen. Dafür wird das PID-Net-Modell auf ei­nem Datensatz für die Semantische Segmentierung trainiert, der nicht von den Erfindern der PID-Net-Architektur ver­wendet wur­de. Dabei wird die Zeiteffizienz und die Ef­­fek­­ti­vi­tät des PID-Net-Modells erhoben. Die eigens erhobenen Daten werden dann mit den von den Erfindern der PID-Net-Architektur erhobenen Daten verglichen.

Für das Erreichen des besagten Ziels ist die vorliegenden Projektarbeit wie folgt ge­glie­dert. In Kapitel 2 der vorliegenden Projektarbeit wird zunächst genauer auf die Pro­blem­stel­­lung eingegangen, die von PID-Net gelöst werden soll. Dabei wird genauer be­schrie­ben, was Segmentierung ist, welche Arten der Segmentierung es gibt und wie dieser Auf­ga­bentyp sich von anderen unterscheidet.

In Kapitel 3 wird das Untersuchungsobjekt der vorliegenden Projektarbeit beschrieben. Es wird die Idee hinter der PID-Net-Architektur und ihre Umsetzung erläutert.

In Kapitel 4 wird der Untersuchungsbereich festgelegt, in dem das Untersuchungsobjekt be­trachtet wird. Es wird geklärt, was einen Datensatz für die Semantische Segmentierung aus­­macht und welcher Datensatz konkret für die vorliegende Projektarbeit verwendet wird.

Im Abschließenden Kapitel 5 wird kurz beschrieben, wie die Untersuchung durchgeführt wur­de. Anschließend werden ausführlich die Ergebnisse der Untersuchung vorgestellt.

# Semantische Segmentierung von Bilddaten

In dem Forschungsfeld der KI gibt es eine Reihe von Aufgabentypen. Die be­­kanntesten die­­ser Aufgabentypen sind die Assoziation, die Klassifizierung, die Re­gression und die Seg­­­mentierung. Nachfolgend werden diese vier Aufgabentypen kurz beschrieben.[[6]](#footnote-7)

1. Assoziation

Als Assoziation wird das Herstellen einer Korrelation zwischen Objekten bezeichnet. Da­durch können Paare oder Gruppen von Objekten gebildet werden, die eine semantische Be­­ziehung zueinander haben.[[7]](#footnote-8) Beispielsweise könnten Artikel gruppiert werden, die häu­fig zusammen gekauft werden.[[8]](#footnote-9)

1. Klassifizierung

Bei einer Klassifizierungs-Aufgabe werden z. B. Bilder aus einer Gesamtmenge in klar dif­fe­­ren­­zierte Teilmengen aufgeteilt. Dabei werden nicht mehrere Objekte in einem Bild se­pa­rat erkannt, sondern das Bilder als Ganzes klassifiziert.[[9]](#footnote-10) Ein Beispiel für eine Klas­si­fi­zie­rungs­aufgabe ist die Einteilung von Kunden in Kre­dit­würdigkeits-Klassen.[[10]](#footnote-11)

1. Regression

Mittels Regression kann aus Daten der Vergangenheit zukünftiges Verhalten prog­nos­ti­ziert werden. Dafür modellieren KI-Systeme einen funktionalen Zusammenhang.[[11]](#footnote-12) Bei­spiel­haft für solche Prognosen ist das „Predictive Maintenance“, bei dem vorausberechnet wird, wann eine Maschine gewartet werden muss.

1. Segmentierung

Bei einer Segmentierungs-Aufgabe werden z. B. Bilder aus einer Gesamtmenge in klar dif­fe­­­ren­­zierte Teilmengen aufgeteilt. Allerdings werden im Gegensatz zur Klassifizierung nicht die Bilder als Ganzes klassifiziert. Stattdessen werden Bilder in Segmente aufgeteilt und jedes dieser Segmente einer Klasse zugeordnet. Dabei können die Segmente in einem Bild unterschiedlichen Klassen zugeordnet werden.[[12]](#footnote-13) Segmentierung wird z. B. häufig beim au­­tonomen Fahren eingesetzt, um relevante Objekte wie z. B. Verkehrsschilder, die Straße oder andere Verkehrsteilnehmer im Verkehr zu erkennen.[[13]](#footnote-14)

Die vorliegende Projektarbeit befasst sich mit einer Segmentierungsaufgabe. Daher wird nach­­­folgend nur noch dieser Aufgabentyp betrachtet. Segmentierungs-Aufgaben treten in ver­schiedenen Anwendungsfeldern auf. Besonders viel Aufmerksamkeit erhält dabei die Bild­­er­ken­nung. In diesem Anwendungsfeld werden KI-Systeme ein­­­gesetzt, um den Inhalt von Bildern oder Videos zu bestimmen. Besonders häufig werden in diesem An­wen­dungs­feld „Künstlichen Neuronalen Netze“ (KNN) eingesetzt.[[14]](#footnote-15) In der vorliegenden Pro­­jekt­ar­beit wird ausschließlich das Anwendungsgebiet der Bildverarbeitung be­trach­tet.

Seg­men­tie­rungs-Aufgaben im Anwendungsfeld der Bilderkennung werden in drei Un­ter­typen aufgeteilt: Die „Semantische Segmentierung“, die „In­­stanz Segmentierung“ und die „Pa­noptische Seg­men­tie­rung“.[[15]](#footnote-16) Nachfolgend werden diese drei Varianten be­schrie­ben. In Abbildung 1 werden die Unterschiede zwischen den drei Varianten her­vor­ge­ho­ben. Da­bei werden die Klassen, denen Objekte zugeordnet werden durch verschiedene Farben re­präsentiert.

1. Semantische Segmentierung

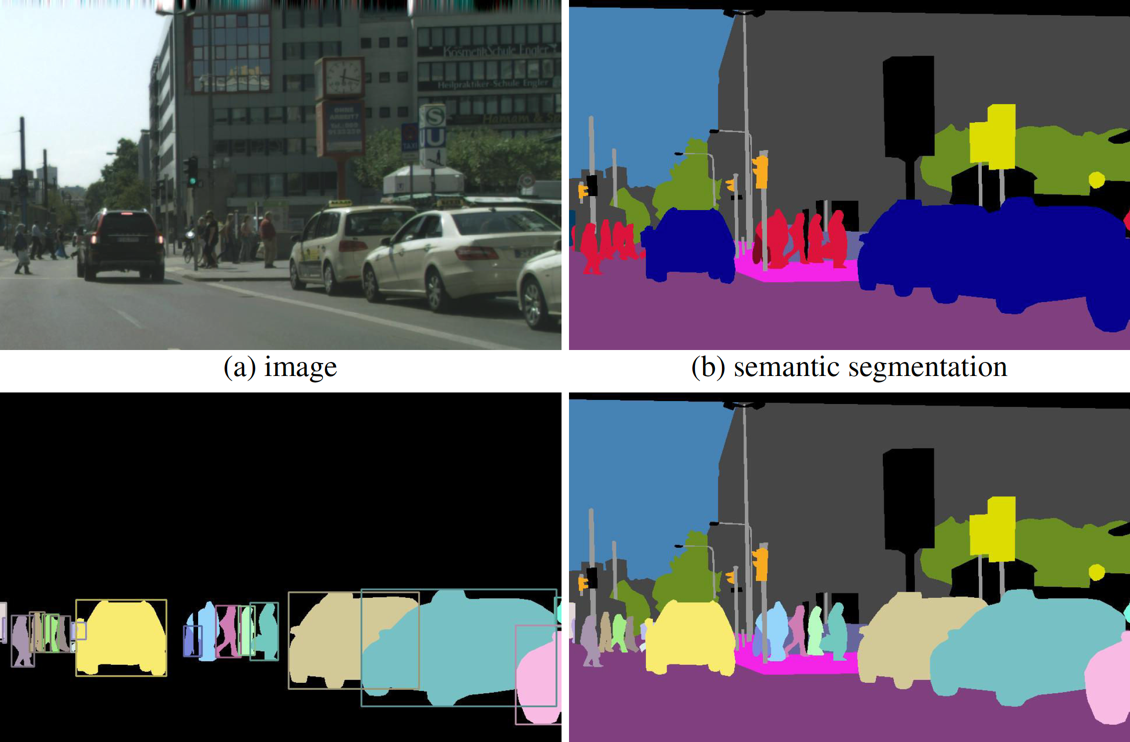
Bei der Semantischen Segmentierung wird das gesamte Bild in Segmente unterteilt. Diese Seg­mente werden ausschließlich nach der semantischen Zugehörigkeit zu einer Klasse klas­­sifiziert. Es wird nicht zwischen zwei Segmenten mit der gleichen Klas­sen­zu­ge­hö­rig­keit unterschieden.[[16]](#footnote-17) So würden z. B. zwei überlappend dargestellte Menschen (z. B. hin­ter­­ei­nanderstehend) nicht durch Se­man­tische Segmentierung unterschieden werden (sie­he Abbildung 1B).

1. Instanz Segmentierung

Im Unterschied zur Semantische Segmentierung, werden bei einer Instanz Segmentierung le­diglich zählbare Objekte in einem Bild erkannt. Diese Segmente werden nach der se­man­­ti­schen Zugehörigkeit zu einer Klasse und der abgebildeten Instanz der Klasse klas­si­fiziert. Es wird im Unterschied zur Semantischen Segmentierung zwischen zwei Seg­men­ten mit der gleichen Klassenzugehörigkeit unterschieden.[[17]](#footnote-18) So würden z. B. zwei über­­lap­pend dar­ge­stellte Menschen (z. B. hintereinanderstehend) durch Instanz Seg­men­tie­­rung un­ter­schie­den werden (siehe Abbildung 1C).

1. Panoptische Segmentierung

Die Panoptische Segmentierung vereint die Semantische und die Instanz Segmentierung. Bei der Panoptischen Segmentierung wird das gesamte Bild in Segmente aufgeteilt. Alle Seg­­mente werden nach der se­man­­ti­schen Zugehörigkeit zu einer Klasse klassifiziert. Seg­men­te, die zählbare Objekte repräsentieren, werden zusätzlich nach der abgebildeten In­stanz der Klasse klas­si­fiziert (siehe Abbildung 1D).[[18]](#footnote-19)



(A) Original Bild

(B) Semantische Segmentierung

(C) Instanz Segmentierung

(D) Panoptische Segmentierung

Abb. 1: Drei Varianten der Segmentierung[[19]](#footnote-20)

Alle drei Varianten der Segmentierung werden Anfang 2023 hauptsächlich mit „Künst­li­chen Neuronalen Netzen“ (KNN) umgesetzt. Allerdings unterscheiden sich die Struktur und die Parameter dieser KNNs je nach Variante.[[20]](#footnote-21) In der vorliegenden Projektarbeit wird ei­ne Semantische Segmentierungs-Aufgabe betrachtet. Daher wird nachfolgend nur noch die­se Variante betrachtet.

# Untersuchungsobjekt: PID-Net

A

# Untersuchungsumgebung: SUIM-Datensatz

A

# Untersuchungsergebnisse

a

Literaturverzeichnis

1. Aggerwal, Charu: Neural Networks and Deep Learning, Springer International Pub­lishing AG (Hrsg.), 2018, https://link.springer.com/book/10.1007/978-3-319-94463-0, abgerufen am 14.01.2023.
2. Cakir, Senay et al.: Semantic Segmentation for Autonomous Driving: Model Eva­lu­ation, Dataset Ge­neration, Perspective Comparison, and Real-Time Capability, Hoch­schule Esslingen (Hrsg.), 26. Juli 2022, https://arxiv.org/pdf/2207.12939. pdf, abgerufen am 14.01.2023.
3. Gholami, Amir et al.: A Survey of Quantization Methods for Efficient Neural Net­work Inference, University of California, Berkeley (Hrsg.), 21. Juni 2021, https:// ar­xiv.org/pdf/2103.13630.pdf, abgerufen am 14.01.2023.
4. Kalliomäki, Roger: Real-time object detection for autonomous vehicles using deep learning, Uppsala Universität – Department of Information Technology (Hrsg.), Juni 2019, https://uu.diva-portal.org/smash/get/di­va2:1356309/FULL­TEXT01.pdf, abgerufen am 14.01.2023.
5. Kirillov, Alexander: Exploring Aspects of Image Segmentation: Diversity, Global Rea­soning, and Panoptic Formulation, Combined Faculty for the Natural Sciences and Mathematics – Universität Heidelberg, https://archiv.ub.uni-hei­del­berg.de/ voll­textserver/25750/1/thesis\_kirillov.pdf, abgerufen am 13.01.2023.
6. O. V.: Real-Time Semantic Seg­men­tation on Cityscapes test, paperswithcode.com (Hrsg.), https://paperswithcode.com/sota/real-time-semantic-segmentation-on-city­scapes, abgerufen am 14.01.2023.
7. O. V.: Real-Time Semantic Seg­men­tation on CamVid, paperswithcode.com (Hrsg.), https://paperswithcode.com/sota/real-time-semantic-segmentation-on-camvid, abgerufen am 14.01.2023.
8. Xu, Jiacong; Xiong, Zixiang; Bhattacharyya, Shankar: PIDNet: A Real-time Se­man­tic SegmentationNetwork Inspired from PID Controller, Texas A&M University (Hrsg.), 10. Juni 2022, https://arxiv.org/pdf/2206.02066v2.pdf, abgerufen am 14.01.2023.
9. a

1. Vgl. Aggerwal, Charu: Neural Networks and Deep Learning, Springer International Publishing AG (Hrsg.), 2018, https://link.springer.com/book/10.1007/978-3-319-94463-0,

   abgerufen am 14.01.2023, S. 3. [↑](#footnote-ref-2)
2. Vgl. Kalliomäki, Roger: Real-time object detection for autonomous vehicles using deep learning, Upp­sala Universität – Department of Information Technology (Hrsg.), Juni 2019, https://uu.diva-por­tal.org/smash/get/diva2:1356309/FULLTEXT01.pdf, abgerufen am 14.01.2023, S. 1f [↑](#footnote-ref-3)
3. Vgl. Kalliomäki, Roger: Real-time object detection for autonomous vehicles using deep learning,   
   a. a. O., abgerufen am 14.01.2023, S. 65. Vgl. auch Gholami, Amir et al.: A Survey of Quantization Methods for Efficient Neural Network Inference, University of California, Berkeley (Hrsg.), 21. Juni 2021, https://arxiv.org/pdf/2103.13630.pdf, abgerufen am 14.01.2023, S. 1. [↑](#footnote-ref-4)
4. Vgl. Cakir, Senay et al.: Semantic Segmentation for Autonomous Driving: Model Evaluation, Dataset Ge­neration, Perspective Comparison, and Real-Time Capability, Hochschule Esslingen (Hrsg.), 26. Juli 2022, https://arxiv.org/pdf/2207.12939.pdf, abgerufen am 14.01.2023, S. 1. [↑](#footnote-ref-5)
5. Vgl. Xu, Jiacong; Xiong, Zixiang; Bhattacharyya, Shankar: PIDNet: A Real-time Semantic Segmentation Net­work Inspired from PID Controller, Texas A&M University (Hrsg.), 10. Juni 2022, https://arx­iv.org/pdf/2206.02066v2.pdf, abgerufen am 14.01.2023. Vgl. auch o. V.: Real-Time Semantic Seg­men­tation on Cityscapes test, paperswithcode.com (Hrsg.), https://paperswithcode.com/sota/real-ti­me-semantic-segmentation-on-cityscapes, abgerufen am 14.01.2023. Vgl. auch o. V.: Real-Time Se­man­tic Seg­men­tation on CamVid, paperswithcode.com (Hrsg.), https://paperswithcode.com/sota/ real-time-semantic-seg­mentation-on-camvid, abgerufen am 14.01.2023. [↑](#footnote-ref-6)
6. Vgl. a [↑](#footnote-ref-7)
7. Vgl. a [↑](#footnote-ref-8)
8. Vgl. a [↑](#footnote-ref-9)
9. Vgl. a [↑](#footnote-ref-10)
10. Vgl. a [↑](#footnote-ref-11)
11. Vgl. a [↑](#footnote-ref-12)
12. Vgl. a [↑](#footnote-ref-13)
13. Vgl. a [↑](#footnote-ref-14)
14. Vgl. a [↑](#footnote-ref-15)
15. Vgl. Kirillov, Alexander: Exploring Aspects of Image Segmentation: Diversity, Global Reasoning, and Pa­noptic Formulation, Combined Faculty for the Natural Sciences and Mathematics – Universität Hei­del­berg, https://archiv.ub.uni-heidelberg.de/volltextserver/25750/1/thesis\_kirillov.pdf, abgerufen am 13.01.2023, S. 73f. [↑](#footnote-ref-16)
16. Vgl. Kirillov, Alexander: Exploring Aspects of Image Segmentation: Diversity, Global Reasoning, and Pa­noptic Formulation, abgerufen am 13.01.2023, S. 73f. [↑](#footnote-ref-17)
17. Vgl. Kirillov, Alexander: Exploring Aspects of Image Segmentation: Diversity, Global Reasoning, and Pa­noptic Formulation, a. a. O., abgerufen am 13.01.2023, S. 73f. [↑](#footnote-ref-18)
18. Vgl. Kirillov, Alexander: Exploring Aspects of Image Segmentation: Diversity, Global Reasoning, and Pa­noptic Formulation, a. a. O., abgerufen am 13.01.2023, S. 74. [↑](#footnote-ref-19)
19. Kirillov, Alexander: Exploring Aspects of Image Segmentation: Diversity, Global Reasoning, and Pa­noptic Formulation, abgerufen am 13.01.2023, S. 74. [↑](#footnote-ref-20)
20. Vgl. a [↑](#footnote-ref-21)