

# AI-Based 3D-Detection of Parked Vehicles on a Mobile Mapping Platform using Edge Computing

JONAS MEYER<sup>1</sup>

*Zusammenfassung: Parkstatistiken bilden eine wichtige Grundlage für die Definition und Anpassung der Parkpolitik im städtischen Raum. Am Institut Geomatik der Fachhochschule Nordwestschweiz wurde das erste System entwickelt, welches basierend auf Low-Cost Sensoren die zuverlässige und kosteneffiziente Erstellung von Parkstatistiken für ganze Städte mit einer hohen Wiederholfrequenz erlaubt. Diese Lösung beruht jedoch vollständig auf zeit- und rechenintensivem Postprocessing und erfordert die Erfassung, Übertragung und Anonymisierung großer Datenmengen. In dieser Thesis wurde die bestehende Erfassungssoftware um die 3D-Fahrzeugdetektion erweitert. Dazu wurde unter Berücksichtigung der Low-Cost Komponenten eine passende 3D-Objektdetektionsmethode evaluiert und ein Workflow für die Datenvorverarbeitung entwickelt. Darüber hinaus wurde die 3D-Objektdetektionsmethode für die relevanten Fahrzeugklassen neu trainiert. Die weiterentwickelte Erfassungssoftware eliminiert sämtliche Einschränkungen bezüglich des Datenschutzes, reduziert das generierte Datenvolumen pro Bild um Faktor 60'000 von 15MB auf 0.25 KB und übertrifft mit einer Precision von 100% und einem Recall von 98% sämtliche bekannten Arbeiten für die Erstellung von Parkstatistiken auf der Basis von Mobile Mapping Daten.*

## 1 Einleitung

Wir erleben derzeit einen Wandel der städtischen Mobilität vom motorisierten Individualverkehr hin zu einer zunehmenden Vielfalt an multimodalen Mobilitätsangeboten. Durch Parkplätze belegter Straßenraum gerät aufgrund des ansteigenden Langsamverkehrs zunehmend unter Druck. Parkstatistiken bieten Behörden und politischen Entscheidungsträgern eine fundierte Grundlage für die Überprüfung und Anpassung der verfügbaren Parkplätze, der Parkregeln und -preise sowie der Parkpolitik im Allgemeinen. Die Erstellung von Parkstatistiken für Stadtteile oder ganze Städte ist jedoch nach wie vor ein größtenteils manueller Prozess. Die Parkstatistiken für die Stadt Basel wurden beispielsweise mithilfe von GoPro-Videos in Kombination mit der manuellen Interpretation durch menschliche Interpretation erstellt (RAPP TRANS AG BASEL-STADT 2019).

MATHUR et al. (2010), BOCK et al. (2015) und FETSCHER (2020) haben erfolgreich die Machbarkeit zuverlässiger Parkstatistiken mit Daten von Mobile Mapping Systemen (MMS) bewiesen. Allerdings sind diese Lösungen entweder abhängig von High-End MMS mit hochwertigen LiDAR und Kamera-Sensoren und daher teuer oder sie beschränken sich auf einen spezifischen Parkplatztyp. NEBIKER et al. (2021) entwickelten die erste Lösung zur zuverlässigen und kosteneffizienten Erstellung stadtweiter Parkstatistiken mit hoher Wiederholfrequenz (Abb. 1). Das System besteht aus einem elektrischen Dreirad, zwei RGB-D-Kameras *Intel RealSense D455* (INTEL CORPORATION 2020) (Abb. 1, a), einer GNSS/INS basierten Navigationseinheit (Abb. 1, b) und dem Em-

---

<sup>1</sup> FHNW Fachhochschule Nordwestschweiz, Institut Geomatik, Hofackerstrasse 30, CH-4132 Muttenz, E-Mail: jonas.meyer@fhnw.ch

bedded-System Modul *Nvidia Jetson TX2* (NVIDIA DEVELOPERS, 2022) (Abb. 1, c). Mithilfe künstlicher Intelligenz werden Fahrzeuge, in den aus Tiefenkarten gewonnenen Punktwolken, dreidimensional detektiert und daraus in einer GIS-Analyse Parkstatistiken abgeleitet. Diese Lösung beruht jedoch vollständig auf zeit- und rechenintensivem Postprocessing und erfordert die Erfassung, Übertragung und Anonymisierung großer Datenmengen. Um die Datenübertragung zu reduzieren, Einschränkungen aufgrund von Datenschutzbestimmungen zu beseitigen und geringe Latenzzeiten zu ermöglichen, werden Daten zunehmend in der Nähe der Erfassungsgeräte – an der Edge – verarbeitet. In den letzten Jahren wurden eigens dafür spezielle Hardware-Module mit einem geringen Energiebedarf entwickelt. Aufgrund der damit einhergehenden eingeschränkten Rechenressourcen, stellt die Kombination von Edge Computing und rechenintensiven Algorithmen wie die 3D-Objektdetektion (OD) gegenwärtig eine erhebliche Herausforderung dar.

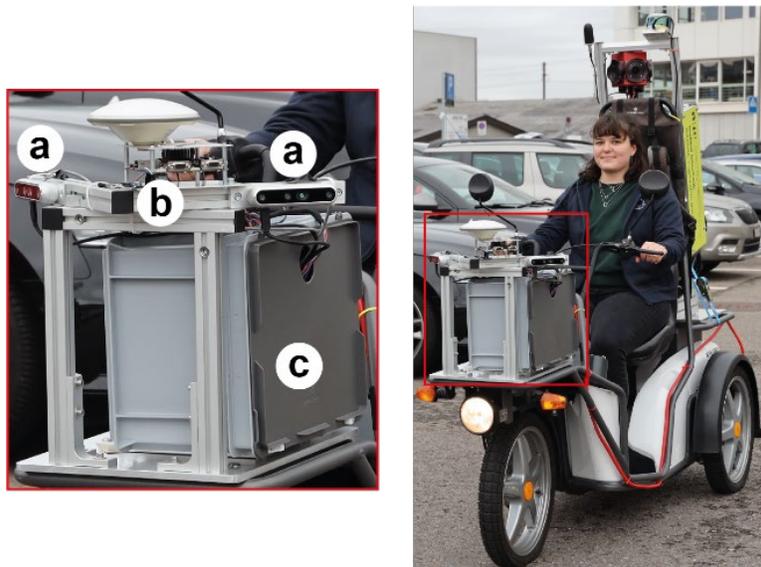


Abb. 1: Elektrisches Mobile Mapping System mit Low-Cost Sensoraufbau (Nebiker et al. 2021).

## 2 Materialien und Methodik

### 2.1 3D-Objektdetektor und Eingabedaten

Die Integration der 3D-Fahrzeugdetection direkt in das Erfassungssystem setzt nebst einer hohen Detektionsgenauigkeit eine tiefe Inferenzzeit voraus. NEBIKER et al. (2021) zeigten jedoch, dass moderne, auf dem KITTI Datensatz (GEIGER et al., 2012) trainierte 3D-OD-Methoden, angewendet auf die Punktwolken aus den RealSense (RS) Kameras, signifikant schlechtere Detektionsergebnisse und Inferenzzeiten erzielten als in den Originalpublikationen. Vergleiche beider Datensätze zeigten, dass die RS-Punktwolken ein massiv höheres Rauschen und teilweise große Datenlücken aufweisen sowie um Faktoren zwischen 200 und 400 dichter sind als die KITTI-Punktwolken (NEBIKER et al., 2021). Aus diesen Erkenntnissen wurden zwei Annahmen formuliert: 1) Eine Reduktion der Punktwolkendichte führt zu tieferen Inferenzzeiten und 2) durch Adaption der Charakteristiken der Trainingsdaten auf die RS-Punktwolken kann die Detektionsgenauigkeit gesteigert

werden. In zwei unabhängigen Experimenten wurden sieben Punktwolken-basierte 3D-OD-Methoden (Part-A<sup>2</sup> Net Anker-basiert und Anker-frei (SHI et al., 2020b), PointPillars (LANG et al. 2019), PointRCNN (SHI et al. 2019), PV-RCNN (SHI et al. 2020a), SECOND (YAN et al. 2018) und Voxel R-CNN (DENG et al. 2021)) hinsichtlich Detektionsgenauigkeit und Inferenzzeit evaluiert. Die Kombination der Resultate aus beiden Experimenten führte zur geeignetsten Methode sowie den optimalen Parametern für die Datenvorverarbeitung.

Für die Experimente mussten die RS-Punktwolken sowohl ausgedünnt als auch das Rauschen darin minimiert werden. Dazu wurde ein effizienter Datenvorverarbeitungsworkflow entwickelt, der anstelle von Punktwolken direkt die von den RS-Kameras erfassten Tiefenkarten verwendet (Abb. 2, links). Glättungsfilter können auf die als Bilddaten repräsentierten Tiefenkarten angewendet werden. Anschließend werden spezifische Pixel in der Tiefenkarte ausgewählt und via bekannte innere Orientierung der Kamera in 3D-Punkte umgerechnet. Ein regelmäßiges Gitter mit beliebiger horizontaler und vertikaler Schrittweite (Abb. 2, Mitte) selektiert die zu transformierenden Punkte. Dieser Ansatz erlaubt eine effiziente Generierung ausgedünnter und geglätteter Punktwolken (Abb. 2, rechts) bei gleichzeitiger Beibehaltung der Geometrieinformation und ohne aufwendige Szeneninterpretation.

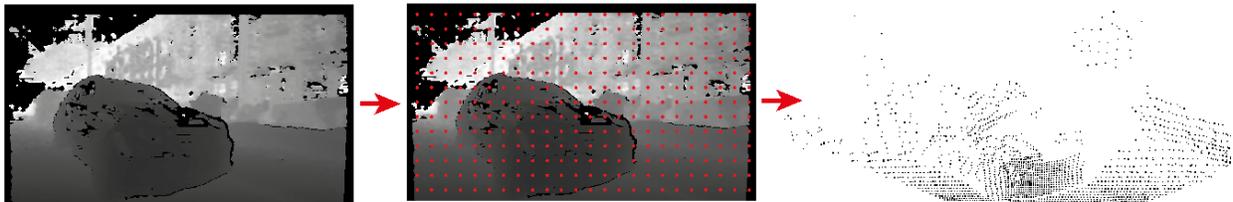


Abb. 2: Datenvorverarbeitungsworkflow von der Tiefenkarte zur aufbereiteten Punktwolke.

## 2.2 Training Objektdetektor

Für die Erstellung vollständiger Parkstatistiken wurde der 3D-Objektdetektor auf den Klassen „Car“, „Van“ und „Truck“ des KITTI Datensatzes (GEIGER et al. 2012) neu trainiert. Das neu trainierte Modell wurde auf den Testdaten des KITTI Datensatzes evaluiert und erreichte die in Tab. 1 dargestellten Average Precision (AP)-Werte bei einem IoU-Schwellwert von 0.7. Während die AP-Werte der Klasse „Car“ im selben Bereich wie jene des vortrainierten Modells liegen, deuten die AP-Werte der Klassen „Van“ und „Truck“ auf schlechte Trainingsresultate hin was aufgrund der stark unterrepräsentierten Trainingsdaten in diesen beiden Klassen zu erwarten war.

Tab. 1: Evaluationsresultate des neu trainierten Objektdetektors. AP<sub>70</sub>-Werte ausgewiesen in % pro trainierte Klasse und Schwierigkeitsstufe (einfach, moderat und schwierig).

Car			Van			Truck		
89.01	78.59	77.82	48.37	35.12	30.54	9.41	9.09	9.09

## 2.3 Erfassungssoftware

Die Erfassungssoftware von NEBIKER et al. (2021) beruht auf dem Graphen-basierten Robotik Framework Robot Operating System (ROS) (QUIGLEY et al. 2009) und verfügt über ein modulares Design. Die 3D-Fahrzeugdetektion konnte daher unkompliziert als neuer eigenständiger Knoten hinzugefügt werden (Abb. 3, rote Ellipse). Die weiterentwickelte Erfassungssoftware besteht aus

drei Knoten. Der *constrained-trigger* Knoten und der *realsense2-camera* Knoten (DORODNICOV & HIRSHBERG 2021) sind für das Triggering respektive die Publikation der Daten zuständig, während der *vehicle-detection* Knoten die 3D-Fahrzeugdetektion vollzieht (Abb. 3).

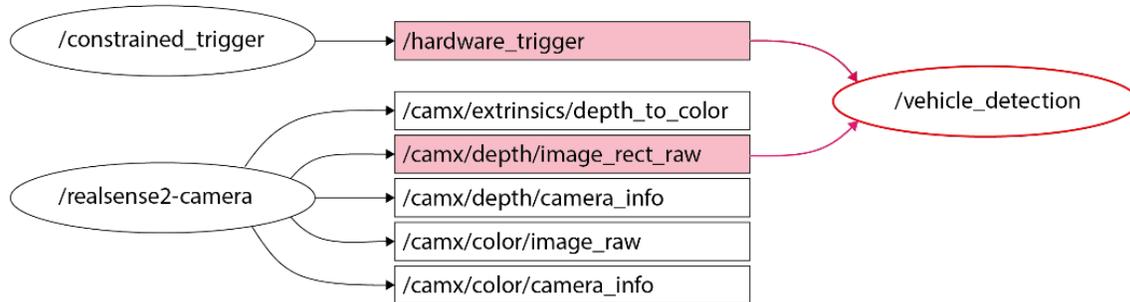


Abb. 3: ROS-Graph der erweiterten Erfassungssoftware. Knoten sind als Ellipsen und Topics als Vierecke dargestellt

Im *vehicle-detection* Knoten werden mit dem entwickelten Datenvorverarbeitungsworkflow und den evaluierten optimalen Parametern ausgedünnte Punktwolken aufbereitet, der neutrainierte Objektdetektor darauf angewendet und die Detektionsresultate gespeichert. Mit der Erweiterung der Erfassungssoftware entfällt die Notwendigkeit große Datenmengen zu speichern, zu übertragen und RGB-Bilder zu anonymisieren, wodurch sich die Anzahl der verbleibenden Nachbearbeitungsschritte auf die folgenden drei reduziert. 1) Berechnung der Bildposen mittels direkter Sensororientierung, 2) Transformation der Detektionsresultate via Bildposen ins übergeordnete Koordinatensystem, 3) GIS-Analyse zur Erstellung der Parkstatistik.

### 3 Experimente und Resultate

#### 3.1 3D-Objektdetektor und Eingabedaten

Mit dem eingeführten Datenvorverarbeitungsworkflow wurden insgesamt 83 Punktwolken-Datensätze mit unterschiedlich starken Glättungs- und Ausdünnungsfaktoren aufbereitet. Die Testdaten stammen aus einer Befahrung mit dem MMS von NEBIKER et al. (2021) und enthalten 593 RGB-D Bilder einer Quartierstrasse in der Stadt Basel mit 34 geparkten Autos. Sämtliche Datensätze wurden mit den sieben 3D-OD-Methoden auf einer Workstation ausgestattet mit einer performanten Grafikkarte prozessiert. Tab. 2 zeigt die erzielten Inferenzzeiten pro Methode und Ausdünnungsfaktor in Bilder pro Sekunden (FPS). Die Resultate zeigen deutlich, dass unabhängig von der 3D-OD-Methode spärlichere Punktwolken schneller prozessiert werden.

Tab. 2: Inferenzzeiten in FPS mit originalen und reduzierten Punktwolken (Reduktionsfaktor 196 & 400).

Methode	Part-A <sup>2</sup> Net Anker	Part-A <sup>2</sup> Net	PointPillars	PointRCNN	PV-RCNN	SECOND	VOXEL R-CNN
Original	7.6	3.6	22.3	3.5	1.7	15.5	7.2
196	11.4	13.0	44.5	4.6	10.9	28.9	20.9
400	11.6	13.2	45.1	4.9	11.1	29.0	22.3

Erste Detektionsresultate zeigten, dass eine vorgängige Glättung der Tiefenkarte keine signifikante Verbesserung der Detektionsresultate bewirkt. Außerdem benötigt die Glättung angesichts der geplanten Verarbeitung an der Edge zu viel Rechenleistung, weshalb die Reduktion des Rauschens in dieser Arbeit nicht weiterverfolgt wurde. Tab. 3 zeigt die besten Detektionsresultate pro 3D-OD-Methode gemessen am F1-Wert für die originale sowie die beste ausgedünnte Punktwolke zusammen mit dem entsprechenden Reduktionsfaktor.

Tab. 3: Beste Detektionsresultate auf ausgedünnten und originalen Punktwolken. Metrik F1-Wert

Methode	Part-A <sup>2</sup> Net Anker	Part-A <sup>2</sup> Net	PointPillars	PointRCNN	PV-RCNN	SECOND	VOXEL R-CNN
Original	0.2	0.3	-	<b>0.94</b>	0.56	0.46	0.34
Ausgedünnte PW (Red. Faktor)	<b>0.99</b> (180)	<b>0.99</b> (320)	0.76 (245)	<b>0.99</b> (500)	0.97 (320)	0.87 (245)	0.97 (180)

Basierend auf den erzielten Inferenzzeiten (Tab. 2) und Detektionsresultaten (Tab. 3) wurde der Objektdetektor Part-A<sup>2</sup> Net für den weiteren Verlauf dieser Arbeit verwendet.

### 3.2 Training Objektdetektor

Die in NEBIKER et al. (2021) publizierte Parkstatistik wurde mit dem neu trainierten Objektdetektor Part-A<sup>2</sup> Net wiederholt. Sie umfasst 350 Parkplätze, von denen 283 belegt waren. Die Fahrzeugdetektionen wurden analog zu NEBIKER et al. (2021) manuell ausgezählt und verifiziert. Tab. 4 zeigt die Ergebnisse aufgeschlüsselt nach Parkplatztyp. Der neu trainierte Objektdetektor erzielte über alle Parkplatz- und Fahrzeugtypen eine Precision von 100% und einen Recall von 98% und übertraf somit den von NEBIKER et al. (2021) verwendeten Objektdetektor um 11%. Es zeigte sich zudem, dass Fahrzeuge der Klassen „Van“ und „Truck“ zwar an der richtigen Stelle detektiert wurden, die geschätzten Dimensionen der 3D-Boundingboxen allerdings größtenteils falsch waren, was ebenfalls durch die schlechten AP-Werte in Tab. 31 angedeutet wird.

Tab. 3: Resultate der Parkstatistik unter der Verwendung des neu trainierten Objektdetektors. Resultate sind analog zu NEBIKER et al. (2021) nach Parkplatztyp aufgeschlüsselt.

	Parallel	Schräg	Senkrecht	2 x 2	Total
TP / TN / FP / FN	181 / 35 / 0 / 0	26 / 5 / 0 / 0	65 / 26 / 0 / 4	4 / 1 / 0 / 3	<b>276 / 67 / 0 / 7</b>
Precision	1.00	1.00	1.00	1.00	<b>1.00</b>
Recall	1.00	1.00	0.94	0.57	<b>0.98</b>

### 3.3 Erfassungssoftware

Die weiterentwickelte Erfassungssoftware wurde hinsichtlich Latenzzeit und generiertes Datenvolumen untersucht. Dazu wurde eine als ROS-Bag-Datei gespeicherte Kampagne erneut abgespielt und durch die Erfassungssoftware verarbeitet. Es konnte eine Latenzzeit von 0.53 s pro Bild (1.9 FPS) auf dem Embedded-System Modul Nvidia Jetson TX2 erreicht werden. Weiter wurden pro Aufnahme Daten von 0.25 KB generiert. Verglichen mit dem generierten Datenvolumen der alten Erfassungssoftware (ca. 15 MB / Aufnahme) wurde eine Reduktion um Faktor 60'000 erreicht.

## 4 Diskussion

In dieser Arbeit wurde die 3D-Fahrzeugdetektion in die in NEBIKER et al. (2021) beschriebene Erfassungssoftware integriert. Dazu wurden in verschiedenen Experimenten die einzelnen Komponenten evaluiert sowie das Gesamtsystem auf dessen Performance getestet. Die Experimente zeigten, dass die Eingabedaten einen erheblichen Einfluss sowohl auf die Detektionsgenauigkeit als auch die Inferenzzeit von 3D-OD-Methoden haben. Durch Ausdünnung der RS-Punktwolken konnten die Inferenzzeiten aller sieben evaluierten 3D-OD-Methode verringert werden (Tab. 2). Weiter konnten durch Adaption der Eigenschaften der Trainingsdaten auf die eigenen Daten deutlich bessere Detektionsresultate erzielt werden. Fünf von sieben 3D-OD-Methoden erreichten mit den ausgedünnten RS-Punktwolken einen F1-Wert von über 95%, währenddem nur eine von sieben Methoden auf den ursprünglichen Punktwolken einen F1-Wert von über 95% erreichte (Tab. 3). Hinsichtlich Detektionsgenauigkeit und Inferenzzeit erwies sich der Objektdetektor Part-A<sup>2</sup> Net als die beste 3D-OD-Methode für unseren Anwendungsfall.

Durch das Neutrainieren des Objektdetektors konnte eine sehr hohe Detektionsgenauigkeit (Precision 100%, Recall 98%) für alle Fahrzeugtypen erzielt werden erreicht (Tab. 4). Allerdings war die Schätzung der Fahrzeugdimensionen für die Klassen „Van“ und „Truck“ ungenügend. Als Hauptgründe werden hierfür die geringe Anzahl an Trainingsdaten sowie die sehr heterogenen Geometrien (Differenzen in Länge und Höhe von bis zu 2m) dieser beiden Klassen gesehen.

Die erweiterte Erfassungssoftware konnte die Einschränkungen des Vorgängersystem weitestgehend eliminieren und das generierte Datenvolumen um eindruckliche 60'000-Mal reduzieren. Allerdings zeigte die Performanceevaluation ein großes Verbesserungspotenzial bei der Latenzzeit, die im Durchschnitt bei 0.53 Sekunden (1.9 FPS) liegt. Die Latenzzeit ist stark von den verfügbaren Rechenressourcen des Nvidia Jetson TX2 abhängig und sollte daher mit einem neuen, leistungsfähigeren Hardwaremodul deutlich reduziert werden können. Darüber hinaus wäre es möglich, die Inferenzzeit des Objektdetektors durch Modellkomprimierung und Beschleunigungstechniken zu verringern, wie sie beispielsweise in NOUSIAS et al. (2021) demonstriert wurden.

## 5 Fazit und Ausblick

Durch die Detektion geparkter Fahrzeuge basierend auf den neutrainierten Objektdetektor Part-A<sup>2</sup> Net direkt auf dem Erfassungssystem und dem dazugehörigen Auswertungsworkflow konnten alle anderen bekannten Arbeiten zur Erstellung von Parkstatistiken (BOCK et al. 2015; FETSCHER 2020; GRASSI et al. 2017; MATHUR et al. 2010; NEBIKER et al. 2021) hinsichtlich der Detektionsgenauigkeit, der Anwendbarkeit auf verschiedene Parkplatztypen und einer hohen Wiederholfrequenz übertroffen werden. Darüber hinaus konnten die Einschränkungen des Vorgängersystem weitestgehend eliminiert werden. Für die effiziente Erstellung von Parkstatistiken muss die Erfassungssystem allerdings mittels neuerer Hardware noch weiter beschleunigt werden. Außerdem sollte die Detektion von Fahrzeugen, insbesondere die Schätzung der Dimensionen, verbessert werden, indem zusätzliche Trainingsdaten direkt mit unserem MMS gesammelt werden. Alternativ könnten auch andere öffentliche Trainingsdatensätze verwendet werden, was jedoch eine Anpassung der

evaluierten Parameter für die Datenvorverarbeitung erfordern würde. Schließlich würde eine robuste Echtzeit-Sensororientierung, die eine Position im Submeterbereich in herausfordernden städtischen Umgebungen sicherstellt, dass Postprocessing gänzlich überflüssig machen und eine Vielzahl zusätzlicher Anwendungsfälle ermöglichen.

## 6 Literatur

- BOCK, F., EGGERT, D. & SESTER, M., 2015: On-street Parking Statistics Using LiDAR Mobile Mapping. Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spanien, 2812-2818. <https://doi.org/10.1109/ITSC.2015.452>.
- DENG, J., SHI, S., LI, P., ZHOU, W., ZHANG, Y & LI, H., 2021: Voxel R-CNN: Towards High Performance Voxel-based 3D Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence, **35**(2), 1201-1209.
- DORODNICOV, S. & HIRSHBERG, D., 2021: realsense2\_camera. [http://wiki.ros.org/realsense2\\_camera](http://wiki.ros.org/realsense2_camera), letzter Zugriff 17.10.2022.
- FETSCHER, S., 2020: Automatische Analyse von Streetlevel-Bilddaten für das digitale Parkplatzmanagement. Bachelorarbeit, Fachhochschule Nordwestschweiz FHNW, Muttenz, Schweiz (unveröffentlicht).
- GEIGER, A., LENZ, P. & URTASUN, R., 2012: Are we ready for autonomous driving? The KITTI vision benchmark suite. Proceedings of the Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 3354-3361, <https://doi.org/10.1109/CVPR.2012.6248074>.
- GRASSI, G., JAMIESON, K., BAHL, P. & PAU, G., 2017: Parkmaster: An in-vehicle, edge-based video analytics service for detecting open parking spaces in urban environments. Proceedings of the Second ACM/IEEE Symposium on Edge Computing, San Jose, CA, USA, 1-14, <https://doi.org/10.1145/3132211.3134452>.
- INTEL CORPORATION, 2020: Intel®RealSense - Product Family D400 Series: Datasheet. <https://www.intelrealsense.com/wp-content/uploads/2020/06/Intel-RealSense-D400-Series-Datasheet-June-2020.pdf>, letzter Zugriff am 17.10.2022.
- LANG, A. H., VORA, S., CAESAR, H., ZHOU, L. & YANG, J., BEJBOM, O., 2019: Pointpillars: Fast encoders for object detection from point clouds. Proceedings of the Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 12689-12697, <https://doi.org/10.1109/CVPR.2019.01298>.
- MATHUR, S., JIN, T., KASTURIRANGAN, N., CHANDRASEKARAN, J., XUE, W., GRUTESER, M. & TRAPPE, W., 2010: ParkNet. Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services - MobiSys '10, San Francisco, CA, USA, 123, <https://doi.org/10.1145/1814433.1814448>.
- NEBIKER, S., MEYER, J., BLASER, S., AMMANN, M. & RHYNER, S., 2021: Outdoor Mobile Mapping and AI-Based 3D Object Detection with Low-Cost RGB-D Cameras: The Use Case of On-Street Parking Statistics. Remote Sensing, **13**(16), 3099, <https://doi.org/10.3390/rs13163099>.
- NOUSIAS, S., PIKOULIS, E.-V., MAVROKEFALIDIS, C., LALOS, A. S. & MOUSTAKAS, K., 2021: Accelerating 3D scene analysis for autonomous driving on embedded AI computing platforms.

- International Conference on Very Large Scale Integration (VLSI-SoC), Singapur, 1-6, <https://doi.org/10.1109/VLSI-SoC53125.2021.9606990>.
- NVIDIA DEVELOPERS, 2022: Jetson TX2 Module. <https://developer.nvidia.com/embedded/jetson-tx2>, letzter Zugriff 17.10.2022.
- QUIGLEY, M., CONLEY, K., GERKEY, B., FAUST, J., FOOTE, T., LEIBS, J., BERGER, E., WHEELER, R. & NG, A., 2009: ROS: an open-source Robot Operating System. ICRA, **3**(3.2) 5.
- RAPP TRANS AG BASEL-STADT, 2019: Erhebung Parkplatzauslastung Stadt Basel 2019. Basel, Schweiz.
- SHI, S., GUO, C., JIANG, L., WANG, Z., SHI, J., WANG, X. & LI, H., 2020a: PV-RCNN: Point-voxel feature set abstraction for 3D object detection. Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 10526-10535, <https://doi.org/10.1109/CVPR42600.2020.01054>.
- SHI, S., WANG, X. & LI, H., 2019: PointRCNN: 3D object proposal generation and detection from point cloud. Proceedings of the Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 770-779, <https://doi.org/10.1109/CVPR.2019.00086>.
- SHI, S., WANG, Z., SHI, J., WANG, X. & LI, H., 2020b: From Points to Parts: 3D Object Detection from Point Cloud with Part-aware and Part-aggregation Network. IEEE Transactions on Pattern Analysis and Machine Intelligence, **1**, <https://doi.org/10.1109/tpami.2020.2977026>.
- YAN, Y., MAO, Y. & LI, B., 2018: Second: Sparsely embedded convolutional detection. Sensors, **18**(10), 3337, <https://doi.org/10.3390/s18103337>.