

Ein Ansatz zur automatisierten Erstellung von Trainingsdaten unter Verwendung von HD-Karten und Mehrfachbefahrungen

Frank Bieder* Haohao Hu[†] Johannes Schantz[‡] Oguzahn Kirik[§]
Florian Ries[¶], Martin Haueis^{||} und Christoph Stiller^{**}

Zusammenfassung:

In diesem Beitrag wird ein Gesamtsystem zur skalierbaren Erstellung von Trainingsdaten für das maschinelle Lernen im Kontext des automatisierten Fahrens vorgestellt. Unter Verwendung einer hochgenauen Lokalisierung in verifizierten HD-Karten und mithilfe von Sensor-Projektionsmodellen können semantische Kartenmerkmale in die verschiedenen Sensordomänen rückprojiziert werden. Hierdurch entsteht eine Korrespondenz von statischen Kartenmerkmalen und den entsprechenden Messaufnahmen, welche für das Training von neuronalen Netzen verwendet werden können. Durch Mehrfachbefahrungen und eine Multi-Drive-Kartierung kann die Anzahl der Trainingsdaten nach Belieben skaliert werden. Durch die Methode lassen sich statische Kartenmerkmale wie Fahrbahnlinsen und Straßenmarkierungen automatisiert annotieren. Es werden Verfahren vorgestellt, um sowohl mit statischen Verdeckungen, wie Pfosten, als auch mit dynamischen Verdeckungen, wie Fahrzeuge oder Fußgänger, umzugehen.

Für die Umsetzung der Methode wurde mit dem Versuchsfahrzeug BerthaOne ein Datensatz in den Städten Karlsruhe und Sindelfingen aufgenommen. Dieser besteht aus mindestens 20 Kilometern Strecke und umfasst 4 Befahrungen pro Strecke. Anhand dieses Datensatzes wird das System und die Qualität der erzeugten Trainingsdaten qualitativ evaluiert.

Schlüsselwörter: Automatisiertes Fahren, skalierbare Erstellung von Trainingsdaten, Deep Learning, HD Karten, Mehrfachbefahrungen

1 Einleitung

Für den sicheren und erfolgreichen Einsatz von hoch-automatisierten Fahrfunktionen wird ein umfassendes Verständnis der Umgebung benötigt. Hierbei werden immer mehr Aufgaben des Umgebungsverständnisses durch tiefe neuronale Netze gelöst. Dadurch ist es möglich, die Umgebung in Echtzeit einzuschätzen und entsprechende Handlungen daraus abzuleiten. Für die performante Anwendung von tiefen neuronalen Netzen werden große

*Frank Bieder ist Doktorand am FZI Forschungszentrum Informatik, E-mail: bieder@fzi.de.

[†]Haohao Hu ist Doktorand am MRT Institut für Mess- und Regelungstechnik, KIT.

[‡]Johannes Schantz ist CTO von enabl Technologies GmbH, zuvor Masterand am MRT, KIT.

[§]Oguzahn Kirik ist Doktorand an der TU Berlin, zuvor Masterand am MRT, KIT.

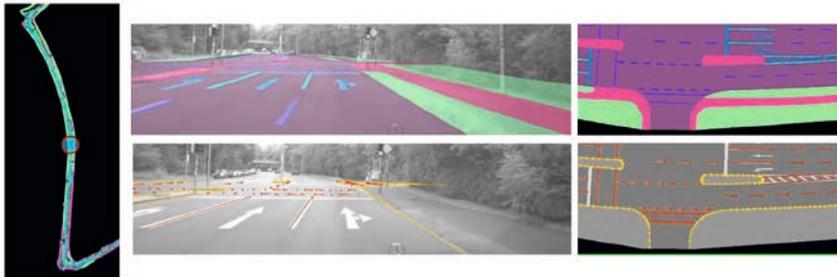
[¶]Florian Ries ist Entwicklungsingenieur bei der Mercedes-Benz Group AG.

^{||}Martin Haueis ist Teamleiter bei der Mercedes-Benz Group AG.

^{**}Christoph Stiller ist Professor am MRT Institut für Mess- und Regelungstechnik, KIT.



(a) Beispiel für eine der verwendeten HD-Karten in der Vogelperspektive. Der Kartenabschnitt zeigt eine Kreuzung in Karlsruhe und ist in drei verschiedenen Zoom-Stufen dargestellt. Für die Luft- und Satellitenbilder wurde Esri World Imagery [1] verwendet.



(b) Generierung von Trainingsdaten. Gegeben einer beliebigen 6D-Pose aus einer Befehrsung (links) können Kartenelemente in die verschiedenen Sensordomänen rückprojiziert werden. In dem Beispiel werden Trainingsdaten für semantische Segmentierung und für die Detektion von instanzbasierten Liniensegmenten generiert - jeweils für die Frontkamera- und Vogelperspektive.

Abbildung 1: Visualisierung der Datenbasis und exemplarische Darstellung der vorgestellten Methode für verschiedene Merkmalstypen und Sensordomänen.

und diverse Datensätze benötigt, welche möglichst genau auf das vorhandene Sensor-Setup und die Zieldomäne passen. Das Aufnehmen und Erstellen dieser Datensätze ist ein aufwendiger und kosten-intensiver Prozess. Hinzu kommt, dass sich durch den Wechsel des Sensor-Setups oder der Zieldomäne eine Lücke zwischen der Domäne der Sensordaten und der Domäne des Datensatzes bildet. Wird diese Lücke zu groß, muss der Prozess der Datensatzerstellung wiederholt werden. Die vorliegende Arbeit stellt eine Methode vor, um Datensätze für das maschinelle Lernen effizient und skalierbar durch die Verwendung von HD-Karten und einer Multi-Drive-Kartierungs-Methode automatisiert zu annotieren. Dabei werden Karten-Merkmale der HD-Karte in die Sensoraufnahmen rückprojiziert und dadurch die Ground Truth für die jeweiligen Sensoraufnahmen erzeugt. Die statischen Informationen einer Szene können aus verschiedenen Perspektiven erfasst und zur mehrfachen Annotation der Sensordaten verwendet werden. Durch die Datenrepräsentation in Form einer Karte reicht es also aus, Informationen einmal zu annotieren, um Trainingsdaten für unterschiedliche Sensoren und unterschiedliche Verkehrssituationen zu erstellen. Abbildung 1 stellt die Methode für verschiedene Merkmalstypen und Sensordomänen dar.



Abbildung 2: Auszüge der kartierten Datenbasis. Links und in der Mitte sind zwei Strecken aus Karlsruhe. Die rechte Strecke wurde in Sindelfingen kartiert. Im Datensatz sind unterschiedliche Szenen von verschiedener Komplexität wie zum Beispiel ein- oder mehrspurige Straßen im städtischen Umfeld mit Fahrradspuren und Fußgängerüberwegen und Kreuzungen enthalten, aber auch am Stadtrand gelegene Straßen, welche Landstraßen ähneln.

2 Stand der Technik

Die manuelle Annotation eines Datensatzes ist sehr zeitintensiv. Dieser hohe zeitliche Aufwand steigt mit der zunehmenden Komplexität der zu annotierten Merkmale und der zunehmenden Größe der benötigten Daten für moderne Modelle des maschinellen Lernens. Im Folgenden wird ein Überblick über die verschiedenen Ansätze zur Erstellung von Datensätzen im Kontext des automatisierten Fahrens gegeben.

Für die pixel-weise semantische Annotation eines Bildes des Cityscapes Datensatzes wurden laut [2] im Durchschnitt 1,5 Stunden benötigt. Hierbei muss jedoch beachtet werden, dass dies die Qualitätskontrolle der Annotation beinhaltet und der Datensatz zu seiner Zeit neue Maßstäbe an Annotations-Qualität gesetzt hat. [3] verweist auch auf eine Annotationszeit von 60 Minuten pro Bild im manuellen Fall und schlägt daher ein Verfahren vor, welches 2D-Labels im 3D-Raum erzeugt. Hierbei wird der Labelvorgang mithilfe von Formprimitiven direkt in der 3D-Punktwolke durchgeführt. Die Informationen werden anschließend in die 2D-Domäne projiziert. In [8] wird ein Ansatz vorgestellt, welcher eine akkumulierten LiDAR Punktwolke semantisch annotiert. Die semantischen Informationen können anschließend sowohl auf einzelne LiDAR Punktwolken als auch auf Kamerabilder projiziert werden. Inzwischen gibt es viele weitere Ansätze, welche versuchen, die Annotation von Trainingsdaten teil- zu automatisieren. Eine Möglichkeit hierfür ist die Verwendung von Methoden des maschinellen Lernens, um den Annotationsprozess zu unterstützen. [4] beschreibt ein Verfahren, bei welchem ein Mensch durch das ziehen einer Bounding Box oder das Verschieben von Polygonpunkten Bedingungen setzt, welche von einem Graph Neural Network interpretiert werden. Das Modell schlägt anschließend in jedem Iterationsschritt eine feinkörnigere Annotation vor. Dieser Vorgang wird so lange wiederholt, bis die erstellte Annotation von der annotierenden Person als korrekt akzeptiert wird. Synthetische Simulationen können auch dazu verwendet werden, mit verhältnismäßig wenig Aufwand eine große Menge an Trainingsdaten zu generieren. GTA5 [5] oder SYNTHIA [6] sind Beispiele für synthetische Datensätze. CARLA [7] hingegen bietet eine vollständige Simulationsumgebung um Trainingsdaten nach den eigenen Wünschen zu

generieren. Bei synthetischen Daten bleibt jedoch das Problem der Domän-Unterschiede zwischen realer Welt und synthetischer Welt bestehen, wodurch die Übertragbarkeit der Modelle in die reale Welt eingeschränkt ist. Neben [8] stellen auch die Datensätze Nusences [10], TuSimple und Argoverse [9] semantische Informationen in HD-Karten oder kartenähnlichen Datenstrukturen dar. Nusences liefert lediglich 2D Karten-Merkmale, wodurch eine Projektion in das Kamerabild ohne weiteres nicht möglich ist. TuSimple und Argoverse liefern 3D Kartenmerkmale sind allerdings auf Fahrstreifenränder konzentriert. Keiner dieser Datensätze liefert feingranularere Merkmale wie Straßenmarkierungen oder verfügt über mehrere Befahrungen mit einer Multi-Drive-Kartierung.

3 Automatisierte Erstellung von Trainingsdaten

Im Folgenden werden die Systemkomponenten vorgestellt, welche im Zusammenspiel die skalierbare Erstellung von Trainingsdaten ermöglichen.

Versuchsfahrzeug und Sensor-Setup

Die Messdaten wurden mit dem Versuchsfahrzeug BerthaOne [12] aufgezeichnet. Der Datensatz umfasst Daten, die von einem auf dem Dach montierten Velodyne Alpha Prime LiDAR, drei BlackFly PGE-50S5M-Kameras hinter der Front- und Heckscheibe und einem Ublox C94-M8P GNSS-Empfänger aufgenommen wurden. Der Alpha Prime LiDAR besitzt 128 Schichten und tätigt mit 10 Hz 360°-Aufnahmen. Die Kameras nehmen jeweils mit 10 Hz Bilder auf und der GNSS-Empfänger lieferte Messungen mit 5 Hz. Alle Sensoren wurden gemeinsam unter Verwendung der in [13] und [14] vorgestellten Ansätze intrinsisch und extrinsisch kalibriert. Der Datensatz besteht aus 7 Strecken mit einer Gesamtlänge von über 20 km, wobei 5 Stecken in Karlsruhe und 2 in Sindelfingen aufgenommen wurden. Drei Strecken sind in Abbildung 2 dargestellt. Hierbei wurden für alle Strecken mindestens 4 Befahrungen aufgenommen.

Multi-Drive-Kartierung und Mehrfachbefahrungen

In der vorliegenden Arbeit wird eine kamerabasierte SLAM-Methode [15] verwendet, welche in der Lage ist einen 6D-Posen-Graphen aus mehreren Befahrungen derselben Strecke zu erstellen. Jede weitere Befahrung erhöht die verfügbare Anzahl an potenzielle Trainingsdaten, einschließlich Kamerabildern und LiDAR-Punktwolken. Ein weiterer Vorteil der Multi-Drive-Kartierung ist die Verbesserung der Genauigkeit des resultierenden 6D-Posen-Graphs. Zusätzlich ermöglicht die Multi-Drive-Kartierung eine größere Bandbreite an Verkehrssituationen sowie Wetter- und Lichtverhältnissen derselben Szenerie. Dies erhöht die Diversität der Daten und kann damit die Robustheit der trainierten Modelle erhöhen. Je nach Anwendungsspezifikation und Charakteristik des urbanen Umfeldes kann die Anzahl der Befahrungen und damit die Skalierung der Trainingsdaten variiert werden. Hierbei ist zu beachten, dass der Mehrwert einer zusätzlichen Befahrung mit steigender Anzahl der Befahrungen abnimmt.

Manuelle Annotation von semantischen Kartenmerkmalen

Im Anschluss an die Multi-Drive-Kartierung kann unter Verwendung von Stereobildern ein Oberflächenabbild der Straße in der Vogelperspektive erstellt werden. Im Zuge dieser Arbeit wurde so anhand der ersten Befahrung einer Strecke eine Rasterkarte mit einer Auflösung von 2cm/Pixel erzeugt. Auf den dadurch entstandenen Rasterkarten sind Straßenmarkierungen und Straßenbegrenzungen sehr gut erkennbar und können manuell annotiert werden. Für weiteren Kontext können sowohl die entsprechenden Kamerabilder als auch Luft- oder Satellitenbilder hinzugezogen werden. Da die Höhe der Kartenmerkmale direkt aus den Stereobildern geschätzt wird, müssen 3D Merkmale auf der Straßenoberfläche lediglich in 2D annotiert werden. Im Zuge dieser Arbeit wurden 6 verschiedene Klassen von Kartenmerkmalen annotiert. Hierbei wurde zwischen drei Arten von Straßenmarkierungen und auch drei verschiedenen Flächen-Elementen unterschieden.

Kartenformat

Als Kartenformat wurde das Lanelet2-Framework [16] verwendet, welches speziell für den Einsatz im Bereich des automatisierten Fahrens entwickelt wurde. Es zeichnet sich unter anderem durch die Einsatzmöglichkeiten in verschiedenen Anwendungen wie Lokalisierung oder Bewegungsplanung aus. Bei der Entwicklung von Lanelet2 wurden Genauigkeit, Vollständigkeit, Verifizierbarkeit und Erweiterbarkeit als Anforderungen an Karten für das automatisierte Fahren identifiziert und das Kartenformat entsprechend konzipiert. Die in dieser Arbeit erstellten HD-Karten enthalten semantische Informationen, welche in 3D-Raum als offene oder geschlossene Linienzüge definiert werden.

Repräsentation der Trainingsdaten

Aus dieser allgemeinen Kartenrepräsentation können Trainingsdaten in einer Vielzahl von Merkmals-Darstellungen erzeugt werden, einschließlich Linien und Flächen. Daher ist die Daten-Generierung nicht auf eine bestimmte Merkmalsrepräsentation limitiert. Es werden zwei Repräsentationen vorgestellt, welche in dieser Arbeit verwendet werden:

- Semantische Segmentierung: Die Zuordnung einer semantischen Klasse zu jedem Pixel bzw. jeder Zelle einer Rasterkarte.
- Detektionsverfahren für Liniensegmente nach [17]: Hierbei handelt es sich um ein instanzbasiertes Verfahren, welches Liniensegmente klassifiziert und lokalisiert.

Hierbei können die Kartenmerkmale in verschiedenen Sensordomänen rückprojiziert werden, um Trainingsdaten für verschiedene Sensortypen zu generieren. In dieser Arbeit werden zwei Sensor-Projektionsmodelle zur Erzeugung von Trainingsdaten verwendet, Frontview-Kamera und die Birds-Eye-View(BEV)-Domäne für Lidar-Rasterkarten.

Verdeckungen bei der Rückprojektion von Sensordaten

Bei der Erzeugung der Ground Truth für Kamerabilder besteht eine weitere wichtige Aufgabe darin Verdeckungen zu erkennen. Dies sollte verhindern, dass bei der Rückprojektion von Kartenmerkmalen in die Sensordaten nicht-sichtbare Kartenmerkmale als Elemente der Ground Truth gelabelt werden. Hierbei kann zwischen zwei Arten der Verdeckung unterschieden werden.

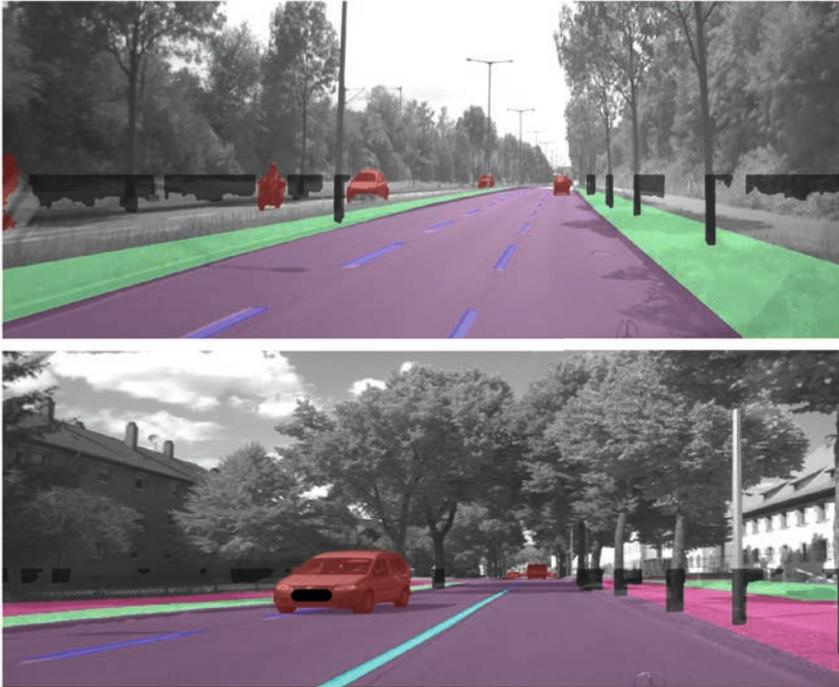


Abbildung 3: Qualitative Evaluation der automatisch generierten Trainingsdaten in Hinblick auf dynamische und statische Verdeckungen für den Fall der semantischen Segmentierung. In den Beispielen wurden alle statischen Klassen direkt aus einer HD-Karte rückprojiziert. Hingegen sind die dynamischen Objektklassen von einem neuronalen Netz geschätzt, um dynamische Verdeckungen zu identifizieren. Die grauen Flächen wurden als statische Verdeckungen mithilfe einer Stereo-Kamera identifiziert.

- **Dynamische Verdeckung:** Als dynamische Verdeckung werden sowohl Personen als auch Fortbewegungsmittel wie LKWs, Busse, Züge, Fahrräder, Motorräder und Autos bezeichnet.
- **Statische Verdeckung:** Als statische Verdeckungen werden unbewegliche Objekte bezeichnet, welche den Blick auf ein rückprojiziertes Kartenmerkmal verdecken können.

Da in beiden Fällen der Ansatz auf Sensordaten des aktuellen Zeitpunktes basiert, können nicht erkannte dynamische Verdeckungen auch durch den Ansatz für statische Verdeckung identifiziert werden. Allerdings ist der Ansatz für dynamische Verdeckungen besonders effektiv. Hierbei wird auf öffentliche Datensätze für semantische Segmentierung zurückgegriffen (Cityscapes [2], Mapillary [19]), welche Annotationen für die oben

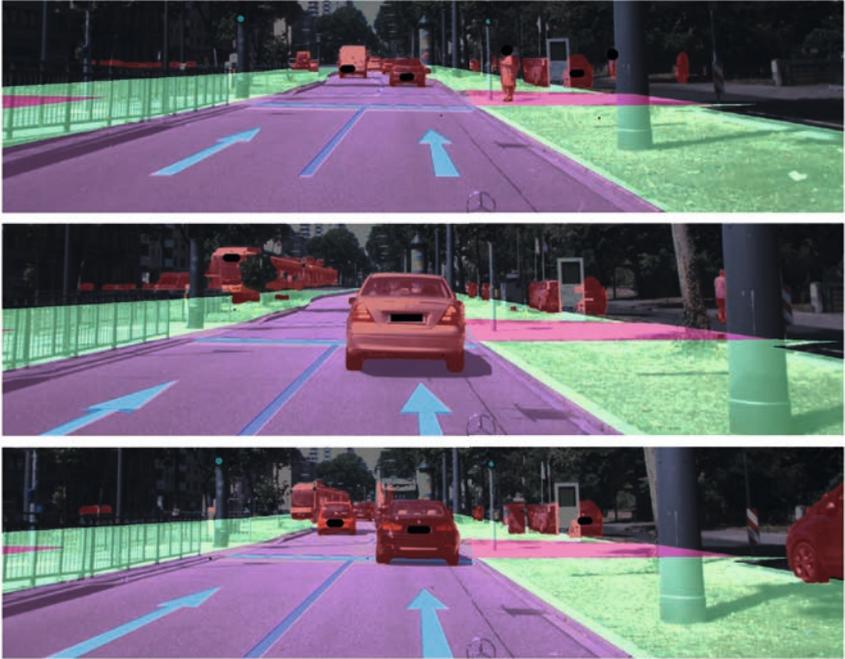


Abbildung 4: Qualitative Evaluation der Güte der Multi-Drive-Kartierung für den Fall der semantischen Segmentierung. Die Aufnahmen stammen aus drei verschiedenen Befahrungen derselben Strecke und wurden zu unterschiedlichen Zeitpunkten aufgenommen. Die Kartenmerkmale einer HD-Karte können in beliebig viele Befahrungen rückprojiziert werden, sofern es die Güte des 6D-Posen-Graphen zulässt.

genannten dynamischen Objektklassen besitzen. Unter Verwendung von großen, nicht-online fähigen Netzarchitekturen, wie beispielsweise SeamSeg: Seamless Scene Segmentation [18], werden diejenigen Pixel identifiziert, welche zu einem dynamischen Objekt gehören. Hieraus werden binäre Masken generiert, welche über die Annotation von rückprojizierten Kartenelementen entscheiden. Anschließend kann anhand von Tiefendaten aus der Stereokamera abgeschätzt werden, ob eine statische Verdeckung vorliegt. Hierbei wird nach homogenen Flächen im Tiefenbild gesucht, welche sich in der Sichtlinie eines Kartenobjektes befinden und eine deutlich geringere Distanz zu der Kamera aufweisen als das Kartenelement vermuten lässt. Hierbei können beispielsweise Pfosten, Bäume oder Häuserecken an Kreuzungen erkannt werden und die Rückprojektion entsprechend angepasst werden. Dieses Verfahren ist stark von der Qualität des Tiefenbildes abhängig und hat sich als weniger robust als das Verfahren für dynamische Verdeckungen erwiesen.

4 Fazit

In diesem Beitrag wurde ein Gesamtsystem vorgestellt, welches in der Lage ist anhand von HD-Karten, einer hochgenauen Lokalisierung und Mehrfachbefahrungen skalierbar und automatisiert Trainingsdaten zu generieren. In Abbildung 3 ist in zwei Beispielen die Rückprojektion der Kartenmerkmale unter Berücksichtigung der statischen und dynamischen Verdeckung dargestellt. Abbildung 4 zeigt hingegen, wie durch die Multi-Drivekartierung die Menge der automatisch generierten Trainingsdaten je nach Bedarf skaliert werden kann. Die Qualität der Annotation hängt von verschiedenen Komponenten des Systems ab: (1) Genauigkeit der HD-Karte inklusive der 3D-Position der Kartenelemente, (2) Genauigkeit der 6D-Pose des Fahrzeugs in der HD Karte, (3) Genauigkeit der Kalibrierung der Sensoren und Rückprojektion in die Sensordomäne und (4) Genauigkeit der Verdeckungserkennung.

Um die Stärke des Ansatzes besser nutzen zu können, sollten Befahrungen mit einer großen Bandbreite von Verkehrssituationen, Wetter- und Lichtverhältnissen derselben Szenerie verwendet werden, um die Diversität des Datensatzes zu erhöhen. Aus der damit verbundenen zeitversetzten Aufnahme der Mehrfachbefahrungen ergeben sich Anforderungen an die Langzeitstabilität der Lokalisierungsmerkmale und die Möglichkeit der HD-Kartenverifikation.

Eine weitere Anwendung der Methode ist die Möglichkeit, semantisches Wissen aus einzelnen Sensordomänen automatisiert zu kartieren und durch das Lernen aus Karten das semantische Wissen auf das neue Sensor-Setup eines Fahrzeuges zu übertragen. Hierdurch könnte der Aufwand für die Erstellung von Trainingsdaten für neue Fahrzeugkonfigurationen stark reduziert werden.

Literatur

- [1] Esri *World Imagery*, https://\{switch:services,server\}arcgisonline.com/arcgis/rest/services/World_Imagery/MapServer/tile, © Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swiss-topo, and the GIS User Community. Accessed 15. August 2023 via JOSM.
- [2] C. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth und B. Schiele, *The Cityscapes Dataset for Semantic Urban Scene Understanding*, International Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [3] J Xie, M Kiefel, M T Sun und A Geiger, *Semantic Instance Annotation of Street Scenes by 3D to 2D Label Transfer*, International Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [4] H. Ling, S. Fidler, *Fast Interactive Object Annotation with Curve-GCN*, International Conference on Computer Vision (ICCV), 2019.
- [5] S. R. Richter, V. Vineet, S. Roth und V. Koltun, *Playing for Data: Ground Truth from Computer Games*, European Conference on Computer Vision (ECCV), 2016.

- [6] G. Ros, L. Sellart, J. Materzynska, D. Vazquez und A. M. Lopez, *The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes*, International Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [7] A. Dosovitskiy, G. Ros, F. Codevilla, A. M. Lopez und V. Koltun, *CARLA: An Open Urban Driving Simulator*, CoRR, 2017.
- [8] Y. Liao, J. Xie und A. Geiger, *KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D*, Pattern Analysis and Machine Intelligence (PAMI), 2022.
- [9] B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. Kaesemodel Pontes, D. Ramanan, P. Carr und J. Hays, *Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting*, arXiv, 2023.
- [10] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan und O. Beijbom, *nuScenes: A multimodal dataset for autonomous driving*, International Conference on Computer Vision (ICCV), 2019.
- [11] TuSimple, *TuSimple benchmark*, <https://github.com/TuSimple/tusimple-benchmark>. accessed August 2023.
- [12] Ö. Ş. Taş, N. O. Salscheider, F. Poggenhans, S. Wirges, C. Bandera, M. R. Zofka, usw. *Making Bertha Cooperate—Team AnnieWAY’s Entry to the 2016 Grand Cooperative Driving Challenge*, Transactions Intelligent Transportation Systems, 2018.
- [13] T. Strauß, J. Ziegler, J. Beck, *Calibrating Multiple Cameras with Non-Overlapping Views using Coded Checkerboard Targets*, Intelligent Transportation Systems Conference (ITSC), 2014.
- [14] J. Kümmerle, T. Kühner, M. Lauer, *Automatic Calibration of Multiple Cameras and Depth Sensors with a Spherical Target*, International Conference Intelligent Robots and Systems (IROS), 2018.
- [15] M. Sons, C. Stiller, *Efficient Multi-Drive Map Optimization towards Life-long Localization using Surround View*, International Conference on Intelligent Transportation Systems (ITSC), 2018.
- [16] F. Poggenhans, J.-H. Pauls, J. Janosovits, S. Orf, M. Naumann, F. Kuhnt und M. Mayr, *Lanelet2: A high-definition map framework for the future of automated driving*, International Conference on Intelligent Transportation Systems (ITSC), 2018.
- [17] A. Meyer, P. Skudlik, J.-H. Pauls und C. Stiller, *YOlinO: Generic Single Shot Polyline Detection in Real Time*, International Conference on Computer Vision Workshops, 2021.
- [18] L. Porzi, S. Rota Bulò, A. Colovic und P. Kotschieder, *Seamless Scene Segmentation*, International Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

- [19] G. Neuhold, T. Ollmann, S. Rota Bulò und P. Kotschieder, *The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes*, International Conference on Computer Vision (ICCV), 2017.