DOI: 10.58895/ksp/1000174544-13 erschienen in:
Proceedings - 34. Workshop Computational Intelligence: Berlin, 21.-22. November 2024
DOI: 10.58895/ksp/1000174544 | https://www.ksp.kit.edu/site/books/e/10.58895/ksp/1000174544/

Modalübergreifende Wissensdestillation für die Radar-Objekterkennung

Patrick Palmer, Martin Krüger und Torsten Bertram

Lehrstuhl für Regelungssystemtechnik
Technische Universität Dortmund
Otto-Hahn-Straße 8, 44227 Dortmund
E-Mail: patrick.palmer@tu-dortmund.de

1 Einleitung

Die Umgebungswahrnehmung stellt das erste Modul der Informationsprozesskette in jedem automatisierten Fahrsystem dar, wobei Sensormodalitäten wie Lidar, Kamera und Radar verwendet werden. Gegenwärtig werden vornehmlich Lidar und Kamera eingesetzt, da sie eine hochauflösende Repräsentation der Fahrzeugumgebung liefern. Lidar-Systeme sind jedoch mit hohen Kosten verbunden, während Kamerasysteme eine starke Abhängigkeit von der Umfeldausleuchtung aufweisen. Radar-Sensoren weisen diese Limitierungen nicht auf und bieten zudem Vorteile wie die direkte Messung der relativen Radialgeschwindigkeit. Die Einführung bildgebender 3+1D[1] Radar-Sensoren, welche gegenüber herkömmlichen Radaren eine höhere Auflösung erreichen, ermöglichen eine rein radarbasierte Wahrnehmung.

Alle aktuell verfügbaren Datensätze zur Radar-Objektdetektion enthalten zusätzlich zum Radar- auch einen Lidar-Sensor [1–3]. Zumeist dient der Lidar in diesen Datensätzen als Referenz, um Ground-Truth Label zu erstellen. Dies eröffnet die Möglichkeit, den Lidar im Trainingsprozess maschineller Lernverfahren einzusetzen, wobei zur Interferenz weiterhin nur der Radar-Sensor verwendet wird. Für den Transfer zwischen Sensormodalitäten werden zumeist Methoden der Wissensdestillation eingesetzt [4]. Aufgrund der identischen

[1] 3 räumliche Dimensionen: Entfernung, Azimut, Elevation, + relativen Radialgeschwindigkeit

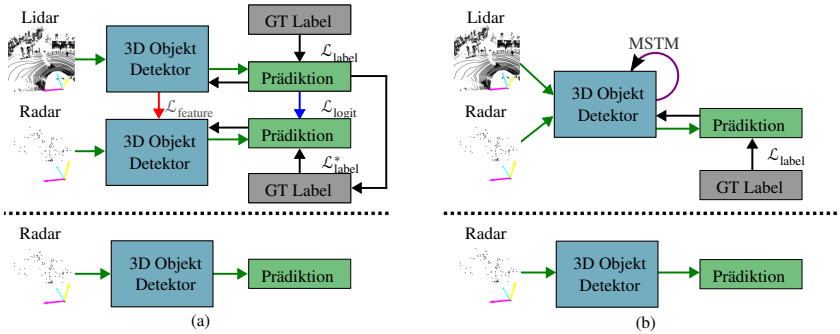

Bild 1: Überblick über die (a) auf Wissensdestillation basierte Methode und (b) die mehrstufige Trainingsmethode zur Nutzung von Lidar-Daten für das Training von Radar-Objektdetektoren. Der Trainingsprozess ist oberhalb der gestrichelten Linie, die Inferenz unterhalb dargestellt.

Datenrepräsentation als Punktwolke eröffnen sich für den Transfer zwischen Lidar- und Radar-Detektion neue Möglichkeiten, da die gleiche Basisarchitektur auf Basis neuronaler Netze verwendet werden kann.

2 Methodik

Im Rahmen dieser Arbeit werden zwei Ansätze untersucht, Wissen aus der Lidar-Objektdetektion während des Lernprozesses auf die Radar-Objektdetektion zu übertragen. Es handelt sich hierbei um einen Ansatz der Wissensdestillation (WD), welcher aus dem Bereich des rechenzeiteffizienten Netzarchitekturoptimierung abgeleitet ist, und eine mehrstufige Trainingsmethode (MSTM). Beide Ansätze sind in Bild 1 dargestellt.

2.1 Mehrstufige Trainingsmethode (MSTM)

Der MSTM-Ansatz wird wie in [5] beschrieben verwendet. Das Netzwerk wird zunächst ausschließlich auf Lidar-Daten bis zur Konvergenz trainiert. Anschließend wird eine mehrstufige Optimierung der Modellparameter durchgeführt. Die Dichte der Lidar-Punktwolke wird in jeder Optimierungsstufe halbiert. Dadurch nähert sich die Lidar-Punktwolke der Charakteristik der Radar-Punktwolke weitestgehend an; was das Netzwerk konditioniert, auf niedrig aufgelösten

Punktwolken Objekte zu detektieren. Im vorletzten Optimierungsschritt werden die Lidar- und die Radar-Punktwolke eingesetzt. Im letzten Optimierungsschritt wird dann nur noch die Radar-Punktwolke verwendet. Der MSTM-Ansatz ist schematisch in Bild 2 dargestellt. [5] hat gezeigt, dass es vorteilhaft ist die Radar-Punktwolke in jedem Trainingsschritt mit der Lidar-Punktwolke zu kombinieren, weshalb diese Kombination auch hier angewendet wird.

Für die Ausdünnung der Lidar-Punktwolke werden drei unterschiedliche Methoden untersucht. Eine zufällige Ausdünnung, eine k-nächste Nachbarn Ausdünnung, welche die Lidar-Punkte mit dem größten euklidischen Abstand zum nächsten Radar-Punkt entfernt, und eine Voxel-basierte Ausdünnung, welche die Anzahl der Lidar-Punkte in jedem Voxel zufällig reduziert, wobei jedoch eine Mindestanzahl an Punkten pro Voxel erhalten bleibt.

2.2 Wissensdestillation (WD)

Für die Wissensdestillation wird ein Detektor nur auf Lidar-Daten trainiert; dieser wird als Lehrer bezeichnet. Anschließend werden für das Training des Radar-Detektors, welcher als Schüler bezeichnet wird, unterschiedliche Kostenfunktionen $\mathcal{L}$ bezogen auf das Lidar-Modell berechnet. In dieser Arbeit werden drei Kostenfunktionen, wie in [6] beschrieben, betrachtet. $\mathcal{L}_{\text{logit}}$ und $\mathcal{L}^*_{\text{label}}$ bilden die Differenz der Lehrer und Schüler-Prädiktion. Der Unterschied beider besteht darin, dass für $\mathcal{L}^*_{\text{label}}$ nur maximale Detektionen verwendet werden. $\mathcal{L}_{\text{feat}}$ bildet die Differenz der Feature-Dimension des Lehrers und Schülers. In allen Fällen wird das Schüler-Modell zunächst mit den Parametern des Lehrer-Models initialisiert, da dies zu einer Verbesserung gegenüber einer zufälligen Initalisierung führt [6].

3 Auswertung

Alle Experimente werden auf dem View-of-Delft Datensatz [1] durchgeführt. Dieser beinhaltet einen 64-Schichten Lidar-Sensor sowie einen 3+1D hochauflösenden Radar-Sensor. Zur Evaluation wird die durchschnittliche Genauigkeit (engl. mean average precision (mAP)), wie durch [7], [8] definiert, verwendet.

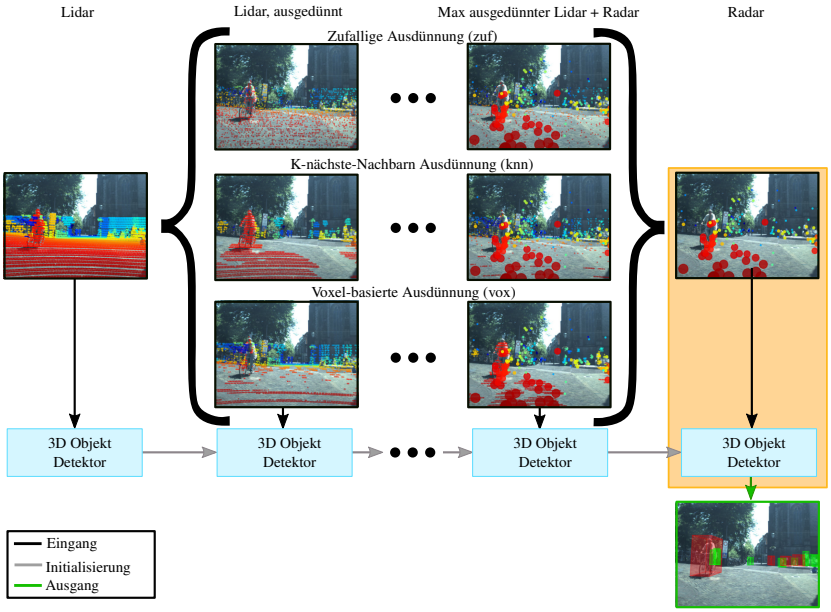

Bild 2: Schema der MSTM. Bei der Inferenz wird nur der orange schattierte Teil ausgeführt. Kleine Punkte stellen Lidar-Punkte dar, große Punkte Radar-Punkte. Die Farbe der Punkte entspricht dem Abstand zum Ego-Fahrzeug. Das Kamerabild wird nicht als Eingabe verwendet, sondern dient nur der Visualisierung.

Für eine Auswertung in unterschiedlichen Weitenbereichen wird die Auswertung in einen kurzen Messbereich (KM): 0-30 m und einen weiten Messbereich (WM): 30-50 m unterteilt [9] . Als Objektdetektor wird ein auf Radar-Daten angepasstes PointPillars [10] verwendet.

Tabelle 1, stellt die Ergebnisse der MSTM für unterschiedliche Ausdünnungsmethoden dar. Es ist zu erkennen, dass insbesondere die MSTM mit vox und knn Ausdünnung zu einer Verbesserung der Detektionsleistung führen. Wobei die Leistung von der Verkehrsteilnehmerklasse und dem Messbereich abhängt.

In Tabelle 2 sind die Ergebnisse für unterschiedliche WD-Methoden dargestellt. Zusätzlich zum Training des Lehrers auf Lidar-Daten (L) wird ein Training auf einer Punktwolke aus Radar- und einer auf ¼ der ursprünglichen Punktdichte ausgedünnten Lidar-Punktwolke ($RL_{¼\,/\,vox}$) betrachtet. Es ist zu erkennen, dass alle Methoden zu einer Verbesserung im WM führen. Die Detektionsleistung, die mit einem WD-Verfahren erzielt werden kann, hängt von dem verwendeten

Tabelle 1: 3D Objektdetektionsergebnisse mit PointPillars als Detektor und der MSTM. Das beste und zweitbeste Ergebnis ist jeweils in **Fett** und Unterstrichen gekennzeichnet. R: Radar, LR: Lidar + Radar, 1-1/16: Genutzte Lidar Ausdünnungsschritte $1 \rightarrow 1/2 \rightarrow 1/4 \rightarrow 1/8 \rightarrow 1/16$.

	mAP ↑		Auto		Fußgänger		Radfahrer	
Trainingsmethode	KM	WM	KM	WM	KM	WM	KM	WM
R (Vergleichsbasis)	36,7	11,9	45,2	18,1	17,1	7,4	47,7	10,2
$RL^{MSTM}_{1\text{-}1/16\ /\ zuf} \rightarrow R$	35,6	14,9	44,1	20,2	17,8	7,2	44,9	**17,3**
$RL^{MSTM}_{1\text{-}1/16\ /\ knn} \rightarrow R$	38,2	14,7	45,5	**23,9**	**18,8**	8,2	50,3	11,9
$RL^{MSTM}_{1\text{-}1/16\ /\ vox} \rightarrow R$	**39,7**	**15,4**	**45,9**	22,5	18,4	**9,7**	**54,7**	13,9

Tabelle 2: 3D Objektdetektionsergebnisse mit PointPillars als Detektor und WD. Das beste Ergebnis pro WD-Methode ist **Fett** gekennzeichnet.

	Nur Init		Logit-WD		Feature-WD		Label-WD	
Lehrer-Daten	KM	WM	KM	WM	KM	WM	KM	WM
R (Vergleichsbasis)	36,7	11,9	-	-	-	-	-	-
L	34,9	13,5	32,8	12,6	34,8	13,6	**36,6**	**13,7**
$RL_{1/4\ /\ vox}$	35,2	15,3	35,6	13,3	**37,1**	**15,8**	35,3	14,3

Lehrer-Datensatz ab. Die Feature-WD profitiert von den Lehrer-Daten die strukturell nah an den Schüler-Daten sind, während die Label-WD von Lehrer-Daten profitiert die eine hohe Detektionsleistung ermöglichen.

4 Zusammenfassung

Diese Arbeit präsentiert zwei unterschiedliche Methoden, um Lidar-Daten im Trainingsprozess von Radar-Objektdetektoren zu nutzen. Die Auswertung zeigt, dass beide Methoden zu einer Verbesserung der Detektionsleistung führen. Die MSTM mit Voxel-basierter Ausdünnung führt dabei zum besseren Ergebnis und kann die Detektionsleistung um bis zu 3,5 Prozentpunkte steigern. Die Leistungsfähigkeit der MSTM hängt von der verwendeten Ausdünnungsmethode ab, während die Leistungsfähigkeit der WD-Methoden von dem verwendeten Lehrer-Datensatz bestimmt wird.

Literatur

[1] A. Palffy , E. Pool, S. Baratam, J. F. Kooij, D. M. Gavrila: „Multi-Class Road User Detection With 3+1D Radar in the View-of-Delft Dataset“. In: *IEEE Robotics and Automation Letters*. 2022.

[2] L. Zheng et al.: „TJ4DRadSet: A 4D Radar Dataset for Autonomous Driving“. In: *25th International Conference on Intelligent Transportation Systems (ITSC)*. 2022.

[3] X. Zhang et al.: „Dual Radar: A Multi-modal Dataset with Dual 4D Radar for Autononous Driving“. In: *arXiv preprint arXiv:2310.07602*. 2023.

[4] M. Klingner et al.: „X3KD: Knowledge Distillation Across Modalities, Tasks and Stages for Multi-Camera 3D Object Detection“. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023.

[5] P. Palmer, M. Krüger, S. Schütte, R. Altendorfer, G. Adam, T. Bertram: „LEROjD: Lidar Extended Radar-Only Object Detectiont“. In: *European Conference on Computer Vision (ECCV)*. 2024.

[6] J. Yang et al.: „Towards Efficient 3D Object Detection with Knowledge Distillation“. In: *Advances in Neural Information Processing Systems*. 2022.

[7] A. Geiger, P. Lenz, R. Urtasun: „Are we ready for autonomous driving? the kitti vision benchmark suite“. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012.

[8] P. Palmer, M. Krüger, R. Altendorfer, G. Adam, T. Bertram: „Reviewing 3d object detectors in the context of high-resolution 3+1d radar“. In: *Workshop on 3D Vision and Robotics at the Conference on Computer Vision and Pattern Recognition*. 2023.

[9] G. Zamanakos et al.:„A comprehensive survey of lidar-based 3d object detection methods with deep learning for autonomous driving“. In: *Computers and Graphics*. 2021.

[10] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, O. Beijbom: „PointPillars: Fast Encoders for Object Detection from Point Clouds“. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019.