

3D-Deep Learning zur Erkennung kleiner Objekte in LiDAR-Daten auf On-Board-Systemen

Yousef Abdullah | Referentin: Prof. Dr. Eleke Hergenröther | Korreferentin: Prof. Dr. Thomas März

Darmstadt University of Applied Sciences

Motivation



Abbildung 1: Formula-Student-Rennfahrzeug des FASTDA-Projekts.

Im FASTDA-Projekt muss ein autonomes Formula-Student-Rennfahrzeug eine durch Verkehrshütchen markierte Strecke selbstständig erkennen und befahren. Die zuverlässige Detektion dieser Hütchen ist daher entscheidend für die Trajektorienplanung und den sicheren Fahrzeugbetrieb. LiDAR-Sensoren liefern präzise 3D-Informationen und ermöglichen eine direkte räumliche Lokalisierung. Bei kleinen Objekten wie Verkehrshütchen nimmt jedoch mit zunehmender Entfernung die Punktdichte stark ab. Dadurch werden die Objekte häufig nur durch wenige Messpunkte repräsentiert, was die Merkmalsextraktion und Detektion deutlich erschwert. Diese Arbeit untersucht daher, welche 3D-Deep-Learning-Modelle kleine, punktarme Objekte zuverlässig erkennen können und gleichzeitig die Echtzeitanforderungen eines ressourcenbeschränkten On-Board-Systems erfüllen.

Forschungsfragen

- Wie zuverlässig können verschiedene 3D-Deep-Learning-Ansätze kleine, punktarme Objekte wie Verkehrshütchen in LiDAR-Punktwolken unter Echtzeitbedingungen detektieren?
- Welches Verhältnis zwischen Erkennungsleistung und Rechenaufwand weisen diese Modelle beim Einsatz auf ressourcenbeschränkten On-Board- Systemen autonomer Rennfahrzeuge auf?

Methodik

Zur Beantwortung der Forschungsfragen wurde folgende Vorgehensweise durchgeführt:

1. **Datenerhebung:** Aufnahme eines eigenen LiDAR-Datensatzes mit einem Ouster OS2-128 Sensor unter realen Bedingungen des FASTDA-Projekts.
2. **Annotation:** Manuelle Annotation der Verkehrshütchen und Konvertierung des Datensatzes in das KITTI-Format.
3. **Transfer Learning:** Feinjustierung von sechs vortrainierten 3D-Objektdetektoren (PointPillars, 3DSSD, PV-RCNN, DeLiVoTr, LION und VoxelNeXt) auf dem eigenen Datensatz.
4. **Evaluation:** Vergleich der Modelle hinsichtlich Precision, Recall, F1-Score, AP@0.5 sowie Inferenzgeschwindigkeit (FPS) zur Bewertung der Echtzeiteignung.

Table 1. Kennzahlen des im Rahmen dieser Arbeit erhobenen und annotierten Datensatzes.

Kennzahl	Anzahl
Gesamtanzahl Frames	169
Instanzen <i>small_cone</i>	1.642
Instanzen <i>large_cone</i>	49
Gesamtanzahl Instanzen	1.691
Trainingsframes	135 (80 %)
Testframes	34 (20 %)

Ergebnisse

Table 2. Vergleich der evaluierten Modelle für die Klasse *small_cone*.

Modell	Precision (%)	Recall (%)	F1 (%)	AP@0.5 (%)	FPS	Echtzeit
PointPillars	61,2	58,8	60,0	50,7	48,3	✓
3DSSD-24k	68,77	73,98	71,9	50,7	22,0	✓
PV-RCNN	57,1	64,6	60,7	51,5	10,2	✓
DeLiVoTr	93,6	59,3	72,6	50,4	4,7	×
LION	92,0	84,6	88,2	83 ,2	5,4	×
VoxelNeXt	89,1	84,6	86,8	79,6	55,0	✓

Anmerkung: ✓ = Echtzeitanforderung erfüllt (≥ 10 FPS), × = nicht erfüllt.

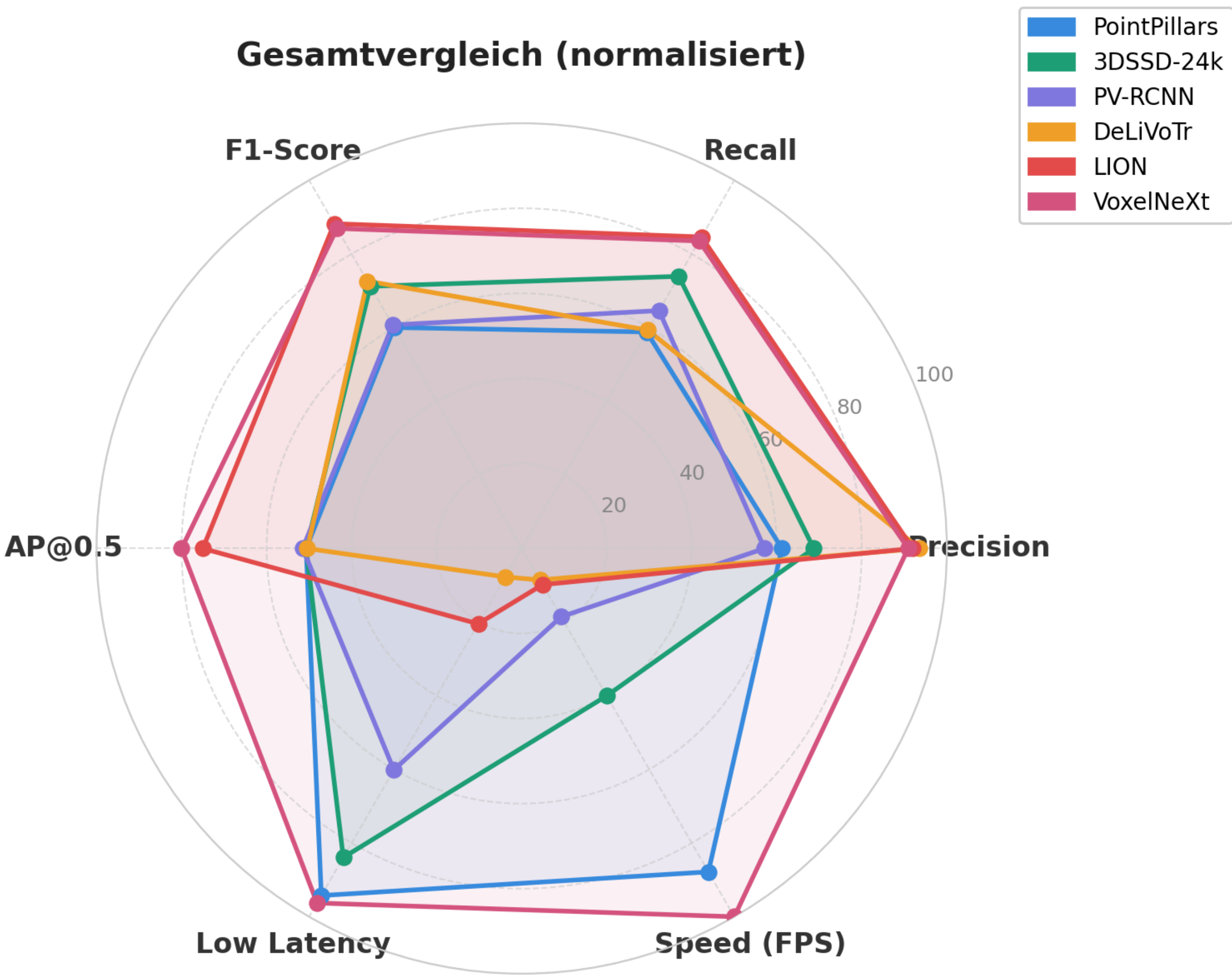


Abb.: Gesamtvergleich aller evaluierten Modelle. Speed und Low Latency: höher = besser (invertiert).

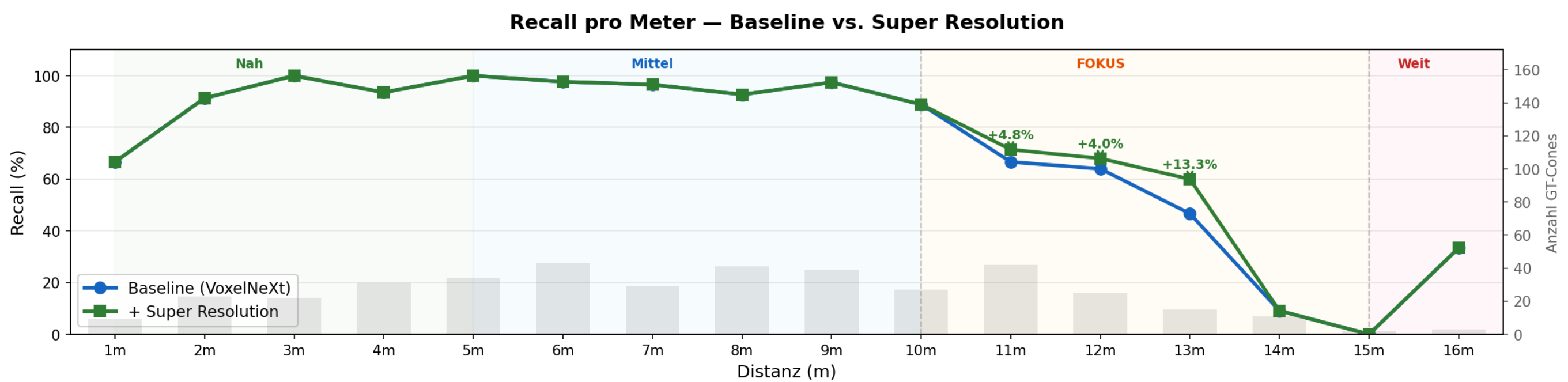
Die Ergebnisse zeigen, dass moderne 3D-Deep-Learning-Verfahren kleine und punktarme Objekte grundsätzlich zuverlässig detektieren können. Die Detektionsleistung hängt jedoch maßgeblich von der gewählten Repräsentationsstrategie der Punktwolke, der Entfernung zum Sensor sowie der zugrunde liegenden Modellarchitektur ab. LION erzielt mit einem F1-Score von 88,2 % die höchste Detektionsleistung, verfehlt jedoch mit 5,4 FPS die Echtzeitanforderung. VoxelNeXt erreicht mit 86,8 % F1 eine vergleichbare Erkennungsleistung, verarbeitet die Punktwolken jedoch mit 55 FPS und erfüllt damit als einziges Modell sowohl die Genauigkeits- als auch die Echtzeitanforderungen des FASTDA-Projekts. Eine gemeinsame Limitation aller untersuchten Modelle ist die distanzabhängige Abnahme der Detektionsleistung. Mit zunehmender Entfernung sinkt die Punktdichte der LiDAR-Daten deutlich, sodass kleine Objekte nur noch durch wenige Messpunkte repräsentiert werden und die Erkennung insbesondere ab etwa 10 m Entfernung deutlich erschwert wird.

Verbesserung der Erkennungsleistung

Zur Verbesserung der Detektion kleiner und punktarmer Objekte im kritischen Fernbereich ab etwa 10 m wurden zwei Optimierungsansätze untersucht.

- **Shape Priors:** Integration zusätzlicher geometrischer Forminformationen zur Unterstützung der Merkmalsextraktion bei spärlichen Punktwolken. Aufgrund begrenzter Sequenzdaten, hoher Architekturkomplexität und geringer Wirksamkeit bei sehr geringer Punktdichte wurde dieser Ansatz nicht weiterverfolgt.
- **Super Resolution:** Datengetriebene Verdichtung der LiDAR-Punktwolke durch Interpolation im Range Image, um die Informationsdichte kleiner Objekte vor der Detektion zu erhöhen.

Die Evaluation zeigt, dass Super Resolution den Recall insbesondere im kritischen Distanzbereich von 10–13 m verbessert. Der deutliche Leistungsabfall bei größeren Entfernungen kann jedoch nicht vollständig kompensiert werden, da kleine Objekte dort nur noch durch wenige LiDAR-Messpunkte repräsentiert werden.



Fazit

Die Ergebnisse zeigen, dass kleine Verkehrshütchen in LiDAR-Punktwolken mit modernen 3D-Deep-Learning-Modellen grundsätzlich zuverlässig erkannt werden können. Für den realen Einsatz auf einem autonomen Rennfahrzeug ist jedoch die Balance zwischen Genauigkeit und Recheneffizienz entscheidend.

Während LION die höchste Detektionsleistung erreicht, erfüllt VoxelNeXt als einziges Modell gleichzeitig die Echtzeitanforderungen des FASTDA-Projekts. Damit stellt VoxelNeXt die geeignetste Architektur für den praktischen On-Board-Einsatz dar.

Die größte Herausforderung bleibt die sinkende Punktdichte mit zunehmender Entfernung. Besonders ab etwa 10 m nimmt die Erkennungsleistung deutlich ab, sodass die Verbesserung der Informationsdichte spärlicher Punktwolken ein zentraler Ansatzpunkt für zukünftige Arbeiten ist.

References

- [1] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11618–11628, 2020.
- [2] Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, and Jiaya Jia. Voxelnext: Fully sparse voxelnet for 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21674–21683, 2023.
- [3] Ying Li, Lingfei Ma, Zilong Zhong, Fei Liu, Michael A. Chapman, Dongpu Cao, and Jonathan Li. Deep learning for lidar point clouds in autonomous driving: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8):3412–3432, 2021.
- [4] Georgios Zamanakos, Lazaros Tsochatzidis, Angelos Amanatiadis, and Ioannis Pratikakis. A comprehensive survey of lidar-based 3d object detection methods with deep learning for autonomous driving. *Computers & Graphics*, 99:153–181, 2021.