
Integration von Hyperspektraldaten in RandLA-Net zur Klassifikation in 3D Punktwolken

Isaak MITSCHKE und Thomas WIEMANN

Zusammenfassung

In den letzten Jahren wurden viele Deep Learning Verfahren entwickelt, um 3D-Punktwolken automatisch zu segmentieren. Als Eingangsdaten werden dabei üblicherweise Punktwolken, welche zusätzlich mit weiteren Informationen wie Reflektionsintensitäten und RGB Farbwerten versehen sein können, verwendet. Mithilfe von Hyperspektraldaten lassen sich detaillierte Aufschlüsse über die chemische Zusammensetzung unterschiedlicher Materialien gewinnen. In vorhergehenden Arbeiten haben wir untersucht, ob sich klassische Verfahren des maschinellen Lernens direkt auf solche Daten übertragen lassen. Es zeigte sich, dass dies in Prinzip funktioniert, die Klassifikationsgenauigkeit aber noch stark verbessert werden kann. In diesem Beitrag kombinieren wir den vorherigen Ansatz mit einem Verfahren, das neben den Spektraldaten auch die geometrische Verteilung der Messpunkte berücksichtigt. Als Grundlage dient das neuronale Netz RandLA-Net, welches auf aktuellen Benchmark-Datensätzen wie Semantic3D und Semanti-KITTI sehr gut abschneidet. Untersucht wird dabei, wie RandLA-Net auf einem selbsterzeugten und segmentierten Datensatz mit und ohne Hyperspektralwerten abschneidet.

1 Einleitung

Für mobile Roboter ist eine semantische Karte hilfreich, damit sie eigenständig Aktionen planen und ausführen können. Ein wichtiger Bestandteil semantischer Karten ist eine räumliche 2D- oder 3D-Darstellung der Umgebung. Diese können als Punktwolken, Voxelgrids oder Dreiecksnetze bereitgestellt werden. Um den mobilen Systemen weitere Informationen über die Umgebung bereitzustellen, können relevante Bereiche in der Umgebungsdarstellung markiert werden. So lassen sich 3D-Punktwolken in verschiedene semantische Klassen einteilen, indem jedem Punkt ein entsprechendes Label zugewiesen wird. In einem urbanen Kontext sind Autos, Gebäude, Straßen und Bürgersteige Beispiele für mögliche Klassen. Die manuelle Zuordnung von Punkten zu semantischen Klassen ist ein zeitaufwändiger und mühsamer Prozess. Insbesondere bei Großvolumigen Punktwolken ist das händische Segmentieren unpraktikabel, daher ist die Entwicklung von Verfahren zur Automatisierung der Segmentierung von 3D-Punktwolken erstrebenswert. Dank der steigenden Beliebtheit von Deep-Learning-Methoden wurden in den letzten Jahren vermehrt Ansätze zur Segmentierung von Punktwolken entwickelt, die auf diesen Methoden basieren (GUO, et al., 2021). Derzeit ist das RandLA-Net (HU, et al., 2020) eines der besten neuronalen Netze für die Segmentierung von Punktwolken. Es erzielt bei aktuellen Benchmark-Datensätzen die besten Ergebnisse. Die Verwendung von RandLA-Net ermöglicht es, großvolumige Punktwolken zu segmentieren, die mit terrestrischen Laserscannern aufgenommen wurden, wie beispielsweise Semantic3D (HACKEL, et al., 2017). Darüber hinaus bietet RandLA-Net auch die Möglichkeit,

neben den reinen 3D-Daten zusätzliche Punktmerkmale wie Farb- oder Reflektanzwerte als ergänzende Informationen für die Klassifizierung zu verwenden. Ziel dieser Arbeit ist es, zu untersuchen, ob Hyperspektraldaten als Ergänzung zu 3D-Daten die Klassifizierungsergebnisse beim Einsatz von RandLA-Net verbessern. Durch die Aufnahme von Hyperspektraldaten ist es möglich, detailliertere Informationen über Materialien zu erhalten als mit herkömmlichen RGB-Farbwerten. Hyperspektralkameras verwenden ein Dispersionselement (z. B. ein Prisma), um das Licht räumlich entlang der Wellenlängen aufzuteilen. Das so in seine einzelnen Spektralbereiche aufgeteilte Licht wird dann auf einen CCD-Chip geleitet, der die individuellen Intensitäten der einzelnen Wellenlängenbereiche misst. Moderne Hyperspektralkameras sind in der Lage, mehrere hundert Kanäle gleichzeitig zu messen. Die Auswertung von Hyperspektraldaten wird vor allem im Bereich der Fernerkundung eingesetzt. Dort werden hyperspektrale Bilder, die von Satelliten, Flugzeugen oder Drohnen aufgenommen wurden, genutzt, um Rückschlüsse über die aufgenommenen Gebiete zu gewinnen. Mit Hilfe von Hyperspektralbildern kann zum Beispiel der Zustand der Vegetation gemessen oder die Materialzusammensetzung von Oberflächen analysiert werden (GOETZ, 2009). Um Punktwolken mit Hyperspektraldaten zu annotieren, müssen die Kamera und der Laserscanner gegeneinander kalibriert werden. Zunächst wird die extrinsische Kalibrierung, die die Transformation vom Laserscanner zur Kamera beschreibt, verwendet, um die Weltkoordinaten in Kamerakoordinaten umzuwandeln. Anschließend werden mithilfe der intrinsischen Kalibrierung der Kamera, die Koordinaten in Pixelkoordinaten transformiert. So kann ein entsprechender Farbwert oder, im Fall von Hyperspektralkameras, jedem 3D-Punkt ein entsprechendes Spektrum zugeordnet werden. IGELBRINK et al. (2018) entwickelten eine Methode zur Ad-hoc-Kalibrierung eines terrestrischen 3D-Laserscanners und einer Hyperspektralkamera. Das Kalibrierungsverfahren erfolgt automatisiert und benötigt keine in der Szene platzierten Referenzmuster.

2 Aufbau des Datensatzes

Um Klassifikatoren zu trainieren, wird eine große Menge an Trainingsdaten benötigt. Für alle Klassen sollte eine große Anzahl von Beispielen zur Verfügung stehen, damit die Klassifikatoren eine gute Übertragbarkeit auf bisher ungeschene Daten gewährleisten. Da es derzeit keine frei verfügbaren Datensätze mit hyperspektralen Punktwolken gibt, haben wir zu diesem Zweck einen eigenen Datensatz aufgebaut. Als zu erlernendes Umgebungsszenario für die Klassifikatoren wurde eine städtische Umgebung gewählt. In diesem Szenario sollen die Klassifikatoren in der Lage sein, zwischen Straßen, Gehwegen, Autos, Vegetation und Gebäuden zu unterscheiden. Da die reflektierten Spektren von den äußeren Lichtverhältnissen beeinflusst werden, sollte der Datensatz Aufnahmen zu verschiedenen Zeitpunkten und unter verschiedenen Wetterbedingungen enthalten. Zur Verbesserung der Genauigkeit und Stabilität der Vorhersagemodelle sollen Beispiele verwendet werden, die unter verschiedenen Beleuchtungsbedingungen erfasst werden. Zu diesem Zweck wurden zwei Straßenabschnitte in der Stadt Osnabrück ausgewählt, an denen Laserscans und Hyperspektralbilder an mehreren Positionen, zu zwei verschiedenen Zeitpunkten und unter unterschiedlichen Lichtverhältnissen aufgenommen wurden. Eine Ansicht des Datensatzes ist in Abbildung 1 zu finden. Es wurden an insgesamt 23 Positionen Daten aufgenommen. Die Messabstände der einzelnen Positionen betragen 15-50 Meter. Jeder Scan besteht aus durchschnittlich 45 Millionen Punkten. Insgesamt beinhaltet der Datensatz etwa

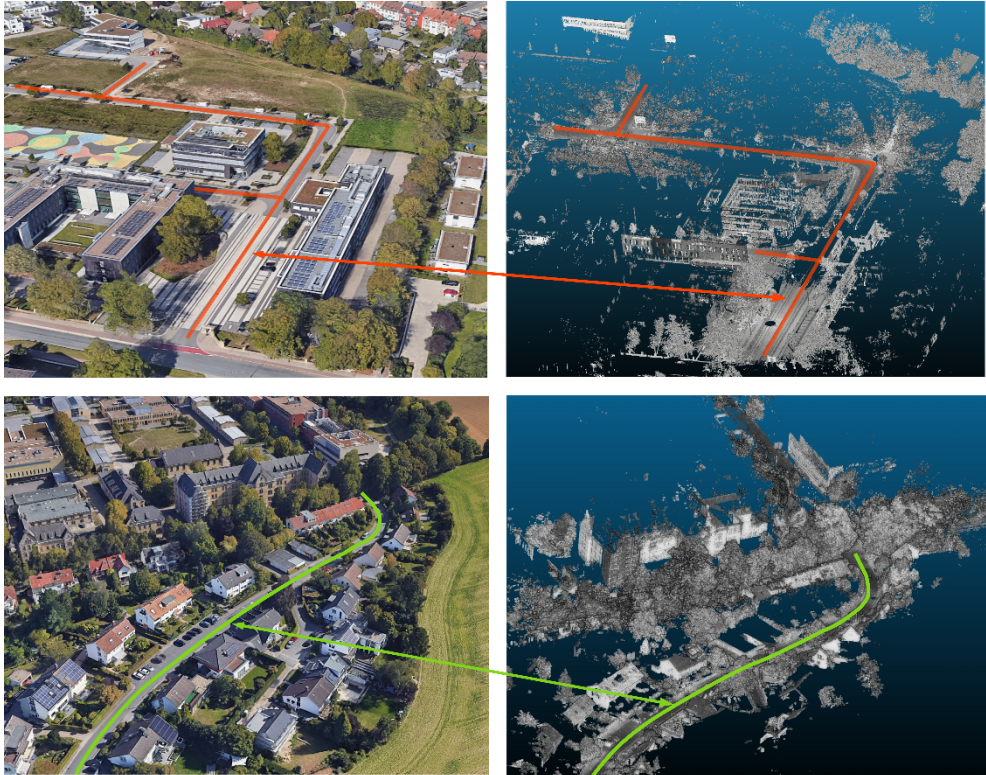


Abb. 1: Aufbau des Scansystems. Die Hyperspektralkamera ist auf den Laserscanner montiert. Mithilfe eines zusätzlichen Mini-PCs wird die Aufnahme gestartet und die Daten synchronisiert.

eine Milliarde Punkte. Neben den 3D-Koordinaten erfasst der Laserscanner auch einen entfernungsunabhängigen Reflektanzwert für jeden Punkt. Dieser Datensatz wurde mit dem terrestrischen Laserscanner Riegl VZ-400i aufgenommen, auf dem eine hyperspektrale Zeilenkamera (Resonon Pika L) montiert ist. Da in den rohen 3D-Punktwolken noch viele Ausreißer und Scan-Artefakte vorhanden sind, werden diese vor der weiteren Verarbeitung mit dem Statistical Outlier Removal Verfahren (RUSU, MARTON, BLODOW, DOLHA, & BEETZ, 2008) herausgefiltert. Die Hyperspektraldaten werden auch durch Rauschen beeinträchtigt, weshalb die einzelnen Spektren mit dem Savitzky-Golay-Filter (SAVITZKY & GOLAY, 1964) geglättet werden. Weitere negativen Einflüsse, welche durch Lichtstreuungseffekte auftreten, werden durch Standard Normal Variate (Formel 1) Transformation (BARNES, DHANOA, & LISTER, 1989) minimiert, bei der jedes Spektralpixel um Null zentriert und mit seiner Standardabweichung skaliert wird:

$$snv_i = \frac{x_i - \bar{x}_i}{SDev(x_i)} \quad (1)$$

Die Verwendung von Hyperspektraldaten zum Erlernen von Klassifikatoren ist auch wegen der hohen Dimensionalität der Daten problematisch. Die Pika L Hyperspektralkamera erfasst

beispielsweise 300 Spektralkanäle pro Pixel. Wenn alle Kanäle verwendet werden, steigt die Komplexität der zu trainierenden Modelle enorm. Um die Dimension der Daten zu reduzieren, werden die Spektren daher vor der Klassifizierung mittels Hauptkomponentenanalyse auf eine Größe von 64 verkleinert.

Die hyperspektralen Punktwolken wurden dann von Hand segmentiert, um eine Referenz für das Lernen von RandLA-Net zu erhalten. Die Auswahl der Klassen wurde so zusammengestellt, dass möglichst viele Bereiche von Punktwolken, die in städtischen Gebieten aufgenommen wurden, annotiert werden können. Die folgenden 17 Klassen wurden dabei definiert: Straße, Bürgersteig, Gebäude, Fahrzeug, Straßenlaterne, Zaun, Schild, Schildermast, Poller, Einfahrt, niedrige Vegetation, Busch, Baumkrone, Baumstamm, Absperrung, Verteilerkasten und Scan-Artefakt. Einige wenige Klassen wie "Zaun" oder "Verteilerkasten" sind nicht in allen, aber zumindest in zwei Scans vorhanden. Außerdem wurde die Klasse "Baum" in "Baumkrone" und "Baumstamm" aufgeteilt. Hintergrund hierfür ist, dass Baumkronen in Laserscans oft unerwünscht sind. Insbesondere dann, wenn aus den Punktwolken eine Polygonnetz-Darstellung erzeugt werden soll. Dünne Äste und Blätter können zu unerwünschten Artefakten führen. Diese Trennung soll es ermöglichen, Baumkronen später zu erkennen und auszuschneiden. Da es nicht möglich ist, alle Scan-Artefakte mit dem SOR-Filter zu entfernen, wurden diese ebenfalls markiert, damit RandLA-Net sie später erkennt und man sie anschließend leichter entfernen kann.

3 Integration von Hyperspektraldaten in RandLA-Net

Die TensorFlow-Implementierung von RandLA-Net, welche in der Softwarebibliothek Open3D-ML (ZHOU, PARK, & KOLTUN, 2018) bereitgestellt wird, bildet die Grundlage für unsere Experimente. RandLA-Net ist ein neuronales Netz basierend auf einer Encoder/Decoder-Architektur, wie sie bei Segmentierungsmethoden häufig angewandt wird. RandLA-Net erhält n Punkte mit entsprechenden Merkmalen (z.B. Farbwerte) als Eingabe. Die Encoder-schichten reduzieren die Anzahl der Daten, bei gleichzeitiger Erhöhung der Feature-Größe. Die Reduktion der Daten erfolgt dabei in jeder Schicht durch eine zufällige Auswahl an Punkten. Die neuen Features einer Schicht werden mithilfe einer Kombination aus den Features aus der vorherigen Schicht sowie der nächsten Nachbarschaft der Punkte generiert. Die Dekodierungsschichten kehren den Prozess um und skalieren die Daten mit jeder Schicht hoch und weisen den Punkten am Ende die geschätzte Klassenwahrscheinlichkeit zu. Bei der Konfiguration von RandLA-Net haben wir uns an dem Beispiel der Klassifizierung von Semantic3D orientiert, da unser Datensatz und der von Semantic3D eine ähnliche Struktur aufweisen. Das Netz wurde so parametrisiert, dass es 2^{16} Punkte mit zugehörigen Hyperspektraldaten sowie Punktintensitäten, als Eingabe erhält. Diese werden dann durch fünf Encoder- und Decoderschichten gegeben. Die von uns konfigurierte RandLA-Net Architektur ist in Abbildung 3 dargestellt. Um eine vollständige Punktwolke zu segmentieren, wird diese in viele sich überlappende Regionen unterteilt, so dass jede dieser Bereiche genau 2^{16} Punkte enthält. Diese werden dann nacheinander klassifiziert beziehungsweise trainiert. Den Eingabedaten wird zusätzliches Rauschen hinzugefügt und die Punktwolken werden um einen zufällig gewählten Winkel vertikal rotiert und zufällig gespiegelt. Diese Ergänzung wird häufig verwendet, um mehr Trainingsdaten zu erhalten und das Modell robuster zu machen. RandLA-Net wird auf dem Trainingsdatensatz ohne Hyperspektraldaten trainiert, um eine

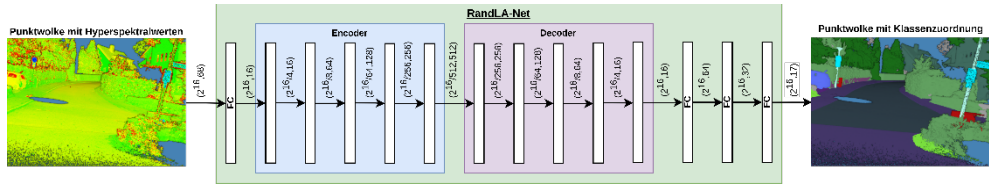


Abb. 3: RandLA-Net Architektur, wie sie von uns verwendet wird. Die Ausgabedimensionalitäten der einzelnen Schichten ist jeweils über den Verbindungen zu finden.

Grundlage für den Vergleich mit der Segmentierung hyperspektraler Punktwolken zu schaffen. Anschließend werden die Punktwolken des Validierungsdatensatzes segmentiert und die Metriken zur Bestimmung des Klassifizierungsergebnisses berechnet. Ein weiteres RandLA-Net-Modell wird dann mit den Punktwolken, die die vorverarbeiteten Spektraldaten enthalten, trainiert und evaluiert. Um die Ergebnisse der verschiedenen Klassifikatoren mit den Validierungsdaten zu vergleichen, werden die folgenden Metriken anhand der Vorhersagen der einzelnen Modelle berechnet: Accuracy, Precision, Recall und Intersection over Union.

4 Ergebnisse

In diesem Abschnitt werden die Ergebnisse der Segmentierung von hyperspektralen Punktwolken vorgestellt und diskutiert. Zu diesem Zweck werden zunächst die Klassifizierungsergebnisse der beiden Modelle vorgestellt, gefolgt von einem Vergleich und einer Diskussion der Unterschiede anhand der berechneten Bewertungsmetriken.

Tabelle 1: Vergleich der Klassifizierungsergebnisse für das Modell mit und ohne Hyperspektraldaten, unter Verwendung der Standardmetriken: IoU (Intersection over Union), F1-Score, Accuracy, Precision und Recall.

Hyperspektral	Acc.	IoU (Micro)	IoU (Macro)	F1 (Micro)	F1 (Macro)	Prec. (Micro)	Prec. (Macro)	Recall (Micro)	Recall (Macro)
Nein	0,83	0,39	0,72	0,46	0,83	0,45	0,83	0,54	0,83
Ja	0,89	0,52	0,80	0,60	0,89	0,66	0,89	0,61	0,89

Die Ergebnisse werden auch visuell überprüft, um weitere Unterschiede zu identifizieren. In Tabelle 1 sind die Durchschnittswerte der Bewertungsmetriken aufgeführt. Da die Anzahl der Stichproben der verschiedenen Klassen im Datensatz teilweise unausgewogen ist, wurden die Durchschnittswerte sowohl mit der Makro- als auch mit der Mikromethode berechnet. Um den Makro-Durchschnitt einer Metrik zu ermitteln, wird der Durchschnitt für jede Klasse einzeln berechnet und dann durch die Anzahl der Klassen geteilt. Der Mikro-Durchschnitt wird anhand der True Positives (TPs), True Negatives (TNs), False Positives (FPs) und False Negatives (FNs) Ergebnisse aller Klassen berechnet. Beispielsweise:

$$Precision_{Micro} = \frac{(TP_1 + \dots + TP_n)}{(TP_1 + \dots + TP_n + FP_1 + \dots + FP_n)} \quad (2)$$

Wie aus der Tabelle hervorgeht, erreichte das Modell, das zusätzliche Hyperspektraldaten erhalten hatte, bei allen Metriken bessere Ergebnisse. Die Klassifikationsergebnisse, nach Klassen aufgeschlüsselt, sind in Tabelle 2 aufgeführt. Einige Klassen, die im unteren Teil der Tabelle aufgeführt sind, wie z. B. Poller, Schildermast, Schranke und Zaun, werden in einigen Fällen sehr schlecht klassifiziert, weil die Anzahl der Beispiele im Datensatz zu gering ist. Der Anteil dieser Klassen im Datensatz beträgt weniger als 1 %. Bei der Untersuchung der Ergebnisse der einzelnen Klassen wird deutlich, dass die Verwendung zusätzlicher Hyperspektraldaten die Qualität des Klassifikators verbessern kann. Für die Klassen "Scan Artefakt", "Straße" und "Fahrzeug" führen die Hyperspektraldaten zu weniger Falschpositiven (erkennbar an den höheren Präzisionswerten). Scan-Artefakte werden mit und ohne Hyperspektraldaten erfolgreich erkannt, wobei die Falsch-Positiv-Rate ohne Hyperspektraldaten sehr hoch ist. Die Modelle können also auch als Filter dienen, um unerwünschte Scan-Artefakte zu entfernen. Auch die Klassifizierung von Baumstämmen wird verbessert. Ohne Hyperspektraldaten werden Baumstämme oft fälschlicherweise als Straßenlaternen klassifiziert (siehe Abbildung 4). Als mögliche Ursache kommt in Betracht, dass die beiden Klassen geometrisch mit einem Zylinder verwandt sind, aber aus unterschiedlichem Material bestehen. Die Hyperspektraldaten ermöglichen daher eine bessere Unterscheidung zwischen ähnlichen Objekten mit unterschiedlicher Materialzusammensetzung

Tabelle 2: Klassenweise Auswertung der Klassifizierungsmodelle. Die Ergebnisse werden als Intersection over Union (IoU), Precision und Recall aufgeführt.

Klasse	Ohne Hyperspektraldaten			Mit Hyperspektraldaten		
	IoU	Precision	Recall	IoU	Precision	Recall
Scan-Artefakt	0,15	0,16	0,99	0,86	0,89	0,97
Straße	0,82	0,83	0,98	0,89	0,90	0,98
Fahrzeug	0,32	0,32	0,96	0,85	0,86	0,98
Schild	0,82	0,94	0,86	0,74	0,91	0,79
Gebäude	0,63	0,98	0,64	0,94	0,99	0,95
Bürgersteig	0,59	0,76	0,72	0,59	0,87	0,64
Einfahrt	0,03	0,08	0,04	0,12	0,17	0,29
Baumstamm	0,17	0,25	0,34	0,55	0,71	0,70
Baumkrone	0,90	0,90	0,99	0,90	0,90	0,99
Busch	0,85	0,97	0,87	0,75	0,97	0,76
Niedrige Vegetation	0,79	0,84	0,93	0,56	0,57	0,97
Straßenlaterne	0,43	0,49	0,79	0,70	0,78	0,88
Poller	0,00	0,00	0,00	0,01	0,02	0,01
Schildermast	0,00	0,00	0,00	0,01	1,00	0,01
Verteilerkasten	0,01	0,04	0,03	0,13	0,34	0,17
Absperrung	0,00	0,00	0,00	0,00	0,00	0,00
Zaun	0,01	0,02	0,04	0,15	0,25	0,27

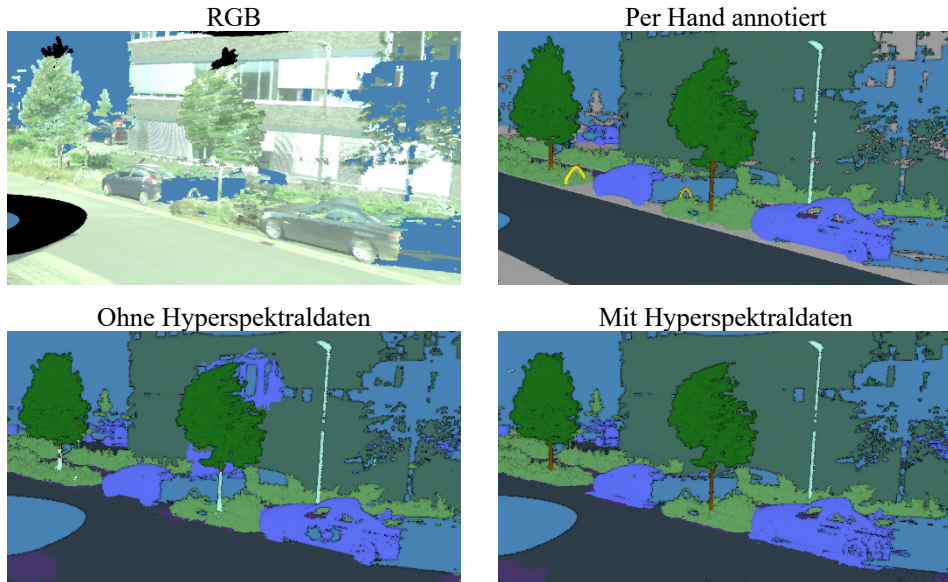


Abb. 4: Segmentierungsergebnisse einer Punktwolke. Die Mehrzahl der Punkte wird sowohl mit als auch ohne Hyperspektraldaten gut klassifiziert. Es fällt jedoch auf, dass ohne Hyperspektraldaten Fehler auftreten, wenn ähnliche Klassen wie Baum (braun) und Straßenlampe (hellblau) klassifiziert werden (ohne Hyperspektraldaten). Diese kann durch die zusätzlichen Hyperspektraldaten reduziert werden.

5 Zusammenfassung und Ausblick

Insgesamt konnten wir die Segmentierung von Punktwolken unter Verwendung zusätzlicher Hyperspektraldaten verbessern. Mit geringem Aufwand können bestehende Punktwolken-segmentierungsverfahren wie RandLA-Net, welche die Einbeziehung zusätzlicher Punktmerkmale zulassen, zur Segmentierung hyperspektraler Punktwolken verwendet werden. Da bisher keine Datensätze zur Segmentierung von hyperspektralen Punktwolken frei verfügbar sind, haben wir einen eigenen Datensatz erstellt, der manuell gelabelte Punkte enthält, um damit verschiedene Deep Learning Ansätze zu testen. In Zukunft soll der Datensatz weiter verbessert werden, um eine bessere Generalisierung der Deep-Learning-Modelle zu erreichen. Insbesondere sollen Punktwolken von Bereichen aufgenommen werden, die eine größere Anzahl der seltenen vorkommenden Klassen enthalten. Außerdem sollten ähnliche Klassen kombiniert werden, z. B. Busch und Baumkronen oder Barriere und Poller. Es wäre auch von Vorteil, Aufnahmen zu verschiedenen Jahreszeiten und bei unterschiedlichen Lichtverhältnissen zu machen, da die spektrale Signatur von Materialien stark von der Umgebungsbeleuchtung abhängig ist. In zukünftigen Untersuchungen möchten wir auch bestehende neuronale Netze wie RandLA-Net erweitern, um Hyperspektraldaten effektiver nutzen zu können, da aktuelle Modelle oft nur für eine kleine Anzahl von Punktmerkmalen ausgelegt sind, Hyperspektraldaten aber regelmäßig aus mehreren hundert Kanälen bestehen. Die Analyse

hyperspektraler Punktwolken ist auch für andere Anwendungsbereiche von Interesse, z. B. für die Landwirtschaft, wo sie für die automatische Kartierung verschiedener Pflanzenarten genutzt werden könnte. Mit dieser Arbeit haben wir eine erste Grundlage geschaffen und gezeigt, dass Hyperspektraldaten als zusätzliche Informationen neben räumlichen Daten die Segmentierungsergebnisse verbessern können.

Literatur

- Goetz, A. F. (2009). Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sensing of Environment*, 113, S.°5–S16.
- Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2021). Deep learning for 3D point clouds: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), S.°4338–4364.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., & Pollefeys, M. (2017). Semantic3d. net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*.
- Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., ... & Markham, A. (2020). Rاندلاندnet: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, S.°11108-11117.
- Igelbrink, F., Wiemann, T., Pütz, S., & Hertzberg, J. (2019). Markerless Ad-Hoc Calibration of a Hyperspectral Camera and a 3D Laser Scanner. In *Intelligent Autonomous Systems*, Springer International Publishing, S.° 748–759.
- Rusu, R. B., Marton, Z. C., Blodow, N., Dolha, M., & Beetz, M. (2008). Towards 3D point cloud based object maps for household environments. *Robotics and Autonomous Systems*, 56, S.°927–941.
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), S.°1627–1639.
- Zhou, Q. Y., Park, J., & Koltun, V. (2018). Open3D: A modern library for 3D data processing. *arXiv preprint arXiv:1801.09847*.

Bitte geben Sie auf einer separaten Seite zu allen genannten Autoren folgende Informationen:

Mitschke, Isaak, M. Sc.
Universität Osnabrück
Institut für Informatik
AG Wissensbasierte Systeme
imitschke@uos.de

Wiemann, Thomas, Priv. Doz. Dr.
Universität Osnabrück
Institut für Informatik
AG Autonome Robotik
twiemann@uos.de