



ulm university universität
uulm

Fakultät für Ingenieurwissenschaften, Informatik und Psychologie
Institut für Mess-, Regel- und Mikrotechnik

Segmentierung von Punktwolken mit neuronalen Netzen

Masterarbeit

von

Tarik Enderes

31.12.2001

Betreuer:	Dr.-Ing. Vasileios Belagiannis
1. Prüfer:	Prof. Dr.-Ing. Brian F. Smith
2. Prüfer:	Prof. Rivera

Hiermit versichere ich, dass ich die vorliegende Arbeit mit dem Titel

Segmentierung von Punktwolken mit neuronalen Netzen

bis auf die offizielle Betreuung selbstständig und ohne fremde Hilfe angefertigt habe und die benutzten Quellen und Hilfsmittel vollständig angegeben sind. Aus fremden Quellen direkt oder indirekt übernommene Gedanken sind jeweils unter Angabe der Quelle als solche kenntlich gemacht.

Ich erkläre außerdem, dass die vorliegende Arbeit entsprechend den Grundsätzen guten wissenschaftlichen Arbeitens gemäß der „Satzung der Universität Ulm zur Sicherung guter wissenschaftlicher Praxis“ erstellt wurde.

Ulm, den 31.12.2001

Tarik Enderes

Inhaltsverzeichnis

1	Theorie	1
1.1	Segmentierung	1
1.1.1	Semantische Segmentierung	1
1.1.2	Instanz-Segmentierung	1
1.1.3	Panoptische Segmentierung	2
1.2	Technologien in DeepLab	2
1.2.1	Deep Convolutional Neural Networks für Semantische Segmen- tierung	3
	Convolutional Neural Networks	3
	Anpassungen für Semantische Segmentierung	3
1.2.2	Atrous Convolution	3
1.2.3	Atrous Spatial Pyramid Pooling	4
1.2.4	Fully-Connected Conditional Random Fields	4
1.2.5	Residual Networks	4
1.3	Kamerakalibrierung	4
2	Aufgabenstellung	7
2.1	Ziele und Anforderungen	7
2.2	Ähnliche Projekte	7
2.2.1	PointNet	7
2.2.2	UPNet	7
3	Arbeitsmethodik	9
3.1	Integration von DeepLab	9
3.2	Backbones	9
3.2.1	Xception	9
3.2.2	MobileNetV2	9
3.3	Datensätze	9
3.3.1	Cityscapes	9
3.3.2	KITTI	9
4	Experimente	11
4.1	Technische Daten des für die Experimente verwendeten Rechners . . .	11

4.2	Backbones	11
4.2.1	Xception	11
4.2.2	MobileNetV2	11
4.3	Auswirkungen der Trainingsdauer	11
4.4	Experimente mit Verfeinerung	11
4.4.1	Verfeinerung mit KITTI	11
4.5	Aufgetretene Probleme und Lösungen	11
4.5.1	False Positives	11
4.5.2	Overfitting	11
5	Ergebnisse	13
5.1	Ergebnisse auf Bildern	13
5.2	Ergebnisse auf Punktwolken	13
6	Zusammenfassung	15
	Literaturverzeichnis	17

1 Theorie

1.1 Segmentierung

Segmentierung bezeichnet einen Vorgang, bei dem ein Bild nach bestimmten Homogenitätskriterien in inhaltlich zusammenhängende Regionen einzuteilen. Von den verschiedenen Ansätzen, die das erreichen sollen, befasst sich diese Arbeit mit pixelbasierten Verfahren, bei denen jedem Pixel in einem Bild eine Klasse zugeordnet wird. Man unterscheidet, wie in [ups] beschrieben, semantische Segmentierung, Instanz-Segmentierung und panoptische Segmentierung.

1.1.1 Semantische Segmentierung

Bei der Semantischen Segmentierung soll jeder Pixel eine valide Klasse erhalten. Es wird dabei nicht zwischen unterschiedlichen Instanzen einer Objektklasse unterschieden. Wenn beispielsweise auf einem Bild zwei Fahrzeuge zu sehen sind und bei der Segmentierung die Klasse "Fahrzeug" geteilt werden soll, erhalten die Pixel beider Fahrzeuge das Label "Fahrzeug". Die Anzahl valider Klassen bleibt somit bei jedem prozessierten Bild gleich.

1.1.2 Instanz-Segmentierung

Im Gegensatz zur semantischen Segmentierung werden bei der Instanz-Segmentierung nur zählbare Objekte betrachtet und deren Instanzen berücksichtigt. Übertragen auf vorheriges Beispiel würden die Pixel des einen Fahrzeug ein Label wie "Fahrzeug1" und die des anderen analog "Fahrzeug2" erhalten.

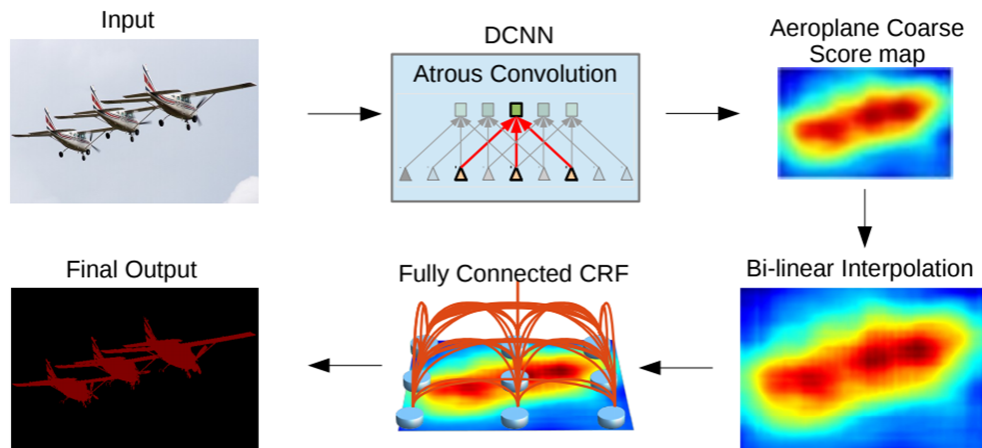


Abbildung 1.1: Grundsätzliche Funktionsweise von DeepLab

1.1.3 Panoptische Segmentierung

Die panoptische Segmentierung stellt eine Kombination der vorherigen Segmentations-Arten dar. Zählbare Objekte werden demnach nach dem Prinzip der Instanz-Segmentierung und amorphe nach dem der semantischen Segmentierung segmentiert. Die Ergebnisse beider Verfahren werden anschließend kombiniert.

1.2 Technologien in DeepLab

DeepLab ist ein von Google entwickeltes, 2015 in [dl1] vorgestelltes Modell für semantische Segmentierung. Bei der in [dl2] vorgestellten Methode wird ein Deep Convolutional Neural Network (DCNN) zum Erzeugen einer Score Map benutzt, die anschließend mit einem Conditional Random Field (CRF) zur endgültigen Ausgabe weiterverarbeitet wird. Das Verfahren wird in Abbildung 1.1 grob dargestellt.

1.2.1 Deep Convolutional Neural Networks für Semantische Segmentierung

Convolutional Neural Networks

Anpassungen für Semantische Segmentierung

Klassische DCNNs haben Eigenschaften, die sie für die Verwendung zur Bildsegmentierung nicht ideal machen.

- Der Einsatz von Downsampling führt zu verringerter Auflösung, die bei Klassifizierungsaufgaben nicht ins Gewicht fällt, für die Segmentierung aber essentiell ist.
- Neuronale Netze sind in der Regel gut geeignet, um Objekte unterschiedlicher Größe zu erkennen, wenn solche in der Lernphase präsentiert werden. Die Eigenschaften der Faltung, insbesondere dem begrenzten Sichtbereich beim Berechnen eines einzelnen Pixels ist allerdings für diese Problematik ungünstig.
- Der wiederholte Einsatz von Convolutional Layers führen zu einem Verlust an Ortsinformation. Infolgedessen produzieren DCNNs bei Segmentierungsaufgaben verschwommene, oft verrauschte Ergebnisse ohne klare Kanten.

Um diese Probleme zu lösen erhält das von DeepLab verwendete DCNN einige Anpassungen. Zunächst werden alle Fully Connected Layers durch Convolutional Layers ersetzt, um ein Fully Convolutional Network zu bilden. Noch dazu wird anstatt von Pooling Layers in den unteren Schichten Atrous Convolution eingesetzt, womit die Auflösung der Ausgabe erhöht wird. In den höheren Schichten werden auch hier Pooling Layers eingesetzt, um Speicherbedarf und Rechenzeit zu verbessern. Um die Größeninvarianz zu verbessern wird bei den unteren Schichten Atrous Spatial Pyramid Pooling verwendet.

1.2.2 Atrous Convolution

Atrous Convolution, auch Dilated Convolution genannt, beschreibt eine Technik bei der eine Matrix mit einem spärlich bestückten Kernel gefaltet wird, wie in Abbildung 1.2 illustriert.

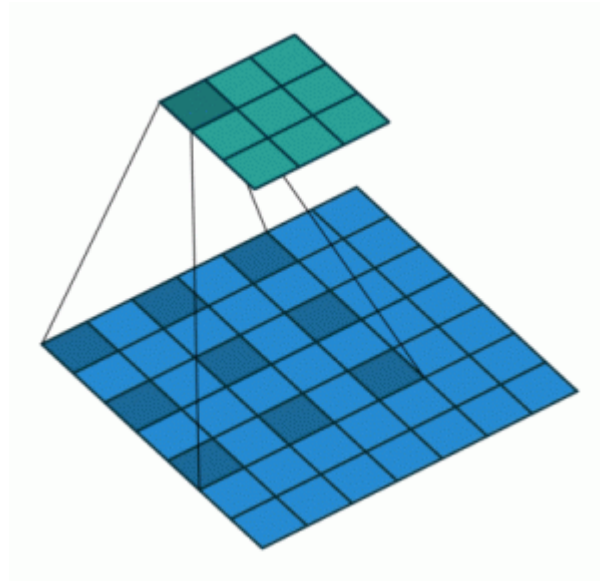


Abbildung 1.2: Prinzip von Atrous Convolution

Die Abstände der zu berücksichtigenden Werte in der Matrix wird dabei durch die s.g. Dilation Rate bzw. Erweiterungsrate (kurz Rate) festgelegt. Das tatsächliche Sichtfeld des Filters wird also festgelegt durch die Größe des Kernels und die Rate bestimmt. Ein Filter mit einem Kernel der Größe 3x3 und einer Rate von 2, was dem Einfügen einer leeren Zeile und Spalte zwischen den Werten entspricht, hat demnach ein Sichtfeld der Größe 5x5. Dadurch wird das effektive Sichtfeld des Filters erhöht und es kann eine höhere Auflösung bei gleichem Rechenaufwand erreicht werden. Die Vorteile der Verwendung von Atrous Convolution für Bildsegmentierung sind in Abbildung 1.3 dargestellt.

1.2.3 Atrous Spatial Pyramid Pooling

Beim Atrous Spatial Pyramid Pooling werden mehrere parallele Convolutional Layers, die Atrous Convolutional Layers mit unterschiedlicher Rate verwenden, in das DCNN eingebaut. Das Prinzip ist in Abbildung 1.4 dargestellt.

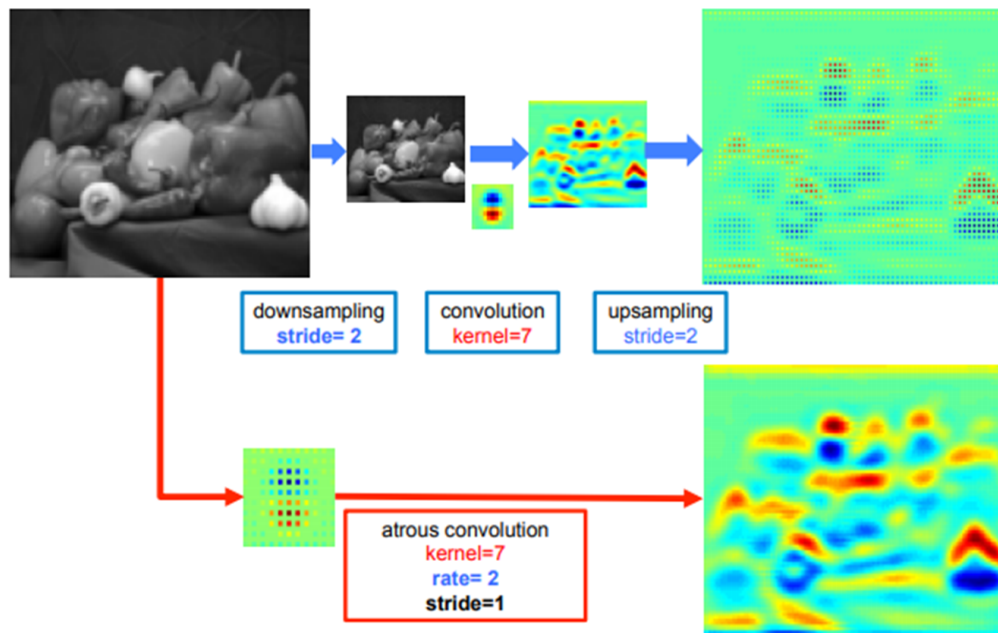


Abbildung 1.3: Beispielhaft dargestellte Vorteile von Atrous Convolution

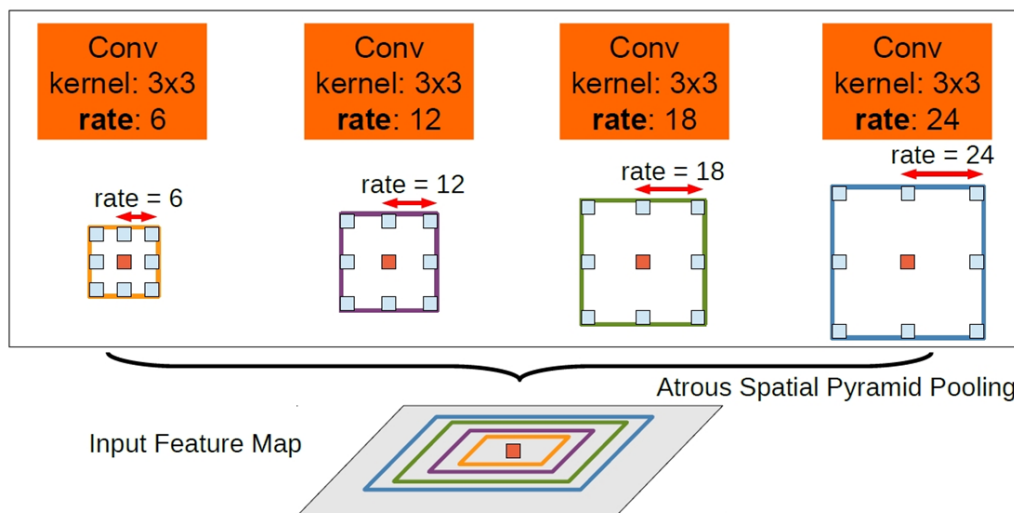


Abbildung 1.4: Beispielhaft dargestellte Vorteile von Atrous Convolution

1.2.4 Fully-Connected Conditional Random Fields

1.2.5 Residual Networks

1.3 Kamerakalibrierung

2 Aufgabenstellung

2.1 Ziele und Anforderungen

2.2 Ähnliche Projekte

2.2.1 PointNet

2.2.2 UPSNet

3 Arbeitsmethodik

3.1 Integration von DeepLab

3.2 Backbones

3.2.1 Xception

3.2.2 MobileNetV2

3.3 Datensätze

3.3.1 Cityscapes

3.3.2 KITTI

4 Experimente

4.1 Technische Daten des für die Experimente verwendeten Rechners

4.2 Backbones

4.2.1 Xception

4.2.2 MobileNetV2

4.3 Auswirkungen der Trainingsdauer

4.4 Experimente mit Verfeinerung

4.4.1 Verfeinerung mit KITTI

4.5 Aufgetretene Probleme und Lösungen

4.5.1 False Positives

4.5.2 Overfitting

5 Ergebnisse

5.1 Ergebnisse auf Bildern

5.2 Ergebnisse auf Punktwolken

6 Zusammenfassung

Literaturverzeichnis

- [dl1] Chen, Liang-Chieh; Papandreou, George; Kokkinos, Iasonas; Murphy, Kevin und Yuille, Alan L.: *Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs*. URL: <https://arxiv.org/pdf/1412.7062.pdf> (zuletzt besucht am 24.07.2019).
- [dl2] Chen, Liang-Chieh; Papandreou, George; Kokkinos, Iasonas; Murphy, Kevin und Yuille, Alan L.: *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*. URL: <https://arxiv.org/pdf/1606.00915.pdf> (zuletzt besucht am 24.07.2019).
- [ups] Xiong, Yuwen; Liao, Renjie; Zhao, Hengshuang u. a.: *UPNet: A Unified Panoptic Segmentation Network*. URL: <https://arxiv.org/pdf/1901.03784.pdf> (zuletzt besucht am 24.07.2019).