Automatisierte semantische Modellierung von Innenräumen aus Bildern und abgeleiteten Punktwolken basierend auf Deep Learning Methoden

LARS S. OBROCK [1] & EBERHARD GÜLCH[1]

Zusammenfassung: In diesem Paper präsentieren wir eine Weiterentwicklung unseres bisherigen Ansatzes, photogrammetrisch erzeugte Punktwolken mit aus Bildern extrahierten semantischen Informationen anzureichern, um eine Automatisierung der BIM-Modellierung zu ermöglichen.
Zur semantischen Segmentierung von Bildern verwenden wir nun das Neuronale Netz DeepLabv3+, um Bauteile und Objekte von Innenräumen zu extrahieren. Während der photogrammetrischen Rekonstruktion projizieren wir die segmentierten Kategorien in die Punktwolke. Dabei auftretende Interpolationen korrigieren wir automatisiert und erreichen bei der Klassifizierung der Punktwolke eine mIoU von 51,9 %. Diese Informationen nutzen wir, um die Punktwolke auszurichten, den Maßstab zu korrigieren und weitere Informationen zu extrahieren.
Durch unsere Untersuchungen sehen wir die Grundlagen für eine automatisierte BIM-Modellierung basierend auf einer Kombination von Photogrammetrie und Deep Learning bestätigt.

1 Einführung

Die Digitalisierung des Bausektors schreitet immer weiter voran und vollzieht mit dem Building Information Modeling (BIM) den Schritt von zweidimensionalen Plänen auf Papier zu umfassenden, dreidimensionalen digitalen Gebäudemodellen. Diese BIM-Modelle sind das zentrale Element und bilden den gesamten Zyklus eines Gebäudes ab, von der Planung über den Betrieb bis zum Abriss. In ihnen sind neben den dreidimensionalen Bauteil- und Objektgeometrien auch die wesentlichen semantischen Informationen hinterlegt. Jedoch vollzieht sich die Einführung von BIM derzeit quasi nur bei der Planung von Neubauten. Eine großflächige Erfassung von bereits bestehenden Gebäuden als BIM-Modell ist aufgrund der Komplexität von Aufnahme und Auswertung bisher ein untergeordnetes Thema.
Der große Nachteil liegt hierbei besonders darin, dass die geometrischen und die für ein BIM-Modell essentiell wichtigen semantischen Informationen derzeit getrennt ermittelt und manuell zu einem Modell zusammengeführt werden müssen. An dieser Stelle setzen wir an, um eine (Teil-) Automatisierung der Modellierung dieser in ihrem „As-Build“ bzw. „As-Is“ Zustand zu entwickeln und dadurch ihre Erstellung zu vereinfachen. Besonders die Verbindung von Photogrammetrie und Bildanalyseverfahren bieten aufgrund der großen Informationstiefe der zugrundeliegenden Bilder ein sehr hohes Potential, diese Extraktion von geometrischen und semantischen Informationen automatisiert zu ermöglichen.

[1] Hochschule für Technik Stuttgart, Labor für Interpretation und Messung bildgebender Sensoren (LIMES), Schellingstraße 24, D-70174 Stuttgart
E-Mail: [Lars.Obrock, Eberhard.Guelch]@hft-stuttgart.de

In OBROCK & GÜLCH (2018) wurden hierzu erste Ansätze veröffentlicht. Diese basierten auf der Anwendung von Deep Learning zur Segmentierung von acht Innenraumbauteilen und Objekten unter Verwendung eines Fully Convolutional Networks (SHELHAMER et al. 2016). Die auf diese Weise segmentierten Bilder wurden als Ersatz für den blauen Kanal in das Originalbild eingefügt. Basierend auf diesen Falschfarbenbildern wurde durch photogrammetrische Verfahren eine Punktwolke erstellt. Jedoch trat in der Punktwolke durch den Verlust an Tiefeninformationen im blauen Kanal eine erhöhte Zahl von Fehlern auf. Trotzdem wurde dabei die Umsetzbarkeit dieses Ansatzes bestätigt, da die zuvor segmentierten Objektkategorien in den Farbwerten der Punktwolke enthalten sind und somit die für BIM so wichtigen semantischen Informationen. In dieser Untersuchung fokussieren wir uns auf die Verbesserung und Erweiterung des zuvor entwickelten Ansatzes.
Aufbauend auf verbesserten Deep Learning Architekturen zur semantischen Segmentierung, einer Anpassung des Transfers der extrahierten Informationen in eine automatisiert erstellte Punktwolke und eine anschließende Analyse und Bearbeitung dieser sind uns weitere Schritte hin zu einer Automatisierung der BIM-Modell Erstellung gelungen.

2 Verwandte Arbeiten

Der in der Baubranche stattfindende Wandel hin zu einer Digitalisierung der Arbeitsschritte vollzieht sich rasant durch die Nutzung von Building Information Modeling und dem hierbei zentralen BIM-Modell.
In Deutschland wurden hierfür mehrere Richtlinien, Leitfäden und Untersuchungen veröffentlicht, welche mehrheitlich die Einführung und Umsetzung von BIM in unterschiedlichen Fachbereichen behandeln, z.B. (EGGER et al. 2013; ESCHENBRUCH et al. 2014; KADEN et al. 2017; BORRMANN et al. 2015, 2015; BRAMANN et al. 2015b). Mit dem „Stufenplan Digitales Planen und Bauen“ führte das Bundesministerium für Verkehr und Digitale Infrastruktur die BIM-Methode stufenweise in die Planungsprozesse der öffentlichen Infrastruktur ein (BRAMANN et al. 2015a). Die Auswirkungen von BIM auf die Durchführung von Infrastrukturprojekten wird in der Einführungsphase von diesem ausführlich untersucht.
Das Modellieren von Bestandsgebäuden basiert derzeit auf ihrer Aufnahme durch geodätische Messgeräte. Mögliche Vorgehensweisen werden unter anderem in (BORRMANN et al. 2015; CLEMEN & EHRICH 2014; KADEN et al. 2017) beschrieben. Der gesteigerte Bedarf an geometrischen und semantischen Informationen für BIM-Modelle bleibt jedoch unbeachtet. Im Umkehrschluss führt dies dazu, dass die Aufnahme und Auswertung der Messdaten sehr komplex ist und entsprechend längere Zeit in Anspruch nehmen kann.
Im gesamten Feld des Deep Learning und auch speziell bei den für die Computer Vision und Bildanalyse sehr gut geeigneten Convolutional Neural Networks (CNN) wurden seit (KRIZHEVSKY et al. 2012) gewaltige Fortschritte gemacht. So steigerten sich die Genauigkeiten bei der Bildklassifizierung durch die Verwendung von weiterentwickelten Netzwerkarchitekturen stark (z.B. SZEGEDY et al. 2014; SIMONYAN & ZISSERMAN 2015; HE et al., 2015; XIE et al. 2017; HAUNG et al. 2018). Entwickelt wurde neben der reinen Klassifikation auch eine Lokalisierung von Objekten in den Bildern. Zu diesem Zweck werden in der Objekt Detektion genannten Bildanalyse Bounding Boxen zur Unterteilung in kleinere Bereiche gesucht, deren Inhalt klassifiziert und somit

eine Lokalisation geschaffen (GERSHICK et al. 2014; GERSHICK 2015; REN et al. 2016). Die semantische Segmentierung setzt im Gegensatz dazu auf eine pixelgenaue Klassifizierung des gesamten Bildes. Diese eignet sich für die Projektion der semantischen Informationen in die Punktwolke. Eines der ersten für die semantische Segmentierung entwickelten Neuronalen Netze stammt von (SHELHAMER et al. 2016) und heißt Fully Convolutional Network (FCN). Dieses bildet die Grundlage für viele weitere Architekturen, mit denen die erreichten Genauigkeiten mehr und mehr gesteigert wurden (JÉGOU et al. 2017; ZHAO et al. 2017; CHEN et al. 2016). In dieser Untersuchung verwenden wir in DeepLabv3+ (CHEN et al. 2018) eine Architektur, welche in Benchmarks sehr gute Genauigkeiten erreicht.
In (BOULCH et al. 2017) werden von vermaschten Punktwolken RGB- und Tiefenbilder erstellt, diese mit einem CNN segmentiert und anschließend über eine Projektion auf das Mesh wieder in die Punktwolke übertragen.

3 Methoden

3.1 Bildsegmentierung

Eine qualitativ hochwertige und semantisch angereicherte Erfassung in drei Dimensionen ist von essentieller Bedeutung für die automatisierte, vollständige Rekonstruktion aller wichtigen Bauteile und Objekte eines Bestandsgebäudes als BIM-Modell.
Aufbauend auf den in (OBROCK & GÜLCH 2018) beschriebenen Methoden bilden Innenraumbilder die Basis der Untersuchung. Bei den damals verwendeten Kategorien handelte es sich nur um einen Grundstock an wichtigen Objekten. Um ein möglichst vollständiges BIM-Modell zu erzeugen, erweiterten wir diese auf insgesamt 24 für Innenräume wichtige Bauteile und Objekte.
Die Auswahl umfasste die am häufigsten in Innenräumen vorkommenden Objekte und richtete sich nach den Erfahrungen der zuvor erfolgten Untersuchung. Die Kategorien lassen sich grob in „Raum formende Bauteile“, „Verbindende Bauteile“, „Feste Objekte von Interesse“ und „Bewegliche Objekte von Interesse“ aufteilen. Eine Übersicht über die getroffene Auswahl findet sich in Tabelle 1.

Tab. 1: Übersicht über die für Innenräume ausgewählten wichtigen Bauteile und Objekte.

Raum formende Bauteile	Verbindende Bauteile	Feste Objekte von Interesse		Bewegliche Objekte von Interesse
Boden	Tür	Lichtschalter	Steckdose	Poster
Wand	Fenster	Lampe	Stütze/Säule	Feuerlöscher
Decke		Heizung	Rohre	Teppiche
		Treppenstufen	Geländer	Schrank
		Waschbecken	Toilette	Regal
		Kabelkanal	Feuermelder	Tisch
		Rauchmeldersirene		Stuhl

Allerdings konnte mit der „Wand“ eines der Hauptmerkmale eines Innenraums nicht berücksichtigt werden, da eine Unterscheidung zwischen ihr und der „Decke“ kaum möglich war und sie

einen großen Anteil in den Bildern einnimmt. Dies führte zu einer zu starken Gewichtung dieser Klasse und resultierte deshalb in einem negativen Einfluss auf die Qualität der Segmentierung der anderen Kategorien.
Basierend auf diesen Kategorien wurde für das spätere Training des Neuronalen Netzes ein Trainingsdatensatz erstellt. Dieser besteht aus ca. 300 Bildern, von welchen ein großer Teil selber aufgenommen und der Rest aus dem Internet bezogen wurde. Zu diesen wurde manuell das „ground truth"-Bild segmentiert und so jedem Pixel eine eindeutige Kategorie zugewiesen. Da dies für das Training eines Neuronalen Netzes wenige Bilder sind, wurde der Datensatz mit Hilfe von Daten Augmentierung (data augmentation) auf fast 18.000 Bilder erweitert. Hierfür wurden die Bilder durch Zerschneiden, Drehen, Spiegeln, Helligkeitsänderung und zufälligem Rauschen verändert. Der so erzeugte Datensatz ist für ein Neuronales Netz noch immer recht klein, ermöglicht aber durch fine-tuning eines vortrainierten Modells (pre-trained model) das Training.
Die rapiden Fortschritte im Bereich des Deep Learning nehmen wir zum Anlass, eine andere Struktur als Neuronales Netz heranzuziehen. Anstelle eines Fully Convolutional Networks wurde das im Jahr 2018 erschienene und stark weiterentwickelte DeepLabv3+ (CHEN et al. 2018) als Grundlage für das Training verwendet. Dieses baut ebenfalls auf einer Encoder-Decoder Struktur auf. Zur Feature Extraktion in der Encoder Phase setzt es auf die Xception Architektur und erweitert diese mit Atrous Spatial Pyramid Pooling zur besseren Einbeziehung der erweiterten Nachbarschaft des untersuchten Neurons. Durch Verwendung des Decoders wird die zuvor reduzierte Ausdehnung der Layer, basierend auf aus dem Encoder abgeleiteten Werten wieder hochskaliert. So ermöglicht dieses Neuronale Netz eine gute Qualität der Segmentierungen und präzise Kanten zwischen unterschiedlichen Bereichen.
Zur semantischen Segmentierung der Bilder wurde das DeepLabv3+ basierend auf der Xception 65 Architektur herangezogen. Zu dieser Architektur wurde das zugehörige auf ImageNet und COCO Daten vortrainierte Modell genommen. Aufbauend auf diesem erfolgte das Training durch fine-tuning.

3.2 Klassifizierte Punktwolke

Zur automatisierten Erstellung eines BIM-Modells werden dreidimensionale, semantisch angereicherte Informationen benötigt. Durch die Verwendung von Photogrammetrie und digitaler Bildzuordnung können die Bilder als Grundlage zur Erstellung einer dreidimensionalen Punktwolke verwendet werden.
Als zentrales Element eines Gebäudes steht die Rekonstruktion von Innenräumen im Fokus der Untersuchungen. Als unabhängiges Vergleichsobjekt wurde zur generellen Darstellung eines solchen ein anderes, typisches Büro in einem öffentlichen Gebäude ausgewählt. Dabei wurde darauf geachtet, dass keines der Bilder im Trainingsdatensatz aus dem Inneren dieses Büros stammt, um die Möglichkeit eines Overfittings zu diesem und damit dessen Auswirkungen auszuschließen. Von diesem Büro wurde unter photogrammetrischen Gesichtspunkten eine Vielzahl an Bildern aufgenommen. Diese Bilder werden durch das trainierte Modell segmentiert, um pixelgenau die in ihnen enthaltenen Objekte verorten zu können.In einem eigens entwickelten Prozessschritt werden kleinste Flächen, die keine realen Objektteile darstellen können, aus den erzeugten Bildern herausgefiltert.

In OBROCK & GÜLCH (2018) wurden vor der Erzeugung der Punktwolke die segmentierten Bilder anstelle des Blauen Kanals in die Originalbilder eingefügt. Die dabei entstehenden Falschfarbenbilder waren noch immer geeignet, eine Punktwolke zu erzeugen, in welcher die segmentierten semantischen Informationen enthalten waren. Jedoch sind durch den Informationsverlust der Tiefeninformationen dieses Kanals einige Fehler in der Punktwolke aufgetreten. Außerdem waren sie durch die Komprimierung der Klasseninformationen im Blauen Kanal für den Menschen nicht mehr einfach visuell zu unterscheiden, da sie alle in leicht unterschiedlichen Lilatönen dargestellt wurden. Um diesen Einschränkungen zu begegnen, haben wir für diese Untersuchung unsere Vorgehensweise angepasst. So werden den einzelnen Kategorien keine IDs mehr zugewiesen, um diese in den Blauen Kanal einzufügen. Die durch DeepLabv3+ segmentierten Bilder werden als RGB-Bilder gespeichert und die Kombination aller drei Farbkanäle steht eindeutig für eine Kategorie.

Als Grundlage für die nächsten Schritte erstellen wir unter Verwendung der Originalbilder eines Innenraums in Agisoft Metashape (AGISOFT 2020) automatisiert und ohne Platzierung von Passpunkten eine Punktwolke. Um die Kategorien in diese zu überführen, ist es möglich, die Originalbilder durch die segmentierten Bilder zu ersetzen und die Punktwolke auf Grundlage der zuvor ermittelten Kameraposition und Rotation einfärben zu lassen. Eine Übertragung findet nicht mehr direkt bei der Punktwolkenerstellung aus einem angereicherten Farbkanal statt, sondern aus der Projektion der segmentierten Bilder auf die Punktwolke. Somit findet der Informationstransfer der semantischen und geometrischen Informationen nicht mehr in einem, sondern in zwei getrennt aber direkt aufeinander aufbauenden Schritten statt. Die hierdurch ermöglichte Reduktion des Informationsverlustes bringt viele Vorteile.

Bei der Generierung der Punktwolke durch Photogrammetrie stehen alle Farbkanäle verlustfrei zur Verfügung. Dies ermöglicht eine bessere Qualität beim Extrahieren und Verknüpfen einzelner prägnanter Bildbereiche und erlaubt auf diese Weise eine bessere Rekonstruktion der gesamten Punktwolke. Außerdem lassen sich die Farbwertkombinationen der einzelnen Kategorien in größerem Abstand zueinander platzieren. Diese sind sehr gering, wenn 24 Klassen auf 256 mögliche Farbwerte eines Kanals aufgeteilt werden sollen. Im Gegensatz dazu ermöglicht die Kombination von Werten aus drei Kanälen deutlich größere euklidische Abstände zwischen den Farbwerten der Kategorien, da sie als Punkte im dreidimensionalen Raum betrachtet werden können. Diese Verteilung auf alle Farbkanäle birgt auch für den Betrachter den größten Vorteil, da alle Farben deutlicher voneinander zu unterscheiden und die segmentierten Klassen einfach festzustellen sind.

Beim Transfer der semantischen Informationen werden die Farbwerte der Pixel durch die Bilder bestimmt, aus denen sie generiert wurden. Ist die in diesen vorgenommene Segmentierung an den sich überschneidenden Stellen nicht übereinstimmend oder sind die Kameraausrichtungen im dreidimensionalen Raum nicht perfekt zueinander, treten durch Interpolation abweichende Farbwerte in den erzeugten Punkten auf.

Deshalb passten wir die RGB-Farbwerte für die Kategorien bereits in den segmentierten Bildern so an, dass deren minimaler euklidischer Abstand untereinander und zu möglichen interpolierten Farben maximal groß ist. Auf diese Weise wird eine einigermaßen gleichmäßige Verteilung der Farbwerte im dreidimensionalen Raum erreicht und die eindeutige Zuordnung der Kategorien auch

bei Interpolationen verbessert. In der Punktwolke auftretende Ausreißerpunkte werden herausgefiltert.
In einem weiteren Arbeitsschritt müssen die aufgetretenen Interpolationen eliminiert werden. Durch die Betrachtung der Farbewerte der einzelnen Punkte im dreidimensionalen Raum und der aus diesen abgeleiteten euklidischen Distanz zu den Farbwerten der Kategorien wird eine eindeutige Zuweisung ermöglicht.Auch werden die Distanzen der Farbwerte eines Punktes zu den Farbwerten errechnet, welche durch die Interpolation einer Kategorie mit Background-Pixeln entstehen würden. Liegen diese aus den Farbwerten ermittelten Distanzen innerhalb eines festgelegten Maximalwerts, erfolgt eine direkte Zuordnung zu der entsprechenden Kategorie.
Im Fall, dass sich keine direkte Zuordnung ergibt, wird für die Umgebung der Punkte untersucht. Erreicht eine Kategorie bei diesen eine deutliche Mehrheit, wird sie für den untersuchten Punkt übernommen. Falls dies nicht der Fall ist, wird eine Kombination der beiden Herangehensweisen gewählt. Dafür wird untersucht, ob die Distanz der Farbwerte des untersuchten Punktes zu der am häufigsten in der Umgebung vorkommenden Kategorie in eines erweiterten Maximalwerts zu finden sind. Sind durch diese Distanz und Nachbarschaftsuntersuchungen keine Kategorien ermittelbar, werden die untersuchten Punkte als unbestimmte Background-Punkte geführt.
In der auf diese Weise erstellten Punktwolke ist jedem Punkt eine eindeutige Kategorie zugeordnet. Somit wurde die Verknüpfung von dreidimensionalen und semantischen Informationen hergestellt.

3.3 Automatisierte Bearbeitung und Analyse einer semantisch angereicherten Punktwolke

Aufbauend auf dieser erzeugten Punktwolke erfolgen nun weitere automatisierte Analyse- und Bearbeitungsschritte.Da die Punktwolke voll automatisiert und ohne die Verwendung von Passpunkten und Marken generiert wurde, entsprechen weder ihre Position und Rotation im Raum noch ihr Maßstab den realen Gegebenheiten. Da dies jedoch für eine spätere Überführung in ein BIM-Modell benötigt wird, müssen diese Parameter nun ermittelt und mit ihnen die Punktwolke angepasst werden. Die dafür notwendigen Informationen lassen sich aus der Punktwolke unter Verwendung der in ihr enthaltenen und für diesen Schritt essentiell wichtigen semantischen Informationen extrahieren.
Als Ausgangsbasis werden die als Boden klassifizierten Punkte verwendet, da diese neben der Rotation auch eine sinnvolle Translation in die waagrechte XY-Achsenebene ermöglichen. Zu diesem Zweck wird eine Best-Fit Ebene des Bodens bestimmt und die Punktwolke in einem iterativen Prozess rotiert, bis sich die Bodenebene im lokalen Koordinatensystem nahezu in der Waagerechten befindet und die Z-Achsen annähernd übereinstimmen. Nachdem die Punktwolke so ausgerichtet wurde, wird die Deckenebene durch die zu ihr gehörenden, segmentierten Punkte bestimmt.
Basierend auf der ausgerichteten Punktwolke lassen sich nun auch Wandebenen und Wandpunkte bestimmen, welche bei der Segmentierung durch Deep Learning in den Bildern nicht berücksichtigt werden konnten. Dies geschieht mit Hilfe einer aus den unbestimmten Punkten der Punktwolke erstellten, zweidimensionalen Heatmap, welche senkrecht von oben in einem feinen Raster die

Dichte der vorhandenen Punkte darstellt. Besonders an senkrechten Bauteilen ist bei dieser Betrachtung in zwei Dimensionen eine Häufung an Punkten zu erwarten, durch die vor allem Wände gut auszumachen sind.

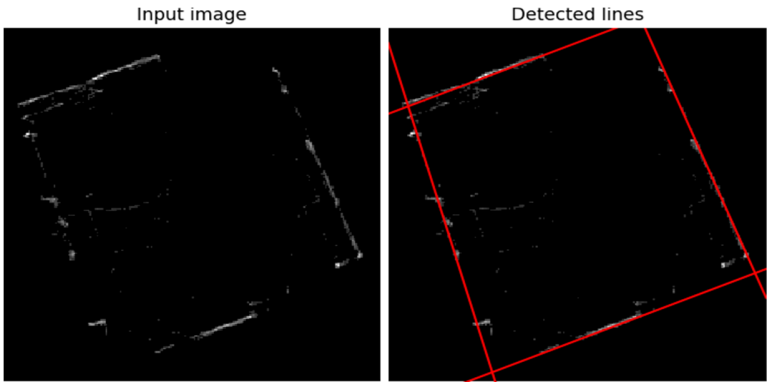

Abb. 1: Abgeleitete Heatmap des Raums (links) und die durch Hough-Linien-Transformation ermittelten Winkel der Wände (rechts)

Aus den erstellten Heatmaps werden mit einer Hough-Linien-Transformation die wahrscheinlichen Winkel der Wände ermittelt, welche im Anschluss wieder in die dritte Dimension übertragen werden. Bis dahin nicht kategorisierte Punkte, welche sich in geringem Abstand zu diesen befinden, werden ausgewählt, um aus ihnen die tatsächlichen Best-Fit Ebenen zu bestimmen. Die diesen Ebenen nahe gelegenen Punkte werden als Wandpunkte klassifiziert.
Zur Ermittlung eines Maßstabs für die Punktwolke wird das Verhältnis der Dimensionen eines Objekts in der Realität zum selben Objekt in der Punktwolke benötigt. Da sie als verbindendes Bauteil ein zentrales Element eines jeden Innenraums ist, wurde die Objektkategorie „Tür“ als Vergleichsobjekt ausgewählt. Zur Ermittlung des Maßstabs wird in der Punktwolke aus den als „Tür“ klassifizierten Punkten die größte zusammenhängende Türfläche ausgewählt. Als Vergleichsgröße wird die Höhe herangezogen. Sie ist sowohl in der Realität als auch in der Punktwolke ohne Schwierigkeiten zu ermitteln. So lässt sich die Höhe der Türfläche in der Punktwolke durch die Differenz der maximalen und minimalen Werte in der Z-Achse bestimmen, da ihre Ausrichtung zuvor korrigiert wurde. Durch den Vergleich mit der an der realen Tür mit einfachen Werkzeugen gemessenen Höhe lässt sich der Maßstab ermitteln. Basierend auf diesem wird eine Skalierung der Punktwolke durchgeführt.
Die Überführung der semantischen Informationen in die automatisiert und ohne Passpunkte generierte Punktwolke ermöglicht somit die Korrektur ihrer Rotation und Translation und erlaubt auch die Extraktion von Best-Fit Ebenen. Aufbauend auf diesen lassen sich weitere Informationen, wie Ebenen und dazugehörige Punkte extrahieren. Gemeinsam mit den vollständig aus den Daten abgeleiteten Wandebenen bildet die Boden- und Deckenebene die Raumgeometrie in ihren Grundzügen ab.
Eine Skalierung anhand eines Maßstabs durch den Vergleich der Höhe eines klassifizierten Objekts mit seinem realen Vorbild ist möglich, da die Ausrichtung der Punktwolke korrigiert wurde. Die Türe bietet sich hierbei als ein in jedem Raum vorhandenes Objekt an und hat auch auf Grund ihrer Größe gute Aussichten, in der Punktwolke in hoher Qualität vorhanden zu sein.

Die korrekt positionierte, rotierte und skalierte Punktwolke steht in Zukunft für weitergehende Untersuchungen hin zu einer Automatisierung der BIM-konformen Modellierung zur Verfügung.

4 Ergebnisse und Evaluation

4.1 Semantische Segmentierung

Aus Zeitgründen wurde darauf verzichtet, einen weiteren Datensatz für die Validierung des trainierten DeepLabv3+ Modells zu erstellen. So lassen sich für die segmentierten aktuell noch Bilder keine Aussagen über die statistische Genauigkeit treffen. Rückschlüsse auf diese lassen sich aber auch aus der in Kapitel 4.2 beschriebene Untersuchung der korrekten Farbwerte in der Punktwolke ziehen. Durch eine visuelle Bewertung der Bilder lässt sich dennoch ein Überblick über die Qualität der Segmentierungen schaffen.

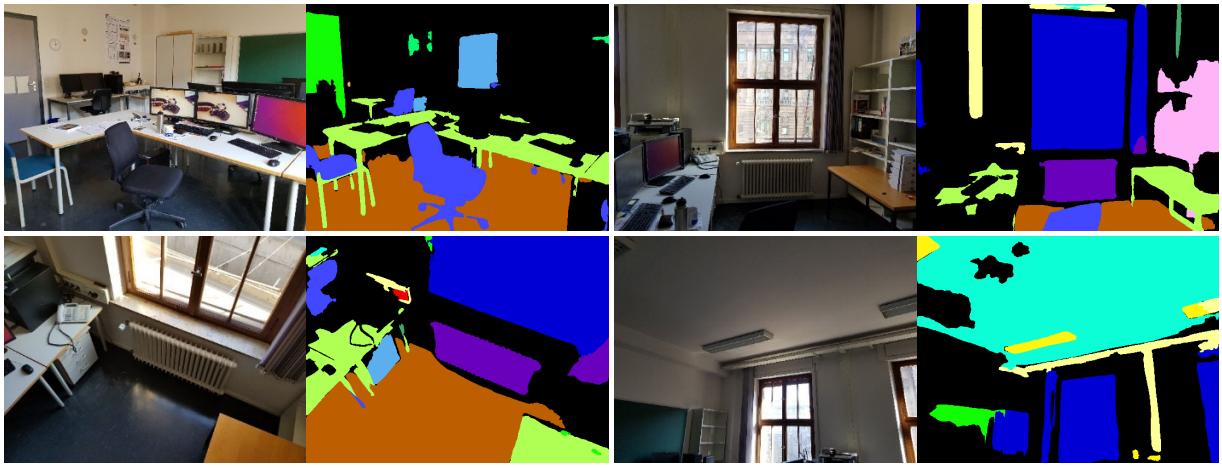

Abb. 2: Vergleich zwischen Originalbildern und mit unserem DeepLabv3+ Modell segmentierten Bildern.

Die eigentlich beim Training angegebene mIoU (mean Intersection over Union) von ca. 90% wird in den vollständig unbekannten Bildern des Büros nicht erreicht. Hierfür sind in den Bildern zu viele fehlerhafte Flächen vorhanden. Dies deutet auf ein Overfitting des Neuronalen Netzes auf die Trainingsdaten hin und lässt sich zu einem großen Teil auf die noch immer sehr kleine Datenbasis zurückführen.
Dennoch sind in den segmentierten Bildern die einzelnen Objekte zumeist korrekt erkannt worden. Hierbei wird die durch den Wechsel der Architektur des Neuronalen Netzes zu erwartende Verbesserung in der Segmentierung deutlich, da zumindest visuell deren Übereinstimmung eine höhere Qualität vermittelt als zuvor. Dies war wegen der gesteigerten Komplexität durch die Erweiterung der Kategorien nicht unbedingt zu erwarten gewesen.
Auffällig ist, dass Objekte mit relativ großer Fläche wie „Tische“ und „Stühle“, „Tür“, „Fenster“, „Boden“ und „Decke“ und besonders die „Heizung“ häufig gut segmentiert wurden. Bei ihnen ist erkennbar, dass Bereiche mit guter Segmentierung zumeist auch mit sehr präzisen Kanten ausgestattet sind. Allerdings scheinen Lücken innerhalb von segmentierten Flächen hin und wieder vorzukommen. Für diese sind besonders große, gleichförmige Bereiche wie die Decke anfällig.

Eine Ausnahme bildet die Kategorie „Regal“, bei der die Bereiche mit Ordnern häufig korrekt erkannt wurden, während Bereiche ohne diese seltener fehlerfrei segmentiert wurden. Diese Lücken werden oft durch andere Kategorien gefüllt. Kleinere Objekte wie „Lichtschalter“, „Steckdose“ und die mittelgroßen wie „Kabelkanal“, „Poster“ und „Lampe“ wurden in unterschiedlicher Qualität segmentiert. Auf ihre Segmentierungsgenauigkeit scheint der Blickwinkel einen starken Einfluss zu haben. Fehlklassifizierungen treten häufiger auch ohne ersichtlichen Grund auf. Komplett falsch klassifizierte Bereiche finden sich häufig an wiederkehrenden Stellen. So ist es für das Neuronale Netz ein Problem, Stuhl- und Tischbeine auseinanderzuhalten. Eine an der Wand hängende Tafel kennt es aus den Trainingsdaten zwar nicht, erkennt aber immer wieder Ähnlichkeiten zu Türen und segmentiert sie entsprechend. Außerdem wurden Vorhänge fälschlicherweise als „Fenster“ erkannt.
Insgesamt muss gesagt werden, dass in den Bildern große, voluminöse Bereiche besser segmentiert wurden als kleine und feine Flächen.
Aufbauend auf dieser mit DeepLabv3+ durchgeführten Segmentierung, welche viele der wichtigsten Kategorien von Innenräumen insgesamt qualitativ hochwertig extrahiert, ist es uns möglich, die semantischen Informationen in die erstellte Punktwolke zu projizieren.

4.2 Klassifizierte Punktwolke

Die automatisierte photogrammetrische Rekonstruktion einer Punktwolke ohne Passpunkte funktioniert aufbauend auf den Bildern des Büroraums ohne Probleme. Die in die Punktwolke übertragenen Farbwerte der Kategorien sind noch immer deutlich zu erkennen. So treten zwar Interpolationen auf, welche die eigentlichen Farben besonders an den Übergängen zwischen Objekten verändern und abdunkeln, aber insgesamt halten sie sich im Rahmen.

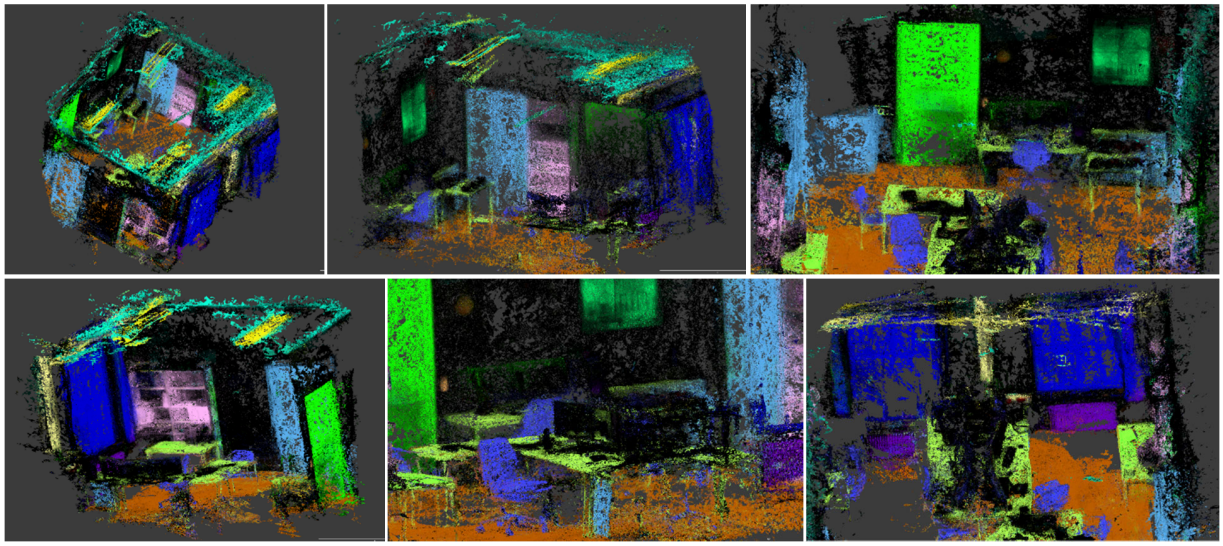

Abb. 3: Bilder der durch Photogrammetrie erzeugten Punktwolke. Aufgrund der durch Interpolation auftretenden Farbveränderungen sind die Kategorien zwar visuell unterscheidbar, aber nicht eindeutig klassifiziert.

Die erzeugte Punktwolke umfasst 16 Millionen Punkte, ist in großen Teilen vollständig und stellt ein einigermaßen gutes Abbild des Raums dar. So sind zwar verschiedene Lücken, durch eine

falsche Zuordnung entstandene falsche Punktcluster und auch einige Ausreißerpunkte vorhanden, doch ein großer Teil der Punktwolke bildet die Realität relativ genau ab.
Bei näherer Betrachtung wird deutlich, dass nicht für alle Bereiche ausreichend Überschneidungen gefunden wurden und Teile daher nur fehlerhaft rekonstruiert werden konnten. Besonders der Boden und die Decke sind hier auffällig. Beide sind großflächig und sehr gleichförmig. Die hierbei auftretenden Schwierigkeiten, Punkte in den Bildern zu extrahieren und mit den homologen Punkten anderer Bilder zu verknüpfen, führten dazu, dass Punkte der Decke fast nur in direkter Nachbarschaft zu anderen Objekten wie den Lampen oder Wänden liegen. Der Boden ist zwar vollständiger, jedoch nicht fehlerfrei rekonstruiert worden. Bei ihm handelt es sich nicht um eine zusammenhängende Fläche. Vielmehr besteht er aus verschiedenen, zusammengestückelten Bereichen, die teilweise unterschiedlich geneigt sind. Die ebenfalls großflächigen Wände sind deutlich besser erfasst worden, dennoch sind auch hier Lücken und nicht rekonstruierte Bereiche in den Ecken des Raums vorhanden. Viele der anderen in diesem Raum vorhandenen Objekte wurden gut, aber nicht immer vollständig und mit der höchsten Genauigkeit erfasst.
In dieser fehlerhaft klassifizierten Punktwolke wurden die in den einzelnen Punkten vorhandenen Farbwerte genutzt, um die durch Interpolation auftretenden Veränderungen zu eliminieren.

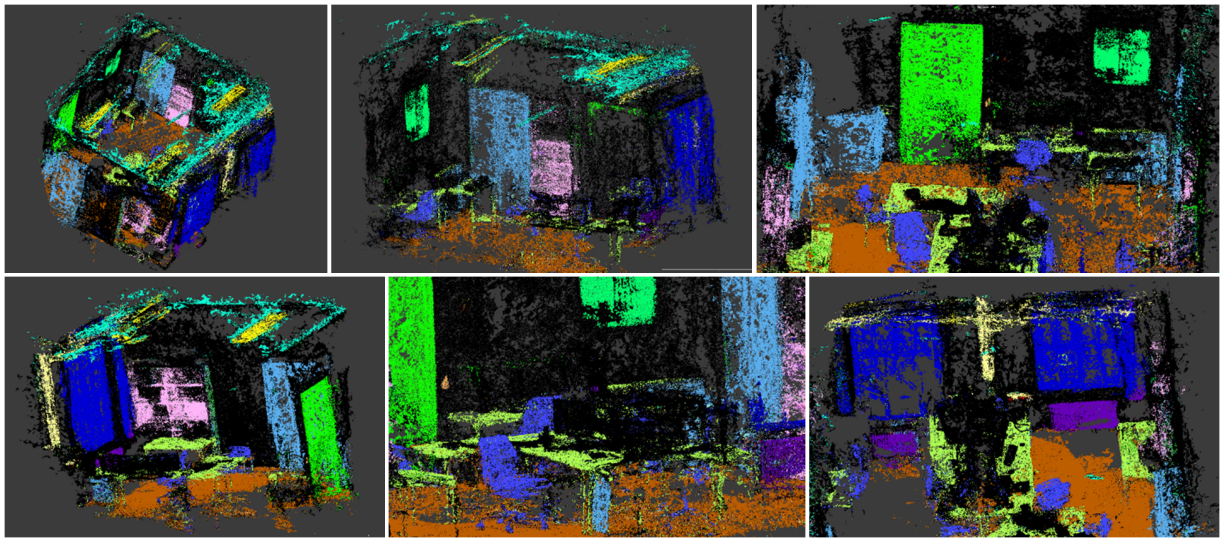

Abb. 4: Bilder der Punktwolke mit eindeutiger Klassifikation jedes einzelnen Punktes

Durch die extrahierten eindeutigen Kategorien sind diese viel besser zu unterscheiden. Es wird ersichtlich, wie gut die Übertragung der semantischen Informationen in vielen Fällen funktioniert. Durch die Überschneidung vieler segmentierter Bilder haben einzelne Fehler in diesen nur selten Einfluss auf die Klassifizierung der Punktwolke. Fehlerhafte Zuweisungen von Kategorien treten dann auf, wenn in den segmentierten Bildern regelmäßige Fehler vorhanden sind oder bei der Erstellung der Punktwolke fehlerhafte Kamerapositionen auftreten.
Basierend auf dieser eindeutig klassifizierten Punktwolke wurde eine Untersuchung zur Übereinstimmung der enthaltenen Kategorien mit den erwarteten Kategorien durchgeführt. Zu diesem Zweck wurde eine Vergleichspunktwolke mit den Soll-Kategorien erzeugt. So können für die in

dem untersuchten Büro vorhandenen Objektkategorien Aussagen über deren Genauigkeit und daraus abgeleitete Rückschlüsse auf die Segmentierung getroffen werden.
In den Genauigkeitsberechnungen wurden während der Interpolationskorrektur entfernte Ausreißerpunkte ebenso wenig berücksichtigt wie fehlerhafte Punkte aus der Vergleichspunktwolke.

Tab. 2: Genauigkeitsuntersuchung der in der Punktwolke enthaltenen Kategorien.

	Mittel	Background	Boden	Tür	Fenster	Lichtschalter	Steckdose	Lampe	Heizung	Poster	Kabelkanal	Schrank	Decke	Regal	Tisch	Stuhl
IoU	**51,9**	60,8	75,1	77,9	75,6	6,3	10,8	59,2	66,4	43,6	41,6	71,4	53,5	38,4	43,3	54,9
Accuracy	**60,4**	93,8	77,1	84,4	88,8	48,2	11,3	62,2	70,5	45,3	43,0	74,2	54,7	38,8	50,0	63,4

Zur Einschätzung der Qualität wird erneut die mIoU herangezogen. Diese beträgt für die Punktwolke gute 51,9 %. Außerdem wird zusätzlich die Genauigkeit der Kategorien, oft auch als Accuracy bezeichnet, berechnet. Ihr Durchschnitt für alle Kategorien beträgt sehr gute 60,4 % und deutet auf eine hohe Übereinstimmung hin. Betrachtet man die einzelnen Kategorien, zeigt sich ein weit gestecktes Feld an Werten. Die beste IoU erreicht die Kategorie „Tür“ mit 77,9 % dicht gefolgt von „Fenster“ mit 75,6 % und „Boden“ mit 75,1 %. Auch die knapp dahinliegenden Kategorien sind noch immer sehr gut. Entsprechend sind viele Objekte in der Punktwolke zu großen Teilen korrekt segmentiert.
Die Klassen, in denen Fehler in der Objektklassifizierung am deutlichsten sichtbar werden, sind die „Poster“, welche teilweise Lücken in der Segmentierung aufweisen, und besonders die „Regale“. Diese sind, wie ja bereits in den Bildern festzustellen war, nur unzureichend segmentiert. Die stattfindende Interpolation hat bei ihnen geholfen, verschieden auftretende falsche Segmentierungen auszugleichen. Dennoch sind die Regale nur unvollständig als solche klassifiziert. Einen positiven Einfluss hatte die Interpolation auch bei der vorhandenen Tafel. So wird diese nur in einem sehr kleinen Bereich als „Tür“ segmentiert.
Außerdem ist auffällig, dass besonders die Kategorien „Lichtschalter“ mit 6,3 % und „Steckdose“ mit 10,8 % nur eine sehr geringe IoU haben. Bei diesen beiden handelt es sich um die Kategorien mit der geringsten Anzahl an erwarteten Punkten (Lichtschalter: 5038, Steckdose: 6104). Entsprechend schnell haben falsch segmentierte Bereiche in den Bildern einen großen Einfluss auf die ermittelten Werte. So sind diese bei der „Steckdose“ oft als Background-Punkte klassifiziert worden. Betrachtet man bei den Lichtschaltern zusätzlich die errechnete Accuracy, wird deutlich, dass die eigentliche Klassifizierung der Vergleichsbereiche der Lichtschalter mit 48,2 % einen deutlich besseren Wert aufweist. Dies hat den Hintergrund, dass kleine Bereiche anderer Klassen, in diesem Fall „Boden“ mit 25741 Punkten, fälschlich als Lichtschalter klassifiziert wurden und einen sehr starken Einfluss auf die IoU dieser Kategorie haben. Die Kategorie „Boden“ weist eine ähnliche, leicht dunklere Farbe auf wie der „Lichtschalter“. So scheint für die schlechte IoU von 6,3% vor allem die bei der Überführung der semantischen Informationen auftretende Interpolation der Farben verantwortlich zu sein. Dies verdeutlicht ungefähr die Dimensionen, in denen die Interpolation nach ihrer Korrektur eine Rolle für die Klassifizierung der Punktwolke spielt. Für die Berechnung

der Qualität des Lichtschalters ist das viel, im Zusammenhang mit der kompletten Punktwolke scheinen diese jedoch keine wesentliche Rolle zu spielen.
Insgesamt ist die Qualität der Punktwolke in Ordnung. Besonders die in ihr segmentierten Kategorien sind ein sehr gutes Abbild der realen Objekte. Die bei der Projektion der semantischen Informationen auftretenden Interpolationen haben wenige negative, im Gegenteil zumeist sogar positive Auswirkung auf die Klassifizierung der Punkte, da einzelne fehlerhafte Segmentierungen aus den Bildern nicht übernommen werden.

4.3 Post-Processing der Punktwolke

Die eindeutig klassifizierte Punktwolke wurde durch Rotation und Translation korrekt ausgerichtet.

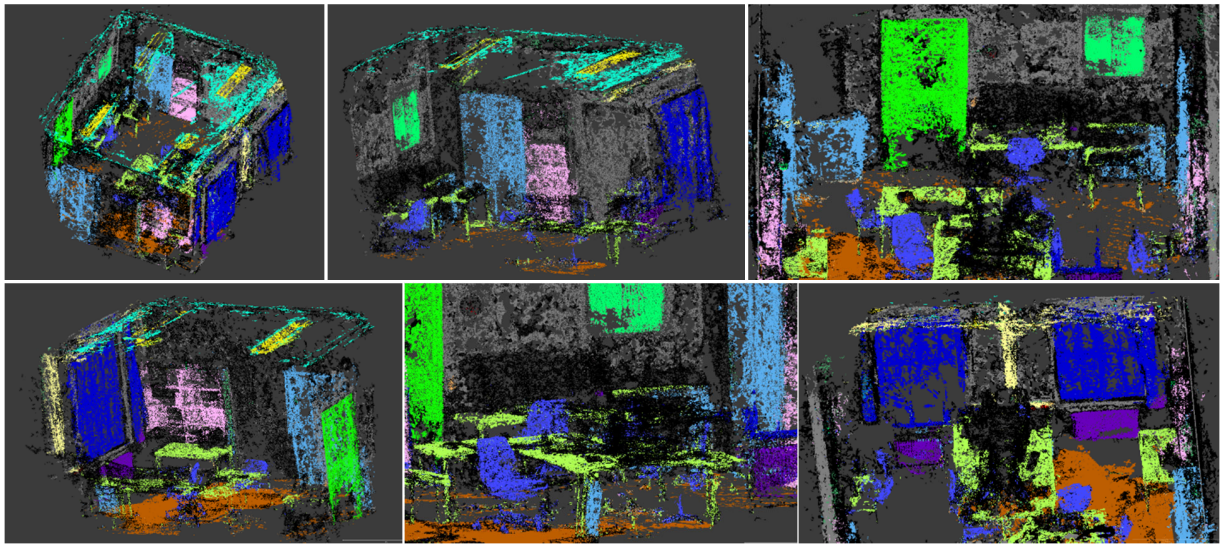

Abb. 5: Durch Rotation und Translation angepasste Punktwolke und zusätzlich in dieser segmentierte Wandpunkte.

Boden und Deckenpunkte wurden auf ihre jeweiligen Ebenen projiziert. Außerdem wurden weitere starke Ausreißerpunkte ebenso wie Punkte unterhalb der Bodenebene und oberhalb der Deckenebene entfernt. Dies hat zwar zur Folge, dass in schlecht rekonstruierten Bereichen mehr Lücken auftreten, die Punktwolke jedoch deutlich klarer aussieht und besser mit der Realität übereinstimmt. Die basierend auf den ermittelten Wandebenen bestimmten Wandpunkte fügen sich gut in das Gesamtbild dieser verbesserten Punktwolke ein und bilden die Wände korrekt ab.
Nach der Skalierung der Punktwolke mit dem ermittelten Maßstab wurden verschiedene Objekte zu Vergleichszwecken gemessen. Bei diesen wurde eine gute Übereinstimmung zwischen Soll- und Istwerten festgestellt. Die Höhe eines Schreibtisches wurde in der Realität auf 72,3 cm gemessen, während in der Punktwolke durchschnittlich 72,1 cm gemessen wurden. Die Breite eines Schranks betrug in der Realität 110 cm, in der Punktwolke durchschnittlich 108,9 cm. Insgesamt zeigt sich, dass die Skalierung der Punktwolke gut funktioniert hat und deren Ausmaße sehr gut mit der Realität übereinstimmen.

Die automatisierte Nachbearbeitung der Punktwolke ermöglichte es uns somit, fehlerhafte Bereiche zu korrigieren, weitere Informationen aus dieser zu extrahieren und sie basierend auf diesen zu optimieren.
Dies verdeutlicht das Potential, welche eine Kombination von semantischen und dreidimensionalen Informationen für die automatisierte Auswertung bietet.

5 Fazit

Die kombinierte Extraktion von geometrischen und semantischen Informationen wurde aufbauend auf der Segmentierung mit DeepLabv3+ und der Projektion in die photogrammetrisch erstellte Punktwolke deutlich verbessert. So sind die gesuchten Objekte sehr gut in der Punktwolke differenzierbar. Diese hohe Genauigkeit wird auch durch die bei der Klassifikation der Punktwolke erreichte mIoU von 51,9 % unterstrichen.
Außerdem wurde gezeigt, dass durch die Analyse und Nachbearbeitung der Punktwolke wichtige Informationen extrahiert werden können, die für eine Automatisierung der Auswertung hin zu einem BIM-Modell von essentieller Bedeutung sind.
Wir sind zuversichtlich, die noch auftretenden Probleme durch eine Optimierung der verwendeten Verfahren minimieren zu können.
In Zukunft ist es angedacht, eine Ausweitung dieses Ansatzes auf die Verknüpfung und gemeinsame Auswertung mehrerer Innenräume herzustellen und dabei auch mobile Laserscanner mit einzubeziehen.
Mit den von uns vorgestellten Methoden haben wir eine solide Grundlage für die Erfassung und Modellierung von semantischen und geometrischen Informationen von Innenräumen für BIM-Modelle hin zu ihrer automatisierten Rekonstruktion geschaffen.

6 Danksagung

Das Projekt „i_city: BIM-konforme Gebäudeerfassung“ wird vom Bundesministerium für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 13FH9E01IA gefördert und vom Projektträger VDI Technologiezentrum GmbH für das BMBF betreut.

7 Literaturverzeichnis

AGISOFT, 2020: Agisoft Metashape. http://www.agisoft.com/.

BORRMANN, A., KÖNIG, M., KOCH, C. & BEETZ, J., 2015: Building Information Modeling - Technologische Grundlagen und industrielle Praxis. VDI-Buch, Springer Vieweg, Wiesbaden, 343-361.

BOULCHA, A., GUERRY, J., LE SAUX, B. & AUDEBERT, N., 2017: SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks. Computers & Graphics **71**, 189-198 https://doi.org/10.1016/j.cag.2017.11.010.

BRAMANN, H., MAY. I. & PLANEN-BAUEN 4.0 – GESELLSCHAFT ZUR DIGITALISIERUNG DES PLANENS, BAUENS UND BETREIBENS MBH, 2015a: Stufenplan Digitales Planen und Bauen.

https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/stufenplan-digitales-bauen.pdf?__blob=publicationFile.

BRAMANN, H., MAY. I. & PLANEN-BAUEN 4.0 – GESELLSCHAFT ZUR DIGITALISIERUNG DES PLANENS, BAUENS UND BETREIBENS MBH, 2015b: Konzept zur schrittweisen Einführung moderner, IT-gestützter Prozesse und Technologien bei Planung, Bau und Betrieb von Bauwerken – Stufenplan zur Einführung von BIM. https://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-stufenplan-endbericht.pdf?__blob=publicationFile.

CHEN, L.-C., PAPANDREOU, G., KOKKINOS, I., MURPHY, K. & YUILLE, A. L., 2016: Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. arXiv:1412.7062.

CHEN, L.-C., ZHU, Y., PAPANDREOU, G., SCHROFF, F. & ADAM, H., 2018: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. arXiv:1802.02611v1.

CLEMEN, C. & EHRICH, R., 2014: Geodesy goes BIM. In: allgemeine vermessungs-nachrichten. (avn) **121**(6), 231-237, https://gispoint.de/artikelarchiv/avn/2014/avn-ausgabe-62014/2552-geodesy-goes-bim.html.

EGGER, M., HAUSKNECHT, K., LIEBICH, T. & PRZYBYLO, J., 2013: BIM-Leitfaden für Deutschland. http://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-leitfaden-deu.pdf?__blob=publicationFile.

ESCHENBRUCH, K., MALKWITZ, A., GRÜNER, J., POLOCZEK, A. & KARL, C. K., 2014: Maßnahmenkatalog zur Nutzung von BIM in der öffentlichen Bauverwaltung unter Berücksichtigung der rechtlichen und ordnungspolitischen Rahmenbedingungen – Gutachten zur BIM-Umsetzung. https://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-massnahmenkatalog.pdf?__blob=publicationFile.

GIRSHICK, R., DONAHUE, J., DARRELL, T. & MALIK, J., 2014: Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv:1311.2524v5.

GIRSHICK, R., 2015: Fast R-CNN. https://arxiv.org/pdf/1504.08083.pdf.

HE, K., ZHANG, X., REN, S. & SU, J., 2015: Deep Residual Learning for Image Recognition. arXiv:1512.03385.

HUANG, G., LIU, Z., VAN DER MAATEN, L. & WEINBERGER, K. Q., 2018: Densely Connected Convolutional Networks. arXiv:1608.06993.

JÉGOU, S., DROZDZAL, M., VAZQUEZ, D., ROMERO, A. & BENGIO, Y., 2017: The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. arXiv:1611.09326.

KADEN, R., CLEMEN, C., SEUß, R., BLANKENBACH, J., BECKER, R., EICHHORN, A., DONAUBAUER, A.,KOLBE, T. H., GUBER, U., DVW – GESELLSCHAFT FÜR GEODÄSIE, GEOINFORMATION UND LANDMANAGEMENT E. V. & RUNDER TISCH GIS E.V., 2017: Leitfaden Geodäsie und BIM. https://rundertischgis.de/images/2_publikationen/leitfaeden/GeoundBIM/Leitfaden%20Geod%C3%A4sie%20und%20BIM_Onlineversion.pdf.

KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E., 2012: ImageNet Classification with Deep Convolutional Neural Networks. https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

OBROCK, L. S. & GÜLCH, E., 2018: First Steps To Automated Interior Reconstruction From Semantically Enriched Point Clouds And Imagery. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **42**(2), 781-787, https://doi.org/10.5194/isprs-archives-XLII-2-781-2018, 2018.

REN, S., HE, K., GIRSHICK, R. & SUN, J., 2016: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497.

SHELHAMER, E., LONG, J. & DARRELL, T., 2016: Fully Convolutional Networks for Semantic Segmentation. arXiv:1605.06211.

SIMONYAN, K. & ZISSERMAN, A., 2015: Very Deep Convolutional Networks For Large-Scale Image Recognition. arXiv:1409.1556.

SONG, S. & XIAO, J., 2014: Sliding Shapes for 3D Object Detection in Depth Images. In: Proceedings of the 13th European Conference on Computer Vision (ECCV2014). http://slidingshapes.cs.princeton.edu/paper.pdf.

SONG, S. & XIAO, J., 2016: Deep Sliding Shapes for Amodal 3D Object Detection in RGB-D Images. 29th IEEE Conference on Computer Vision and Pattern Recognition, 808-81, http://dss.cs.princeton.edu/paper.pdf.

SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCKE, V. & RABINOVICH, A., 2014: Going deeper with convolutions. arXiv:1409.4842.

XIE, S., GIRSHICK, R., DOLLÁR, P., TU, Z. & HE, K., 2017: Aggregated Residual Transformations for Deep Neural Networks. arXiv:1611.05431.

ZHAO, H., SHI, J., QI, X., WANG, X. & JIA, J., 2017: Pyramid Scene Parsing Network. arXiv:1612.01105.