

## IMPLEMENTIERUNG VON KI-BASIERTEN REFERENZPROZESSEN FÜR DIE COMPUTERGESTÜTZTE OBJEKTERKENNUNG IM GEBÄUDE

M. Leonhardt<sup>1</sup>, N. Pauen<sup>1</sup>, L. Kirnats<sup>1</sup>, J.-N. Joost<sup>1</sup>, J. Frisch<sup>1</sup>, C. van Treeck<sup>1</sup>

<sup>1</sup>E3D - Lehrstuhl für Energieeffizientes Bauen, RWTH Aachen, Aachen, Deutschland

### KURZFASSUNG

Im Gebäudesektor und speziell in der Bauwirtschaft resultieren aus der Digitalisierung im Rahmen des Building Information Modeling (BIM) heterogene Datenmengen. Die darin enthaltenen visuellen Informationen können für eine Objekterkennung, die auf dem Konzept der Künstlichen Intelligenz (KI) beruht, genutzt werden. Die vorliegende Entwicklung eines datengetriebenen Referenzprozesses behandelt diese Problemstellung. Unter Zuhilfenahme des KI-basierten Objekterkennungssystems PointNet können innerhalb eines Gebäudes Objekte klassifiziert und Szenen semantisch segmentiert werden. Das ursprünglich für Punktwolken konzeptionierte Objekterkennungssystem wird dahingehend erweitert, dass auch Gebäudedatenmodelle im Format der Industry Foundation Classes (IFC) für die Objekterkennung verwendet werden können. Diese Vorgehensweise führt insbesondere im Gewerk der technischen Gebäudeausrüstung zu einer Erweiterung der vorhandenen Datengrundlage, die für die maschinelle Objekterkennung notwendig ist.

### ABSTRACT

In the building sector and especially in the construction industry, digitization in the context of Building Information Modelling (BIM) results in heterogeneous data. The visual information contained can be used for object recognition based on the concept of Artificial Intelligence (AI). The present development of a data-driven reference process deals with this problem. Using the AI-based object recognition system PointNet, objects within a building can be classified and scenes can be semantically segmented. The object recognition system, originally designed for point clouds, will be extended so that building data models in the Industry Foundation Classes (IFC) format can also be used for object recognition. This approach leads to an extension of the existing data sets, especially with regard to the building technology, which is necessary for AI-based object recognition.

### EINLEITUNG

Die digitale Transformation bezeichnet im Allgemeinen nicht nur die Weiterentwicklung von

technischen Möglichkeiten, sondern vielmehr die Integration von intelligenten Daten in das gesamte, menschliche Handeln (IDC 2018). Die dadurch erzeugten Datenmengen bieten zum jetzigen Zeitpunkt unausgeschöpfte Potenziale, die viele Fragestellungen und Probleme des Alltags lösen könnten (BMW 2018). Auch die Bauwirtschaft bleibt hiervon nicht unberührt. Sie ist in Deutschland aktuell einer der am geringsten digitalisierten Wirtschaftszweige (Accenture Strategy 2016). Dementsprechend groß sind die Entwicklungspotenziale. Der herkömmliche Informationsaustausch innerhalb des Gebäudesektors bei Bau, Umbau oder im Facility Management eines Bauwerks entspricht der Übermittlung von technischen Zeichnungen sowie von Informationen über das Bauwerk in grafischer Form (beispielsweise Grundrisse, Ansichten oder Schnitte). Diese Zeichnungen und Dateien, die große Informationsmengen beinhalten, können vom Computer meist nicht selbstständig verarbeitet und interpretiert werden (Borrmann et al. 2015).

Building Information Modeling (BIM) bietet dabei genau jene Datenbasis für den Gebäudesektor, die für digitalisierte Anwendungen und Prozesse notwendig ist. Informationen werden dabei durch eine logische Verknüpfung verschiedener Datenbanken als 3D-Gebäudemodell digital aufbereitet und im Lebenszyklus an die Beteiligten übergeben (van Treeck et al. 2016). Die Industry Foundation Classes (IFC) (ISO 16379 2013) sind das gängigste Produktdatenmodell zur Weitergabe dieser Informationen (Hausknecht and Liebich 2016). Das Modell basiert auf EXPRESS und ermöglicht mit über 800 Objektdefinitionen und tausenden Eigenschaften im Release IFC4.1 (buildingSMART 2020) einen herstellerneutralen Datenaustausch über alle Phasen des Projektlebenszyklus.

Neben diesen digitalen Gebäudemodellen aus der Planung gibt es eine weitere relevante dreidimensionale Repräsentation eines existierenden Gebäudes. Durch digitalisierte Bestandsaufnahmen werden Punktwolken mit Hilfe von Topometrie, stationären oder mobilen 3D-Laserscannern oder durch Photogrammetrie (Hichri et al. 2013) generiert. Jeder Punkt  $P_i$  dieser Punktwolke ist dabei ein

mehrdimensionaler Vektor und wird mindestens durch die drei Koordinaten  $(x_i, y_i, z_i)^T \in \mathbb{R}^3$  beschrieben. Punktwolken sind dabei nicht nach einzelnen Objekten segmentiert, sondern umfassen ganze Szenen. Das Forschungsgebiet der As-Built-Dokumentation behandelt in diesem Zusammenhang die Verarbeitung und Überführung der Punktwolken in Gebäudedatenmodelle.

Ein Werkzeug zur Nutzung der geometrischen und visuellen Daten aus den digitalen Modellen im Gebäudesektor ist die Computer Vision als Teilgebiet der Künstlichen Intelligenz. Das Gebiet der Computer Vision umfasst die computergestützte Lösung von Aufgabenstellungen, die sich an den Fähigkeiten des menschlichen visuellen Systems orientieren. Dazu gehört unter anderem die Bildklassifizierung, die Objektlokalisierung und die semantische Segmentierung (Süße and Rodner 2013). Diese Aufgaben sind in Abbildung 1 dargestellt.



Abbildung 1: Bildklassifizierung (links), Objektlokalisierung (mitte) und semantische Segmentierung (rechts) (Lin et al., 2013)

Vielversprechende Forschungsansätze verwenden zur Lösung der Objekterkennung das Konzept des Deep Learning. Deep Learning ist eine spezielle Art des überwachten Lernens im Bereich des Machine Learning und ermöglicht es dem Computer, komplexe Konzepte aus einfacheren Konzepten aufzubauen. Weitere Informationen hierzu sind (Goodfellow et al. 2016) zu entnehmen.

Das vorliegende Paper präsentiert einen Referenzprozess, um visuelle und geometrische Daten aus digitalen Gebäudemodellen mithilfe des Computers automatisiert zu verarbeiten und zu interpretieren. Es wird ein existierendes Objekterkennungssystem, das zum aktuellen Zeitpunkt keine Gebäudedatenmodellen verarbeiten kann, als Grundlage verwendet und durch den vorangestellten Referenzprozess erweitert. Das Ergebnis ist die Klassifizierung von Objekten und die semantische Segmentierung von Szenen sowohl in Punktwolken als auch in \*.ifc-Dateien innerhalb eines einzigen Systemes.

## STAND DER FORSCHUNG

Objekterkennungssysteme für 2D-Daten sind zum aktuellen Zeitpunkt wesentlich weiter entwickelt als jene für 3D-Daten, da die Verfügbarkeit der Daten schon länger sichergestellt ist (Su et al. 2015). Die zunehmende Entwicklung von dreidimensionalen Datensätzen ist hauptsächlich durch die fortschreitende Technologie in Hinsicht auf die Hardware begründet und bildet die Grundlage der Entwicklung dreidimensionaler Objekterkennungs-

systeme (Armeni et al. 2017). Bei diesen Systemen muss in einem ersten Schritt differenziert werden, ob sie für Punktwolken oder Oberflächenmodelle konzipiert sind (Wang et al. 2019).

Oberflächenmodelle entsprechen in erster Linie synthetischen Abbildungen der Objekte. In der Literatur existieren zwei unterschiedliche Ansätze, die ansichtsbasierte und die volumenbasierte Objekterkennung. Ansichtsbasierte Systeme reduzieren im Vorhinein die Dimension der Daten durch ein sogenanntes Rendering, sodass die 3D-Form aus den entstehenden Ansichten erlernt werden kann (Han et al. 2019). Diese Ansichten werden dann durch die weitaus höher entwickelten Objekterkennungssysteme für 2D- bzw. 2.5D-Daten weiterverarbeitet.

Volumenbasierte Ansätze extrahieren die Merkmale direkt aus den 3D-Oberflächenmodellen. Sie können dahingehend unterschieden werden, ob sie als Repräsentation Voxel Grids (Sedaghat et al. 2016, Brock et al. 2016), Meshes (Boscaini et al. 2015, Han et al. 2018) oder sonstige Repräsentationen wie Graph Pooling (Dominguez et al. 2018) verwenden. Diese Ansätze können jedoch keine reinen Punktwolken verarbeiten.

Werden bei der Bestandsaufnahme von existierenden Gebäuden ausschließlich die visuellen Gebäudedaten, also Geometrie und Topologie, betrachtet, so wird eine Überführung der Punktwolken in Meshes bzw. Voxel Grids als Oberflächenrekonstruktion bezeichnet. Die Systeme ShapeNet (Wu et al. 2015) und VoxNet (Maturana and Scherer 2015) transformieren gegebene Punktwolken in dreidimensionale Repräsentationen und führen daraufhin einen volumenbasierten Ansatz der Objekterkennung durch. Derzeit steht die Wissenschaft jedoch noch vor dem Problem, die gewonnenen Erkenntnisse auf ganze Szenen anzuwenden. Dies ist aber für die semantische Segmentierung unabdingbar, weshalb eine Datenvorbereitung der Punktwolken durch die Rekonstruktion der Oberflächen auszuschließen ist (Cheraghian and Petersson 2019).

Das System PointNet (Qi et al. 2017a) kann Punktwolken ohne vorherige Vorverarbeitung mit einer hohen Genauigkeit klassifizieren und semantisch segmentieren. Parallel zur Entwicklung von PointNet wurde in (Ravanbakhsh et al. 2016) ein Modell aus Layern beschrieben, das Merkmale aus ungeordneten Mengen extrahiert. Diese Layer liegen allerdings hinsichtlich der Genauigkeit sowie der Anwendungsfälle hinter PointNet. Die darauffolgenden Werke, die Punktwolken ohne Transformation verwenden, basieren und verbessern jeweils einen Teil von PointNet. In (Angelina Uy and Hee Lee 2018, Wang et al. 2018) wird die Aggregation von PointNet verbessert. Generelle Verbesserungen hinsichtlich der Feature Extraction wurden im Rahmen von (Qi et al. 2017b, Li et al. 2018, Wang et

al. 2019) entwickelt. Eine Verbesserung der Extraction von lokalen Features erfolgte in (Xu et al. 2018, Liu et al. 2019). Die veröffentlichte Arbeit 3DCapsule (Cheraghian and Petersson 2019) bietet, basierend auf der Vorverarbeitung von PointNet, einen alternativen Classifier, der nicht auf Fully Connected Layern basiert.

Ein Vergleich der Ansätze am Benchmark ModelNet40 ist in Tabelle 1 gegeben.

*Tabelle 1:  
Vergleich der 3D-Objekterkennung für die  
Objektklassifizierung sowie semantische  
Segmentierung von Punktwolken anhand der  
Gesamtgenauigkeit (Overall Accuracy) in Prozent*

Ansatz	Klassifizierung	Segmentierung
Qi et al., 2017a	89.2	78.62
Wang et al., 2019	93.5	84.1
Qi et al., 2017b	91.9	x
Xu et al., 2018	92.4	x
Wang et al., 2018	89.9	x
Cheraghian and Petersson, 2019	92.7	x
Ravanbakhsh et al., 2016	87	x
Achlioptas et al., 2017	84.5	x
Li et al., 2018	93.4	x

**METHODIK**

**Anforderungen**

Unter Beachtung der verfügbaren Datengrundlage und dem aktuellen Forschungsstand im Bereich der Objekterkennung ergeben sich vier zentrale

Anforderungen an das Objekterkennungssystem und den Referenzprozess:

- I. Die Objekterkennung basiert auf Deep Learning.
- II. Das Objekterkennungssystem kann mit den zentralen 3D-Repräsentationen, also Gebäudedatenmodellen und Punktwolken, umgehen.
- III. Repräsentativ für das BIM-Modell ist eine Abbildung dessen im IFC-Datenformat, also auf physischer Ebene eine \*.ifc-Datei.
- IV. Objekte können klassifiziert und Szenen semantisch segmentiert werden.

Es existiert kein Deep-Learning-basiertes Objekterkennungssystem, das sowohl rohe Punktwolken als auch 3D-Oberflächenmodelle für die Anwendungsfälle der Klassifizierung und der semantischen Segmentierung verarbeiten kann. Mit Ausnahme der Möglichkeit zur Verarbeitung von Gebäudedatenmodellen, repräsentiert durch das IFC-Datenformat, entspricht das Objekterkennungssystem PointNet allen Anforderungen. Zudem basieren alle weiteren Frameworks zur Verarbeitung von Punktwolken auf PointNet. Die Adaptivität hinsichtlich der Systemanpassung von PointNet ist somit als hoch einzuschätzen. PointNet unterstützt außerdem als einziges Framework für Punktwolken sowohl die Klassifizierung als auch die semantische Segmentierung vollumfänglich. Da die Erweiterung von (Wang et al. 2019), die gemäß Tabelle 1 eine höhere Genauigkeit aufweist, zum Zeitpunkt der Entwicklung der Methodik noch nicht veröffentlicht worden war, wird PointNet (Qi et al. 2017a) im Weiteren für die Entwicklung des datengetriebenen Referenzprozesses verwendet. Der Referenzprozess erweitert das Objekterkennungssystem also dahingehend, dass allen genannten Anforderungen entsprochen wird.

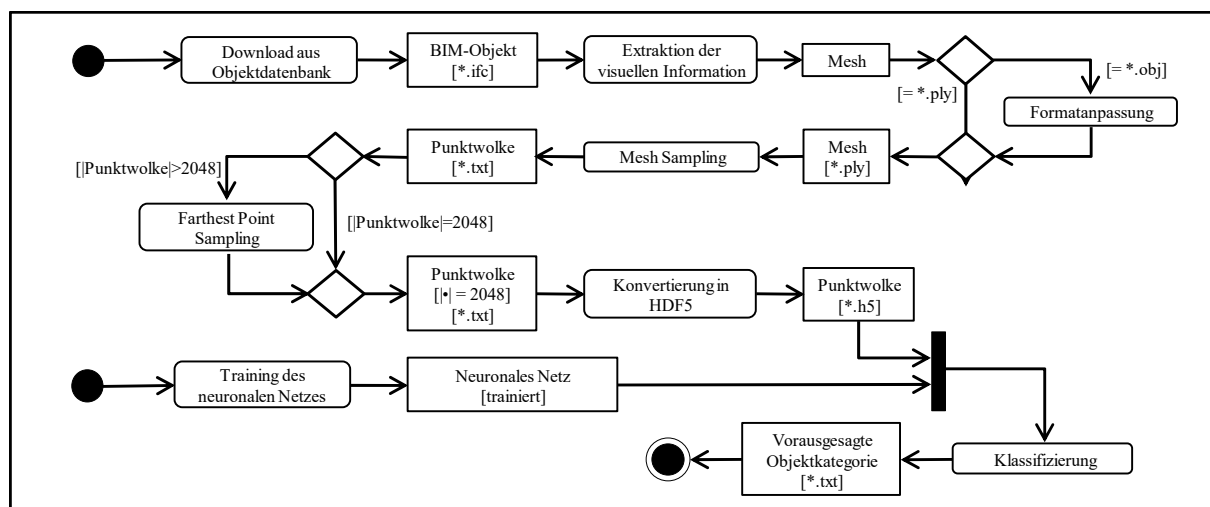


Abbildung 2: Datengetriebener Referenzprozess der Klassifizierung als UML-Diagramm

**Klassifizierung**

Der datengetriebene Referenzprozess für die Klassifizierung ist Abbildung 2 zu entnehmen.

Produktkataloge werden in sogenannten BIM-Objektdatenbanken zusammengefasst und ergänzen bei der Modellierung von Gebäuden die Modellbibliothek in einer CAD-Umgebung (van Treeck et al. 2016). Diese Datenbanken enthalten parametrische Modelle der Bauprodukte in verschiedenen Formaten. Exemplarisch wird die Datenbank bimobject verwendet.

Die Datengrundlagen der Informationsextraktion sind die \*.ifc-Dateien von Objekten. Innerhalb des Resource Layers ist die Geometrie und Topologie des Objektes enthalten. Diese visuellen Informationen müssen aus der physischen \*.ifc-Datei extrahiert und anschließend in einem geeigneten 3D-Geometrieformat abgespeichert werden. Dafür kann das Tool IFCConvert aus der IFCOpenShell Bibliothek (Krijnen 2012) verwendet werden. Das Ergebnis dieses Vorgangs sind die jeweils in einer \*.obj-Datei gespeicherten Meshes der Objekte.

Im nächsten Schritt wird ein Mesh Sampling durchgeführt. In diesem Prozess werden die Meshes in Punktwolken überführt, indem auf den Flächenstücken der Meshes durch Abtast-Algorithmen zufällig erzeugte Punkte generiert werden. Dieses Mesh Sampling kann durch die Verwendung der Python-Bibliothek PyntCloud erfolgen. Das Ergebnis dieses Vorgangs ist jeweils die Punktwolke des Objektes mit der Mächtigkeit von 2.048 oder größer als \*.txt-Datei. Da PointNet als Input Punktwolken mit einer Mächtigkeit von 2.048 fordert, müssen Objekte mit mehr Punkten reduziert werden. Dies geschieht durch einen Farthest-Point-Sampling-Algorithmus gemäß Abbildung 3.

```

1: procedure FPS( $\{p_1, p_2, \dots, p_r \mid p_i = (x_i, y_i, z_i)^T \forall i \in \{1, \dots, r\}\}$ )
2:    $P \leftarrow \{1, 2, 3, \dots, r\}$            ▶ Input: Punktwolke
3:    $P' \leftarrow \emptyset$ 
4:   int i                               ▶ Geforderte Ganzzahligkeit
5:   int j
6:   int nextpoint
7:    $j = \text{generate\_random\_integer}(1, v)$  ▶ Zufällige Zahl zwischen 1 und v
8:    $P = P \setminus \{j\}$                    ▶ j aus P entfernen
9:    $P' = \{j\}$                            ▶ j zu P' hinzufügen
10:  repeat
11:    nextpoint = arg max $_{i \in P, \min_{j \in P'} \|P_i - P_j\|}$ 
12:     $P = P \setminus \{\text{nextpoint}\}$ 
13:     $P' = P' \cup \{\text{nextpoint}\}$ 
14:  until  $|P'| = 2048$                    ▶ Wiederholen bis Punktzahl erreicht
15:  return  $P'$                              ▶ Output: Menge der Positionen der reduzierten Punktwolke
16: end procedure
    
```

Abbildung 3: Farthest-Point-Sampling-Algorithmus

PointNet stellt im Rahmen der Veröffentlichung des Quellcodes eine Sammlung von implementierten Funktionen bereit, welche die für das Training des Klassifizierungsnetzwerkes notwendigen \*.h5-Dateien schreibt und speichert. Das Labeln, das heißt die Zuweisung der Kategorie der Objekte, erfolgt manuell innerhalb des Codes. Die \*.h5-Dateien bilden die Datengrundlage der Klassifizierung. Die softwareseitige Voraussetzung bildet dabei das trainierte neuronale Netz von PointNet. Die Klassifizierung der Objekte erfolgt dann auf Basis der Daten, die das Ergebnis des datengetriebenen Referenzprozesses sind.

**Semantische Segmentierung**

Der datengetriebene Referenzprozess für die semantische Segmentierung ist Abbildung 4 zu entnehmen.

Für die Entwicklung des datengetriebenen Referenzprozesses ist die Verfügbarkeit eines beispielhaften BIM-Modells eines gesamten Gebäudes notwendig. Vorliegend wurde das BIM-Modell des eigenen Institutsgebäudes verwendet. Die physische Austauschdatei besitzt das IFC-Dateiformat und ähnelt in dieser Hinsicht der Datenbasis des bereits erläuterten datengetriebenen Referenzprozesses der Klassifizierung.

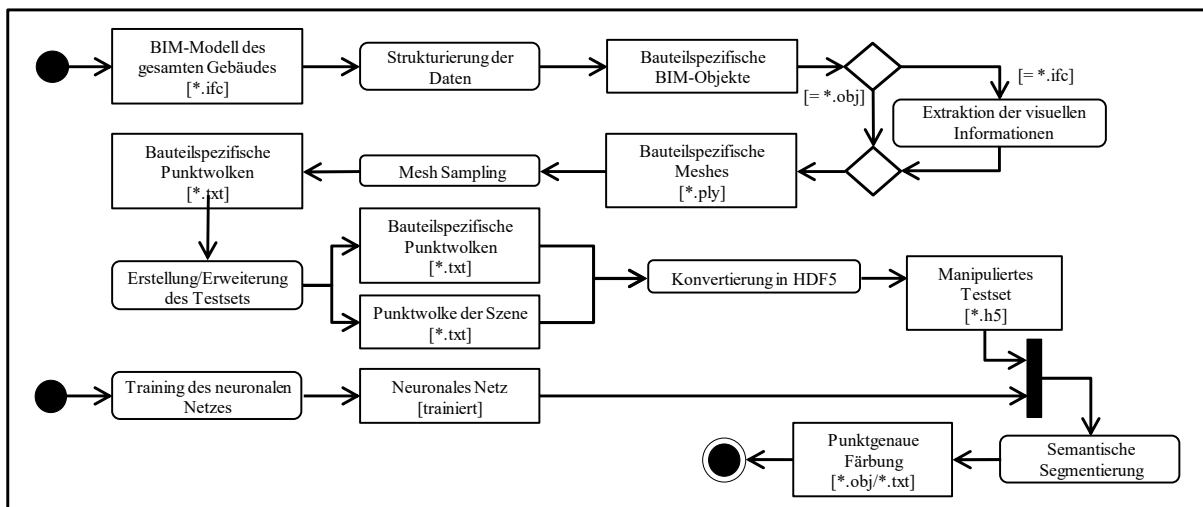


Abbildung 4: Datengetriebener Referenzprozess der semantischen Segmentierung als UML-Diagramm

Das neuronale Netz im Rahmen von PointNet kann Szenen, die einzelne Räume enthalten, semantisch segmentieren. Das gesamte Gebäudemodell muss daher in einzelne Räume untergliedert werden, die wiederum in ihre einzelnen Bestandteile aufgeteilt werden. Für diesen Vorgang wird die Software Desite MD Pro verwendet. Der beispielhafte Raum wird im Folgenden als Referenzszene bezeichnet. Nach der Identifikation der in der Referenzszene vorhandenen relevanten Bauteile werden diese einzeln durch die integrierte Exportfunktion von Desite MD abgespeichert. Die softwareseitig unterstützten und für den datengetriebenen Referenzprozess maßgebenden Dateiformate sind diesbezüglich das IFC- sowie das OBJ-Format. Das Ergebnis dieses Vorgangs sind getrennte \*.obj-Dateien für jedes Bauteil der Szene.

Sollten einzelne Bauteile nicht, wie beim vorangegangenen Vorgang beschrieben, als \*.obj-Datei vorliegen, wird die Extraktion der visuellen Information aus der \*.ifc-Datei des relevanten Bauteils entsprechend des Klassifizierungsprozesses mit IFCConvert durchgeführt.

Das Mesh Sampling wird analog zur Klassifizierungsaufgabe getrennt für jedes Bauteil durchgeführt, allerdings ohne die Vorgabe hinsichtlich der Mächtigkeit der Punktwolke. Dies resultiert aus der Variabilität der Punkte bei der Aufnahme von Punktwolken durch Scanner in der Realität. Daher entfällt die optionale Anwendung eines Farthest-Point-Algorithmus auf die Punktwolken der relevanten Bauteile. Das Ergebnis dieses Vorgangs sind einzelne \*.txt-Dateien, die jeweils die Punktwolke aller relevanten Bauteile der Szene enthalten.

Die Vereinigung der \*.txt-Dateien aller relevanten Bauteile zuzüglich der nicht zuordenbaren Punkte (Clutter) bildet die Punktwolke des gesamten Raumes. Die Reihenfolge, in der die Bauteile der Datei hinzugefügt werden, spielt aufgrund der Eigenschaft einer Punktwolke hinsichtlich der Sortierung keine Rolle. Dieser Vorgang muss prinzipiell für jede Szene wiederholt werden. Die Dateien werden dann dem bereits bestehenden Testset des S3DIS-Datensatzes (Armeni et al. 2017) hinzugefügt. Die Benennung der entstehenden Dateien sowie die Ordnerstruktur erfolgt hierbei gemäß den Vorgaben durch PointNet. Das Ergebnis dieses Vorgangs sind die \*.txt-Dateien aller relevanten Bauteile sowie der gesamten Szene.

Im nächsten Schritt wird eine Formatkonvertierung in HDF5 durch zwei implementierte Transformationsvorgänge durchgeführt. Dazu wird die strukturierte Szene dem ursprünglichen Testset hinzugefügt und die Transformation auf diesen manipulierten Datensatz angewendet. Bei der Erstellung bzw. der Erweiterung des Testsets ist es entscheidend, die Dateinamenskennung von PointNet beizubehalten. Die Dateinamen werden im

Rahmen der HDF5-Konvertierung für das Labeln eines jeden Punktes verwendet. Das händische Labeln der Punkte der gesamten Szene wäre mit einem unverhältnismäßigen Aufwand verbunden. Die Szene ist nun final für die semantische Segmentierung durch das Netzwerk vorbereitet. Diese erfolgt dann auf Basis der Daten, die das Ergebnis des datengetriebenen Referenzprozesses sind.

## ERGEBNISSE

### Klassifizierung

Die beispielhafte Objektkategorie muss bereits im ModelNet40-Datensatz enthalten sein und in den Gesamtkontext des Gebäudes eingebettet werden können, da dieser Datensatz zum Training des Netzwerkes verwendet wurde. Das Referenzobjekt für die Klassifizierung ist vorliegend ein Stuhl.

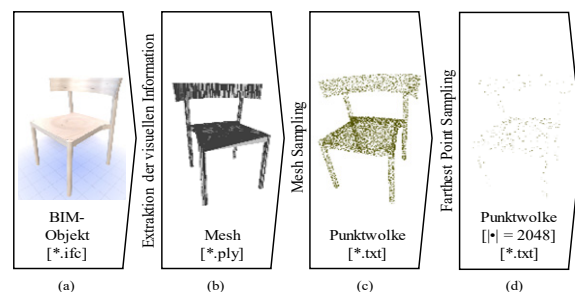


Abbildung 5: Ergebnisse des Referenzprozesses der Klassifizierung

Abbildung 5 (a) zeigt das Referenzobjekt in der Downloadmaske der ausgewählten BIM-Objektdatenbank. Gespeichert wurden die Objekte als Industry Foundation Classes. Die \*.ifc-Dateien selbst sind in ihrer reinen Textform nicht anschaulich, weshalb an dieser Stelle auf eine Visualisierung verzichtet wird. Mit Hilfe des Tools IFCConvert der Open Source Softwarebibliothek IFCOpenShell wurden dann die visuellen Informationen extrahiert und in einer \*.ply-Datei abgespeichert. Das als Mesh visualisierte Ergebnis ist Abbildung 5 (b) zu entnehmen. Der implementierte Mesh-Sampling-Algorithmus, der auf der Bibliothek PyntCloud basiert, transformiert die gegebenen Meshes in Punktwolken. Durch Parameter kann gesteuert werden, wie viele Punkte erzeugt werden sollen. Das Ergebnis ist Abbildung 5 (c) zu entnehmen. Liegen die Punktwolken nicht mit der benötigten Mächtigkeit vor, so ist der Farthest-Point-Sampling-Algorithmus zu verwenden. Das Ergebnis ist in Abbildung 5 (d) dargestellt. Das Netzwerk klassifiziert das Referenzobjekte mit einer dem Netzwerk angemessenen Genauigkeit. Die korrekt vorausgesagte sowie tatsächliche Objektkategorie werden während der Evaluierung in einer \*.txt-Datei gespeichert.

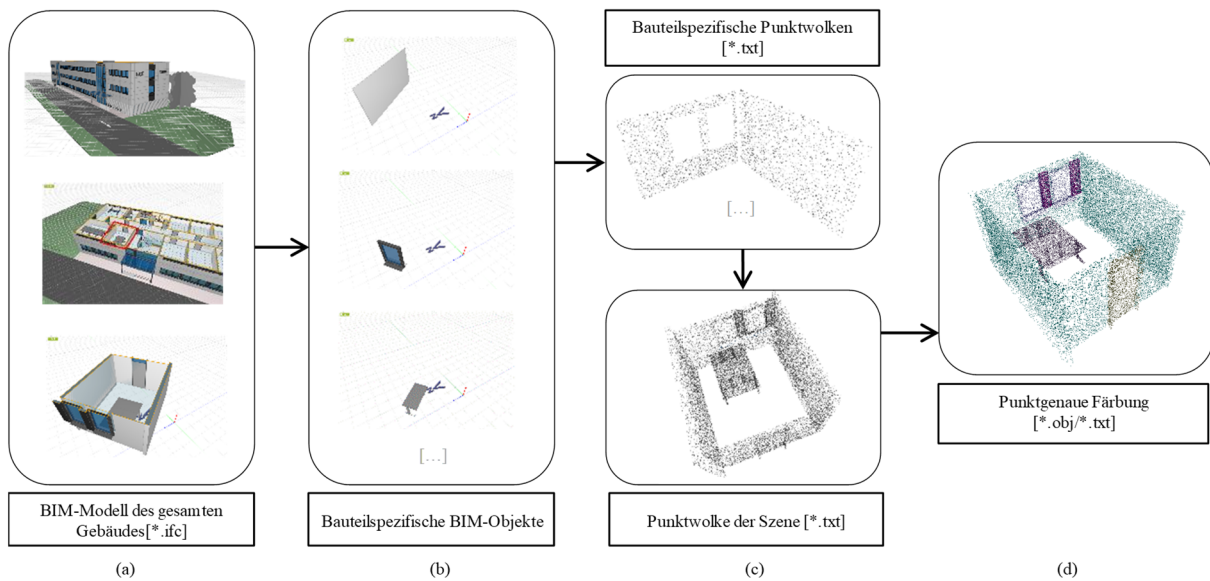


Abbildung 6: Ergebnisse des Referenzprozesses der semantischen Segmentierung

### Semantische Segmentierung

Abbildung 6 (a) zeigt die Visualisierung des gesamten BIM-Modells von außen mittels der Software Desite MD Pro auf Grundlage der \*.ifc-Datei, den Raum gemäß der Markierung im Kontext des Gebäudes sowie den aus dem Gesamtmodell extrahierten Raum, der die Grundlage für den weiteren Prozess darstellt. Dieser Raum wurde dann entsprechend in die einzelnen Bauteile aufgeteilt. Die Darstellung ausgewählter Bauteile in Desite MD Pro ist Abbildung 6 (b) zu entnehmen. Die einzelnen Bauteile wurden dann durch den bereits implementierten Abtast-Algorithmus in Punktwolken transformiert. Die Punktwolken der einzelnen Bauteile der Referenzszene wurden in beliebiger Reihenfolge in einer \*.txt-Datei zusammengefasst. Das visualisierte Ergebnis dieser Zusammenfassung ist Abbildung 6 (c) zu entnehmen. Das Netzwerk segmentiert die Referenzszene mit einer dem Netzwerk angemessenen Genauigkeit. Die semantische Segmentierung jeder Szene wird durch die Färbung der einzelnen Punkte visualisiert und gespeichert. Abbildung 6 (d) zeigt die semantische Segmentierung der Referenzszene.

### LIMITATIONEN

Der datengetriebene Referenzprozess ist durch die Qualität des Objekterkennungssystems limitiert. Obwohl PointNet zu den akkuratesten Systemen bei der Objekterkennung auf Punktwolken-Ebene gehört, kann noch immer keine vollkommene Genauigkeit erzielt werden. Das bedeutet, dass nicht jedes Objekt fehlerfrei klassifiziert und nicht jede Szene fehlerfrei semantisch segmentiert werden kann. Eine händische Nachbearbeitung der identifizierten Falschklassifizierungen kann zum aktuellen Zeitpunkt nicht verhindert werden und kann mit einem Mehraufwand verbunden sein. Bei der Klassifizierung

betrifft dies rund 10 % der Objekte und bei der semantischen Segmentierung rund 20 % aller Punkte. Die Möglichkeit einer Vernachlässigung der Falschklassifizierungen kann immer im jeweiligen Anwendungsfall hinsichtlich der angemessenen Toleranzen geprüft werden.

3D-Repräsentationen eines Gebäudes, die durch das vorliegende Modell verarbeitet werden, sind Gebäudedatenmodelle und Punktwolken. Diese beiden Repräsentationen sind in ihrer Struktur und hinsichtlich der notwendigen Verarbeitung vollkommen konträr. Das Objekterkennungssystem ist zum aktuellen Zeitpunkt durch diese Kontrarietät limitiert. Konkretisiert sind Gebäudedatenmodelle im Rahmen der Planung eine synthetische Abstraktion des realen Bauwerks aus dem Gedankengut der Beteiligten, wohingegen Punktwolken direkt aus dem existierenden Bauwerk generiert werden. Sowohl ein Gebäudedatenmodell als auch eine Punktwolke enthalten immer Ungenauigkeiten. Im vorliegenden Fall sind im BIM-Modell, aus dem die Referenzszene stammt, beispielsweise nur synthetische Farbinformationen gespeichert. Anders verhält es sich bei den für das Training verwendeten Punktwolken, die eine realitätsnahe Farbinformation unter Berücksichtigung der Lichtverhältnisse für jeden einzelnen Punkt enthalten. Außerdem konnten in der Referenzszene weder eine Decke noch ein Boden als Bauteil identifiziert werden. Der Grund dafür ist die etagenübergreifende Modellierung des horizontalen Raumabschlusses im Rahmen von BIM. Dieser liegt somit nicht raumspezifisch vor. Bei der Aufnahme der Punktwolke im realen Gebäude wird der horizontale Raumabschluss augenscheinlich durch den vertikalen Raumabschluss, also die Wände, begrenzt.

## ZUSAMMENFASSUNG UND AUSBLICK

Das vorliegende Paper zeigt, dass es möglich ist, die zentralen dreidimensionalen Repräsentationen eines Gebäudes innerhalb eines KI-basierten Objekterkennungssystem zu verarbeiten. Zu den zentralen Repräsentationen von Gebäuden gehören dreidimensionalen Gebäudedatenmodellen sowie in der Realität aufgenommenen Punktwolken. Das Modell ermöglicht somit die Verknüpfung von visuellen Daten des Planungsprozesses mit denen der As-Built-Dokumentation von Gebäuden. Dazu werden die BIM-Daten derart transformiert, dass sie in ein für Punktwolken konzeptioniertes Objekterkennungssystem eingespeist und korrekt verarbeitet werden können.

Dies ermöglicht beispielsweise die Erweiterung von Trainingsdaten für ein neuronales Netzwerk. Die Objekterkennung im Bestand von Gebäuden kann auf alle Gewerke ausgeweitet werden, für die ausreichend BIM-Objekte vorhanden sind. Dies betrifft insbesondere das Gewerk der technischen Gebäudeausrüstung, für das zum aktuellen Zeitpunkt keine umfassenden Datensätze vorhanden sind. Ein derartiges Objekterkennungssystem bietet Potenziale für die generelle Weiterentwicklung des Indoor Scene Understanding, das heißt einem computergestützten Verständnis des Interieurs von Gebäuden und Räumen. Außerdem kann die Objekterkennung hinsichtlich der Problematik von Intra-Klassen-Variationen sowie der Größe des benötigten Datensatzes durch ein möglichst heterogenes Trainingsset verbessert werden. Vorausgesetzt, das neuronale Netzwerk wird mit den BIM-Daten eines Gebäudes trainiert und eine aufgenommene Punktwolke des gleichen Gebäudes soll im Anschluss semantisch segmentiert bzw. die enthaltenen Objekte klassifiziert werden. Dann resultiert aus dem datengetriebenen Referenzprozess die Hypothese, dass die Ergebnisse der Objekterkennung hinsichtlich der Genauigkeit besser werden, da das neuronale Netz auf ähnlichen Daten trainiert wurde.

Eine weitere Herausforderung der allgemeinen Objekterkennung ist die zumindest teilweise Verdeckung der Objekte durch nahe gelegene bzw. perspektivenabhängige andere Objekte. Durch das BIM-Modell eines Gebäudes ist allerdings der gesamte Kontext bekannt. Für Unternehmen mit einem entsprechenden Bestand an Immobilien kann ferner das Datenmanagement vereinfacht und somit die Produktivität gesteigert werden. Der KI-basierte Ansatz ermöglicht außerdem eine neue Herangehensweise an das Forschungsgebiet der BIM As-Built-Dokumentation.

Zukünftige Forschungsarbeiten müssen beispielsweise die datengetriebene Referenzprozesse automatisieren, um so eine praktische Anwendungsfähigkeit zu erreichen. Ferner sollte evaluiert werden, ob die Qualität der Objekterkennung steigt, wenn das Netzwerk zuvor mit den BIM-Daten des gleichen Gebäudes trainiert wurde. Dann könnten

aufgrund der Reduzierung der Intra-Klassen-Variation weniger Daten für das Training des Netzwerkes notwendig sein.

## LITERATUR

- Accenture Strategy 2016. Digitalisierung entzaubern – wie die deutschen Top500 digitale Blockaden lösen. [https://www.accenture.com/\\_acnmedia/PDF-5/Accenture-Top500-DE-Executive-Summary-Final-Web.pdf](https://www.accenture.com/_acnmedia/PDF-5/Accenture-Top500-DE-Executive-Summary-Final-Web.pdf). Accessed: 2020-04-04.
- Achlioptas, P., Diamanti, O., Mitliagkas, I., and Guibas, L. 2017. Learning representations and generative models for 3d point clouds.
- Angelina Uy, M. and Hee Lee, G. 2018. Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4470–4479.
- Armeni, I., Sax, A., Zamir, A. R., and Savarese, S. 2017. Joint 2d-3d-semantic data for indoor scene understanding.
- BMWi 2018. Den digitalen Wandel gestalten. <https://www.bmwi.de/Redaktion/DE/Dossier/digitalisierung.html>. Bundesministerium für Wirtschaft und Energie, Accessed: 2020-04-04.
- Borrmann, A., König, M., Koch, C., and Beetz, J. 2015. Building Information Modeling: Technologische Grundlagen und industrielle Praxis. Springer-Verlag.
- Boscaini, D., Masci, J., Melzi, S., Bronstein, M. M., Castellani, U., and Vandergheynst, P. 2015. Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. In Computer Graphics Forum, volume 34-5, pages 13–23. Wiley Online Library.
- Brock, A., Lim, T., Ritchie, J. M., and Weston, N. 2016. Generative and discriminative voxel modeling with convolutional neural networks.
- buildingSMART. 2020. IFC 4.1 Specification. [https://standards.buildingsmart.org/IFC/RELEASE/IFC4\\_1/FINAL/HTML/](https://standards.buildingsmart.org/IFC/RELEASE/IFC4_1/FINAL/HTML/).
- Cheraghian, A. and Petersson, L. 2019. 3dcapsule: Extending the capsule architecture to classify 3d point clouds. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1194–1202.
- Dominguez, M., Dhamdhere, R., Petkar, A., Jain, S., Sah, S., and Ptucha, R. 2018. General-purpose deep point cloud feature extractor. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1972–1981.
- Goodfellow, I., Bengio, Y., and Courville, A. 2016. Deep Learning. MIT Press.

- Han, Z., Liu, Z., Vong, C.-M., Liu, Y.-S., Bu, S., Han, J., and Chen, C. P. 2018. Deep spatiality: Unsupervised learning of spatially-enhanced global and local 3d features by deep neural network with coupled softmax. *IEEE Transactions on Image Processing*, 27(6):3049–3063.
- Han, Z., Lu, H., Liu, Z., Vong, C.-M., Liua, Y.-S., Zwicker, M., Han, J., and Chen, C. P. 2019. 3d2seqviews: Aggregating sequential views for 3d global feature learning by cnn with hierarchical attention aggregation. *IEEE Transactions on Image Processing*.
- Hausknecht, K. and Liebich, T. 2016. BIMKompodium. Building Information Modeling als neue Planungsmethode. Stuttgart.
- Hichri, N., Stefani, C., De Luca, L., and Veron, P. 2013. Review of the “as-built bim” approaches. *Proceedings of the 3D-ARCH International Conference*.
- IDC 2018. The digitization of the world from edge to core. <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>. International Data Corporation sponsored by Seagate, Accessed: 2020-04-04.
- ISO 16379. 2013. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries. International Organisation for Standardisation.
- Krijnen, T. 2012. An open source IFC-geometry engine.
- Li, J., Chen, B. M., and Hee Lee, G. 2018. So-net: Self-organizing network for point cloud analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9397–9406.
- Liu, X., Han, Z., Liu, Y.-S., and Zwicker, M. 2019. Point2sequence: Learning the shape representation of 3d point clouds with an attention-based sequence to sequence network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8778–8785.
- Maturana, D. and Scherer, S. 2015. Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928.
- Qi, C. R., Su, H., Mo, K., and Guibas, L. J. 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660.
- Qi, C. R., Yi, L., Su, H., and Guibas, L. J. 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108.
- Ravanbakhsh, S., Schneider, J., and Poczos, B. 2016. Deep learning with sets and point clouds.
- Sedaghat, N., Zolfaghari, M., Amiri, E., and Brox, T. 2016. Orientation-boosted voxel nets for 3d object recognition.
- Su, H., Maji, S., Kalogerakis, E., and Learned-Miller, E. 2015. Multiview convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953.
- Süße, H. and Rodner, E. 2014. Bildverarbeitung und Objekterkennung.
- van Treeck, C., Elixmann, R., Rudat, K., Hiller, S., Herkel, S., and Berger, M. 2016. Gebäude. Technik. Digital.: Building Information Modeling.
- Wang, C., Samari B., and Siddiqi, K. 2018. Local spectral graph convolution for point set feature learning. In *Proceedings of the European conference on Computer Vision (ECCV)*, pages 52–66.
- Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M. and Solomon, J.M. 2019. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 38(5), 1-12.
- Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. 2015. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920.
- Xu, Y., Fan, T., Xu, M., Zeng, L., and Qiao, Y. 2018. Spidercnn: Deep learning on point sets with parameterized convolutional filters. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 87–102