

2. Platz Bauingenieurwesen

Optimierung der semantischen Segmentierung von 3D-Punktwolken durch multimodale Fusion mit 2D-Bildern/ Enhancing 3D Point Cloud Semantic Segmentation Using Multi-Modal Fusion With 2D Images

Changyu Du

Technische Universität München

Die Arbeit beschäftigt sich mit der automatisierten Generierung von BIM-Modellen der gebauten Umwelt mittels Methoden der Künstlichen Intelligenz. Es wird ein neuartiger Ansatz für die Segmentierung vorgestellt, der unimodale Deep-Learning-Algorithmen zur Verarbeitung von 3D-Punktwolken mit multimodalen Fusionsalgorithmen, die 2D-Bilder als Zusatzinformationen integrieren, kombiniert.

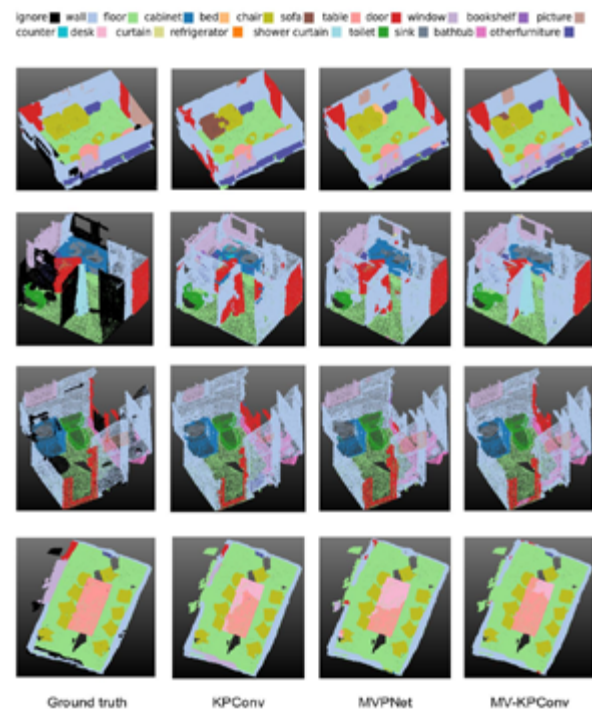
Die Lösung ermöglicht somit die automatisierte Erstellung digitaler Zwillinge der gebauten Umwelt und den automatisierten Abgleich zwischen gebauter Realität und digitalem Zwilling.

Die Scan-to-BIM-Methode zielt darauf ab, BIM-Modelle von Gebäuden durch Scannen von Punktwolken zu erstellen. Da die ursprünglichen 3D-Punktwolken keine semantischen Informationen enthalten, sind Deep-Learning-Algorithmen erforderlich, um die Punktwolken des Gebäudes zu verarbeiten und daraus gültige semantische Daten zu identifizieren. Im Vergleich zu traditionellen unimodalen Deep-Learning-Algorithmen, die direkt 3D-Punktwolken verarbeiten, haben multimodale Fusionsalgorithmen, die 2D-Bilder als ergänzende Informationen für 3D-Szenen verwenden, einen größeren Leistungsvorteil.

Diese Studie optimiert die Performance eines multimodalen Open-Source-Algorithmus, Multi-View Point Regression Networks (MVPNet), bei 3D-Semantik-Segmentierungsaufgaben, indem sie das leistungsfähigere und robustere Kernel Point Convolution (KPCConv) als neues 3D-Backbone-Netzwerk verwendet. Die in dieser Studie vorgestellte MV-KPCConv Methode kombiniert die verschiedenen Module von MVPNet

und KPCConv auf sinnvolle Weise: Die 2D-3D-Lifting-Methode von MVPNet wird verwendet, um Bilder auszuwählen und zu verarbeiten und die 2D-Bild-Features im 3D-Raum auszudrücken. Anschließend wird KPCConv verwendet, um diese Features mit geometrischen Informationen aus Punktwolken zu fusionieren, um 3D-Semantik-Labels vorherzusagen.

Auf einem benutzerdefinierten ScanNet Dataset erreicht das vorgestellte Netzwerk eine Score von 74,40 mIoU bei der 3D-Semantik-Segmentierungsaufgabe und übertrifft damit das ursprüngliche MVPNet. Darüber hinaus werden umfangreiche Ablationsstudien durchgeführt, um die geeignete Fusionsstruktur, das Timing und den Effekt der 3D-Farbe et cetera zu untersuchen.



Visualisierungsergebnisse der Basismodelle und von MV KPCConv. Bildquelle: Du, Changyu; Enhancing 3D Point Cloud Semantic Segmentation Using Multi-Modal Fusion With 2D Images, Technische Universität München, 2021.