

Institut für Photogrammetrie
und GeoInformation



Leibniz
Universität
Hannover

Domain Randomization anhand synthetischer Trainingsdaten für eine semantische Instanz-Segmentierung artikelreiner Behälter im Bereich Logistik

Masterarbeit

Patrick Pirner, B.Eng.

10022725

Leibniz Universität Hannover
Institut für Photogrammetrie und GeoInformation

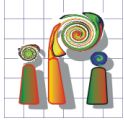
3. Mai 2021

Erstprüfer: Prof. Dr.-Ing. habil. Christian Heipke

Zweitprüfer: M.Sc. Dennis Wittich

Externer Betreuer: M.Sc. Markus Weiß (IGZ)

Die Arbeit wurde angefertigt in Kooperation mit:
IGZ Ingenieurgesellschaft für logistische Informationssysteme mbH



Institut für Photogrammetrie und GeoInformation,
Nienburger Straße 1, 30167 Hannover

Nienburger Straße 1, 30167 Hannover
Fakultät für Bauingenieurwesen und
Geodäsie

Institut für Photogrammetrie
und GeoInformation

Prof. Dr.-habil. Christian Heipke

Thema der Masterarbeit von Herr Patrick Pirner B.Eng.

Domain Randomization anhand synthetischer Trainingsdaten für eine semantische Instanz-Segmentierung artikelreiner Behälter im Bereich Logistik

Tel.+49 511 762-2482
Fax +49 511 762-2483
E-Mail: heipke@ipi.uni-hannover.de

(in Zusammenarbeit der IGZ Ingenieurgesellschaft für logistische Informationssysteme mbH)

25. November 2020

Für eine vollautomatische Kommissionierung müssen Artikel für den Pick-Vorgang innerhalb einer Lagerbox automatisch detektiert werden. In dem hier adressierten Szenario befinden sich in einer Box jeweils nur Artikel desselben Typs, man spricht von artikelreinen Behältern. Ein Ansatz für die automatische Detektion ist die Verwendung eines neuronalen Netzes für bildbasierte Instanzen-Segmentierung, dessen Training üblicherweise eine große Anzahl an gelabelten Trainingsdaten erfordert. Da das manuelle Generieren der Trainingsdaten mit einem immensen Kosten- und Zeitaufwand verbunden ist, scheint die Verwendung von synthetischen Trainingsdaten naheliegend. Publikationen zu verwandten Forschungsbereichen zeigen jedoch, dass Modelle die auf synthetischen Daten trainiert wurden oftmals nur unter Qualitätseinbußen auf reale Daten übertragen werden können. Neben dem Transfer von synthetischen- auf reale Daten ist ein weiteres potentielles Problem, dass nicht alle Lagerartikel synthetisiert werden können. Eine mögliche Lösung hierfür liegt in der sog. „Domain Randomization“. Hierbei werden Trainingsbilder mit vielen unterschiedlichen Parametern generiert, um u.a. die Anzahl und Anordnung von Artikeln, deren Oberflächenbeschaffenheit und Textur, und die Belichtung variieren.

Herr Pirner erhält folglich die Aufgabe zu prüfen, inwiefern synthetische Bilddaten geeignet sind, um ein neuronales Netz zu für die Instanzen-Segmentierung im Rahmen des Pick-Vorgangs zu trainieren und welche Qualität diese Modelle auf realen Szenarien erreichen. Zudem soll untersucht werden, inwiefern Domain Randomization die Erkennung eines breiten Artikelspektrums begünstigt. Für die automatische Generierung synthetischer Trainingsdaten soll Herr Pirner CAD-Modelle von diversen Lagerartikeln aus Open-Source-Datenbanken verwenden bzw. eigene Modelle generieren. Mit Hilfe einer 3D-Engine sollen die Objekte dann in einer virtuellen Lagerbox platziert, die entsprechende Ansichten gerendert und die zugehörigen Referenzen berechnet werden. Für die Evaluation müssen zusätzlich noch einige reale Bilder manuell gelabelt werden. Die entsprechenden Bilder werden von der Firma IGZ bereitgestellt. Im Anschluss soll Herr Pirner eine geeignete Netzarchitektur wählen wobei auf existierende Architekturen zurückgegriffen werden kann. Die letzte Teilaufgabe bezieht sich auf das eigentliche Training der Modelle. Da es in der adressierten Applikation besonders wichtig ist, dass die oberliegenden Objekte korrekt detektiert werden, soll Herr Pirner dies in der Formulierung der Verlustfunktion berücksichtigen, beispielsweise durch eine entsprechende Gewichtung. Für die Validierung der Methode soll Herr Pirner geeignete Kombinationen aus realen und synthetischen Datensätzen für das Training bzw. die Evaluation der Modelle verwenden.

Prof. Dr.-habil. Christian Heipke / M.Sc. Dennis Wittich

Besucheradresse:
Nienburger Straße 1
30167 Hannover
www.ipi.uni-hannover.de

Erklärung

Hiermit versichere ich, Patrick Pirner, dass die Arbeit selbstständig verfasst wurde und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden, dass alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht sind und dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt wurde.

Hannover, den 03. Mai 2021

(Ort, Datum)

(Unterschrift, Stempel)

Kurzfassung

In dieser Masterarbeit wird ein Verfahren zur Generierung synthetischer Trainingsbilder unter Anwendung von Domain Randomization für eine Objektdetektion vorgestellt. Für den Anwendungsfall Bin Picking mit artikelreinen Behältern, bei dem ein Roboter ein beliebiges Objekt aus einer Box greifen soll, können Fehler in der Objektdetektion nach wie vor als eine Hauptfehlerquelle ausgemacht werden. Des Weiteren ist die Beschaffung von realen Trainingsbildern bei wechselnden Lagerartikeln und einer breiten Artikelvielfalt kosten- und zeitintensiv. Durch den Einsatz synthetischer Daten kann die Beschaffung der Trainingsbilder zu einem großen Teil automatisiert werden. Es kann gezeigt werden, dass trotz des Transfers der Trainingsdomäne mit synthetischen Bildern auf die Testdomäne mit realen Kamerabildern viele Instanzen der Objekte erkannt werden. Durch die Verwendung von Mask R-CNN für eine Instanz-Segmentierung werden detektierte Objekte pixelweise von anderen Objekten oder dem Hintergrund abgegrenzt. Diese zusätzliche Information kann die Greifplanung des Roboters gegenüber einer Prädiktion von Bounding Boxen verbessern. Es wird untersucht, welche Auswirkungen auf die Ergebnisse der Instanz-Segmentierung durch verschiedene Randomisierungsparameter bei der Generierung der Trainingsbilder bewirkt werden. In einem weiteren Experiment mit unterschiedlichen Trainingsobjekten werden Ergebnisse für Artikel betrachtet, welche selbst nicht als synthetisches Modell im Trainingsdatensatz vorliegen. Eine erfolgreiche Detektion solcher Artikel birgt den Vorteil, dass nicht alle zu erkennenden Artikel synthetisiert werden müssen, wenn geeignete Trainingsmodelle und eine ausreichende Randomisierung mittels Domain Randomization stattfindet. Obwohl ein Roboter bei artikelreinen Behältern ein beliebiges Objekt greifen kann, wird der Greiferfolg durch das Picken eines oberliegenden und damit nicht verdeckten Objekts erhöht. Dies schließt zudem ein versehentliches Anheben weiterer Objekte aus. Dazu wurde Mask R-CNN dementsprechend modifiziert, dass die Sichtbarkeit eines Objekts als zusätzliche Ausgabegröße prädiziert wird. Die Sichtbarkeit der synthetischen Objekte kann während der Generierung mitberechnet werden und liegt so als zusätzliche Trainingsgröße vor. Durch die Prädiktion der Sichtbarkeit können in synthetischen sowie realen Testbildern frei liegende Artikel von verdeckten unterschieden werden. In Experimenten kann eine Randomisierung der Skalierung und Textur der Objekte bei der Generierung synthetischer Trainingsbilder als wichtiges Kriterium für den Erfolg einer Instanz-Segmentierung bestimmt werden. Weiterhin kann gezeigt werden, dass Artikel in realen Kamerabildern durch Verwendung synthetischer Trainingsbilder und Domain Randomization detektiert werden können, selbst wenn von diesen Artikeln kein synthetisches Modell für das Training vorhanden ist. Es konnte festgestellt werden, dass synthetische Trainingsartikel mit einer ähnlichen Form gegenüber den realen Testartikeln zu einer besseren Detektion führen, als die Anzahl der Modelle im Training mit beliebigen und teils komplexen 3D-Modellen zu erweitern. Durch die Prädiktion der Sichtbarkeit eines Objekts können oberliegende Objekte identifiziert und von verdeckten Objekten abgegrenzt werden, selbst wenn diese nicht als synthetisches Modell im Training vorliegen.

Abstract

In this master thesis, a methodology is presented to generate synthetic images with the concept of domain randomization to train an object detector. A major flaw of bin picking with single product containers, where a robot can pick an arbitrary object, are errors related to object detection. Moreover, the gathering of large training data with real camera images is very time and cost consuming. Synthetic images are a possible solution to solve these issues. The generation of those images can be automated to a large extent. It is shown that a large number of instances can be detected in real images despite the domain gap. The term domain gap refers to the problem that occurs when a different training domain (i.e. synthetic images) and test domain (i.e. real images) is used. In this thesis, Mask R-CNN is used for an instance-aware semantic segmentation to enable a distinction between other objects and the background on a pixel level. Compared to a prediction of bounding boxes for object detection, this results in more useful information for a grasping algorithm. The effects on the result of object detection will be analyzed for different parameters of randomization. In further experiments with various synthetic objects in the training dataset, it is shown that objects can be detected, which do not appear within the training data themselves. A huge benefit of this result is that not all articles which have to be perceived have to be part of the training set if suitable objects and sufficient randomization are used in training. Another point of single product containers is, that a robot can grasp an arbitrary object. Nevertheless, it is desirable to grasp an object on top which is not occluded by other objects. In this case, other objects should not be grasped unintentionally. For this purpose, Mask R-CNN is modified to predict the visibility of an object as a further output. The visibility of a synthetic object can be computed during generation and be used as ground truth for training. With a prediction of the visibility of objects, it is possible to distinguish between occluded and non-occluded objects in synthetic and real images. In experiments, it appears that randomization of an object's scale and texture is crucial for good results of an instance-aware segmentation. Furthermore, it is shown that articles in camera images can be detected while using synthetic data for training and domain randomization, even if no synthetic model of an article is present in training data. A benefit of similar-looking articles in the training set compared to the real articles in the test set can be demonstrated instead of increasing the number of articles used in training with arbitrary or more complex models. Non-occluded objects can be distinguished from occluded objects by predicting the visibility of an object, even if those objects and corresponding ground truth for visibility are not present in the training data.