

Echtzeit-3D-Objekterkennung auf mobilen Geräten mit MediaPipe

Kategorien : [MediaPipe](#), [Open Source](#), [Software](#)

Schlagwörter : [Deep Learning](#), [Künstliche Intelligenz](#), [Machine Learning](#)

Datum : 16. November 2020



Die Objekterkennung ist ein umfassend untersuchtes Problem der Computer-Sicht, aber der Großteil der Forschung hat sich auf 2D konzentriert. Durch die Erweiterung der Vorhersage auf 3D kann man die Größe, Position und Orientierung eines Objekts in der Welt erfassen, was zu einer Vielzahl von Anwendungen in der Robotik, bei selbstfahrenden Fahrzeugen, der Bildgewinnung und der erweiterten Realität führt.

Obwohl die 2D-Objekterkennung relativ ausgereift ist und in der Industrie weit verbreitet ist, stellt die 3D-Objekterkennung ein schwieriges Problem dar.

Google veröffentlichte im März **MediaPipe, eine mobilen Echtzeit-3D-Objekterkennungs-Pipeline für Alltagsgegenstände**. Diese Pipeline erkennt Objekte in 2D-Bildern und schätzt ihre Posen und Größen anhand eines ML-Modells (Machine Learning), das auf einem neu erstellten 3D-Datensatz trainiert wurde.

Implementiert in MediaPipe, einem plattformübergreifenden Open-Source-Framework für den Bau von Pipelines zur Verarbeitung von Wahrnehmungsdaten verschiedener Modalitäten, berechnet Objectron orientierte 3D-Bounding-Boxen von Objekten in Echtzeit auf mobilen Geräten.

-
-
-
-

AR Synthetische Datenerzeugung

Ein beliebter Ansatz ist die Ergänzung von Daten aus der realen Welt durch synthetische Daten, um die Genauigkeit der Vorhersage zu erhöhen. Versuche führen jedoch oft zu schlechten, unrealistischen Daten oder erfordern im Falle einer fotorealistischen Darstellung einen erheblichen Aufwand und Rechenaufwand.

Der neuartige Ansatz, der als **AR Synthetic Data Generation** bezeichnet wird, platziert virtuelle Objekte in Szenen mit AR-Sitzungsdaten, wodurch wir Kamerapositionen, erkannte ebene Oberflächen und geschätzte Beleuchtung nutzen können, um Platzierungen zu generieren, die physikalisch wahrscheinlich sind und mit einer Beleuchtung, die der Szene entspricht.

Dieser Ansatz führt zu hochwertigen synthetischen Daten mit gerenderten Objekten, die die Szenengeometrie respektieren und nahtlos in reale Hintergründe passen. Durch die Kombination von Daten aus der realen Welt und synthetischen AR-Daten können wir die Genauigkeit um etwa 10 % erhöhen.

Eine ML-Pipeline für die 3D-Objekterkennung

Google hat ein einstufiges Modell aufgebaut, um die Pose und die physikalische Größe eines Objekts aus einem einzigen RGB-Bild vorherzusagen. Das Modell-Backbone hat eine Encoder-Decoder-Architektur, die auf [MobileNetv2](#) aufbaut. Google verwendet einen Multi-Task-Learning-Ansatz, bei dem sie gemeinsam die Form eines Objekts mit Erkennung und Regression vorhersagen. Die Shape-Task prognostiziert die Formsignale des Objekts in Abhängigkeit davon, welche Ground-Truth-Annotation verfügbar ist, z.B. Segmentierung. Dies ist optional, wenn keine Shape-Annotation in den Trainingsdaten vorhanden ist.

Objectron-Datensatz

Nun stellt Google den Objectron-Datensatz anderen Forschern kostenlos zur Verfügung.

Der Objectron-Datensatz ist eine Sammlung kurzer, objektorientierter Videoclips, die von AR-Sitzungsmetadaten begleitet werden, zu denen Kamerapositionen, spärliche Punktwolken und die Charakterisierung der planaren Oberflächen in der Umgebung gehören.

In jedem Video bewegt sich die Kamera um das Objekt herum und fängt es aus verschiedenen Winkeln ein. Die Daten enthalten auch manuell kommentierte 3D-Bounding-Boxen für jedes

Objekt, die die Position, Orientierung und Abmessungen des Objekts beschreiben.

Der Datensatz besteht aus 15K annotierten Videoclips, ergänzt durch über 4M annotierte Bilder in den folgenden Kategorien: Fahrräder, Bücher, Flaschen, Kameras, Schachteln, Stühle, Tassen, Laptops und Schuhe.

Um die Geodiversität zu gewährleisten, wird der Datensatz darüber hinaus in 10 Ländern auf fünf Kontinenten gesammelt. Zusammen mit dem Datensatz steht auch eine 3D-Objekterkennungslösung für vier Objektkategorien - Schuhe, Stühle, Tassen und Kameras - zur Verfügung.

Diese Modelle werden anhand dieses Datensatzes trainiert und in MediaPipe veröffentlicht, Googles Open-Source-Framework für plattformübergreifende, anpassbare ML-Lösungen für Live- und Streaming-Medien.

Tutorials

- [Downloading the dataset](#)
- [Hello, World example: Loading examples in Tensorflow](#)
- [Loading data in PyTorch](#)
- [Parsing raw annotation files](#)
- [Parsing the AR metadata](#)
- [Evaluating the model performance with 3D IoU](#)
- [SequenceExample tutorial](#)

Links

<https://mediapipe.dev/>

<https://github.com/google/mediapipe>

Objectron Dataset

<https://github.com/google-research-datasets/Objectron/>

Wie können wir Ihnen helfen?

Sie wollen online richtig verkaufen? Wir unterstützen Sie bei der Umsetzung Ihres Augmented Reality (AR) oder Virtual Reality (VR) Projektes! Egal ob [Produktfotografie](#), [3D Scan Service](#), [3D-Visualisierung](#) oder fertige [3D Modelle für AR/VR](#) – wir beraten Sie persönlich und unverbindlich.

Wo kann ich Anregungen, Lob oder Kritik äußern?

Ihre Meinung ist uns wichtig! Schreiben Sie uns, was Ihnen in Bezug auf unser Angebot bewegt. info@r23.de

R23 — Ihre Digitalagentur für Virtual Reality und interaktive Markenerlebnisse

Wünschen Sie ein individuelles Angebot auf Basis Ihrer aktuellen Vorlagen, nutzen Sie einfach unser [Anfrageformular](#).

<https://blog.r23.de/echtzeit-3d-objekterkennung-auf-mobilen-geraeten-mit-mediapipe/>

Besuchen Sie uns auch auf [Facebook](#) und [Twitter](#).

r23

Thüringenstr. 20

58135 Hagen

Deutschland

Telefon: 02331 / 9 23 21 29

E-Mail: info@r23.de

Ust-IdNr.:DE250502477