

Generated Dataset for Drive Image Segmentation

Studiengang: BSc in Informatik | Vertiefung: Computer Perception and Virtual Reality
Betreuer: Prof. Marcus Hudritsch

Ist es möglich ein Neuronales Netzwerk durch synthetisch generierte Bilder zu verbessern?

Ausgangslage

Um ein visuelles neuronales Netzwerk zu trainieren ist eine enorm hohe Anzahl an von Bildern notwendig. Weiter müssen diese Trainingsbilder eine Ground Truth besitzen. Das bedeutet für jedes Bild müssen alle Pixel markiert werden damit ein neuronales Netzwerk in der Lage ist erfolgreich dessen Klasse wie zum Beispiel Autos, Gebäude oder eine Strasse zu identifizieren. Dieser Prozess wird Labeling genannt und ist mit einem enormen Arbeitsaufwand verbunden.

Ein solches gelabeltes Datenset ist das Cityscape Datenset. Fast 3000 Bilder von deutschen Städten (und Zürich), sind mit der Ground Truth für 32 Klassen bearbeitet. Doch das Problem ist, dass 3000 Bilder immer noch eine sehr begrenzte Anzahl ist, um ein neuronales Netzwerk zu trainieren.

Meine Arbeit stellt sich die Frage, ob es möglich ist, mit synthetisch generierten Bildern, bei denen es möglich ist, die Ground Truth gleichzeitig zu generieren, die Genauigkeit der Klassifizierung zu verbessern.

Vorgehen

Der Gesamtprozess besteht wesentlich aus zwei Teilen.

- **Bilder generieren:** Mit dem 3D-Grafikprogramm Blender ist es möglich eine 3D-Szene zu erstellen, im Fall dieser Arbeit handelt es sich um eine Sicht aus der Frontkamera auf den Verkehr auf der

Strasse. Dabei wird die Szene zufällig generiert. Im Anschluss wird ein Bild gerendert. Die Qualität dieses Bildes ist dabei stark abhängig von der Zeit, die man dafür aufwendet. Ein einzelnes Bild wird in ungefähr 30 Sekunden gerendert.

- **Training:** Mit den generierten Bildern kann nun zusammen mit den bereits bestehenden Cityscape Bildern ein neuronales Netz trainiert werden. Dabei werden die 3000 Trainingsbilder mit einer bestimmten Anzahl an generierten Bildern ergänzt. Das Validierungsset bleibt dabei das Cityscape Datenset, da Verbesserungen auf realen Bildern erzielt werden sollen.



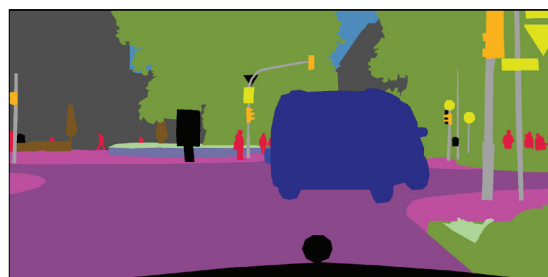
Ueli Bühler

Resultat und Fazit

Mithilfe von Blender lassen sich einwandfrei Bilder mit Ground Truth generieren und es eignet sich somit sehr gut für die synthetische Datengeneration. Dabei ist die Qualität der generierten Bilder stark abhängig von der Zeit, die man dafür aufwenden kann oder will. Das Erkennen von Autos und Motorrädern zu verbessern, stellte sich als sehr schwierig heraus. Die Qualität der verwendeten Assets (Autos, Motorräder) variierte stark und es müssten deutlich mehr Assets zur Verfügung stehen damit eine höhere Variabilität für die Trainingsbilder erzielt werden kann. Für das Verbessern von gewissen Spezialfällen, wie zum Beispiel schlechten Sichtverhältnissen sehe ich klares Potential.



Generiertes Bild von Zwei Autos im Nebel und Nacht



Cityscape Bild mit Ground Truth