

Semantic Segmentation für die Eisenbahnsimulation LOCSIM

Studiengang: BSc in Informatik | Vertiefung: Computer Perception and Virtual Reality
Betreuer: Prof. Marcus Hudritsch
Experte: Dr. Harald Studer
Industriepartner: LOCSIM, Schüpfen

Semantic Segmentation ist ein neues Konzept im Bereich des überwachten Lernens, das es ermöglicht jedes Pixel eines Bildes zu klassifizieren und damit Objekte in einem Bild zu lokalisieren. Angewendet auf die Eisenbahnsimulation der Firma LOCSIM soll Semantic Segmentation die Aufbereitung der Zugstrecken unterstützen, in dem vorgängig in Videobildern Zugsignale formgenau detektiert werden.

Ausgangslage

Die Firma LOCSIM entwickelt vollwertige, video-basierte Eisenbahnsimulatoren. Dazu nehmen sie mit einer Kamera Zugstrecken aus der Sicht eines Zugführerstandes auf. Anschliessend werden die Streckenvideos aufbereitet und für die Simulation der Fahrgeschwindigkeit entsprechend unterschiedlich schnell abgespielt. Die Fahrgeschwindigkeit wird aus vielen physikalischen Parametern berechnet, damit so ein vollwertiger Simulator entsteht.

Die Aufbereitung der Videos ist bis anhin sehr zeitintensiv, da über alle relevanten Signale manuell eine Bitmap gelegt werden muss, um sie unterschiedliche Informationen anzeigen zu lassen.

Im Zuge einer Neuentwicklung ihres Editors bietet sich nun die Möglichkeit, diesen Schritt mittels Semantic Segmentation zu automatisieren und die Signale automatisch zu detektieren.

U-Net

Das U-Net ist eine Convolutional Neuronale Netzwerk Architektur, die sich sehr bewährt hat. Anfänglich für den medizinischen Bereich gedacht, hat sich wegen der hohen Genauigkeit die Architektur schnell auch in anderen Bereichen etabliert.

Auf dem sogenannten Cityscape Datenset, das oftmals für Benchmarks verwendet wird, habe ich erste Erfahrungen gesammelt und mit Pytorch meine Trainingsumgebung entwickelt.

In einem zweiten Schritt habe ich mittels eines eigens entwickelten Converters alte, bereits aufbereitete Streckenvideos von LOCSIM in nutzbare, annotierte Trainingsdaten umgewandelt. Mit einem daraus resultierenden Datenset von ungefähr 24'000 annotierten Bildern habe ich das U-Net auf die Zugsignaldetektion trainiert und durch die Anpassungen der verschiedenen Parameter versucht, eine möglichst hohe Detektionsgenauigkeit zu erreichen.

Trainings Resultate

Die Detektionsgenauigkeit hängt stark von der Beschaffenheit des Bildes ab. Schlecht beleuchtete Bilder, wie zum Beispiel Tunnels, haben sehr schlechte Werte. Weiter müssen die Signale eine gewisse Grösse haben, da aus Performanzgründen, die Bildauflösung von Full HD auf ca 1/9 reduziert wird. Da mein Testdatenset eine sehr grosse Bandbreite an Situationen beinhaltet, fällt die durchschnittliche Genauigkeit nicht sehr hoch aus.

- **Pixel Accuracy: 39.26%**
- **Intersection over Union: 23.33%**

Einzelne Bilder haben aber hervorragende Resultate erzielt (vgl. Abbildung).

Die Integration des trainierten Modells

In der C++ Applikation von LOCSIM konnte das Model über das DNN-Modul von OpenCV einfach importiert werden. Die entwickelte Semantic Segmentation Erweiterung ist nun in der Lage, relevante Signale zu detektieren und zu klassifizieren. Als Input wird einzig das gewünschte Videoframe benötigt und nach der Detektion, ergänzt durch einige weitere Bildverarbeitungsalgorithmen, werden die erkannten Signale in dem von LOCSIM verwendeten Datenformat zurückgegeben und können somit weiterverwendet werden.



David Nathanael Märki



Formgenaue Detektion und Klassifikation von Zugsignalen mit Semantic Segmentation