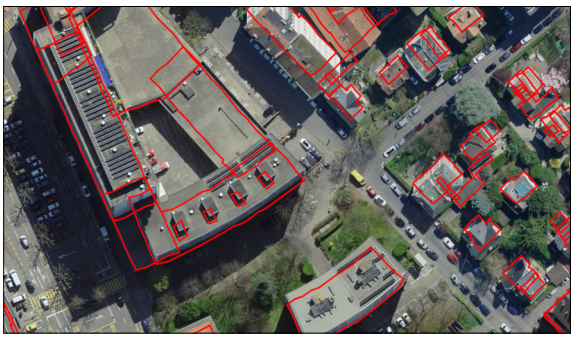

Diego Dolp


Noah Samuel Heuberger

Students	Diego Dolp, Noah Samuel Heuberger
Examiner	Hannes Badertscher
Subject Area	Artificial Intelligence
Project Partner	VRMotion AG, Dübendorf, ZH

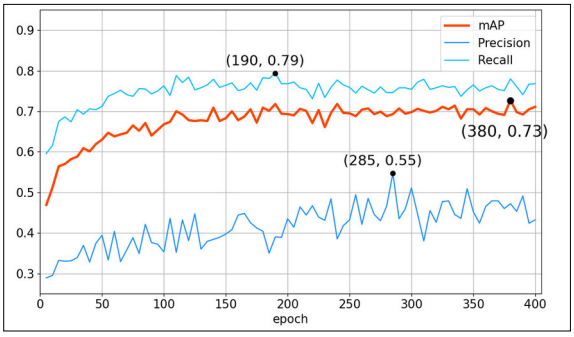
Building Detection on aerial and satellite imagery

A Deep Learning approach using the Mask R-CNN architecture



Aerial image overlaid with ground-truth GIS-shapes
Own presentment

Bounding boxes and prediction masks generated by our Neural Network
Own presentment



mAP, Precision and Recall plotted against the duration of training/number of epochs
Own presentment

Introduction: VRMotion AG is developing a VR-based simulator to train helicopter pilots. For accurate terrain representation, models of buildings need to be generated based on real data. In some countries, extensive Geodata of buildings and landscapes is available in the GIS-format. Where this isn't the case, aerial imagery has to be used to infer the position of buildings. The goal of our project is to automate this process with a Convolutional Neural Network (CNN) for building detection.

Approach: Our solution is based on Matterport's implementation of Mask R-CNN, an object detection architecture using Google's TensorFlow Deep Learning environment. For training- and validation-data, we rely on Switzerland's federal office of topography, using their database of aerial imagery and GIS-data of buildings, from which we generate pixel-by-pixel segmentation masks.

The Mask R-CNN implementation suggests a narrow range of hyperparameters which have proven successful in numerous applications. Following these suggestions after promising initial training runs, we focused on a three-pronged data-driven approach for model-tuning: (1) Merging of overlapping buildings (2) Regularizing our model via data augmentation (By artificially increasing the amount of training data through image processing methods, we effectively multiply the amount of available training data by a factor of 6.) (3) Increasing the amount of training data from initially 5GB to 385GB, containing roughly 125'000 buildings.

To handle this amount of data, training and testing had to be done on OST's deep learning server with multiple trainings running in parallel. The resulting prediction masks generated by the Neural Network were evaluated qualitatively - manually inspecting predictions rendered onto images - and quantitatively - calculating metrics over a test-batch of input pictures. Both inspections confirmed our model's strong performance.

Our Python API contains components for mask creation, data augmentation, training, prediction, statistical evaluation, and mask export. For easier server-handling, the code using Mask R-CNN can be configured via a single text-file.

Conclusion: When trained with augmented data on the full dataset, our Convolutional Neural Network performs with a mean Average Precision (mAP) of at least 0.73 after 380 epochs (IoU > 0.5). Its ability to detect and mask most buildings, in free-standing as well as urban scenarios, provides a solid baseline for the intended VR-application. The remaining False Positives issue from comparatively rare landscape features such as trains, construction sites and boats, suggesting that further improvement will be possible with more diverse training data. Due to its modular design, our model can easily be extended to recognize different types of buildings, as well as merge overlapping prediction masks into contiguous buildings, taking automated landscape generation one step further.