Title
Tree segmentation from AHN4, using a non-end-to-end neural network and random forest

Author
Blokker, Goof

Date
2022-08-24

Abstract
In this thesis effective ways to segment trees from AHN4 will be studied. With the modern day focus on biodiversity, maintaining it or increasing it, a good knowledge of vegetation is especially vital. AHN4 is the fourth edition of a national point cloud covering the entire Netherlands. It is distributed with basic classification, but trees are part of a larger class 'other' which also contains other non-terrain objects like street furniture and cars. The methods considered here are non-end-to-end which means that first features need to be generated which will be consecutively fed to a classification algorithm. Features are created by applying Principal Component Analysis (PCA) with a suitable point neighborhood. A neighbourhood of 20 nearest neighbours turned out to work most effectively. A Neural Network and Random Forest will be tested as classification al- gorithms. The training dataset consisted of a combination of a tree database from the municipality of Delft and AHN4. This training data is imperfect since it contained both false positives and false negatives due to how it was created. For both algorithms a training dataset of 400 000 points was used, to prevent overfitting, with a 50/50 split between tree and non-tree points. Both algorithms were validated against two datasets, Validation Dataset one is a small generic dataset and Validation Dataset two is a dataset consisting of problem scenarios such as cars and construction sites. The Neural Network scored an accuracy of 92.8% and the Random Forest 94.4% against Validation Dataset one. Against Validation Dataset two the Neural Network scored an accuracy of 66.9% and Random Forest 72.4%. Both algorithms had trouble with correctly classifying cars and construction sites as non trees. In the end, the Random Forest algorithm had a higher accuracy than the Neural Network, however this might be the result of the imperfect training data.