Classification of satellite images using Convolutional Neural Networks

Jakob Grönlund, Susanna Larsson, Anna Birgersson,
Linus Eriksson och Sanne Ingvarsson

LINKÖPINGS UNIVERSITET

Problem

Accurate geodata is required when developing tomorrows telecom network, where an increased demand are seen in several fields such as video, 5G and IoT solutions. By using geodata, it is possible to plan and design telecom networks with optimal performance. Telecom networks are often over traced in crowded areas, and by localizing especially crowded areas, it is possible to distribute resources more efficiently. The project aims to classify such areas in form of airports in satellite images, by using two different neural networks.

DeepLabV3+

DeepLabV3+ is based on the previous DeepLabv3 but with a decoder module to improve the segmentation along object boundaries [1]. Encoder-decoder networks typically contain one encoder module that reduces the feature map with convolution layers and pooling, and a decoder module that gradually recovers the spatial information to the encoded image [1]. The segmentation is performed using an adapted Xception model, with depthwise separable convolution in both the ASPP module and decoder module. DeepLabV3+ also uses pre-trained weights from the PASCAL VOC2012 dataset to boost the performance[1].

ResNet50

ResNet50 is a state-of-the art, award winning neural network architecture, and a development of the famous AlexNet and VGGNet. It builds upon the theory that the deeper the network is, the better is its performance. By utilizing residual skips, a concept were the output of two previous layers are fed into the next one, where one layer is skipped, its accuracy is increased. ResNet50 applies Global Average pooling to reduce the number of parameters and letting it train faster [2].

Method

The satellite images were first preprocessed which resulted in numpy-arrays with 5 channels: R, G, B, NVDI and DSM (height map). The preprocessed data was used as training and validation data. The ResNet50 and DeepLabV3+ models were found as open source and tweaked to fit the set requirements. The training was performed using a Google cloud compute engine with GPU support and the models were trained with different learning rates. The trained model was used to to perform classifications, predictions on new images that had not been used in the training. The predicted images were then evaluated using F1-score, which takes both precision and recall into account.

References

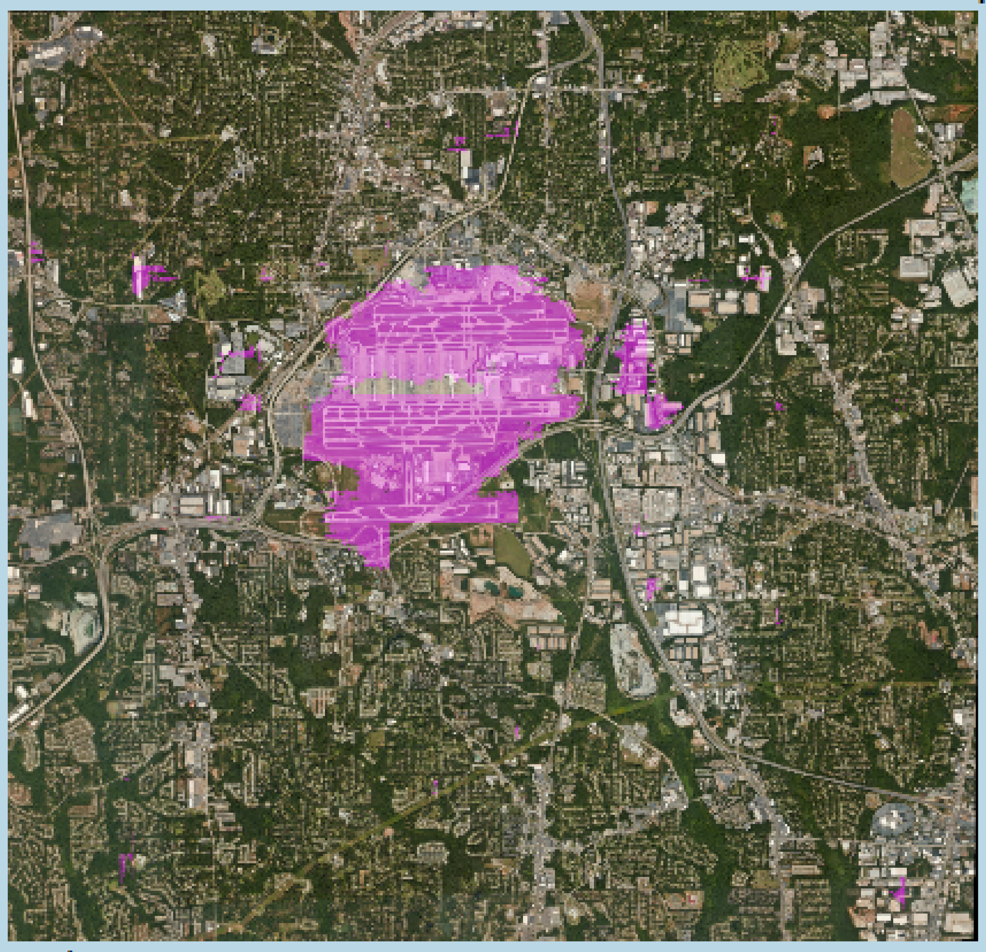
[1] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff and Hartwig Adam: *Encoder-decoder withatrous separable convolution for semantic image segmentation*, arXiv preprint arXiv:1802.02611, 2018.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: *Deep Residual Learning for Image Recognition*, arXiv preprint arXiv:1512.03385, 2015.

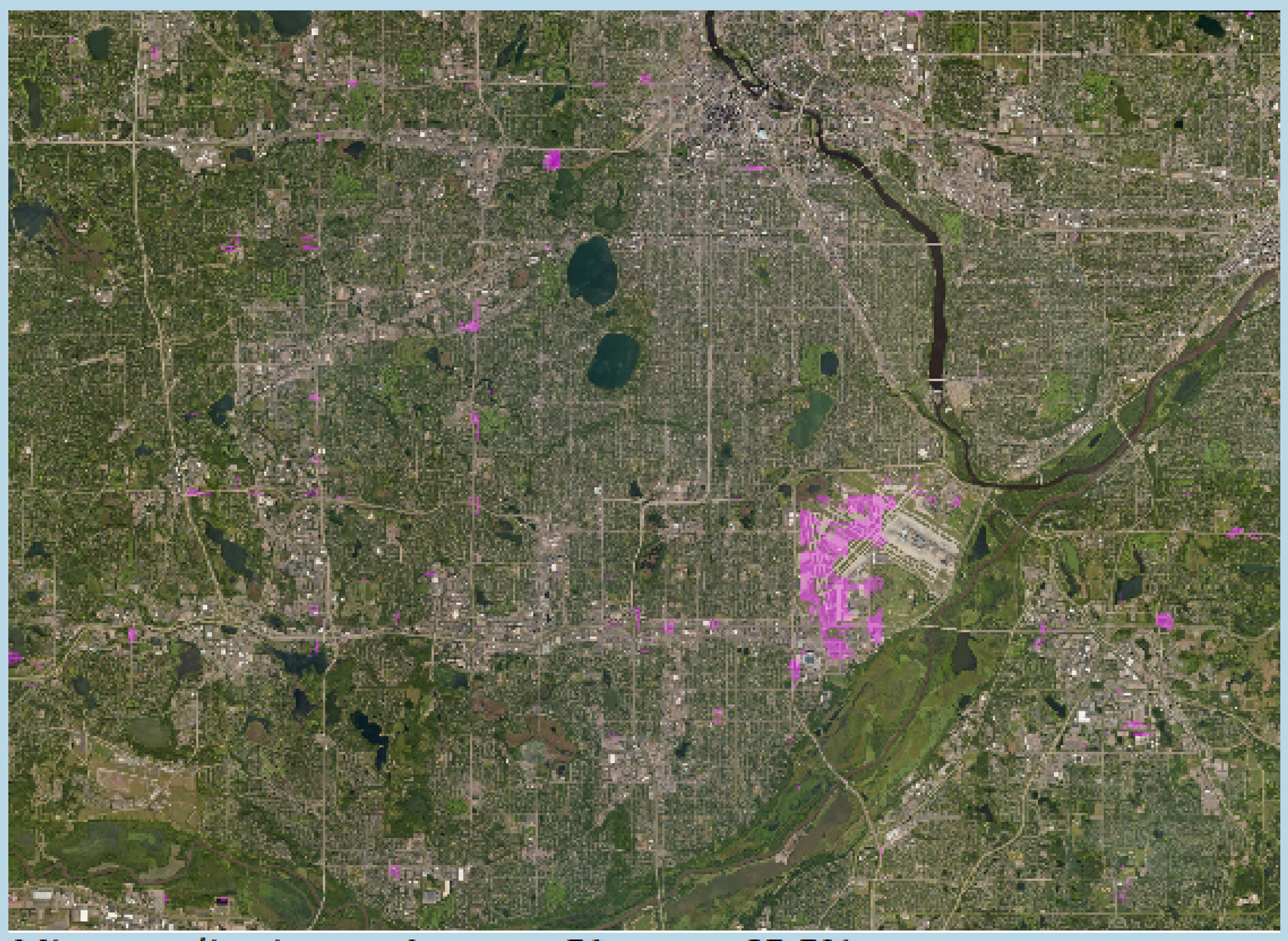
Results

The images below show the result of the prediction using the two different networks - DeepLabV3+ and ResNet50. The pink color corresponds to predicted airport pixels. The average F1-score is an average between the F1-score for airport pixels and non-airport pixels.

DeepLabV3+

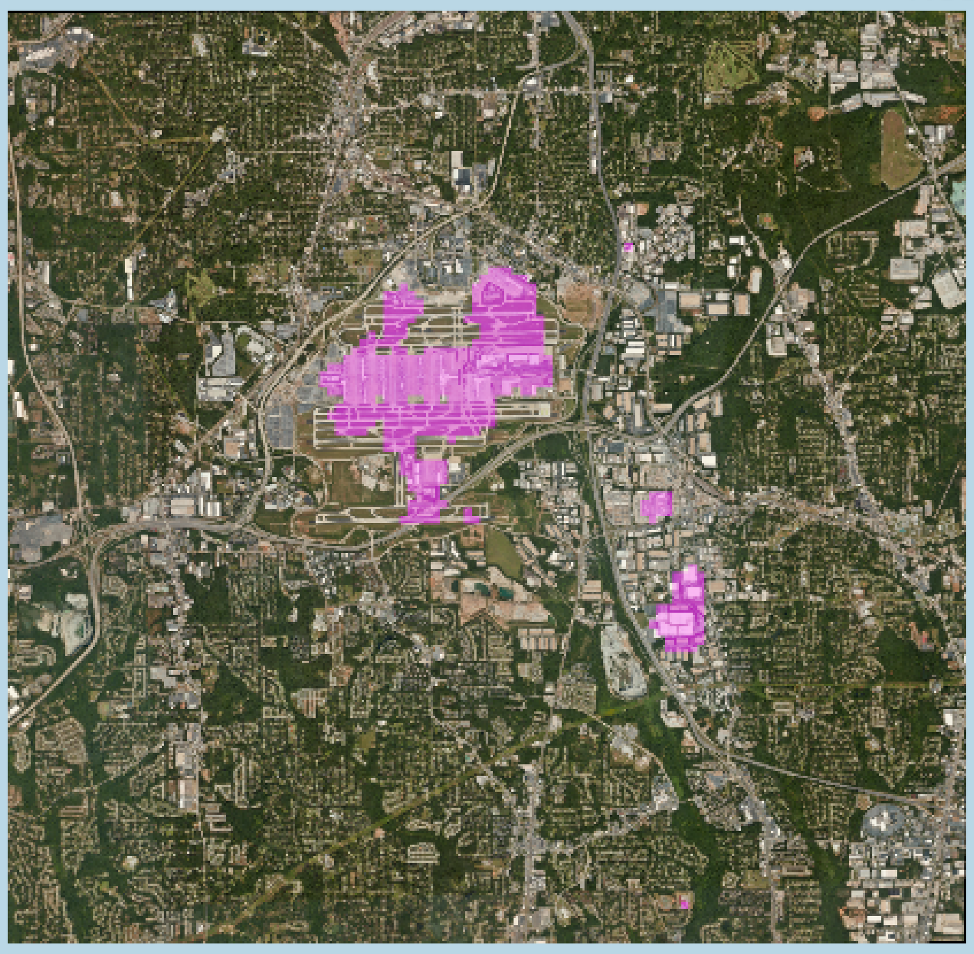


Atlanta airport. Average F1-score: 86.6%

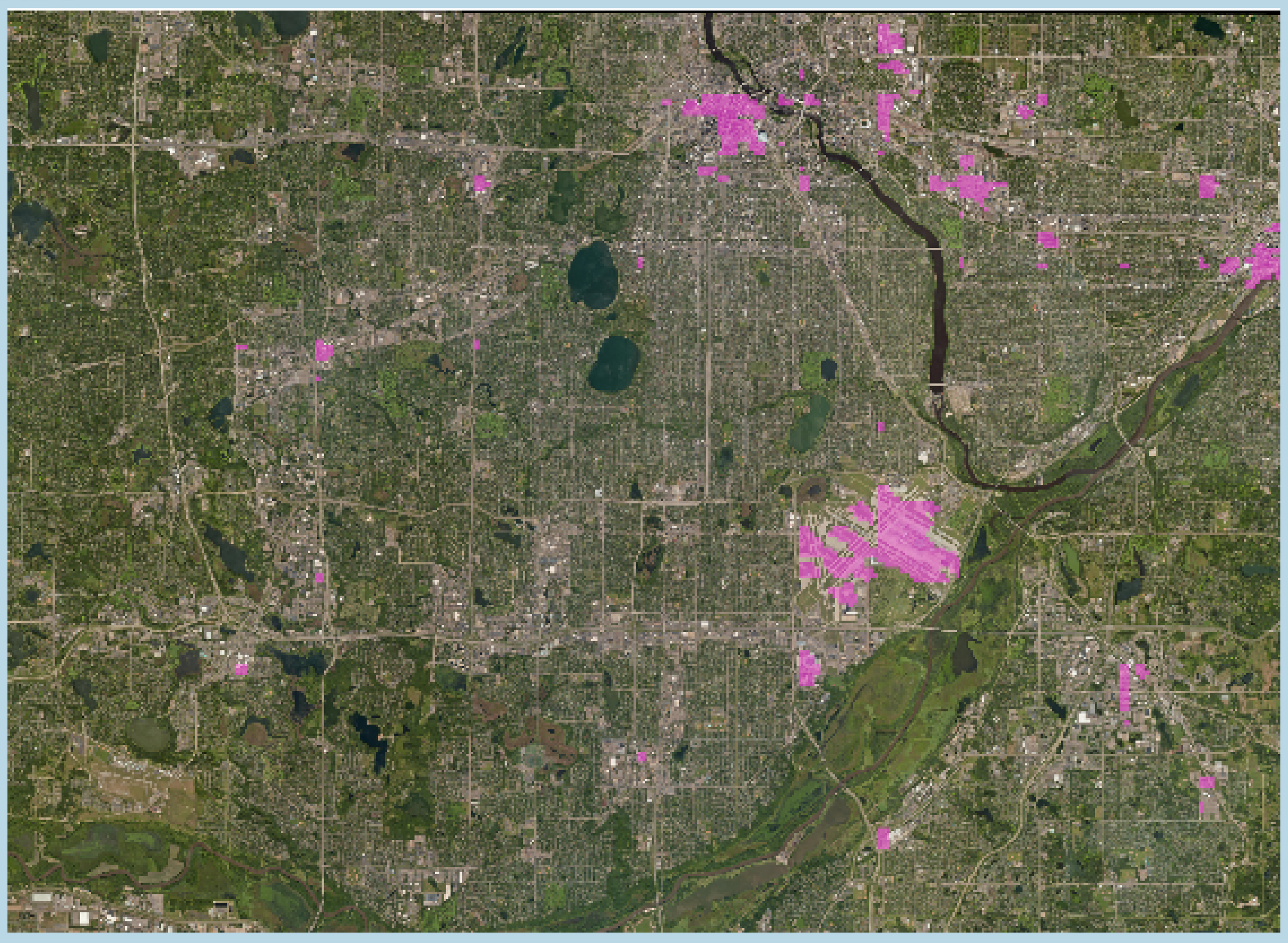


Minneapolis airport. Average F1-score: 65.5%

ResNet50



Atlanta airport. Average F1-score: 78.3%



Minnepolis airport. Average F1-score: 67.7%

Discussion

The network trained with the DeepLabV3+ model produced a good result for both images. It is clear that the prediction of Atlanta airport got a higher average F1-score compared to the prediction of Minneapolis for both of the networks. A reason for this could be that the size of the airport matters and that larger airports are easier to find. The Minneapolis airport image actually contains two airports, one larger that are easily seen and one smaller in the lower left corner, but only the larger airport are predicted as an airport for both networks.

The results shows that the best performance from the DeepLabV3+ implementation is an average F1-score of 86.8% which is achieved on the Atlanta airport with learning rate 10^{-2}. This can be compared to the best performance of the ResNet50 implementation, with an average F1-score of 78.3 %. This is achieved on the Atlanta airport using a learning rate of 10^{-3}

The amount of labeled data is probably the main problem when it comes to classification using a neural network. The labeled data consist of totally 60 satellite images with airports where all airports have different shapes, sizes and appearances. Some airports are bigger than others, some have more grass around or between the runways, they can have multiple runways or only a single one, they can be surrounded by many or almost no buildings, etc.

The conclusion that can be drawn from this project is that airports are a challenging type to classify, possibly because of their variety in appearance as well as similarities to its surrounding areas. The results does however show that it is possible to train networks to find larger airports, but with an insufficient reliability.