

Institute of Photogrammetry and GeoInformation  
Leibniz Universität Hannover

## **Master Thesis**

# **Developing strategies to handle imbalanced class-distributions in the land cover classification based on Convolutional Neural Networks**

Xiangyu Bao

Matriculation Number: 10010132

Hanover, February 2nd, 2021

First Examiner: apl. Prof. Franz Rottensteiner

Second Examiner: M.Sc. Mirjana Voelsen

# Abstract

Land cover classification is a standard task in remote sensing, where each pixel in remote sensing image is assigned a class label to indicate the physical substance on the surface of the object. In order to rapidly obtain the accurate classification results, the recent work has focused on convolutional neural networks (CNN). This is mainly because the CNN provides a framework where the features can be learned from the training dataset, which is different from conventional classifiers that use hand-crafted features. However, one of the main problems in land cover classification is the class imbalance problem, which is caused by the imbalanced distribution of the number of samples in each class. This may lead the CNN perform poorly when training these minority classes.

For that purpose, we propose two developed strategies to solve the class imbalance problem in the existing CNN structure training. The first strategy is from algorithm aspect to modify the loss function, where we implement a method called class-balanced loss to replace the original weighted cross-entropy loss and focal loss in original CNN structure. Our results show that when training with appropriate parameters in class-balanced loss, the performance of classifying minority classes can be slightly improved. On the basis of class-balanced loss, the other strategy from data aspect which change the class distribution is carried out. In this case, we combine the over-sampling method and several data augmentation methods to preprocess the input data. Through the comparison of various data augmentation methods, it can be seen that when the appropriate augmentation method is used, the result can be further improved.

In this CNN model, the training dataset is automatically generated by using the Sentinel-2 image with reference data from geospatial database provided by the German Land Survey Office of Lower Saxony (LGLN).