

Monitoring Temporal Developments using AI Fine-Grained Segmentation

S. Zamarialai (1), Z. Shi (1), B. Foing (2), M.J.L. Perenboom (3), A. Kruijver (3)

1. University of Amsterdam, The Netherlands | 2. European Space Agency - ESTEC, The Netherlands | 3. 52Impact B.V., Rotterdam, The Netherlands

Introduction

Remote sensing (RS) imagery, generated by e.g. cameras on satellites, airplanes and drones, has been used for a variety of applications such as environmental monitoring, detection of craters, monitoring temporal changes on planetary surfaces. In the early days of RS based scene classification, analysis methods were mainly based on handcrafted features. These methods do not generalize well on different scenes, they require a considerable amount of engineering skills and domain expertise to design features like: color, texture, shape and spectral information [1].

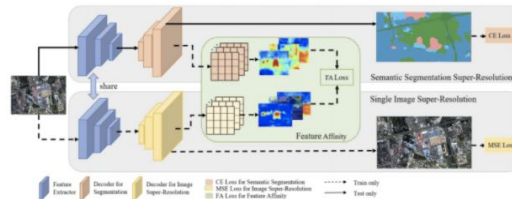
In recent years researchers started applying Computer Vision (CV) methods on RS data. This led to a steady development of remote sensing classification, providing good results on classification and segmentation tasks on RS data [5, 6, 7, 8].

When temporal data is available for a specific region, segmentation allows to classify each pixel of that image for a given class. The temporal changes could be monitored by applying this to multiple images for a given timespan.

However, there are still problems with current approaches for segmentation. Firstly, the main focus is on high-resolution RS imagery. Apart from the fact that this data is not accessible to everyone, the models fail to generalize on lower resolution data [5]. Secondly, the models fail to generalize on more fine-grained classes. For example, models tend to generalize very well on detecting buildings in general, however they fail to distinguish if a building belongs to a fine-grained subclass like residential or commercial buildings. Fine-grained classes often appear very similar to each other, therefore, models have problems to distinguish between them. This problem occurs both in high-resolution and low-resolution RS imagery, however the drop in accuracy is much more significant when using lower resolution data.

Methodology

For the reasons mentioned in the introduction, we propose a Multi-Task Convolutional Neural Network (CNN) with three objective functions for segmentation of RS imagery. This model should be able to generalize on different resolutions and receive better accuracy than state-of-the-art approaches, especially on fine-grained classes. Our method is inspired by [9] and consists of two main components.



The first component is a CNN that transforms the input image to a segmentation map. If the input image is of lower resolution, this segmentation map will miss out on complete structure of the images. The second component is another CNN that aims to build a high-resolution image from the low-resolution input image in order to reconstruct fine-grained structure information. This module essentially guides the model to learn more fine-grained feature representations. The transformed image from this module will have much more details like sharper edges, better color etc. Finally, the two images created by the model are then evaluated by a third objective function that aims to learn the distance of similarity between the segmented input image and the super-high resolution segmentation. After the model is finished with training, the second module should be detached, meaning **high-resolution imagery is only needed during the training phase**.

Results

Our method is applied to the GID dataset [5], this is a large-scale land-cover dataset that contains high-resolution satellite images. These images come with pixel-wise annotations for 15 different classes. Several models with various input sizes have been trained, 1024x1024 had the best performance. The models accuracy is 87%, with a mIoU of 0.694 and fIoU of 0.802. The figure below shows an example of an image that was fed into the model. The upper image is the original RGB followed by the true annotation and the predicted label.

