MONITORING TEMPORAL DEVELOPMENTS FROM REMOTE SENSING DATA USING FINE-GRAINED SEGMENTATION. S. Zamariala[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]. Over the past few years, the availability of open-source remote sensing datasets increased. As a result of this, it is possible to use more advanced data-consuming methods, such as Convolutional Neural Networks (CNN), for classification of remotely sensed imagery. These type of models are able to automatically detect and extract features from the input data and tend to generalize better on different scenes, even when captured by different sensors.

Computer Vision: In the past few years there have been great advances in the field of Computer Vision (CV) mainly due to the developments in Deep Learning. Convolutional Neural Networks have reached beyond-human performance on image classification tasks such as ImageNet [2]. Recently, these models have advanced beyond classification tasks. Models for object recognition and image segmentation have shown promising results [3, 4]. Image segmentation is the process of partitioning an image into multiple classes that represent a certain feature like land cover type (Figure 1).

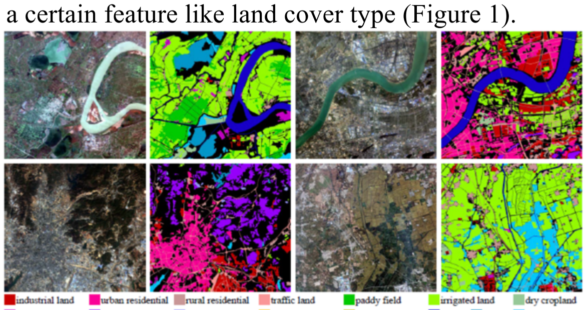

Figure 1 (source: [5]): RS images with their corresponding ground truth segmentation

Applying image segmentation techniques to RS imagery allows to accurately capture the changes over time because it is classifying every pixel in the image. For example, this could be done by taking the difference of two segmentation maps for a specific region.

Computer Vision for Remote Sensing Analysis: In recent years researchers started applying 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].

However, there are still problems with current approaches. 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. For example, in [5] the developed model was not always able to distinguish between rural residential and urban residential classes. 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 these reasons, 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 (Figure 2).

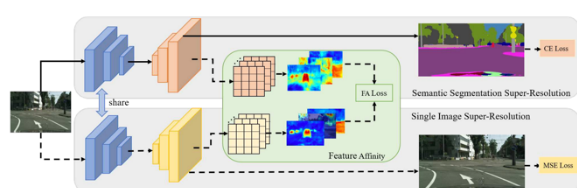

Figure 2 (source: [9]): Proposed model for RS segmentation

The first component is a CNN that transforms the input image to a segmentation map. This module is optimized with a pixel wise Cross-Entropy loss function between the segmentation map of the model and the ground truth annotations. 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. (Figure 3).

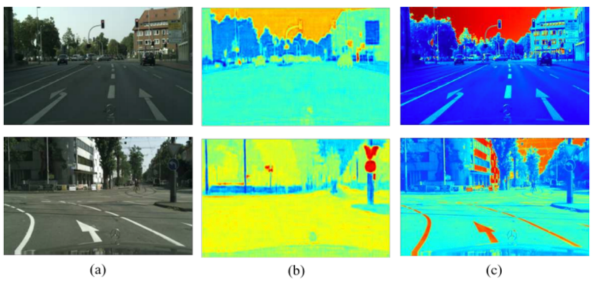

Figure 3 (source: [9]): (a) Input image, (b) Features from segmentation module (first object), (c) Features from Super Resolution module (second object)

This module is optimized with a Mean-Squared-Error loss function between the original high-resolution image and the transformed image. 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. The final objective function consists of a sum of the three objectives mentioned above. After the model is finished with training, the second module should be detached, meaning high-resolution imagery is only needed during the training phase.

Conclusion: This abstract showed a concrete example of this model on the Earth RS domain, however depending on the dataset, it could also be used for other tasks concerning RS imagery. The model could be trained on RS images of the Earth, Moon or other planetary bodies. Afterwards, it is possible to automatically apply segmentation on similar imagery at lower resolution to gather new insights, like monitoring changes in planetary surfaces, detection of variations due to phase angle illumination changes, detection of craters gullies and avalanches.

Acknowledgments: I would like to thank Alexis Guillot and Hans van het Woud for the meaningful discussions.

References:

[1] Rajesh Dhumal, Amol D. Vibhute. Advances in Classification of Crops using Remote Sensing Data

[2] Deng, J. and Dong, W. and Socher, R. and Li, L.-J. and Li, K. and Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database

[3] Wei Xia, Csaba Domokos et al. Semantic Segmentation without Annotating Segments

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition

[5] Xin-Yi Tong, Gui-Song Xia, Qikai Lu, Huanfeng Shen, Shengyang Li, Shucheng You, Liangpei Zhang. Land-Cover Classification with High-Resolution Remote Sensing Images Using Transferable Deep Models

[6] PRIIT ULMAS, INNAR LIIV. Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification

[7] Patrick Helber, Benjamin Bischke, Andreas Dengel, Damian Borth. EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification

[8] Ilke Demir1 (Facebook), Krzysztof Koperski (DigitalGlobe). DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images.

[9] Li Wang, Dong Li, Yousong Zhu, Lu Tian, and Yi Shan. Dual super-resolution learning for semantic segmentation. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020, pages 3773–3782. IEEE, 2020