

# OPERATIONAL PIPELINE FOR A GLOBAL CLOUD-FREE MOSAIC AND CLASSIFICATION OF SENTINEL-2 IMAGES

M. Swaine<sup>1</sup>, C. Smit<sup>1</sup>, S. Tripodi<sup>2</sup>, G. Fonteix<sup>2</sup>, Y. Tarabalka<sup>2</sup>, L. Laureore<sup>2</sup>, J. Hyland<sup>1</sup>

<sup>1</sup>LuxCarta South Africa, Cape Town, South Africa,

<sup>2</sup>LuxCarta Technology, Mouans Sartoux, France

**KEY WORDS:** Deep learning, optical satellite images, cloud detection, U-net, classification

## ABSTRACT:

Global Earth observation from satellite images is an active research topic, driven by numerous applications, such as telecommunications, defence, natural hazard monitoring and urban management. The recently launched twin Sentinel-2 satellites acquire 13-band optical data with a 2-5 day revisit time, freely available for any use, and thus very valuable for global Earth observation. In this paper, we present a completely automatic operational chain for a global cloud-free mosaic and classification of Sentinel-2 images. The proposed pipeline enables producing the world 10-m cloud-free mosaic, and a 6-class landcover classification map.

## 1. INTRODUCTION

Image classification in remote sensing serves as a vital tool for many applications, such as precision agriculture (Neetu, Ray, 2019), urban planning (Maggiori et al., 2017), land resource management and environmental protection. In land-cover (LC) classification, every pixel in a data source is assigned a thematic class, for example, water, trees and grass.

Under the EU Copernicus program, two twin Sentinel-2 satellites (Sentinel-2A and Sentinel-2B) were launched in 2015 and 2017, respectively, providing 13 optical bands at between 10-60 meter spatial resolution data with a 2-5 day revisit time. The data is freely available for use in both research and commercial applications. Sentinel-2 satellites deliver around 10 terabytes of Earth observation data daily (Kempeneers, Soille, 2017); the volume and frequency of data acquisition means that artificial intelligence is a critical tool for gaining value from the data.

Researchers have actively been working on designing efficient models for Sentinel-2 data classification, using both traditional machine learning techniques, such as support vector machines, random forest (Neetu, Ray, 2019), and deep learning (Helber et al., 2019, Pelletier et al., 2019). The main difference between deep learning and traditional machine learning methods is that unlike traditional machine learning, deep learning will automatically learn features such as texture, edges and morphological attributes, therefore there is no need for complicated feature engineering (Zhang et al., 2016). Training convolutional neural networks require the availability of a large training data set. Very recently, a large-scale multi-label training data archive BigEarthNet has been released (Sumbul et al., 2019), which boosted the development of deep neural network models for Sentinel-2 image classification (Sumbul, Demir, 2019, Ulmas, Liiv, 2020). However, this archive provides annotations for image patches and not for every pixel;

it is challenging to use these data for training a model for pixel-wise classification.

In this work, we propose an automatic operational pipeline for a global mosaic and classification of Sentinel-2 images. This proposed chain enables, producing in a completely automatic way, a cloud-free mosaic of 4-band (red, green, blue, and near-infrared) images of the Earth land surface, and a landcover classification map where each pixel is classified to one of the following six classes: water, grass, tree/forest, residential building, industrial/commercial building and barren. One of the important *contributions* of this work is the development of a chain which works automatically on a *global scale*.

## 2. PROPOSED CHAIN

The proposed pipeline consists of the following steps:

- 1) Compile a cloud-free scene list.
- 2) Downloading Sentinel-2 images and creating a full resolution cloud mask for each image.
- 3) Generating a cloud-free mosaic.
- 4) Inferring a 6-class landcover map.

In the following, we describe each step and show experimental results, using Sentinel-2 tiles from different parts of the world.

### 2.1 Compiling a cloud-free scene list

Some zones of the Earth are covered by clouds most of the time, which means that no cloud-free tile can be found within the given date range (typically, the most recent imagery is required to be able to keep both mosaic and landcover map up to date). An initial list of scenes is compiled to ensure a cloud free

composite can be created. The following steps are applied at a preview level:

a) All previews between a user specified date range with a maximum cloud cover percentage are downloaded and steps b) and c) are processed on all previews.

b) A U-net model (Ronneberger, 2015) is trained to infer a cloud mask, specifically on a full preview (Fig. 1 & Fig. 2).



Figure 1: Raw input preview.

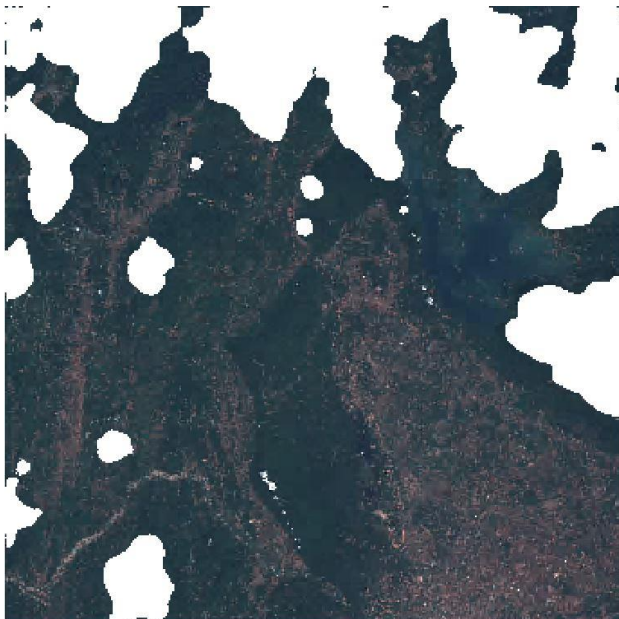


Figure 2: Cloud mask from prediction.

c) Small convolutional neural network (CNN, with 3 convolutional layers) is trained to classify each preview image from the stack of available images as hazy (Fig. 3) or not hazy (Fig.4).

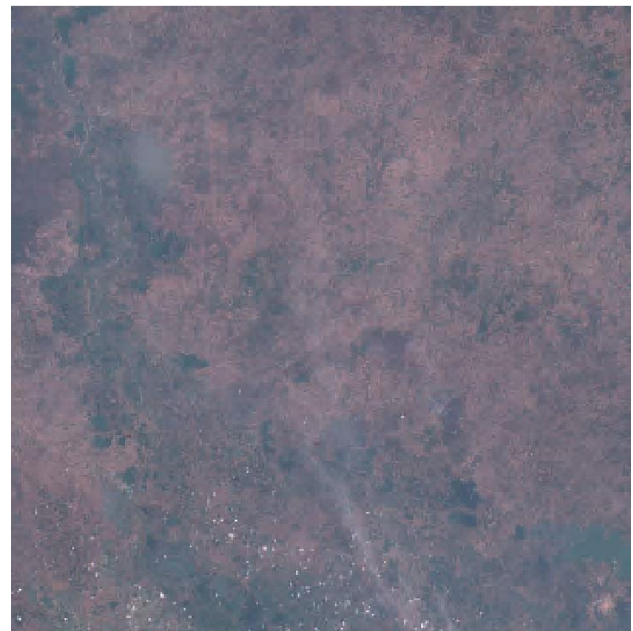


Figure 3: Hazy preview.

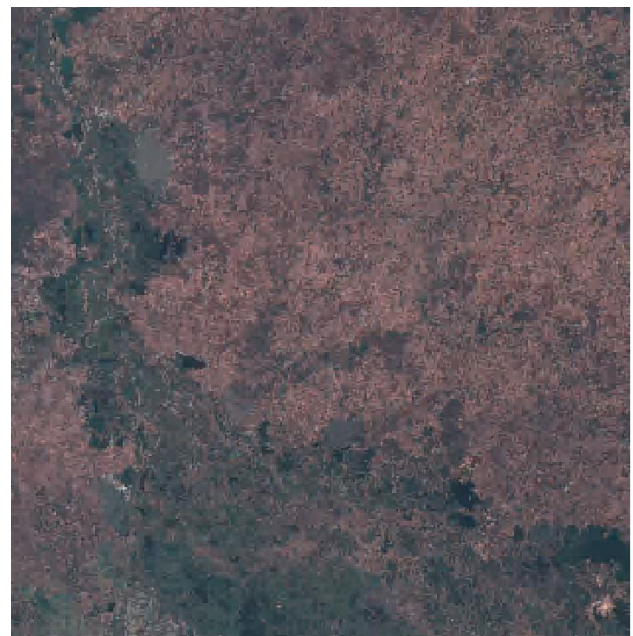


Figure 4: Non-hazy preview.

d) A minimum cloud-free scene list is compiled prioritizing image date per Sentinel-2 scene (see Fig.5 for an example of the selected scenes).



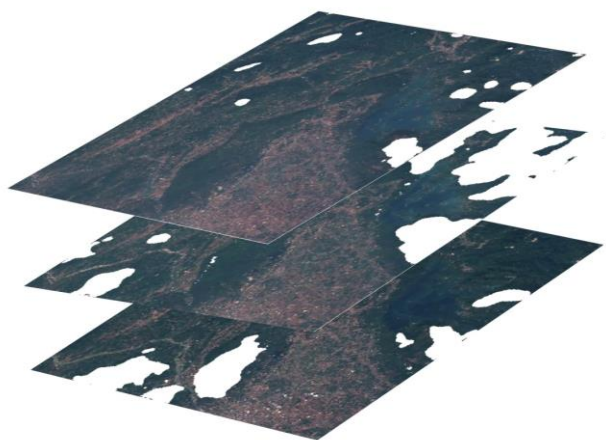


Figure 5: Stacked previews with cloud masks used to compile a minimum scene list.

## 2.2 Downloading Sentinel-2 images and creating a full resolution cloud mask for each image

In order to generate an image for each Sentinel tile which is as recent as possible, cloud-free image parts extracted at several dates can be combined. We apply the following procedure:

a) Images selected in step 1 are downloaded in full 10-m resolution and scaled from 16 to 8 bits. Image enhancement techniques are applied in this process to enhance contrast and retain texture details in both dark and bright areas. A comparison of the True Color Image (8 bit) provided by the European Space Agency (ESA) and the TCI (8 bit) created in LuxCarta's operational pipeline are shown in Fig. 6 and Fig. 7.

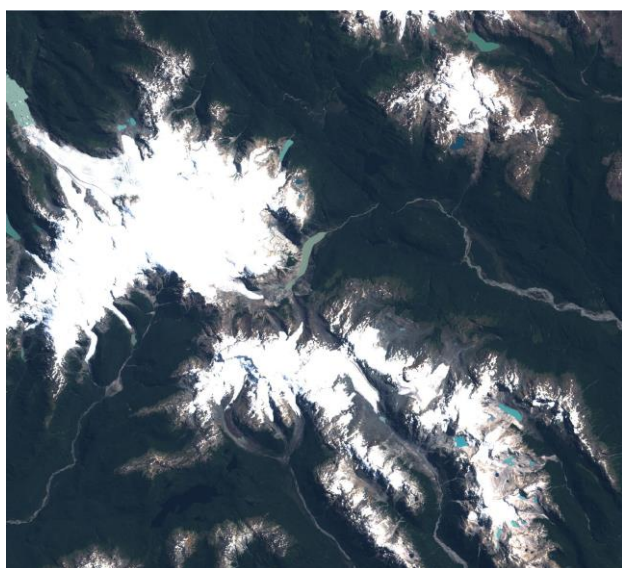


Figure 6: ESA produced 8bit True Color Image.

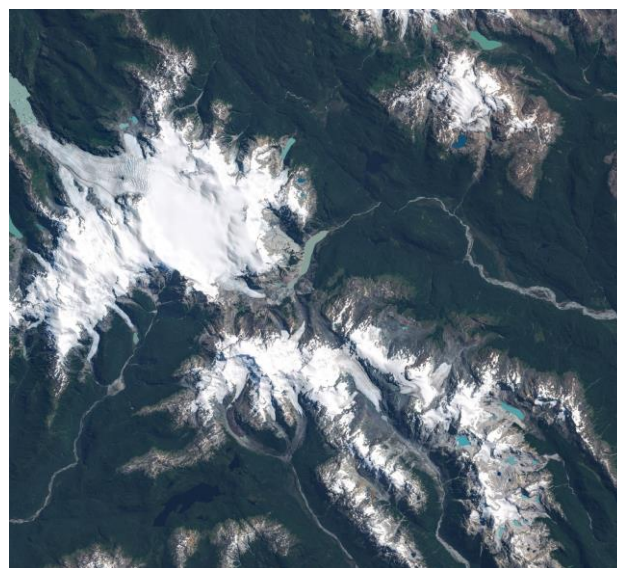


Figure 7: LuxCarta produced 8bit True Color Image.

b) U-net model (see Fig. 14) is trained to infer a cloud mask for every Sentinel image, until all pixels have full cloud-free data coverage (see Fig. 8-9).

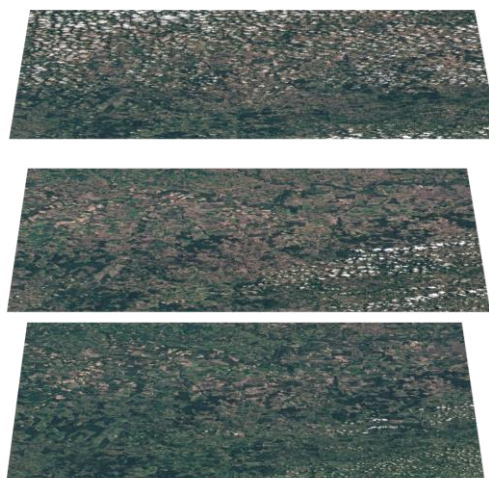


Figure 8: Input imagery stack.

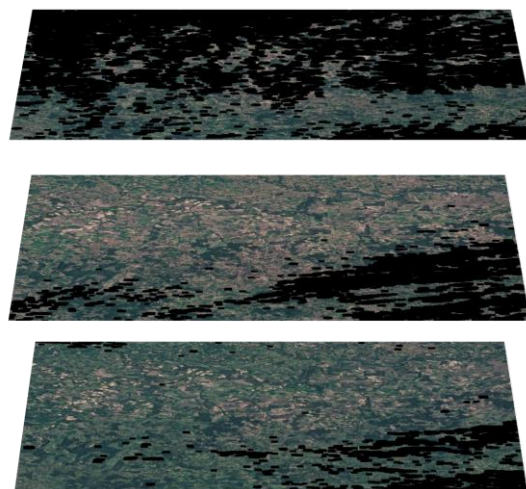


Figure 9: Input imagery stack with cloud mask.





Figure 10: Fused cloud free composite.

### 2.3 Generating a cloud-free mosaic.

The selected images and their associated cloud masks are further fused together to create a cloud free composite scene using a mosaicking algorithm (Fig. 10). The mosaicking process is also applied between the neighboring tiles, so that the global Earth mosaic has no seam lines between tiles. The following steps are followed to blend seam lines:

a) Neighboring tiles are identified via a spatial query with the Sentinel grid system (Fig. 11).

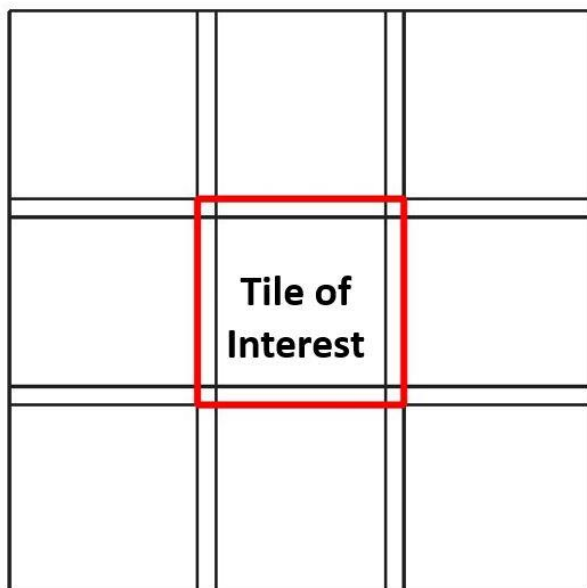


Figure 11: Tile of interest used to identify neighboring tiles

b) The overlap between the tile of interest and neighboring tile area is clipped for all adjacent tiles. (Fig. 12)

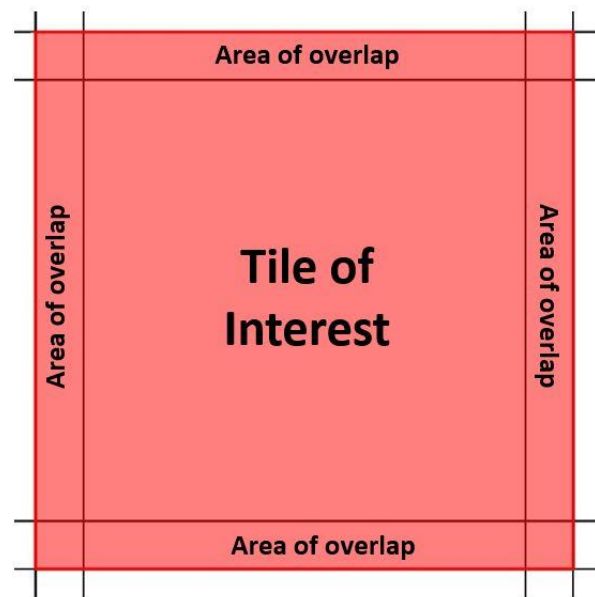


Figure 12: Tile of interest and area of overlap of neighboring tiles which are clipped to be blended with tiles of interest.

c) Clipped out neighbor regions used to blend with the overlapping edge of the tile of interest. Fig. 13 and 15 illustrate the mosaicking result (before and after mosaicking, respectively).



Figure 13: Seam lines evident between four tiles prior to mosaicking.

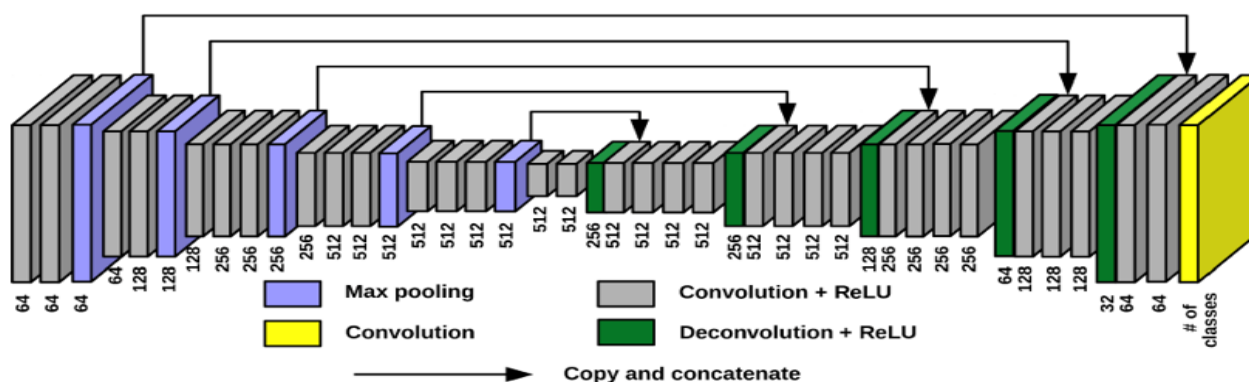


Figure 14: U-net model used for landcover classification.

- Buildings: residential and industrial/commercial.

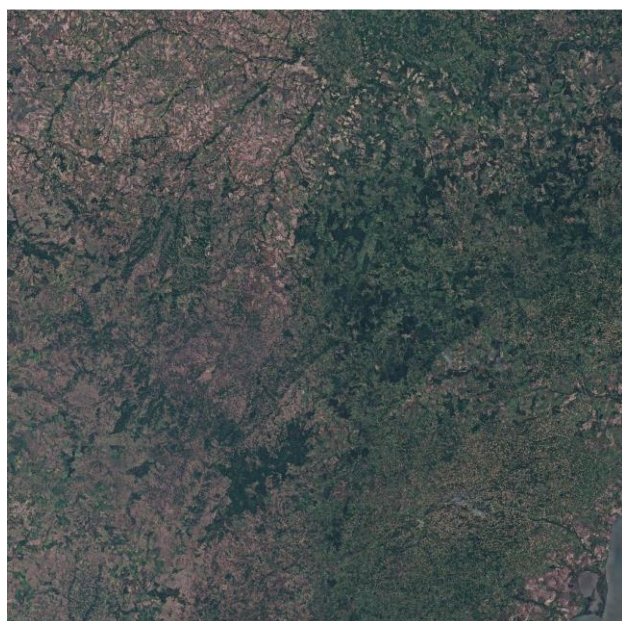


Figure 15: Hard seam lines eliminated between tiles post mosaicking

## 2.4 Inferring a 6-class landcover map

We have adopted a U-net convolutional neural network architecture (Ronneberger, 2015) for landcover classification of Sentinel-2 images. U-net architecture has exhibited the highest performances in several benchmarks involving satellite image classification (Huang et al., 2018, Saux et al., 2019), and is particularly suitable to classify each image pixel to one of the classes of interest, without losing fine image details. The applied network architecture is adopted from (Tasar et al., 2018) and illustrated in Fig. 14, which consists of an encoder that is architecturally the same as the first 13 convolutional layers of VGG16 (Simonyan, Zisserman, 2014), a corresponding decoder, mapping low resolution encoder feature maps to original input image size of outputs, and two center convolutional layers.

We have trained three separate U-net models to classify:

- Water class.
- Vegetation classes: grass, tree/forest.

We have used Red, Green and Blue spectral channels as the input to train/predict building classes; while Infrared, Red and Green channels have been used as the input for the other two models. Predictions of the three models are further combined to infer a final map. We have also trained one model to classify all classes; however, it yields worse performance than the results obtained by the proposed multi-model approach.

Providing geodata for thirty years, LuxCarta ([www.luxcarta.com](http://www.luxcarta.com)) has a huge quantity of very accurate pixel-wise annotated ground-truth data, with the precise date of the source image acquisition. Using these data coupled with the time-corresponding Sentinel images allowed us to train generic models yielding high accuracies all over the world.

To further improve the precision of the inferred landcover map, we have developed a multi-temporal classification approach: A time series of images is classified using the U-net model. The developed probabilistic model is further used to combine the inference results.

Fig. 16 shows the Sentinel-2 10-m image over the city of Hradec Kralove, Czech Republic, extracted from our generated cloud-free mosaic; and the corresponding 6-class landcover map.

The proposed pipeline is currently in operation to generate geodata for our customers. We have generated a cloud-free mosaic for approximately 130 million km<sup>2</sup> of the Earth surface (the whole Earth land surface with the exception of polar regions and very small islands in the ocean) from summer images of 2019. We have recently validated our multitemporal classification approach to produce a landcover map over the 5 million km<sup>2</sup> surface of Canada. As validated by our quality control team, the deep learning models trained on an extensive dataset, coupled with the multitemporal aspect of the method, the proposed automatic classification pipeline yields comparable and, in some cases, more accurate landcover maps when compared to the human annotations.





(a)



(b)

Figure 16: (a) Sentinel-2-derived 10-m image over Hradec Kralove, Czech Republic, and (b) corresponding 6-classes landcover map: water (blue), grass (green), tree/forest (dark green), residential building (orange), industrial/commercial building (brown) and barren (cream).

## REFERENCES

- P. Helber, B. Bischke, J. Hees, A. Dengel, "Towards a Sentinel-2 based human settlement layer," IGARSS 2019.
- B. Huang et al., "Large-scale semantic classification: outcome of the first year of Inria aerial image labelling benchmark, " in IGARSS, 2018.
- P. Kempeneers and P. Soille, "Optimizing Sentinel-2 image selection in big data context," Big Earth data, 1: 1-2, pp. 145-158, 2017.
- E. Maggiori, Y. Tarabalka, G. Charpiat and P. Alliez, "Convolutional neural networks for large-scale remote sensing image classification," IEEE TGRS, 55(2), pp. 645-657, 2017.
- Neetu and S. S. Ray, "Exploring machine learning classification algorithms for crop classification using Sentinel-2 data," Int archives ISPRS, XLII-3/W6, pp. 573-578, 2019.

C. Pelletier, G. Webb, F. Petitjean, "Deep learning for the classification of Sentinel-2 image time series," IGARSS 2019.

O. Ronneberger et al., "U-Net: convolutional networks for biomedical image segmentation," in MICCAI, 2015.

B. Le Saux, N. Yokoya, R. Hansch, and M. Brown, "2019 IEEE GRSS data fusion contest: large-scale semantic 3D reconstruction," IEEE GRSM, pp. 33-36, Dec 2019.

K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv, 2014.

G. Sumbul, M. Charfuelan, B. Demir, V. Markl, "BigEarthNet: a large-scale benchmark archive for remote sensing image understanding," IGARSS 2019.

G. Sumbul, B. Demir, "A novel multi-attention driven system for multi-label remote sensing image classification," IGARSS 2019.

O. Tasar, Y. Tarabalka, P. Alliez, "Incremental learning for semantic classification of large-scale remote sensing data," arXiv, oct. 2018.

P. Ulmas, I. Liiv, "Segmentation of satellite imagery using U-net models for land-cover classification," arXiv, March 2020.

L. Zhang, G.S. Xia, T. Wu, L. Lin and X.C. Tai, "Deep learning for remote sensing image understanding, " Journal of sensors, 2016.