

A new method for geomorphological studies and land cover classification using Machine Learning techniques

Giorgia Miniello^{a,*} and Marco La Salandra^b

^a*Department of Physics, University of Bari Aldo Moro & INFN Bari, Italy*

^b*Department of Earth and Environmental Sciences, University of Bari Aldo Moro, Italy*

E-mail: giorgia.miniello@ba.infn.it, marco.lasalandra@uniba.it

The processing of aerial high-resolution images is key for territorial mapping and change detection analysis in hydro-geomorphological high-risk areas. A new method has been developed in the context of CLOSE (Close to the Earth) project, resulting in a workflow based on open source MicMac photogrammetric suite and on High-Performance Computing. The workflow allowed to process a sequence of more than 1000 drone images captured along a reach belonging to the Basento River in Basilicata (Italy) during one single run. The workflow optimisation aims to extract the orthophotomosaic, the point cloud and the Digital Surface Model (DSM) of selected areas. The high quality of the image details can be used for land cover classification and extrapolating features useful to mitigate the hydro-geomorphological hazard, through Machine Learning models trained with satellite public data. Several Convolutional Neural Networks have been tested using progressively more complex layer sequences, data augmentation and callback techniques for training procedures. The results are given in terms of model accuracy and loss.

*International Symposium on Grids & Clouds 2021, ISGC2021 22-26 March 2021
Academia Sinica, Taipei, Taiwan (online)*

*Speaker

1. Introduction

The last decade has been characterized by a considerable development in geoinformatics domain, especially regarding to new techniques which allow to detect the structure and dynamics of the Earth surface, extrapolating 3D information from overlapping 2D images. In this context, one of the most popular among such techniques is photogrammetry, a low-cost technique, since the only mandatory device is a photcamera which provides very accurate topographic models exploiting only one single acquisition. Currently, this technique is emerging due to the development and the increasing use of UAVs (Unmanned Aerial Vehicles) and SfM (Structure from Motion) algorithms in several application (geology, archeology, etc.).

The UAVs are very powerful tools that allow on-demand acquisition of high-resolution aerial images, useful to identify and monitor active phenomena that result in topographic changes.

SfM is an algorithm that addresses the problem of determining 3D position of image descriptors, regardless of the shooting angles, to estimate three-dimensional structures [1].

By leveraging the potential of these tools, large datasets can be acquired increasing the extension of the surveyed areas and the image overlaps and increasing spatial resolution of the model using larger image formats [2].

An important limit arises when trying to detect extended and hydrogeological high-risk areas of the surface of the Earth, since a large amount of detailed data must be managed and processed to develop high resolution models in a reasonable time. In this case, high-performance hardware is required to run these processes. There are different commercial and free open source software which are able to perform the photogrammetric tasks, but only a few commercial software, leveraging expensive and dedicated hardware, can handle a large amount of data, although taking long periods of time. This is due to the poor performance of a single computer in terms of CPUs availability and memory capacity.

The images acquired and elaborated by the algorithm as orthophotos and orthophotomosaics can be used for territorial classification exploiting Machine Learning (ML) techniques. The use of Deep Neural Networks (DNNs) is considered an essential mean for detection, evaluation and management of land resources [3]. Land cover and land use classification are related both with the different land features on the surface of the Earth and with the different purposes these territories are addressed (e.g., residential, agriculture, industrial areas, etc.). This classification is accomplished by training Neural Networks in order to learn and extract some particular features for each class of land surface, starting from a set of pre-labelled aerial images previously acquired.

In our paper we will describe both the implementation of an original photogrammetric workflow, optimized to be less time consuming with respect to other commercial software (sections 3, 4 and 5), and the setup of two different DNNs for multiple classification of aerial images (section 6). Our results and observations are finally summarized in section 7.

2. The ReCaS-Bari infrastructure

The ReCaS-Bari data center was created by the University of Bari *Aldo Moro* and the National Institute of Nuclear Physics (INFN, Division of Bari - Italy) as part of the ReCaS project (“PON Ricerca e Competitività 2007-2013, Avviso 254/Ric”).

The aim of the data center is to meet the growing need for scientific computing coming from experimental and theoretical groups operating within the INFN Section and the Department of Physics of Bari (Italy). Several services are available at the infrastructure, such as HTC, HPC, IaaS, PaaS.

In our study we deployed the FOSS (Free Open Source Software) photogrammetric workflow on HTC cluster which counts 128 servers for a total of about 8000 CPU core, with 4GB of RAM per core and 4PB of parallel disk space. Each computing server, which contains up to 64 slots, can access the whole ReCaS-Bari disk space at a speed of 10 Gbps. The GPFS distributed file system was used for storage management. The operating system used is CentOS 7, and the queues are managed by the HTCondor batch system, which can run single and multi-core jobs.

In order to develop ML tasks for this study a new cluster hosted in ReCaS-Bari data center was exploited. Currently, the cluster is composed of 5 machines. Each one counts:

- 4 GPU NVIDIA V100 32GB
- 96 CPUs
- 753.5 GB RAM
- 6 TB SSD Disk

Thanks to the high number of cores, the GPUs allow the running of high-performance parallel algorithms reducing the overall execution time. The configured cluster is able to run batch jobs or open interactive environments where users are able to write code and test it in real time. In our study we used interactive mode implementing ML algorithms directly via Jupyter/Tensorflow interface.

3. Method

Photogrammetry is a technique that allows to extrapolate three-dimensional information from overlapping two-dimensional images. In order to perform this process, the Structure from Motion algorithm [4] [5][6] adopts a sequence of fundamental steps:

- detection of key features and tie-points of the images,
- estimation of calibration parameters and camera positions and orientations,
- dense-matching and point cloud generation.

In this study, the UAV data acquisition was done with "DJI Inspire 2", a quadcopter with aluminium-magnesium composite body and carbon fiber arms equipped with optical sensor "Zenmuse X5S" (20.8 MP, supported lens DJI MFT 15mm/1.7 ASPH, sensor CMOS 4/3", FOV 72°, image resolution 5280 per 3956 pixels). The flight altitude was on average 50 m (above ground level of take-off location) in order to acquire the desired high-spatial resolution of about 1 cm/pixels.

Two reaches of the Basento river near Ferrandina (MT), in the Basilicata region of southeastern Italy, were selected as research areas. Two datasets related to different reaches of the river were used: 1139 images (related to 2019), covering an area of approximately 600 per 200 m, and 2190 images

(related to 2020), covering an area of approximately 1160 per 300 m. Our workflow is based on MicMac, GDAL and Orfeo ToolBox open-source libraries and was developed on High-Throughput Computing environment (HTC). Through this approach it was possible to leverage the resources of the ReCaS-Bari computing cluster in order to manage large image datasets and return different output such as the orthophotomosaic, the dense point cloud and the Digital Surface Model (DSM) in a reasonable time lapse.

4. Workflow implementation

In this work a photogrammetric workflow based on SfM was developed. All the steps are summarized in Fig.1, although here we will only go through the main ones in details.

The first step (MicMac “Tapioca Graph”) is characterized by a Scale Invariant Features Transform operation (SIFT), which extrapolates the key features from a set of reference images and find the candidate matching features of images creating a list of all overlapping image pairs.

In the second step (MicMac “Tapioca File”) the homologous features (known as *tie-points*) [7] [8] are computed and in step 4 (MicMac “Tapas”) the 3D positions of the key features and camera parameters are retrieved, so that internal orientation and external orientation parameters are generated.

The step 6 (MicMac “Malt Ortho”) generates the orthophoto for each image of the dataset and step 8 (OTB “otbcliMosaic”) blends them to generate the final orthophotomosaic (using a feathering method, i.e. a mean convolution filter that replaces each pixel value with the mean value of its neighbours). Finally, step 10 (MicMac “C3DC”) generates the dense point cloud, computing a depth map for each image (grey-scale image that contains information relating to the distance of the surfaces of scene objects from a viewpoint) and determining the 3D coordinate of tie points, and step 11 (MicMac “PIMs2Mnt”) generates the Digital Surface Model (DSM) that blends the individual depth maps.

In order to handle a large amount of high-resolution images within a good time range, high computing and data storage resources must be required. High Throughput Computing (HTC) represents a method for handling this computationally intense task, which let the users split a big task into many small independent ones (a.k.a. *jobs*) that are distributed across a computer cluster [9].

5. Test results and performance evaluation of the photogrammetric workflow

In this particular case of study, bottlenecks are represented by computing the tie-points, orthophotos and DEM, whose tasks are split into small jobs which run in parallel on independent worker nodes of the cluster. In detail, each job performs the parallel computing of tie-points on a small subset of the original image dataset and then the results are combined.

A similar approach was developed for orthophotos and DEM, in which each job generates individual orthophoto and depth maps, using a small subset of images and blending the results to generate the final orthophotomosaic and DEM.

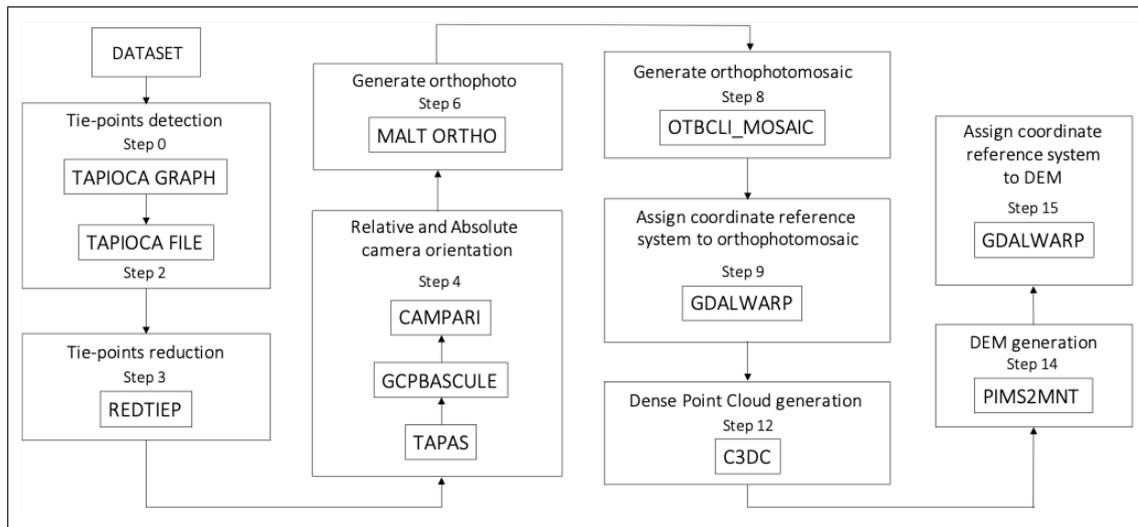


Figure 1: Processing chain of the FOSS photogrammetric workflow

In order to overcome the computational load related to those steps, the following issues must be considered :

- parallelization of the most demanding steps on independent worker nodes (WNs);
- each job must perform the parallel calculation of tie-points on a small subset of the original image dataset and then combine;
- similar approach must be developed for orthophoto and DEM.

The performance evaluation of the photogrammetric workflow was performed with a series of test exploiting different configurations and number of images. In particular, the processing time of a previous workflow configuration was compared to the one currently used. A first version of the photogrammetric workflow involved the use of “micmac-distmatching” command to split the image dataset into tiles and “coeman-par-local” command to generate the orthophoto and DEM for each tile on independent worker node. Finally, in order to combine the results and generate the final orthophotomosaic and DEM, the MicMac commands “Malt Ortho” and “Tawny” were used.

As a first approach, we considered the dataset made of 1139 images. The overall processing time of the workflow using this configuration was about ~37 hours and incorrect radiometric calibration issues have been also found in the final output. Those issues have been solved using a different command configuration, which is the one currently used, that leverages the ability to run the “Malt Ortho” MicMac command in parallel on independent subset of images (50 images per job) and to merge the resulting orthophotos using the Orfeo ToolBox library. In addition, through the “image-matching” operation (“C3DC” MicMac command) performed in parallel on subsets of 50 images, depth maps useful to reduce the processing time of the final DEM (through the MicMac command “PIMs2Mnt”) were generated. Using this approach, a remarkable improvement in the overall processing time (~25 hours) and output quality was observed.

A further improvement was also reached using a different cluster configurations on the best command

configuration, increasing the number of jobs and worker nodes on which they are deployed, the parallel execution and consequently ensuring faster file writing of each node on the File System (GPFS).

As first, the number of worker nodes was increased from 23 to 56, reducing the number of input images per job from 50 to 3 images, recording a significant reduction of the time needed to produce the orthophotomosaic, the dense point cloud and DSM (~15 hours). All these results were obtained using batch jobs but some issues arose related to:

- cluster's queue that manages the execution of user's jobs;
- computing resources and performance of the worker nodes;
- several parallel jobs running on a single node greatly affecting the performance since the photogrammetric workflow is characterized by some multi-thread MicMac command.

Once again, a considerable improvement in the workflow performance was obtained using a dedicated slot on each worker node to run a single job via "pssh", ensuring a higher deployment on different nodes, parallel access to a greater number of nodes, thus overcoming the batch system job management and generating a list of 103 nodes associated with an id number to create a jobs execution defined scheme. The optimization of our workflow implied a significant time reduction of the workflow's distributed steps. Comparing to the 56-nodes test a time reduction of ~67% for the orthophotomosaic generation and of ~37% in the dense point cloud generation were recorded. The overall processing time with this configuration is less than 10 hours. Fig.2 shows the output for the 1139 image dataset.

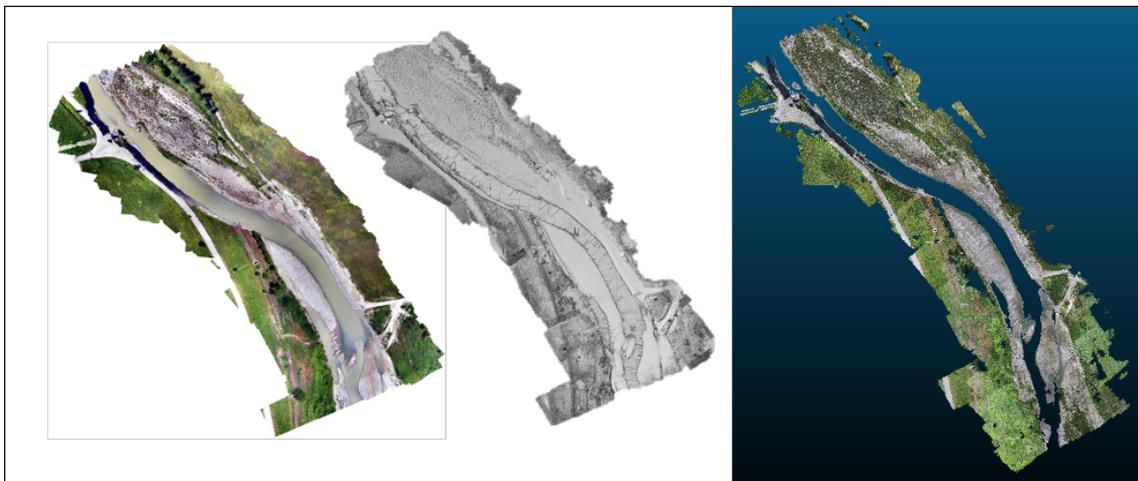


Figure 2: a) Orthophotomosaic (1.3 cm/pixel); b) Digital Surface Model (2.5 cm/pixel); c) Point cloud (~95 000 000 densified points) related to 1139 image dataset

This represents a very good result, especially compared to the processing time of the same dataset using a commercial SfM software (Pix4D) on a single workstation, reaching a significant reduction of the processing time of ~73%.

Finally, the FOSS photogrammetric workflow was processed using a dataset of 2190 images. The total amount of processing time in this case is less 22 hours.

6. Territorial classification using Deep Neural Networks

The images acquired by the drone can be used to create an original dataset of aerial images that can be divided in classes. This part of our work is currently ongoing to get the most suitable arrangement. Comparing our images to those used in other datasets built for the purpose, a considerable resolution downgrade must be applied to use ML techniques on them. For change detection studies, the most important classes to be included for a “minimal classification” must be «Terrain», «Water» and «Vegetation».

Two different models have been tested for territorial classification on the EuroSAT dataset [10] (available at <https://github.com/phelber/eurosat>), which was explored as a first approach for the next step of our studies. The EuroSAT dataset consists of a collection of 27000 Sentinel-2 satellite images made of 13 spectral bands and 10 pre-labelled classes : 'Annual Crop', 'Forest', 'Herbaceous Vegetation', 'Highway', 'Industrial', 'Pasture', 'Permanent Crop', 'Residential', 'River', 'Sea Lake'. In the first model a sequence of pairs of max pooling and convolution layers ending with a dropout layer (30%) and a dense layer was used, setting 150 epochs. A test accuracy of ~90% was reached but the loss function presents huge fluctuations, see Fig.3. Fig.4 shows a summary of the accuracy for each class compared with the overall one.

In the second model we used a different approach adding a convolutional base of the VGG16 Keras Model (pre-loaded weights) to exploit data augmentation technique (i.e horizontal flip, vertical flip and rotation) and improving the results.

In addition to Model Checkpoint used in the previous model, the EarlyStopping and ReduceLROn-Plateau callback functions were added to limit the overfitting and reduce the learning rate if no improvements are seen after a fixed number of epochs. Using this configuration, the training was early stopped after 30 epochs reaching a test accuracy of ~97% and stabilizing the test loss just after a single decade, as can be observed in Fig.5. Fig.6 shows a summary of the accuracy for each class compared with the overall one.

It could be interesting to make some observations about the misclassification of the classes and the percentage of wrong predictions for each class for the best model, the VGG16. As expected, Tab.1 shows a reciprocity in misclassification (e.g., the Permanent Crop class is mostly misclassified for Herbaceous Vegetation and viceversa, the Residential class is mostly misclassified for Industrial and viceversa, etc..) for most classes and Fig.7 shows the number of wrong predictions for each class. The Herbaceous Vegetation turned to be the most misclassified class reaching a percentage of 7.5%.

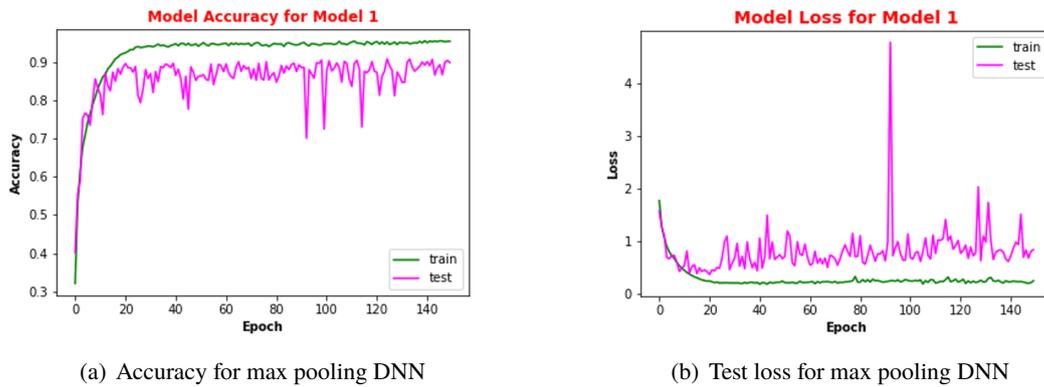


Figure 3: Score for max pooling DNN model

index	y_true	accurate_preds	label_count	class_acc	overall_acc	
0	0	b'AnnualCrop'	1276	1500	0.850667	0.900148
1	1	b'Forest'	1426	1500	0.950667	0.900148
2	2	b'HerbaceousVegetation'	1311	1500	0.874000	0.900148
3	3	b'Highway'	1082	1250	0.865600	0.900148
4	4	b'Industrial'	1176	1250	0.940800	0.900148
5	5	b'Pasture'	912	1000	0.912000	0.900148
6	6	b'PermanentCrop'	999	1250	0.799200	0.900148
7	7	b'Residential'	1406	1500	0.937333	0.900148
8	8	b'River'	1124	1250	0.899200	0.900148
9	9	b'SeaLake'	1440	1500	0.960000	0.900148

Figure 4: Summary of the accuracy for each class for the max pooling DNN model

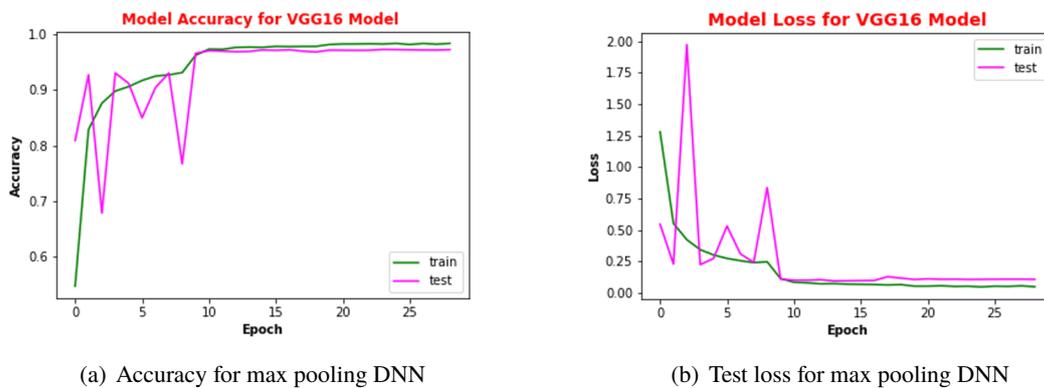


Figure 5: Score for VGG16 DNN model

index	y_true	accurate_preds	label_count	class_acc	overall_acc	
0	0	b'AnnualCrop'	1276	1500	0.850667	0.900148
1	1	b'Forest'	1426	1500	0.950667	0.900148
2	2	b'HerbaceousVegetation'	1311	1500	0.874000	0.900148
3	3	b'Highway'	1082	1250	0.865600	0.900148
4	4	b'Industrial'	1176	1250	0.940800	0.900148
5	5	b'Pasture'	912	1000	0.912000	0.900148
6	6	b'PermanentCrop'	999	1250	0.799200	0.900148
7	7	b'Residential'	1406	1500	0.937333	0.900148
8	8	b'River'	1124	1250	0.899200	0.900148
9	9	b'SeaLake'	1440	1500	0.960000	0.900148

Figure 6: Summary of the accuracy for each class for the VGG16 DNN model

Table 1: Misclassification table for each class for VGG16 DNN model

Class Number	Class Name	Mostly misclassified for class...
1	Permanent Crop	Herbaceous Vegetation
2	Residential	Industrial
3	Herbaceous Vegetation	Permanent
4	Pasture	Forest
5	River	Highway
6	Industrial	Residential
7	Forest	Pasture
8	Annual Crop	Permanent Crop
9	Highway	River
10	Sea Lake	Annual Crop/Forest

7. Conclusions

An original FOSS photogrammetric workflow to process a large dataset of geotagged high-resolution images in a single run was presented. Processing time has been optimized distributing the computationally most expensive steps on cluster nodes. A comparison of the processing time with different configurations both of the workflow commands and of the WNs was presented. Results showed that increasing the number of the jobs (thus reducing their workload) and the number of WNs, the processing time can be drastically reduced, ensuring parallel execution and faster file writing performed by each node on the File System (GPFS).

Two datasets made of 1139 and 2190 high-resolution (1.09 cm/pixel) images respectively have been processed in a relative short time, generating the orthophotomosaic (1.3 cm/pixel), the dense point cloud (~95 000 000 and ~200 000 000 densified points) and DEM (2.5 cm/pixel) of the detected areas. All these objects are useful to perform detailed hydro-geomorphological analysis of the investigated area.

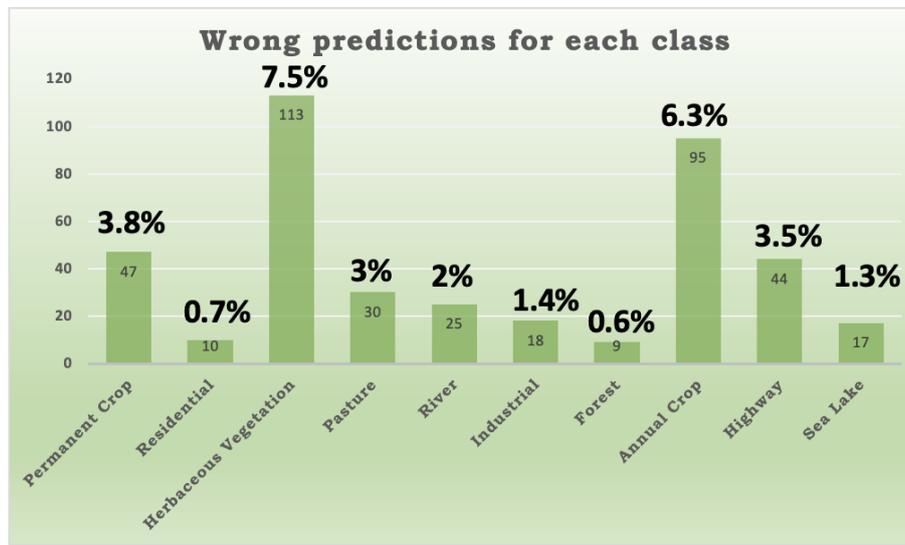


Figure 7: Wrong predictions for each class using the VGG16 DNN model

Further improvements could be expected using more dedicated slot of all worker nodes of the cluster to perform jobs, distributing them and further increasing their parallel execution. More populated dataset will be used and a re-calibration on the workflow will be considered for future studies. From the orthophotos an original dataset will be generated to perform territorial classification for next analyses.

We applied DNNs to the EuroSAT dataset (27000 satellite images) to test two different models for land cover classification. Our best model was a VGG16 Keras Model (pre-loaded weights) in which we used data augmentation technique reaching an overall accuracy of 97% and also reducing test loss to less than 10%.

8. Acknowledgments

This work has been developed in the context of the **Close to the Earth** and **RPASInAir** projects Call: “Avviso MIUR n. 1735 del 13/07/2017, AVVISO PER LA PRESENTAZIONE DI PROGETTI DI RICERCA INDUSTRIALE E SVILUPPO SPERIMENTALE NELLE 12 AREE DI SPECIALIZZAZIONE INDIVIDUATE DAL PNR 2015-2020”. Both the projects are cofunded by European Union – SIF, Ricerca e Innovazione 2014-2020.

The **Close to the Earth** project aims to build a technological prototype opening the access to the missions at Very Low Earth Orbit (a.k.a. VLEO, orbits lower than 250 km from the earth surface) and designing a low mass vehicle (LOW MASS= below 500 kg, including the propulsion system and payload) with an operating life of at least three years. Its final goal is to obtain the elements that make it possible to carry out a mission at VLEO .

The **RPASInAir** aims to enable innovative service for land monitoring through the employment of data collected by RPAS (Remotely Piloted Aircraft Systems), fulfilling the need to manage new categories of critical events directly linked to the flight of these systems, the loss of datalink between air platform and pilot station, the loss of ATM-pilot station connection and the loss of vehicle

cognitive capacities.

We would like to thank our supervisors prof. G. Maggi, prof. D. Capolongo and prof. S. Stramaglia for scientific supports and advices constantly provided for this work.

We want also to thank G. Donvito, A. Italiano, S. Nicotri and G. Vino for the technical support and implementation without which we could't get the improvements we had in our results.

The computational work has been executed on the IT resources of the ReCaS-Bari data center, which have been made available by two projects financed by the MIUR (Italian Ministry for Education, University and Research) in the "PON Ricerca e Competitività 20072013" Program: ReCaS (Azione I - Interventi di rafforzamento strutturale, PONa3_00052, Avviso 254/Ric) and PRISMA (Asse II - Sostegno all'innovazione, PON04a2_A).

The Aerospace Technology District (DTA Scarl) is the leading proponent for both the Close to the Earth and the RPASInAir projects.

The photogrammetry study activities described are part of the objectives of a PhD project in Geosciences "Application of UAV system and SfM techniques to assess the hydrogeological hazard of a fluvial system" at the Department of Earth and Environmental Sciences of Bari and of the "RPASinAir - Integrazione dei Sistemi Aeromobili a Pilotaggio Remoto nello spazio aereo non segregato per servizi", PON ricerca e innovazione 2014-2020.

References

- [1] Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., Reynolds, J.M., 2012. *Structure-from-Motion photogrammetry: A low-cost, effective tool for geoscience applications*. *Geomorphology*, Vol 179: 300–314. <https://doi.org/10.1016/j.geomorph.2012.08.021>
- [2] Martinez-Rubi, O., Nex, F., Pierrot-Deseilligny, M. et al., 2017, *Improving FOSS photogrammetric workflows for processing large image datasets*. *Open geospatial data, softw. stand.* 2, Article number: 12. <https://doi.org/10.1186/s40965-017-0024-5>
- [3] Vandana S., 2020, *emphLand Cover Classification using Machine Learning Techniques - A Survey*, *International Journal of Engineering and Technical Research* V9(06), DOI - 10.17577/IJERTV9IS060881
- [4] Ullman S., 1979, *The interpretation of structure from motion*, *Proceedings of the Royal Society of London, Series B Biological Sciences*, 203(1153), 405-426.
- [5] Snavely, N., Seitz, S. M., Szeliski, R., 2008. *Modeling the world from internet photo collections*, *International journal of computer vision*, 80(2), 189-210.
- [6] Wang, X., Rottensteiner, F., Heipke, C., 2019. *Structure from motion for ordered and unordered image sets based on random kd forests and global pose estimation*, *ISPRS Journal of Photogrammetry and Remote Sensing*, 147, 19-41: <https://doi.org/10.1016/j.isprsjprs.2018.11.009>.

- [7] Lowe, D. G., 1999. *Object recognition from local scale-invariant feature*, Proceedings of the seventh IEEE international conference on computer vision, Vol. 2, pp. 1150-1157. doi: 10.1109/ICCV.1999.790410
- [8] Lowe, D.G., 2004. *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision 60, 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [9] Erickson R.A., Fienen M.N., McCalla S.G., Weiser, E.L., Bower, M.L., Knudson, J.M., Thain, G. 2018. *Wrangling distributed computing for high-throughput environmental science: An introduction to HTCondor*. *PLoS computational biology*, 14(10): e1006468. <https://doi.org/10.1371/journal.pcbi.1006468>.
- [10] P. Helber, B. Bischke, A. Dengel, D. Borth, 2019, *Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.
- [11] K. Simonyan, A. Zisserman, 2015, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, International Conference on Learning Representations, abs/1409.1556.