The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W1-2022
Free and Open Source Software for Geospatial (FOSS4G) 2022 – Academic Track, 22–28 August 2022, Florence, Italy

DEVELOPMENT OF A GRAPHICAL USER INTERFACE TO SUPPORT THE SEMI-AUTOMATIC SEMANTIC SEGMENTATION OF UAS-IMAGES

A. Masiero[1,*], I. Cortesi[1], G. Tucci[1]

[1] Dept. of Civil and Environmental Engineering, University of Florence,
via di Santa Marta 3, Florence 50139, Italy - (andrea.masiero, irene.cortesi, grazia.tucci)@unifi.it

KEY WORDS: Semantic segmentation, Graphical User Interface, Machine Learning, Multispectral Imagery

ABSTRACT:

The development of remote sensing techniques dramatically improved the human knowledge of natural phenomena and the real time monitoring and interpretation of the events happening in the environment. The recently developed terrestrial, aerial and satellite remote sensors caused the availability of huge amount of data. The large size of such data is leading the research community to the search for efficient methods for real time information extraction, and, more in general, understanding the collected data. Nowadays, this is typically done by means of artificial intelligence-based methods, and, more specifically, usually by means of machine learning tools. Focusing on semantic segmentation, which is clearly related to a proper interpretation of the acquired remote sensing data, supervised machine learning is often used: it is based on the availability of a set of ground truth labeled data, which are used in order to properly train a machine learning classifier. Despite the latter, after a proper training phase, usually allows to obtain quite effective segmentation results, the ground truth labeled data production is usually a very laborious and time consuming task, performed by a human operator. Motivated by the latter consideration, this work aims at introducing a graphical interface developed in order to support semi-automatic semantic segmentation of images acquired by a UAS. Certain of the potentialities of the proposed graphical are shown in the specific case of plastic litter detection in multi-spectral images.

1. INTRODUCTION

Image semantic segmentation focuses on the problem of properly separating and classifying different regions in an image depending on their specific meaning or use, e.g. belonging to the same object (Pal and Pal, 1993). It is worth to notice that in general segmentation is a ill posed problem: it is not possible to provide a unique solution to such problem, different solutions can typically be acceptable, depending on the segmentation criterion which is applied. Nevertheless, regularization techniques are typically used to reduce the issues related to ill posedness, hence ensuring the computability of a unique solution (Marroquin et al., 1987). In the case of semantic segmentation, ill posedness is also reduced by the specific data and object interpretation that shall be included in the semantic part of the data.

It is also worth to notice that image semantic segmentation tools can be useful in many several applications (Galarreta et al., 2015, Lianos et al., 2018), related both to the interpretation of images themselves, but also of other entities related to such images. The latter is for instance the case of a point cloud, whose objects and areas are also described by some images. In this case, a proper image semantic segmentation could be back projected from the images to the point cloud, in such a way to exploit such process to properly segment the point cloud itself (Pellis et al., 2022, Stathopoulou and Remondino, 2019).

Automatic image semantic segmentation is a quite challenging problem that nowadays is usually handled by taking advantage of the use of artificial intelligence tools, such as deep learning based neural networks (Simonyan and Zisserman, 2014, Long et al., 2015, Noh et al., 2015).

The availability of reliable image segmentation datasets plays a key role in the training phase of any artificial intelligence and machine learning tool based on the image analysis: indeed, despite artificial intelligence tools can currently be considered as the state of the art method in terms of recognition and segmentation ability, they do require a huge size learning dataset in order to ensure reliable segmentation results.

The developed graphical user interface aims at supporting the semi-automatic semantic segmentation of images, hence easing and speeding up the generation of a ground truth segmentation database. Then, such database can be of remarkable importance for properly training any machine or deep learning based classification and segmentation method.

Despite the development of the proposed graphical user interface has been originally motivated by the need of easing the process of producing a ground truth segmentation and classification of plastic objects in maritime and fluvial environments, within a project aiming at reducing plastic pollution in rivers (Cortesi et al., 2021, Cortesi et al., 2022), the developed tool can actually be used in contexts that are more general.

Indeed, the interface supports in particular two types of quite specific operations: 1) segmenting and identifying objects in a single image, 2) exporting previously obtained results in new images, while also enabling the computation of certain related parameters (e.g. navigation related, such as tracking the same object over different data frames). Different types of images are supported: standard RGB, and multispectral images (already available as TIFF (Tagged Image File Format) images).

For what concerns the semantic segmentation of a single image, several alternative segmentation options are supported, starting from manual and going to semi-automatic segmentation methods. First, the manual segmentation of the objects is ensured by means of properly inserted polylines. Then, intensity based and graph based methods are implemented as well.

* Corresponding author

On the semi-automatic side, two tools are being developed: a) a machine learning based method, exploiting few click choices by the user (implementing a rationale similar to that in (Majumder et al., 2020), i.e. aiming at minimize the user input), b) when images are periodically acquired by an Unmanned Aerial System (UAS), at quite high frequencies, two successive frames are expected to be not that different from each other. Consequently, the system aims at determining the camera motion between different frames, and using machine learning tools to properly extend and generalize the results in the previous image to those of the new one.

The latter method opens to a wider scenario, where some more information may come by the availability of consecutive frames. In particular, such additional information that could be determined by properly analyzing consecutive frames could be used to: assess and track the UAS movements while acquiring the video frames, increase the automation in the segmentation and classification process of an object. Despite not fully implemented yet, it is expected to be integrated in the interface in the interface versions.

The interface, implemented as a Matlab app, which will be downloadable from the website of the GeCo (Geomatics and Conservation) laboratory of the University of Florence (Italy) https://www.geomaticaeconservazione.it/downloads/, will be described in the next Sections, focusing on certain examples in the plastic litter detection in fluvial environment case study, which is shortly introduced in the next Section just to provide a contextualization of the considered examples.

2. CASE STUDY USED TO TEST THE INTERFACE

Plastic litter detection in litter environments could play a remarkable role to reduce the amount of litter in seas, and, more in general, in the natural environment. To such aim, a proper machine learning approach, based on (Belgiu and Drăguţ, 2016), has been previously developed (Cortesi et al., 2022), deploying multi-spectral imagery, collected by the MAIA-S2 camera, mounted on a DJI Matrice 300 UAS (Figure 1). In the examples considered in this work, the UAS flew over a portion of the Arno river (Prulli, Reggello, Florence, Italy), shown as a rectangle in Figure 2. A circle shows where a set of plastic samples, introduced in the river during the test, were anchored, in order to be recollected after the data collection end.

Table 1 shows the bands of the MAIA-S2 camera (SAL Engineering and EOPTIS, 2018) (the same bands of the Sentinel-2 Satellite).

Band	Start WL [nm]	Stop WL [nm]	Color
S1	433	453	Violet
S2	457.5	522.5	Blue
S3	542.5	577.5	Green
S4	650	680	Red
S5	697.5	712.5	Red Edge 1
S6	732.5	747.5	Red Edge 2
S7	773	793	NIR 1
S8	784.5	899.5	NIR 2
S9	855	875	NIR 3

Table 1. Wavelength (WL) intervals of MAIA S-2 bands.

Automatic radiometric and geometric correction and coregistration of the bands were computed by means of the MAIA image-processing software before analyzing the images with the proposed interface.

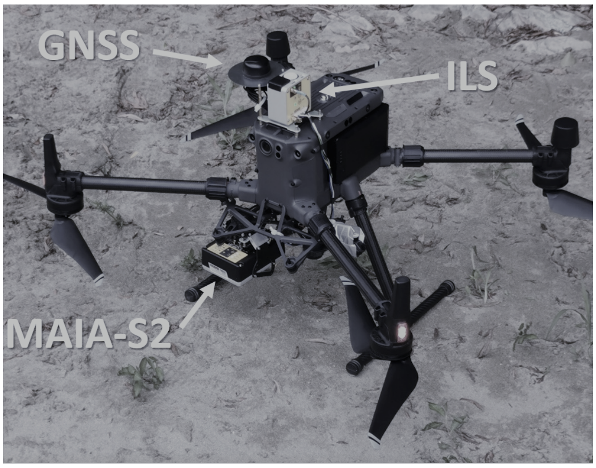


Figure 1. MAIA-S2 multispectral camera and Irradiance Light Sensor mounted on DJI Matrice 300 UAS.

Figure 2. Study area: Arno river in the locality of Prulli (Italy).

3. GRAPHICAL USER INTERFACE

This Section aims at providing an overview of the proposed graphical interface, whose general look is shown in Figure 3 in View mode.

Three working modes can be identified: View, Segmentation, Classification, depending on the aim, as described in the following subsections.

3.1 View mode

View mode is just dedicated to properly viewing the available imagery, the segmented objects, and their classification, if available.

As a first step, imagery folder, and processed data (segmented and/or classified) folders, if available, should be set (on the top left of the window, Figure 3).

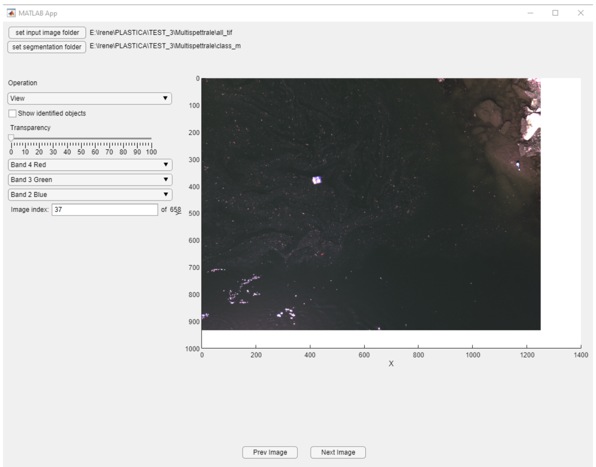


Figure 3. Graphical user interface window in View mode.

Standard operations, such as navigating among the available imagery, zooming and shifting the visible part of each image have been implemented.

When certain objects have already been identified (or such information has been previously stored), they can be highlighted (e.g. red contour as in Figure 4).

The interface maps three bands of the visualized image on the RGB channels, to make them properly visible by the user. The user is allowed to select among the available image bands, if more than three, in order to improve the view of certain objects, for instance to make them more distinguishable.

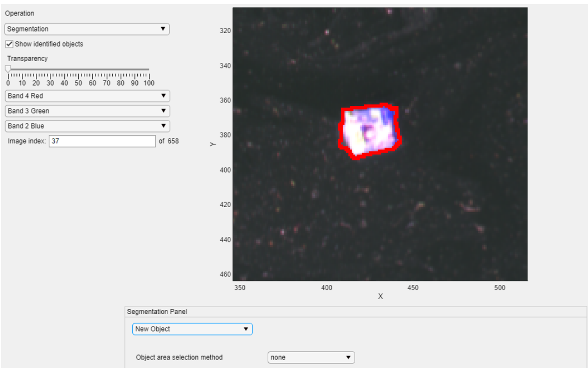


Figure 4. Example of object visualization and segmentation.

The transparency of the object contours, if their visibility is set to on, can also be changed (Figure 5).

Directly jumping to a specific image of the imagery dataset folder can also be done, by inputting its progressive index in the image index box.

3.2 Segmentation mode

Segmentation mode probably represents the core of the interface: once selected, the segmentation panel is activated, as shown in Figure 6 (overall view of the interface in segmentation mode) and Figure 7.

The segmentation panels allows either to insert a new object or to edit an existing one, as shown in Figure 8. Navigation among the detected objects is allowed in the latter case.

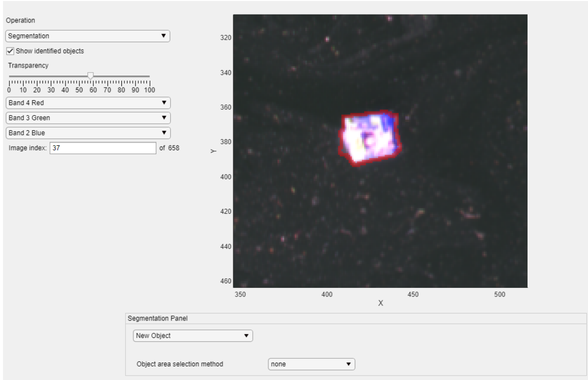


Figure 5. Example of changing transparency.

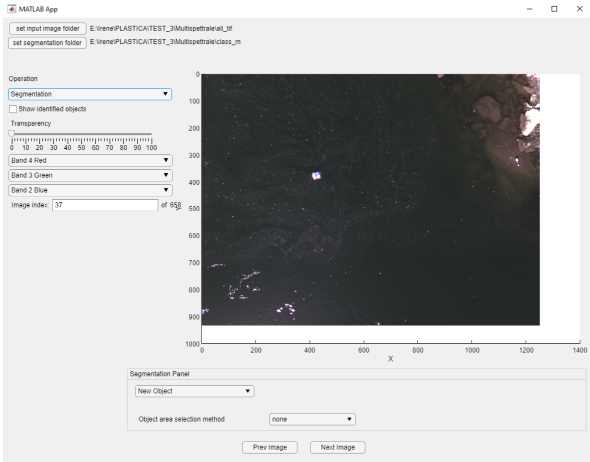


Figure 6. Graphical user interface window in Segmentation mode.

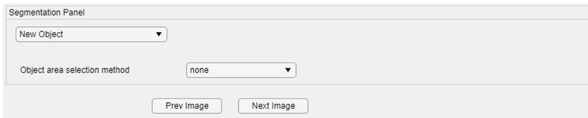


Figure 7. Segmentation panel.

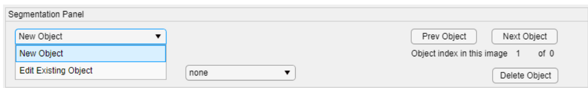


Figure 8. Segmentation panel: new object/existing object selection.

Independently of the new/existing object choice, the tools available for the object area selection (Figure 9) are quite similar. Examples of the use of certain of such tools will be shown in the segmentation of the case shown in Figure 10.

To be more specific, the following alternative tools are available for selecting an object area:

- Polygonal selection
- Pixel selection
- Freehand area selection
- 1-click segmentation

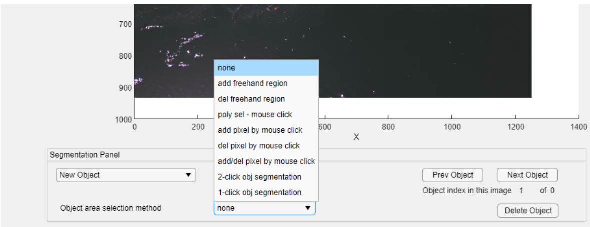


Figure 9. Segmentation panel: object area selection tools.

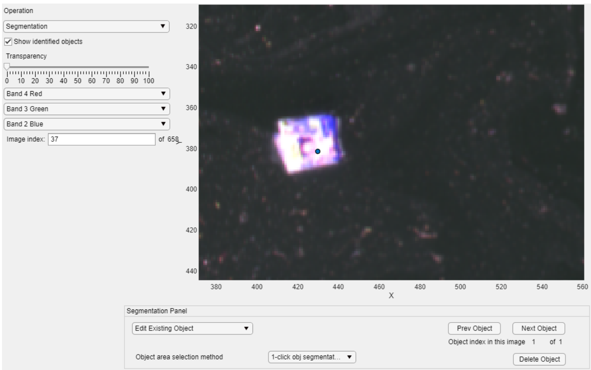


Figure 10. Example of plastic object to be segmented.

- 2-click segmentation

The meaning of the first options is rather intuitive. Polygonal selection allows to to select the vertices of a polygonal area, corresponding to the considered object. Pixel selection can be used to add (or remove) any pixel set (Figure 11), each of them uniquely identified with a mouse click, to an object.

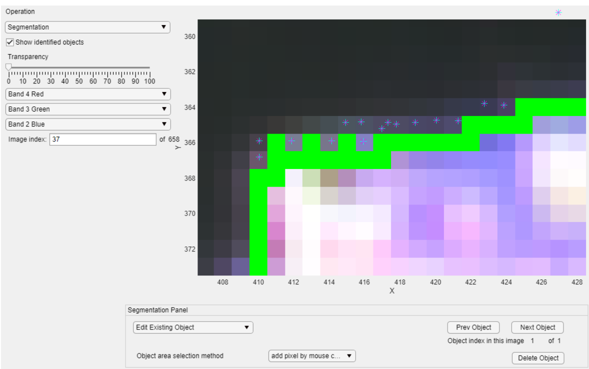
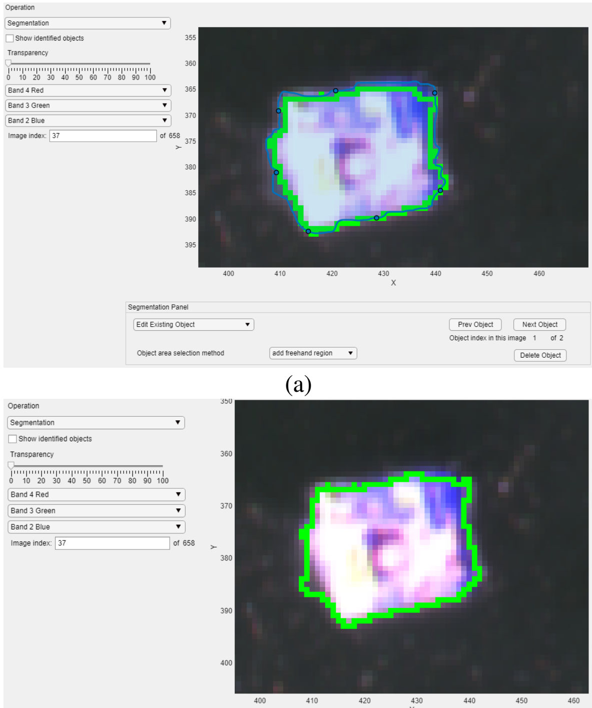


Figure 11. Pixel by pixel selection.

Instead, the freehand tool can be used to add (or remove) an area to an object by a freehand selection of such area, which is then converted into a dense set of point coordinates, describing the borders of the selected area. This tool may also be considered for improving the pixel selection on the borders of an already approximately determined object region (Figure 12).

Differently from the previous methods, the 1-click and 2-click segmentation options aim at exploiting the peculiar spectral characteristics of an object in order to reduce the human operator effort in its identification. The rationale is that if the object to be identified is characterized by a different spectral signature with respect to its neighborhood in the considered image, then an approximation of such spectral signature should enable the detection of its borders with a reasonable accuracy.



(a)

(b)

Figure 12. (a) Example of freehand area object selection tool, and of the result (b).

In the 1-click segmentation option, the operator is required to select with a mouse click a pixel inside of the object to be segmented. Once that such mouse click is done, the values of all the bands of the corresponding image pixel are extracted, and assumed to be similar to those characteristics of the object. Then, Otsu's method (Otsu, 1975, Otsu, 1979) is applied on all the image bands in a rectangular region around the selected pixel in order to determine initial classification hypotheses for each pixel in the region.

To be more specific, initially the content in each band is separately analyzed. Since the object of interest is assumed to be characterized by a different spectral signature than its neighborhood area, then the values of the pixels inside the object should be distinguishable on certain bands. The Otsu's method is used to automatically determining a proper threshold for detecting the object pixels, where such threshold is determined in such a way to optimize the discrimination between two classes of pixels in the considered region. Among the two, the object class is determined as the one determined with a comparison with the threshold. Finally, a decision on each pixel in the area is made based on the majority of the class votes from the image bands on such pixel.

Figure 13 shows an example of the results obtained with the 1-click segmentation option.

The 2-click segmentation option works similarly to the 1-click one: the main difference in this case is that the user has to select both a pixel inside the object and one outside of the object. Then, the Otsu's method is applied to each image band as previously explained, however, in this case if the two selected pixels fall in the same class identified by the Otsu's method such band is excluded from the class voting procedure, i.e. such band is expected to not provide reliable/usable information for distinguishing the object pixels.

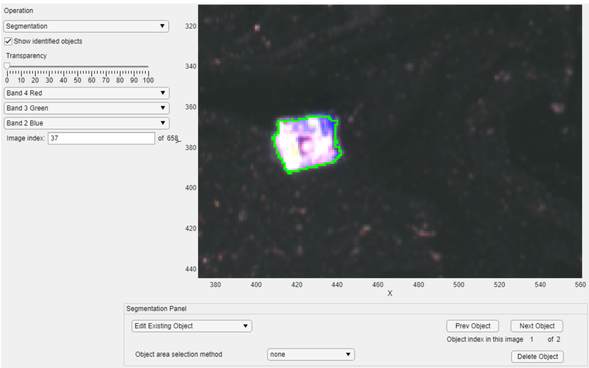


Figure 13. Example of 1-click segmentation.

3.3 Classification mode

The classification mode aims at classify each previously segmented object. If the available object folder has been set, then the software automatically allows to show in the bottom left the different object alternative options, otherwise the objects are simply identified with different colors, associated to the corresponding object identification index. If the list of objects is available, the object selection is conveniently done by using a dropdown menu (Figure 14).

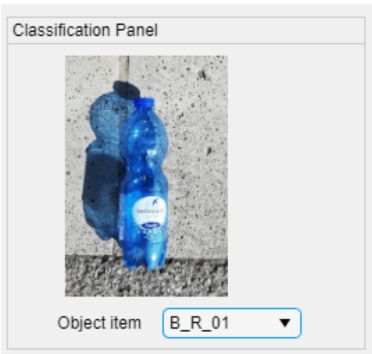


Figure 14. Classification panel.

4. CONCLUSIONS

This paper presents a graphical user interface for easing the segmentation and classification of objects, in particular on images collected from UASs. The interface will be freely available for download from the website of the GeCo (Geomatics and Conservation) laboratory of the University of Florence (Italy) https://www.geomaticaeconservazione.it/downloads/ .

It is worth to notice that the interface has been implemented as a Matlab app: despite Matlab is a commercial programming software it is very well known, used and available to many users in the World, hence the availability of such interface is expected to be useful even in this case.

Full autonomous image semantic segmentation would clearly be of interest, however its development is to be quite challenging, hence, overall, the developed graphical user interface is expected to be useful to support the semi-automatic identification of objects, where the semi-automatic tools available in the interface should play a remarkable role in easing the manual object identification. The interface has been tested on the case study of plastic litter detection in a fluvial environment, showing a reasonable ability in semi-automatically detecting plastics in multi-spectral images.

In addition to the already available semi-automatic segmentation tools, based on the use of the Otsu's method, the availability of a tool for automatically extending the segmentation results in successive images is foreseen in the future versions of the interface. The availability of such tool shall increase the segmentation procedure automation level.

The implementation of different segmentation methods, such as level set-based (Sethian, 1999, Osher and Fedkiw, 2003, Masiero et al., 2015), is also foreseen in the future versions of the interface.

REFERENCES

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114, 24–31.

Cortesi, I., Masiero, A., De Giglio, M., Tucci, G., Dubbini, M., 2021. Random Forest-Based River Plastic Detection with a Handheld Multispectral Camera. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 9–14.

Cortesi, I., Masiero, A., Tucci, G., Topouzelis, K., 2022. UAV-based river plastic detection with a multispectral camera. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2022, 855–861.

Galarreta, J. F., Kerle, N., Gerke, M., 2015. UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning. *Natural Hazards and Earth System Sciences*, 15(6), 1087.

Lianos, K.-N., Schönberger, J. L., Pollefeys, M., Sattler, T., 2018. Vso: Visual semantic odometry. *Proceedings of the European Conference on Computer Vision (ECCV)*.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.

Majumder, S., Khurana, A., Rai, A., Yao, A., 2020. Multi-stage fusion for one-click segmentation. *DAGM German Conference on Pattern Recognition*, Springer, 174–187.

Marroquin, J., Mitter, S., Poggio, T., 1987. Probabilistic solution of ill-posed problems in computational vision. *Journal of the american statistical association*, 82(397), 76–89.

Masiero, A., Guarnieri, A., Pirotti, F., Vettore, A., 2015. Semi-Automated Detection of Surface Degradation on Bridges Based on a Level Set Method. *ISPRS - International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3), 15–21.

Noh, H., Hong, S., Han, B., 2015. Learning deconvolution network for semantic segmentation. *Proceedings of the IEEE international conference on computer vision*, 1520–1528.

Osher, S., Fedkiw, R., 2003. *Level set methods and dynamic implicit surfaces*. 153, Springer Science & Business Media.

Otsu, N., 1975. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296), 23–27.

Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(1), 62-66.

Pal, N. R., Pal, S. K., 1993. A review on image segmentation techniques. *Pattern recognition*, 26(9), 1277–1294.

Pellis, E., Murtiyoso, A., Masiero, A., Tucci, G., Betti, M., Grussenmeyer, P., 2022. 2D to 3D label propagation for the semantic segmentation of heritage building point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2022, 861–867.

SAL Engineering and EOPTIS, 2018. Maia, the multispectral camera. (12 january 2022). https://www.spectralcam.com/maia-tech/.

Sethian, J., 1999. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Stathopoulou, E., Remondino, F., 2019. Semantic photogrammetry-boosting image-based 3d reconstruction with semantic labelling. *8th Intl. Workshop 3D-ARCH "3D Virtual Reconstruction and Visualization of Complex Architectures"*, 42number 2/W9, 685–690.