Evaluation of Deep Learning models applied to remotely sensed imagery for object detection and segmentation tasks in territorial monitoring.

AIT Series

Trends in earth observation

Volume 4

Smart Earth Observation for a Sustainable Future

Edited by

Associazione Italiana di Telerilevamento (AIT)

Smart Earth Observation for a Sustainable Future

Edited by

Associazione Italiana di Telerilevamento (AIT)

AIT Series: Trends in earth observation

The guest editors of this volume are:

Enrico Borgogno Mondino (AIT President)
Mirco Boschetti
Claudia Giardino
Lucia Maria Laurenza

All contributions published in the Volume Smart Earth Observation for a Sustainable Future were subject to blind peer review by independent reviewers. The revision of the contributions was meticulously carried out by the guest editors and the following researchers, to whom we extend our sincere gratitude

Claudia Arcidiacono
Anna Balenzano
Cristiana Bassani
Piero Boccardo
Luigi Boschetti
Mariano Bresciani
Sandro Brivio
Elena Candigliota
Francesco Casu
Gherardo Chirici
Roberto Colombo
Giacomo De Carolis
Maria Antonietta Dessena
Biagio Di Mauro
Francesco Fava
Federico Filipponi
Giacomo Fontanelli
Paolo Gamba
Marco Gianinetto
Fabio Giulio-Tonolo
Antonino Maltese
Maria Teresa Melis
Giuseppe Modica
Fernando Monterroso
Simonetta Paloscia
Monica Pepe
Pietro Picuno
Monica Pinardi
Serena Sapio
Giuseppe Satalino
Pasquale Striano
Daniela Stroppiana
Paolo Tagliolato
Deodato Tapete
Emiliana Valentini
Paolo Villa
Maria Virelli
Cristina Vittucci

The volume editing and typsetting was carried out by Patrizia Rossi.

Volume 4 - Published in June 2026
Edited by Associazione Italiana di Telerilevamento (AIT)
Published on behalf of the Associazione Italiana di Telerilevamento (AIT)
Via Lucca 50
50142 Firenze, Italy

ISSN 2612-7911
ISBN 978-88-944687-3-1
DOI: 10.978.88944687/31

AIT Series "Trends in earth Observation" is indexed in Scopus and Web of Science

All contributions published in the Volume "Smart Earth Observation for a Sustainable Future" were subject to blind peer review from independent reviewers.
Publication Ethics and Publication Malpractice Statement
Editors, Reviewers and Authors agreed with Publication Ethics and Publication Malpractice Statement

Preface

Earth Observation is a new and effective way to mean image-based remote sensing. Something has drastically changed since the first Landsat missions, and now we are experiencing a new era, where the ordinary satellite missions are sharing the space and the market with the various LEO (Low Earth Orbit), aerial and drone platforms. This new situation would require a significant reinforcement of users' skills and, consequently, a higher level of data and data processing consciousness. This reasonable expectation has been paradoxically disattended, and a new type of users seems to be, now, on. They appear to be well skilled in using tools, but poorly conscious about the inner meaning of data where the information they are looking for, comes from. The main responsible of this unexpected effect has been the huge technological advance, especially in the ICT context (computational power has enormously increased in the last decades), that has made possible the development of processing tools that can be applied globally, providing the users ready-to-use products and services accepted as they come. Scientific community has therefore to recover a central role, not only in proposing applications for those products/services, but also, once they have been eventually found, defining guidelines and rules for a proper use of data/methods/tools.

AIT strongly believe that the greatest and most urgent new challenge is a proper reverse engineering of the tools, products and services that are continuously exposed on the web. This must be intended for (i) recognizing the level of explainability of the processing chain, (ii) making explicit (if present) the calibration/validation metrics associated to the products/services in terms of both meaning and recurrency, (iii) eventually proposing (if lacking) guidelines and operational standards. Legal actions should be ignited to support this process, aiming at recognizing accredited controllers able to ensure an adequate level of **officiality** to data, products and services, to guarantee robustness and comparability of deductions. This need is, for example, at the basis of making deductions from EO data to be assumed as probatory within legal concerns (e.g. environmental crimes). This is furtherly true today, when we are experiencing the introduction of AI in many of Geomatics-related topics. Robustness and reliability of deductions from EO products/services, beyond whatever commercial exploitation, has still to be demonstrated and supported. **Each new map provided** (and EO is continuously producing maps wherever around the world) should undergo an adequate, recognized and official **validation procedure certified by accredited subjects**. This is not, now. Whatever standard procedure (for calibration and validation) requires the adoption of well-defined criteria and tools/procedures aimed at testing their satisfaction, that, once defined, should remain steady for a reasonable time. Such requirement, if considered in the framework of the present technology transfer, suggest to change the paradigm: one must stop running behind the latest new tool/algorithm and embrace slowness of science as the only reasonable way to reliability. Technological advance cannot, obviously, be stopped. Its application, differently, can. Proper time steps, consistent with the definition and application of standards for products/services utilizations by final users, have to be defined and respected (e.g. standards refer to the state-of-the-art of time t and are updated every 3-5 years). In the meantime, the scientific community has to monitor, and possibly drive, the ongoing technological advance preparing new standards/procedures for the next updating step. It is AIT conviction that scientific associations can be the place where this process can be ignited and eventually managed, having no direct political or commercial involvements in the technology transfer process. And academy (Bachelor, Master and PhD) the place where a new generation of experts should be properly shaped.

It is within this vision that we have filled this volume, longing for a mosaic of works showing different points of views (from scientists, institutions, service providers, professionals, etc.), that could make it profitable for the future advance of Earth Observation.

AIT President (2023-2026)
Enrico Borgogno Mondino

Contents

Introduction

Smart Earth Observation for a Sustainable Future

The IV volume of the TEO collection presents a curated selection of studies originally presented at the XII AIT International Conference, held in Milan from November 12 to 14, 2025.

While the conference brought together the Earth Observation (EO) community to discuss how innovation in remote sensing, data science, and environmental monitoring can contribute to sustainability and societal resilience, this specific volume includes **38 peer-reviewed papers organized into 5 dedicated thematic sections**:

- Urban Applications: Focuses on urban monitoring, exploring Local Climate Zones , Urban Heat Islands, surface temperature simulations, and solar energy deployment.
- Environment, Climate Risk: Addresses hazards and climate indicators, including post-fire erosion, multi-temporal SAR flood mapping, glacier dynamics and regional heat stress.
- Agriculture & Forestry: demonstrates multi-source EO applications for precision farming, crop traits/ yield estimation, tillage mapping and post-fire forest recovery
- Evapotranspiration, Soil Moisture, and Irrigation: Advances high-resolution soil water content retrieval via multi-band SAR, evapotranspiration modelling and Deep Learning for agricultural drought monitoring and irrigation management.
- Algorithm, Tool, and New Mission: showcases new space programs (SBG-TIR, FLEX, EarthCARE) alongside innovations in PRISMA pre-processing tools and GeoAI deep learning frameworks for territorial monitoring.

The central theme of this volume, “Smart Earth Observation for a Sustainable Future,” reflects the growing importance of advanced remote sensing technologies in addressing environmental and societal challenges. The scientific contributions collected in this volume demonstrate significant progress in the integration of multi-sensor data, emerging spaceborne EO missions, and advanced computational methodologies - particularly GeoAI - to enhance environmental monitoring and risk mitigation.

We warmly thank all authors and reviewers for their high-quality contributions, making this TEO volume a valuable document to share ideas, build collaborations, and advance the role of Earth Observation in supporting a more sustainable future.

Urban application

Evaluation of Deep Learning models applied to remotely sensed imagery for object detection and segmentation tasks in territorial monitoring.

A. Demartis* and F. Giulio Tonolo

Politecnico di Torino, Department of Architecture and Design (DAD), Laboratory of Geomatics for Cultural Heritage (G4CH), Viale Pier Andrea Mattioli 39, 10125, Torino (TO)

KEY WORDS: Deep Learning, GeoAI, Satellite Imagery, RPAS Imagery, Wildfires, Floods

ABSTRACT:

Remote sensing applications such as agro-forestry monitoring, land classification, environmental surveillance, and border supervision increasingly rely on automated solutions to reduce manual interpretation. Artificial intelligence (AI), particularly deep learning (DL) applied to image analysis, represents the state of the art and has fostered the development of GeoAI, integrating geospatial analysis and AI to extract information from Earth Observation (EO) data. Satellite remote sensing is a primary EO data source due to its broad spatial coverage, high temporal resolution, and availability of open-access datasets such as the Copernicus Sentinel missions. Applications requiring very high spatial detail, including the detection of objects related to environmental crimes, benefit from data acquired by Remotely Piloted Aircraft Systems (RPASs), which provide flexible, centimeter-level resolution imagery. Despite the growing availability of EO data, DL workflows remain complex for non-experts. To address this limitation, pre-trained DL models and commercial software with graphical user interfaces have been developed. In object detection (OD), these include traditional class-specific approaches and prompt-based models (PBMs) that enable zero-shot detection through textual prompts, eliminating task-specific training. This study evaluates both categories, focusing on PBMs, using RPAS imagery for OD within the EU Horizon Europe EMERITUS project and Sentinel-2 data for land-cover classification.

1. INTRODUCTION

Remote Sensing based on satellite and aerial imagery plays a fundamental role in natural disaster monitoring, environmental surveillance, and land management. The increasing availability of multi-spectral and high-resolution data has significantly enhanced the capability to detect, classify, and monitor phenomena across spatial and temporal scales. In the context of GeoAI, DL has become the state of the art for image-based analysis. Despite the proven effectiveness of DL methods, their practical implementation usually remains challenging, especially in operationally oriented frameworks, such as the ones involving local law enforcements. Training, fine-tuning, and inferencing models typically require annotated datasets, computational resources, and coding expertise. To lower this barrier, commercial Geographic Information System platforms have integrated a DL interface through Graphical User Interfaces (GUIs), enabling inference directly within the software environment. These solutions allow end users to train, fine-tune, or inference DL pre-trained models without specific coding skills. In the context of the available pre-trained models, recent advancements in tasks such as OD have introduced PBMs that combine vision architectures with Large Language Models, enabling zero-shot detection through textual prompts. Despite the growing availability of pre-trained and prompt-based models in commercial GIS environments, systematic evaluations of their performance in operational EO scenarios remain limited. This study presents an exploratory analytical evaluation of traditional and prompt-based pre-trained DL models for object detection and semantic segmentation, using exclusively the inference capabilities available within the ESRI ArcGIS Pro DL environment. The research can be divided into two main parts: i) an analysis based on PBMs conducted on RPASs imagery acquired within the EU Horizon Europe "EMERITUS" project for environmental crime detection, and ii) an analysis based on available pre-trained DL models trained with Sentinel-2 imagery, including flooded areas and wildfires burn scars segmentation. The objective is to assess the effectiveness and practical applicability of these models in real-world remote sensing workflows.

* corresponding author

2. MAIN BODY

2.1 Materials

This OD and semantic segmentation tests were conducted in ArcGIS Pro (versions 3.4.2 to 3.5.3) using a workstation equipped with an Intel Core i9-14900KF CPU @ 3.20 GHz, 64 GB RAM, an NVIDIA GeForce RTX 4070 GPU with 12 GB VRAM, and Windows 11 Pro. The RPAS imagery used as input test data was collected within the EMERITUS Project, and consists of three orthoimages with centimetric resolution acquired in a semi-controlled environment in the period between October 2023 and June 2025. This environment, a former NATO base in Italy, was used to set up ad-hoc scenarios of illegal dumping by scattering waste objects, mainly tyres and barrels. The multispectral satellite imagery used as input test data was collected from the website of the Copernicus Mission Sentinel-2, selecting data acquired during flood and wildfire events for emergency related models, while an image of the Piemonte region was used for the other analyses. The data for the emergency related models were sensed in May and March 2025 for the floods of Emilia Romagna and Tuscany respectively, while data related to the wildfires of Sardinia and Pantelleria were sensed in the Augusts of 2020 and 2022 respectively. Regarding the Piemonte region, the sensing date is the 27th of October. The pre-trained model selection was performed through an exploration of the Living Atlas, Esri's platform for imagery, maps, and pre-trained DL models with different purposes.

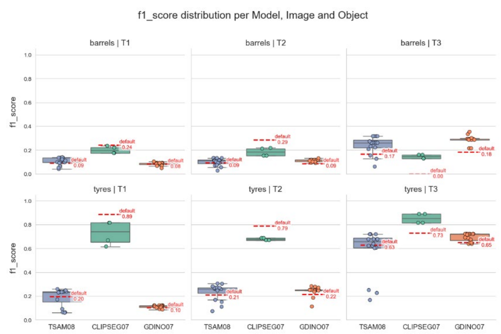
For the RPAS-based analysis, Prompt-Based Models (PBMs) were selected, namely models that should be able to detect any object in the image by means of a user-defined text prompt. This is possible thanks to the usage of a Text Encoder (TE), an architecture that, through tokenization, can convert words into features. This type of encoder is usually trained using large text datasets, which allow them to retrace the features of the target object from text. The input image, instead, goes through a Feature Encoder (FE), that extracts feature representations from the image. Subsequently, the features extracted by the TE and the FE are projected into a shared embedding space where, if overlap is found, an output of the model is generated. The PBMs selected for this research are TextSAM (Spiegler *et al.*, 2025), GroundingDINO (Liu *et al.*, 2024), and CLIPSeg (Lüddecke e Ecker, 2021), which perform Instance Segmentation, Object Detection, and Semantic Segmentation respectively. These models were chosen since their output can be interpreted within an object detection framework. Ground truth (GT) labels for the RPAS analyses were manually digitized from the orthoimagery for each object category under investigation.For the RPAS analyses, model performance was assessed using Precision, Recall, and F1-score metrics, given the object detection-oriented nature of the task. For the satellite-based analysis, four models compatible with Sentinel-2 imagery were selected. These include a Corine Land Cover (CLC) classification model and a Human Settlements segmentation model, representing traditional remote sensing applications. In addition, two fine-tuned versions of the Prithvi Foundation Model (Jain, 2023) were adopted for flood and burn scar segmentation tasks. For the satellite imagery applications, the CLC 2021 dataset (European Environment Agency, 2024) was used as GT for both the CLC classification model and the human settlements segmentation model, considering either the complete set of classes or a subset restricted to built-up areas, depending on the model configuration. For the evaluation of the fine-tuned Prithvi models, two flood events and two wildfire events were selected from the Copernicus Emergency Management Service archive. For the satellite imagery analyses, which correspond to pixel-level classification tasks, performance was evaluated using a confusion-matrix-based approach, computing Producer and User Accuracy.

2.2 Methods

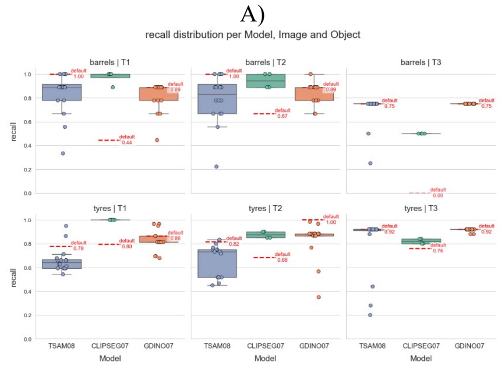
All the tested models could be inferenced with different tools in ArcGIS Pro through GUIs. Specifically, TextSAM and GroundingDINO could be inferenced using the “Detect Object using Deep Learning” tool, while all the other models could be inferenced using the “Classify Pixel using Deep Learning” tool.
As previously explained, the PBMs require as input, along with the orthoimage, a user defined text prompt. For the sake of this research, the prompt was kept as simple as possible, specifically “tyres” or “barrels”. Along with the text prompt, it is possible to change some other hyperparameters. This part of the research focused on understanding if, changing the available hyperparameters, these models could be used for unsupervised monitoring applications. To do so, numerous inferences of the models with different hyperparametrization were launched across all orthoimages, obtaining around 220 outputs from the three PBMs. The target hyperparameters for the analysis were padding, tile size, and batch size. This last hyperparameter was chosen to investigate the interaction with padding, as ArcGIS Pro processes tiles in a square group, applying padding to the outer boundary of the groups. Consequently, identical padding values may lead to different effective contexts depending on the batch size. The models were also inferenced with default hyperparameters for comparison. After the inferences, all the outputs underwent a post-processing workflow, specifically TextSAM and GroundingDINO masks underwent a Non-Maximum-Suppression, a Dissolve and an Area Filter (0.2 m^2), while CLIPSeg rasters, after conversion into polygon underwent the same Area Filter. For satellite applications only the default settings of the related pre-trained models were considered. To evaluate the outputs the “Compute accuracy for object detection” tool was used for PBMs paired with previously described manually extracted labels. A deliberately low IoU=0.1 was adopted due to the different natures of the masks generated by the outputs. This choice was necessary to accommodate shape discrepancies between prediction types. For CLIPSeg, raster outputs were converted to polygons before object-level evaluation. For satellite applications the “Compute accuracy for pixel classification” tool was used paired with the vector layers provided by the CEMS activations in the case of emergency related tasks, and with CLC 2021 (or a subselection) for the CLC classification and Human Settlement Segmentation models. The bounding area of the accuracy assessment in this case is the area of interest provided by the CEMS.

2.3 Results and Discussion

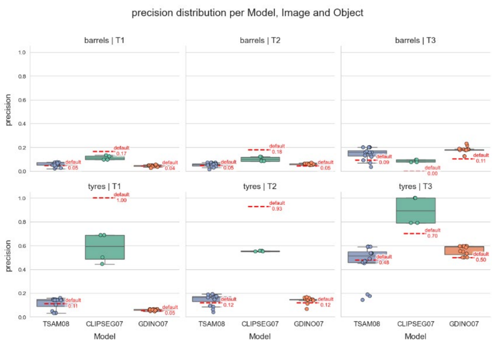
The boxplots in Figure 1. A report the F-1 scores obtained by all the inferences of the PBMs with RPAS imagery, whose examples can be found in Figure 2. The graphs show how the F-1 scores are not significantly improved by a hyperparametrization different from the default one. This is highlighted by the heights of the boxes in the boxplots, which describes the variability of the results. The graphs highlight how the models, on average, perform worse with barrels than with tyres, especially in the third orthoimage. This happens probably because of the ambiguity in the text prompt which is harder for the TE to interpret; the prompt “tyres” can trace up to a consistent range of features (e.g. round, black, holey), while “barrels” could lead to less consistent features (e.g. wooden or metallic). Lastly, for each orthoimage, the model CLIPSeg performs better than the others scoring an F-1 Score of 0.89. In the boxplots in Figure 1.B and C are reported recall and precision respectively. The high recalls and low precision reported in the graphs, especially with barrels, are clear symptom of false positives. While this highly impact the possibility to use these models in unsupervised applications, the high recall could still be useful from an operational point of view in cases where the time constrain for the inspection of the monitored area can be relaxed. Regarding satellite applications, the output of the CLC model is reported in Figure 3. The producer and user accuracies per each class are reported in Table 1. The model returns high Producer Accuracies with the classes Urban fabric, Industrial/commercial/transport units, Arable land, Open Spaces with little to no vegetation, and Water, with values of 0.78, 0.58, 0.90, 0.94, and 0.58 respectively.



A)



B)



C)

Figure 1 Evaluation of the performances of the PBMs subdivided per model, object, and RPAS orthoimage

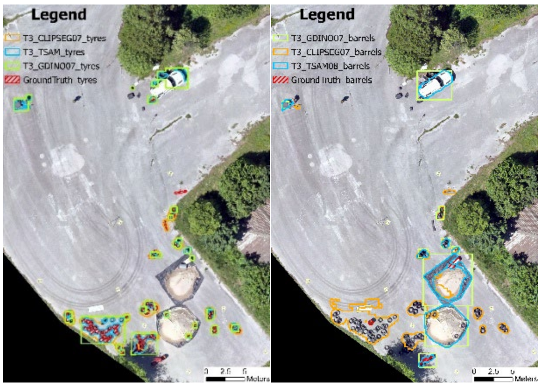


Figure 2 Examples of output of PBMs with GT in the an RPAS orthoimage

These values highlight high recall for these classes, while all the other classes are characterised by very low values, around 0.1. According to User Accuracy, 7 classes out of 11 have values higher than 0.5, with a maximum of 0.78 for Forests. These values, representative of the precision of the model, show that many false positives are generated. The Human Settlement Classification model output is reported in Figure 4, with its accuracies reported in

Table 2. In the test performed in this research, this model performed particularly well, with an overall accuracy of 0.96.

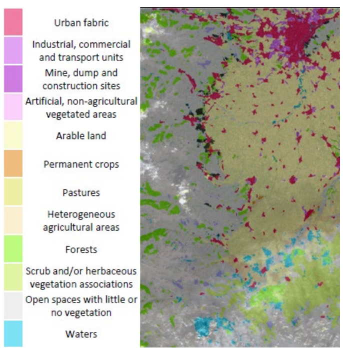


Figure 3 Output of the CLC classification model

Table 1 Producer and User accuracies per each class.

	Producer Accuracy	User Accuracy
Urban fabric	0.78	0.66
Industrial, commercial and transport units	0.58	0.70
Mine, dump and construction sites	0.06	0.75
Artificial, non-agricultural vegetated areas	0.14	0.34
Arable land	0.90	0.69
Permanent crops	0.03	0.59
Pastures	0.00	0.02
Heterogeneous agricultural areas	0.06	0.53
Forests	0.18	0.78
Scrub and/or herbaceous vegetation associations	0.06	0.10
Open spaces with little or no vegetation	0.94	0.35
Waters	0.58	0.04

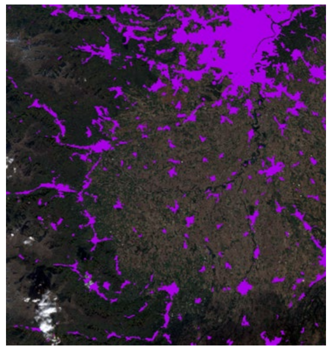


Figure 4 Output of the Human Settlement Classification model

Table 2 Summary accuracies table for the Human Settlement Classification model

	Producer Accuracy	User Accuracy
No Settlement	0.96	0.99
Human Settlements	0.93	0.72

The finetuned version of the Prithvi model for Flood Segmentation was tested on two case studies happened in Emilia Romagna (Left in Figure 5, Table 3) and in Tuscany (Right in Figure 5, Table 4). In both cases user and producer accuracies are close to 1 for the Non-Flooded class, while values around 0.5 reported for the flooded class. It is unclear if the large misclassification is due to the extraction of the GT, labelled using a different image and/or a different type of imagery, or to model’s performances. The same procedure was performed for the finetuned version of the Prithvi model for Burn Area segmentation. The selected case studies are wildfires happened in Sardinia (Left in Figure 6, Table 5) and in Pantelleria (Right in Figure 6, Table 6). The accuracies in this case are slightly different across case study; the inference with the Sardinian wildfire has a close to perfect recall with false positives pixels in both classes, while in the Pantelleria’s wildfire the false positive is mostly related to no Burnscar pixels misclassified as burned areas.

Figure 5 Red: Output of the Flood Segmentation Prithvi model. Blue: GT from CEMS. Left: Emilia Romagna flood. Right: Tuscany Flood.

Table 3 Summary Table of accuracies for the Privthi Flood Segmentation Model with the Emilia Romagna Flood.

	Producer Accuracy	User Accuracy
No Flood	0.99	0.98
Flood	0.48	0.59

Table 4 Summary Table of accuracies for the Privthi Flood Segmentation Model with the Tuscany Flood

	Producer Accuracy	User Accuracy
No Flood	0.97	0.98
Flood	0.71	0.55

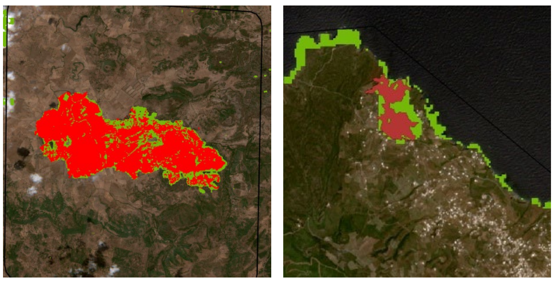


Figure 6 Red: Ground Truth from CEMS. Green: Output of the Burn Scar Segmentation Prithvi model. Left: Sardinia's Wildfire. Right: Pantelleria's Wildfire.

Table 5 Summary Table of accuracies for the Privthi Burn Scar Segmentation Model with the Sardinian Wildfire

	Producer Accuracy	User Accuracy
No BurnScar	0.78	0.99
BurnScar	0.99	0.78

Table 6 Summary Table of accuracies for the Privthi Burn Scar Segmentation Model with the Pantelleria Wildfire

	Producer Accuracy	User Accuracy
No BurnScar	0.98	0.99
BurnScar	0.94	0.3

3. CONCLUSIONS

This study wanted to evaluate the practical operability of the DL open-source models made available by the commercial GISsoftware ESRI ArcGIS Pro through a user friendly GUI, inferencing the pre-trained models available in their online framework. The results for the first part of the research, focused on PBMs and RPAS imagery, shows that the available pre-trained models can be used in operative conditions after preliminary testing with the target object. The models are not particularly sensitive to hyperparametrizations that deviates from the default one, which usually performs better and is therefore the advised one. The high recall returned by the models in almost all the inferences with barrels suggests their use can still be valuable for tasks not strongly related to time for the retrieval of the objects. Overall, satellite-based applications returned better results, which is probably linked to the abundance of training data compared to RPAS data. While the results for the automatic CLC model is not satisfactory for some classes, the tested Human Settlements segmentation model returned satisfactory results, showing how it could be already used in operational workflows. Fine-tuned segmentation models for floods and burned areas demonstrated good overall capacities for a preliminary delineation. The experiments indicate that the commercial GIS software ESRI ArcGIS Pro provides an accessible environment for deploying DL models, but that the accuracy of the outputs strongly depends on the task.

ACKNOWLEDGEMENTS

We would like to thank the DAD-Laboratory of Geomatics for CulturalHeritage and DIATI-Laboratory of GIS and Photogrammetry from Politecnico di Torino for the field data acquisition and the data processing resources. We would like to thank SAFE for their support in the planning of the acquisition campaigns, as well as for their assistance during the data collection activities and the setup of the scenarios. The first part of this work was supported by EMERITUS project, which was funded by the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101073874.

REFERENCES

European Environment Agency (2024) «CLC+Backbone 2021 (raster 10 m), Europe, 3-yearly, Jun. 2024». European Environment Agency. Disponibile su: https://doi.org/10.2909/71FC9D1B-479F-4DA1-AA66-662A2FFF2CF7.

Jain, S. (2023) «Locality Induced Non-Universality for Abelian Symmetries». arXiv. Disponibile su: https://doi.org/10.48550/ARXIV.2306.05435.

Liu, S. *et al.* (2024) «Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection». arXiv. Disponibile su: https://doi.org/10.48550/arXiv.2303.05499.

Lüddecke, T. e Ecker, A.S. (2021) «Image Segmentation Using Text and Image Prompts». arXiv. Disponibile su: https://doi.org/10.48550/ARXIV.2112.10003.

Spiegler, P. *et al.* (2025) «TextSAM-EUS: Text Prompt Learning for SAM to Accurately Segment Pancreatic Tumor in Endoscopic Ultrasound». arXiv. Disponibile su: https://doi.org/10.48550/ARXIV.2507.18082.

Smart Earth Observation for a Sustainable Future

Edited by Associazione Italiana di Telerilevamento (AIT)

Trends in Earth Observation is Directly Edited and published by AIT and it is aimed at hosting peer-review contributions that result as an extension of selected contributions to the biannual AIT Congresses. The Volume will therefore be published every two years.

This volume, organized in 5 thematic chapters, includes 38 contributions

Trends in Earth Observation is indexed by scopus and Web of Science

ISSN 2612-7911

ISBN 978-88-944687-3-1