

RONNY HÄNSCH, CLAUDIO PERSELLO, GEMINE VIVONE, JAVIERA CASTILLO NAVARRO, ALEXANDRE BOULCH, SEBASTIEN LEFEVRE, AND BERTRAND LE SAUX

## The 2022 IEEE GRSS Data Fusion Contest: Semisupervised Learning

Data availability plays a central role in any machine learning setup, especially since the rise of deep learning. Although input data are often available in abundance, reference data used to train and evaluate corresponding approaches are usually scarce due to the high cost of obtaining them. Although this is not limited to remote sensing, it is of particular importance in Earth-observation applications. Semisupervised learning is one approach to mitigate this challenge and leverage the large amount of available input data while relying only on a small, annotated training set.

The semisupervised learning challenge of the 2022 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest (DFC22) (see Figure 1) is organized by the Image Analysis and Data Fusion Tech-

nical Committee (IADF TC) of the GRSS, Université Bretagne-Sud, ONERA, and ESA Φ-lab. It seeks to promote research in automatic land cover classification from only partially annotated training data



Digital Object Identifier 10.1109/MGRS.2022.3144291  
Date of current version: 8 April 2022

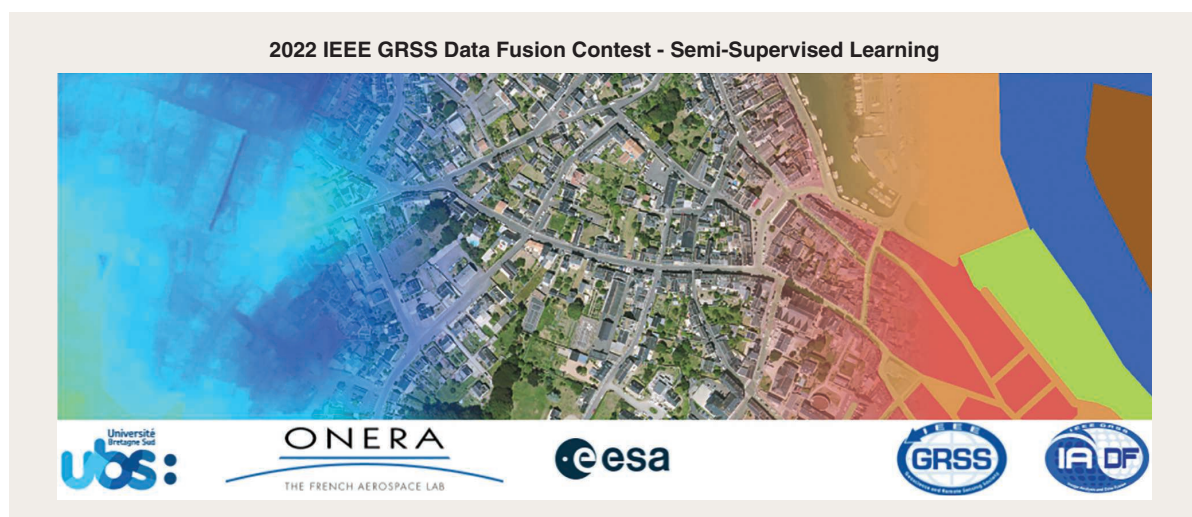


FIGURE 1. The banner image for DFC22.

consisting of very high-resolution (VHR) red, green, blue (RGB) imagery.

To this aim, DFC22 is based on MiniFrance (MF) [1], a data set for semisupervised semantic segmentation. As in real life, Earth-observation applications, MF comprises both labeled and unlabeled imagery for developing and training algorithms. It consists of a variety of classes at several locations with different appearances, which pushes the generalization capacities of the models.

DFC22 extends and modifies the MF data set for training semisupervised semantic segmentation models for land use/land cover mapping. The multimodal, MF-DFC22 data set contains aerial images, digital elevation model (DEM) information, and land use/land cover maps corresponding to 19 conurbations and their surroundings from different regions in France. It includes urban and countryside scenes, residential areas, and industrial and commercial zones, but also fields, forests, seashore, and low mountains. It gathers data from the following three sources:

- ▮ Open data VHR aerial images from the French National Institute of Geographical and Forest Information (IGN) BD ORTHO database (see Figure 2), which are provided as 8-bit RGB tiles roughly  $2,000 \times 2,000$  pixels in size at a resolution of 50 cm/px, namely,  $1 \text{ km}^2$  per tile. The images included in this data set were acquired between 2012 and 2014.
- ▮ Open data DEM tiles from the IGN RGE ALTI database. The DEM data give a representation of the bare ground (bare Earth) topographic surface of the earth. They are provided as 32-bit float rasters approximately  $1,000 \times 1,000$  pixels in size at a spatial resolution of 100 cm/px, i.e., also  $1 \text{ km}^2$  per tile. The altitude is given in meters, with submetric precision in most locations. This database is regularly updated, so images included in the data set were acquired between 2019 and 2020.
- ▮ Labeled class-reference from the UrbanAtlas 2012 database. Fourteen land use classes are considered, corresponding to the second level of the semantic hierarchy defined by UrbanAtlas. The original data are openly available as vector images at the European Copernicus program website and were used to create raster maps that geographically match the VHR tiles from BD ORTHO. They are provided as raster images with index labels roughly  $2,000 \times 2,000$  pixels in size at a resolution of 50 cm/px, i.e.,  $1 \text{ km}^2$  per tile.

The contest is designed as a benchmark competition following previous editions [2]–[7] and will consist of the following two parallel tracks:

- 1) “Track SLM: Semi-Supervised Land Cover Mapping.”
- 2) “Track BNI: Brave New Ideas.”

Track SLM is a traditional benchmarking contest: The estimates of the participants are compared against the reference data via the mean intersection over union score. During the development phase, the UrbanAtlas 2012 data will serve as references. However, during the test phase, we will

use annotations that have been carefully created from the test images.

Due to the uniqueness of DFC22, a second track is created, which explores new ideas more freely without being limited to land cover classification. In this track, all is possible and all is allowed, as long as it is novel and exciting. Even if it would be possible, the results of different submissions will not be compared to each other—neither qualitatively nor quantitatively.

The ranking in Track BNI is based on the evaluation of a methodological description of the approach by the DFC Committee. The participants will receive the same data as described previously for Track SLM. They are free to submit their results to the CodaLab competition website or even join the Track SLM in parallel if the “new idea” fits into its scope. However, evaluation will be based on the submitted description only.

**AS IN REAL LIFE, EARTH-OBSERVATION APPLICATIONS, MF COMPRISES BOTH LABELED AND UNLABELED IMAGERY FOR DEVELOPING AND TRAINING ALGORITHMS.**

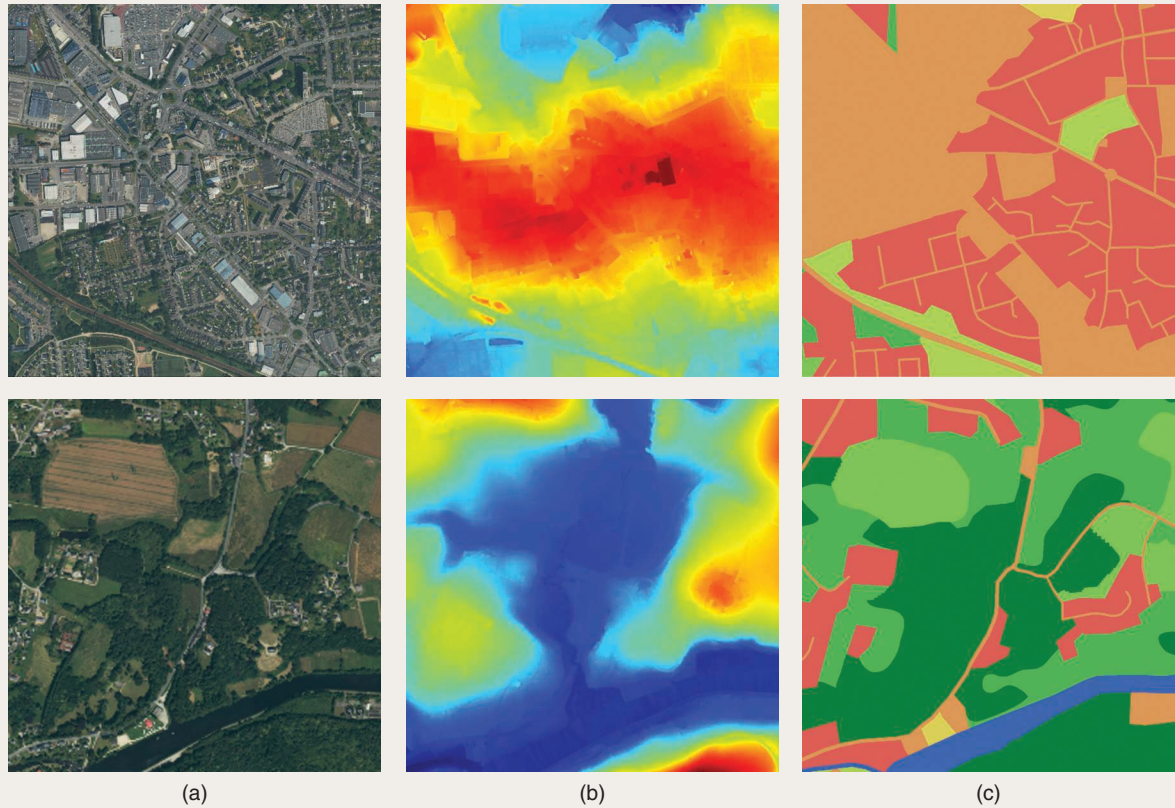
#### GET THE DATA AND ENTER THE CONTEST

The training and validation data sets were made publicly available on 3 January 2022 on IEEE DataPort (<https://ieeedata.org/competitions/data-fusion-contest-2022-dfc2022>). The evaluation server with a public leaderboard was opened on 4 January 2022 so that participants can submit prediction results for the validation set to the CodaLab competition website (<https://codalab.lisn.upsaclay.fr/competitions/880>) to receive feedback on the performance of their approaches. To enter the test phase, participants submit a short description of the used approach by 28 February 2022. The test phase is allocated during 7–11 March. The test phase is kept short to ensure an objective and fair comparison among methods. After the final check of the submitted classification maps compared with the undisclosed ground truth for testing, winners are announced on 25 March 2022. The evaluation server will be reopened after the winners announcement for further development in the field (see “Using the Data and Joining the Image Analysis and Data Fusion Technical Committee”).

More information regarding data download and registration to the evaluation server can be found at the IADF TC website (<https://www.grss-ieee.org/technical-committees/image-analysis-and-data-fusion/>). Questions and comments on the data and the contest can be submitted to the IADF TC LinkedIn group (<https://www.linkedin.com/groups/3678437/>). Updates about the contest will also be published via the IADF Twitter channel @grssiadf.

#### AWARDEES, AWARDS, AND PRIZES

The first-, second-, and third- as well as the first-ranked team of Track SLM and Track BNI, respectively, are declared winners. These four teams will be invited to submit their



**FIGURE 2.** (a) VHR aerial images from the IGN BD ORTHO database. (b) Open data DEM tiles from the IGN RGE ALTI database. (c) A labeled class-reference from the UrbanAtlas 2012 database.

### Using the Data and Joining the Image Analysis and Data Fusion Technical Committee

The data of DFC22 will remain available to the GRSS community for benchmarking algorithms and publishing research works. The data are usable free of charge for scientific purposes, but the Contest Terms and Conditions on the contest webpage remain applicable. Please read them carefully at <https://www.grss-ieee.org/community/technical-committees/2022-ieee-grss-data-fusion-contest/>.

You can contact the IADF TC Chairs at [iadf\\_chairs@grss-ieee.org](mailto:iadf_chairs@grss-ieee.org). If you are interested in joining the IADF TC, please complete the form on our website

at <https://www.grss-ieee.org/technical-committees/image-analysis-and-data-fusion>. Members receive information regarding research and applications on image analysis and data fusion topics, and updates on the annual DFC and on all other activities of the IADF TC. Membership in the IADF TC is free! You can also join the LinkedIn IEEE GRSS Data Fusion Discussion Forum (<http://www.linkedin.com/groups/IEEE-GRSS-Data-Fusion-Discussion-3678437>) and Twitter channel (<https://twitter.com/grssiadf>).

manuscripts to the 2022 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) in Kuala Lumpur, Malaysia. Each manuscript will describe the method used for achieving the winning results. All papers will be presented by the winners in an oral invited session dedicated to the contest at IGARSS 2022. They will also be included into the proceedings of IGARSS 2022.

The four winning teams will be awarded IEEE Certificates of Recognition and will receive a

special prize, namely, a total of US\$20,000 in Microsoft Azure credits. The first- and second-ranked teams in Track SLM and the first-ranked team in Track BNI will coauthor a journal paper, which will summarize the outcome of DFC22 and will be submitted with open access to *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. GRSS will cover the costs of the open access fee. The winners of both tracks receive a total of US\$20,000 in Azure cloud credits as prizes, courtesy of Microsoft's AI for Earth program.

### ACKNOWLEDGMENTS

The IADF TC chairs would like to thank the IEEE GRSS for continuously supporting the annual DFC through funding and resources.

**DUE TO THE UNIQUENESS OF DFC22, A SECOND TRACK IS CREATED, WHICH EXPLORES NEW IDEAS MORE FREELY WITHOUT BEING LIMITED TO LAND COVER CLASSIFICATION.**



## REFERENCES

- [1] J. Castillo-Navarro, B. Le Saux, A. Boulch, N. Audebert, and S. Lefevre, "Semi-supervised semantic segmentation in earth observation: The minifrance suite, dataset analysis and multi-task network study," *Mach. Learn.*, 2021, doi: 10.1007/s10994-020-05943-y. [Online]. Available: <https://link.springer.com/article/10.1007/s10994-020-05943-y#citeas>
- [2] N. Yokoya *et al.*, "Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1363–1377, May 2018, doi: 10.1109/JSTARS.2018.2799698.
- [3] Y. Xu *et al.*, "Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1709–1724, Jun. 2019, doi: 10.1109/JSTARS.2019.2911113.
- [4] S. Kunwar *et al.*, "Large-scale semantic 3-D reconstruction: Outcome of the 2019 IEEE GRSS data fusion contest—Part A," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 922–935, Oct. 2020, doi: 10.1109/JSTARS.2020.3032221.
- [5] Y. Lian *et al.*, "Large-scale semantic 3-D reconstruction: Outcome of the 2019 IEEE GRSS data fusion contest—Part B," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1158–1170, Nov. 2020, doi: 10.1109/JSTARS.2020.3035274.
- [6] C. Robinson *et al.*, "Global land-cover mapping with weak supervision: Outcome of the 2020 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3185–3199, Mar. 2021, doi: 10.1109/JSTARS.2021.3063849.
- [7] Y. Ma *et al.*, "The outcome of the 2021 IEEE GRSS data fusion contest - Track DSE: Detection of settlements without electricity," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12,375–12,385, Nov. 2021, doi: 10.1109/JSTARS.2021.3130446.

GRS



# TAP. CONNECT. NETWORK. SHARE.

## Connect to IEEE—no matter where you are—with the IEEE App.



Stay up-to-date  
with the latest news



Schedule, manage, or  
join meetups virtually



Get geo and interest-based  
recommendations



Read and download  
your IEEE magazines



Create a personalized  
experience



Locate IEEE members by location,  
interests, and affiliations

**Download Today!**

