







The Outcome of the 2021 IEEE GRSS Data Fusion Contest - Track DSE: Detection of Settlements Without Electricity

Yanbiao Ma, Yuxin Li, Kexin Feng, Yu Xia, Qi Huang, Hongyan Zhang , *Senior Member, IEEE*, Colin Prieur, Giorgio Licciardi , Hana Malha, Jocelyn Chanussot , *Fellow, IEEE*, Pedram Ghamisi , *Senior Member, IEEE*, Ronny Hänsch , *Senior Member, IEEE*, and Naoto Yokoya , *Member, IEEE*

Abstract—In this article, we elaborate on the scientific outcomes of the 2021 Data Fusion Contest (DFC2021), which was organized by the Image Analysis and Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society, on the subject of geospatial artificial intelligence for social good. The ultimate objective of the contest was to model the state and changes of artificial and natural environments from multimodal and multitemporal remotely sensed data towards sustainable developments. DFC2021 consisted of two challenge tracks: Detection of settlements without electricity (DSE) and multitemporal semantic change detection. We focus here on the outcome of the DSE track. This article presents the corresponding approaches and reports the results of the best-performing methods during the contest.

Index Terms—Convolutional neural networks, deep learning, image analysis and data fusion, land cover change detection, multimodal, random forests, weak supervision.

Manuscript received October 30, 2021; revised November 15, 2021; accepted November 19, 2021. Date of publication November 24, 2021; date of current version December 15, 2021. This work was partially supported by MIAI @ Grenoble Alpes (ANR-19-P31A-0003) (*Corresponding author: Naoto Yokoya.*)

Yanbiao Ma, Yuxin Li, and Kexin Feng are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: ybmamail@stu.xidian.edu.cn; yxli_12@stu.xidian.edu.cn; 17390953976@163.com).

Yu Xia and Qi Huang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: whuxiayu@whu.edu.cn; qihuang@whu.edu.cn).

Hongyan Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zhanghongyan@whu.edu.cn).

Pedram Ghamisi is with the Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Machine Learning Group, 09599 Freiberg, Germany, and also with the Institute of Advanced Research in Artificial Intelligence, 1030 Vienna, Austria (e-mail: p.ghamisi@gmail.com).

Ronny Hänsch is with the German Aerospace Center (DLR), 82234 Weßling, Germany (e-mail: ronny.haensch@dlr.de).

Naoto Yokoya is with the Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, University of Tokyo, Chiba 277-8561, Japan, and also with the RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan (e-mail: yokoya@k.u-tokyo.ac.jp).

Colin Prieur, Giorgio Licciardi, and Jocelyn Chanussot are with the GIPSA-Lab, Grenoble INP, 38400 Saint Martin, France (e-mail: colin.prieur@grenoble-inp.org; g.a.licciardi@gmail.com; jocelyn.chanussot@grenoble-inp.fr).

Hana Malha is with the AI Strategy and Solutions, Hewlett Packard Enterprise Co., 94089 Grenoble, France (e-mail: hana.malha@hpe.com).

Digital Object Identifier 10.1109/JSTARS.2021.3130446

I. INTRODUCTION

THE 2021 Data Fusion Contest (DFC2021), organized by the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society (GRSS), promotes interdisciplinary research on geospatial artificial intelligence (AI) for social good [1]. IADF TC is an international network of scientists working on earth observation, geospatial data fusion, and algorithms for image analysis. It aims at connecting people and resources, educating students and professionals, and promoting theoretical advances and best practices in image analysis and data fusion.

Since 2006, the IADF TC organizes an annual challenge, i.e., DFC, for fostering ideas and progress in remote sensing, distributing novel data, and benchmarking analysis methods [2]–[16]. The contest follows the ultimate goal of building models to understand the state and changes of artificial and natural environments using multisensor and multitemporal remote sensing data towards sustainable developments. The contest is designed as a benchmark competition following previous editions [13], [14], [16]–[18].

DFC2021 is dedicated to the following real-world social challenges: 1) Analysis of multisensor, multiresolution, and multitemporal data, and 2) learning from low quality labeled samples (weak supervision). The aforementioned problems are major open challenges in a wide range of fields, from Earth observation to computer vision and machine learning [1]. The main feature of the contest is that it tackles directly some of the most challenging social problems such as energy equality and environmental conservation. In other words, the results of the contest will not only lead to technological development, but also develop a tool for solving actual social problems [1]. DFC2021 includes the two following tracks, which were run in parallel as follows.

- 1) *Track DSE*: Detection of settlements without electricity;
- 2) *Track MSD*: Multitemporal semantic change detection.

This article presents the main outcome of Track DSE. This track was co-organized by Hewlett Packard Enterprise, SolarAid, and Data Science Experts, to address the automatic detection of human settlements deprived of access to electricity using multimodal, multiresolution, and multitemporal satellite remote sensing data. Sentinel-1 SAR data, Sentinel-2, Landsat-8, and

- 1: Human settlements without electricity (Region of Interest)
- 2: No human settlements without electricity
- 3: Human settlements with electricity
- 4: No human settlements with electricity

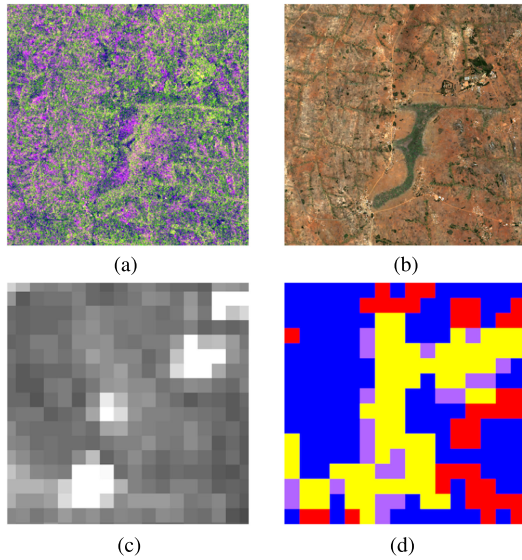


Fig. 1. Visual examples of satellite images (a)–(c) and reference data (d) used in DFC2021 Track DSE. (a) Sentinel-1 image (RGB: VV, VH, VV-VH). (b) Sentinel-2 image (RGB: Bands 4, 3, 2). (c) VIIRS image (9 temporal images are accumulated).

Suomi Visible Infrared Imaging Radiometer Suite (VIIRS) night time images were utilized as input (see Fig. 1). While the original ground sampling distance (GSD) ranges from 10 m to 750 m, all images were resampled at 10 m. Four classes of semantic labels (i.e., settlements with and without electricity, no settlements with and without electricity) were provided at a GSD of 500 m for the training set. Participants are required to submit binary classification maps of settlements without electricity at a GSD of 500 m. The classification accuracy of different solutions was estimated using the F1 score. The main challenge of Track DSE is to develop robust and efficient methods to extract high-level semantic information from such heterogeneous data.

In this article, we describe the datasets used in DFC2021 in Section II, and discuss the overall results of the competition in Section III. Then, we will focus in more detail on the approaches proposed by the first-ranked teams of Track DSE in IV and Section V. Finally, Section VI concludes this article.

II. DATA AND BASELINE OF THE DATA FUSION CONTEST 2021

The data of the DFC2021 mapped the Dezda and Salima sectors in Malawi. It has been split into 98 tiles of 800×800 pixels. The tiles have been distributed, respectively, across the training, validation and test sets as follows: 60, 19, and 19 tiles. Each tile included 98 channels from the below listed satellite images. All the images have been resampled to a GSD of 10 m. Thus, each tile corresponded to a 64 km^2 area.

- 1) *Sentinel-1 polarimetric SAR*: Two channels corresponding to intensity values for VV and VH polarization at a $5 \text{ m} \times$

20 m spatial resolution, scaled to a $10 \text{ m} \times 10 \text{ m}$ spatial resolution from an interferometric wide swath mode of acquisition.

- 2) *Sentinel-2 multispectral dataset*: In total, 12 channels of reflectance data covering VNIR and SWIR ranges at a GSD of 10 m, 20 m, and 60 m. The cirrus band 10 was omitted, as it does not contain ground information, [19].
- 3) *Landsat 8 multispectral dataset*: In total, 11 channels of reflectance data covering VNIR, SWIR, and TIR ranges at a GSD of 30 m and 100 m, and a Panchromatic band at a GSD of 15 m.
- 4) *The Suomi VIIRS night time dataset (Black Marble Suite)*: In 1 channel, the Day-Night Band (DNB) sensor of the VIIRS provided the global daily measurements of nocturnal visible and near-infrared light at a GSD of 750 m. The VNP46A1 product is a corrected version of the original DNB data, and is at a 500 m GSD resolution, [20].

The provided training data was split across 60 folders named TileX, X being the tile number. Each folder included 100 files, 98 files corresponding to the satellite images listed earlier, and two were ground truth images: groundTruthRGB.png provided in RGB for easier visualization, and groundTruth.tif that represented the label reference information for each tile.

The labeling process has been performed as follows.

- 1) *Human settlement*: If a building was present in a patch of $500 \text{ m} \times 500 \text{ m}$, this area was considered to have human settlement;
- 2) *Presence of electricity*: If in a $500 \text{ m} \times 500 \text{ m}$ patch, a fraction of the pixels, regardless of its size, was illuminated, the patch was considered to be fully illuminated.

The reference file (groundTruth.tif) was 16×16 pixels large, with a resolution of 500 m corresponding to the labeling strategy described previously. The pixel values (1, 2, 3, and 4) corresponded to the four following classes.

- 1) Human settlements without electricity (Region of Interest).
- 2) No human settlements without electricity.
- 3) Human settlements with electricity.
- 4) No human settlements with electricity.

Submissions were expected to be provided as a binary classification of the Region of Interest, at a 500 m GSD resolution (i.e. 16×16 pixels), as a GeoTIFF with the Byte (uint8) data type as follows:

- 1) for human settlements without electricity (Region of Interest);
- 2) for any other class.

Particular attention was paid regarding the ease of accessibility to the data sources used in the contest, so that it would be freely available for Solaraid's future work.

III. ORGANIZATION, SUBMISSIONS AND RESULTS

There were 168 unique registrations at the CodaLab competition website¹ during the development phase and 28 teams entered the test phase after screening the descriptions of their

¹[Online]. Available: <https://competitions.codalab.org/competitions/27943>

TABLE I
TOP RANKED TEAMS AND THEIR APPROACHES IN TERMS OF USED DATA, PROBLEM SETTING, AND CLASSIFICATION MODELS

Rank	Team	F1	Data				Problem setting		Classification model	
			S1	S2	L8	VIIRS	Single	Dual	Settlement	Electricity
1	<i>fengkexin</i>	0.8939		✓		✓		✓	SENet	SENet
2	<i>WHU_YuXia</i>	0.8806		✓		✓	✓	✓	GC-CNN	GC-CNN
3	<i>dimartinot</i>	0.8798	✓	✓	✓	✓	✓		Multi-branch CNN with EfficientNet	
3	<i>JIOJIO</i>	0.8798		✓		✓		✓	UNet, LinkNet	Thresholding

approaches submitted by the end of the development phase. In total, 3456 submissions were received during the development phase with active participation from all registered teams. During the test phase, the maximum number of submissions per team was limited to ten, and 170 submissions were received. The final ranking was determined based on F1 score.

The first to fourth ranked teams were awarded as winners of the DFC2021 Track DSE and presented their solutions during the 2021 IEEE International Geoscience and Remote Sensing Symposium. The four winning teams are the following.

- 1) *1st place: fengkexin* team; Y. Ma, Y. Li, K. Feng, and X. Geng from Xidian University, China; Dual-task models based on squeeze-and-excitation networks (SENet) followed by postprocessing based on expert priors [21].
- 2) *2nd place: WHU_YuXia* team; Y. Xia, Q. Huang, and H. Zhang from Wuhan University, China; Ensembling of single-task and dual-task models based on global context convolutional neural networks (GC-CNN) [22].
- 3) *3rd place: dimartinot* team; Thomas D. Martino, M. Lenormand, and E. Colin Koeniguer from ONERA, France; Multibranch CNN with three-dimensional convolutions and EfficientNet [23].
- 4) *3rd place: JIOJIO* team; R. Feng, M. Wang, X. Zhang, and J. Zhang from Xidian University, China; Ensembling of UNet and LinkNet with depthwise over-parameterized convolutional layers [24].

The top four teams commonly employed techniques such as preprocessing, including data normalization and data augmentation, ensembling (or model fusion) to integrate the results of multiple models, and test-time augmentation. All teams utilized recent neural network architectures (or modules/layers), which have been shown to be effective in the field of computer vision, for classification and built models that were well customized to the task of the contest. Since multisensor data were provided and the task was more complex than typical building detection, various approaches were developed in terms of which data to use and what problem setting to solve.

Table I summarizes the differences in the approaches of the top four teams regarding used data, problem setting, and classification models. While all teams used VIIRS data for electricity detection, the use of Sentinel-1, Sentinel-2, and Landsat-8 data for settlement detection was different: three teams used only Sentinel-2 data, while one team (*dimartinot*) utilized all data and showed its effectiveness in the ablation study [23]. For this contest, Sentinel-2 data alone were sufficient for settlement detection, but the use of multi-sensor data is worthy of further investigation in order to scale up more widely and improve generalization performance.

Interestingly, there was no consensus on whether the task should be solved as a single classification problem or as two classification problems, i.e., settlement detection and electricity detection; similar performance was achieved with different approaches. As for the classification models, all teams adopted CNN-based methods for processing relatively high-resolution Sentinel-2 images. On the other hand, different methods such as thresholding, random forest, and CNN were used for electricity detection when the dual task is employed, suggesting that simple methods may be sufficient for processing low resolution images.

IV. FIRST PLACE TEAM OF TRACK DSE

In this section, the winning algorithm in the track DSE will be described. In order to detect settlements without electricity, we design the dual-task models, i.e., we divide the task into settlement detection and electricity detection, as shown in Fig. 2(a). Specifically, we first preprocess the images and propose a novel outlier removal method for the VIIRS night time dataset. Next, SENet [25] is used to train multiple classifiers for the two detection tasks. Then, the classification results of the two tasks are postprocessed separately, which contains model integration and correction of misclassification based on expert priors, etc. Finally, the results of settlement detection and light detection are fused to obtain the final detection result. The experiments show that our preprocessing method can maximize the filtering of interference information and noise in the images, allowing the dual-task models to make full use of the valuable information in the multisource data, and the postprocessing approach further utilizes the priori knowledge to effectively correct the misclassification, which together contribute to the outstanding performance of our model.

A. Data Preprocessing

Data preprocessing is an essential part of our method and plays a crucial role in the performance of the subsequent detection models, especially for the electricity detection. It consists of two important steps.

- 1) Considering that the noise in the data affects the performance of the model, outliers of the multisource images need to be removed.
- 2) Due to the small number of training samples, the true distribution of the data is not well reflected and the feature diversity is not sufficient, so we use data augmentation to enrich the features.

Sentinel-2 Multispectral Dataset: We use 4, 3, and 2 channels to synthesize the true color images. To remove extreme outliers, the pixel values of the three channels are clipped to the interval

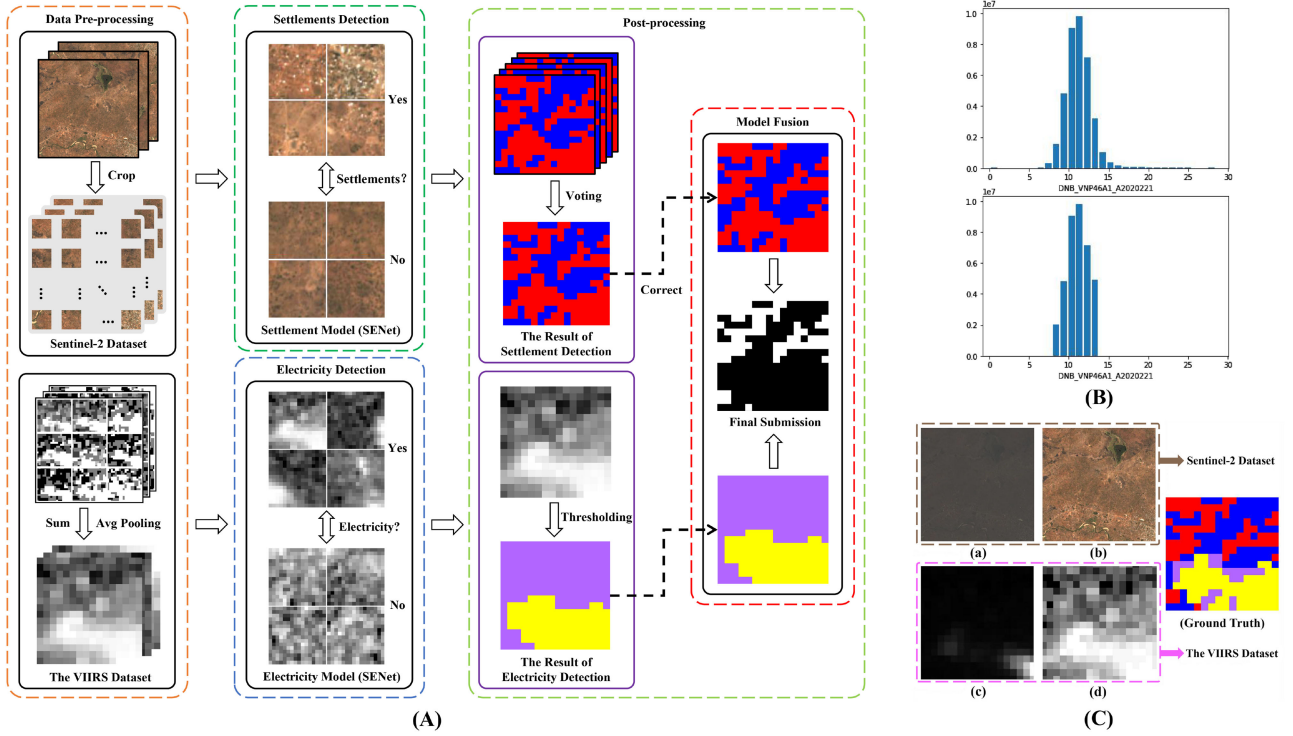


Fig. 2. (a) For ease of observation, in settlement detection, 1 is represented by red and 0 is represented by blue. In electricity detection, 1 and 0 are denoted by yellow and purple. In submission, white means 1 and black means 0. (b). Take channel DNB_VNP46A1_A2020221 as an example, the horizontal axis is the pixel value and the vertical axis is the number of pixels. (c). (a) and (c) are images without outliers removed, and (b) and (d) are images with outliers removed.

[0,2000] [26], and then the synthesized images are normalized to [0,255]. Since the ground sampling distance of the true color image is $10\text{ m} \times 10\text{ m}$ and that of the ground truth is $500\text{ m} \times 500\text{ m}$, it means that for the 800×800 pixels image, it is necessary to distinguish whether there is a human settlement in each 50×50 pixels region. Therefore, each true color image is cropped into 256 smaller images of 50×50 pixels, and they are classified into two categories of “settlements” and “no settlements” based on the labels, thereby yielding a training set that can be used to train a model to classify the presence or absence of settlements in smaller images. To enrich the features, the training set is augmented with five data augmentation methods: 90, 180, and 270 degrees of rotation, horizontal flip, and vertical flip, respectively.

VIIRS Dataset: To obtain cleaner data, we remove the noise in each channel using the proposed formula

$$\text{pixel}c_{i,j} = \begin{cases} L_{c_min}, & \text{pixel}c_{i,j} \leq L_{c_min} \\ \text{pixel}c_{i,j}, & L_{c_min} < \text{pixel}c_{i,j} < L_{c_max} \\ L_{c_max}, & \text{pixel}c_{i,j} \geq L_{c_max} \end{cases} \quad (1)$$

where $\text{pixel}c_{i,j}$ represents the pixel values of the image, c denotes the name of the channel, and $i, j \in [1, 800]$. All pixel values of a total of 60 training images from the same channel c are sorted from smallest to largest, and L_{c_min} and L_{c_max} denote the pixel values located at 5% and 95% of them, respectively. Fig. 2(b) shows the pixel changes after removing the outliers using a channel as an example.

For each tile of the training set, we sum the pixel values of the 9 images after removing outliers and normalize them to the interval [0,255], thereby yielding one night time image corresponding to the tile. Since the ground truth is a coarse label of 16×16 pixels, we use an average pooling of 50×50 pixels instead of max-pooling to generate a new image of 16×16 pixels. The reason for not using max-pooling is that it would cause the original area with light to be overly expanded and more prone to false detection. The images are also augmented using the five data enhancement methods described previously to obtain a night time dataset for training a model to classify the presence or absence of electricity in an image.

Taking tile14 as an example, Fig. 2(c) shows the differences in images due to outlier removal on the two datasets, respectively. It can be seen that our outlier removal method is critical, especially for the electricity detection, which filters out the interference of noise to make the regions with electricity more accurate.

B. Dual-Task Models

Transfer Learning: After preprocessing, two datasets are available for settlement detection and electricity detection, respectively. For both classification tasks, we use SENet154 pretrained on ImageNet for transfer learning. The channel attention mechanism allows the network to learn the weights of channels adaptively, assigning different weights to different channels. Multiple training models can be obtained by setting different hyperparameters. The base settings are batchsize of 32, epoch of 15, and initial learning rate of 0.001, and the

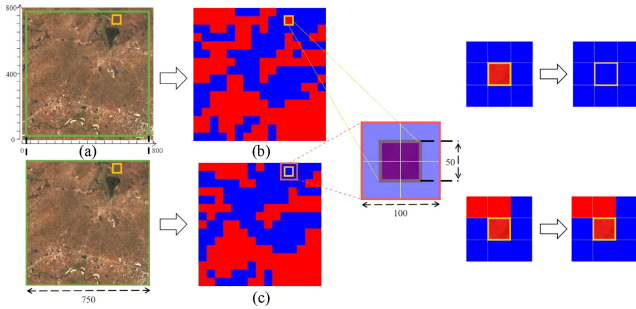


Fig. 3. Correction of misclassification based on expert priors. The regions marked by green lines indicate the image of 750×750 pixels.

learning rate decays to 90% of the original for each epoch during training.

Settlement model and electricity model: After training, multiple models for detecting settlements can be obtained and they are called settlement models. In addition, we define a portion of 50×50 pixels in an image as a region, which means that an image consists of 256 regions. The regions with settlements are denoted as 1, otherwise they are denoted as 0. Similarly, we refer to the model used for electricity detection as the electricity model. It can distinguish for us whether there is an area with electricity in an image.

C. Postprocessing

For further correction to make the predictions more correct, the postprocessing consists of model integration and the proposed approach based on expert priors to correct misclassification for settlement detection, and a threshold approach for electricity detection. The final regions with settlements but no electricity can be obtained by fusing the results of settlement detection and electricity detection.

Settlement detection: The trained models have different preferences for different scenarios due to different settings of hyperparameters. In the testing phase, the predictions of the top-5 best performing classification models on the validation set are integrated. Model integration fuses the outputs of multiple models by voting to improve the settlement model performance.

Correction of misclassification based on expert priors: We use the priori knowledge of the continuity of the building distribution to correct the predictions of the settlement model, thus proposing a correction approach based on expert priors.

In Fig. 3(a) is a true color image, and (B) is a binary map of 16×16 pixels, which is also the prediction result of (A). An 750×750 pixels image is obtained by center cropping (A), and (C) is a binary map of 15×15 pixels corresponding to the image. Select a random region in (A), and take the region marked by the orange line as an example, it can see that the region is predicted to be 1 in (B), but is predicted to be 0 in (C), which means that the two predictions conflict. At this time, we check the pixel values of the other 8 regions around the conflicting region in (B), and if the pixel values are all 0, the 1 in the region will be corrected to 0. Otherwise, based on the priori knowledge, the region will not be corrected.

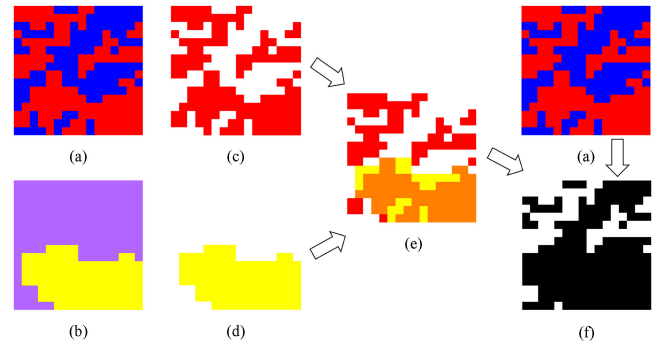


Fig. 4. Fusion strategy for the results of settlement detection and light detection. The orange color represents the overlapping regions.

Electricity detection: The electricity model is used to classify the presence or absence of areas with electricity. If there is an area with electricity, then we use the threshold approach to select specific regions with electricity in an image. The presence of electricity is denoted as 1 and the absence as 0. First, when an area consisting of multiple regions does not contain any independent regions, i.e., when the regions are adjacent to each other, the area is called a contiguous area. Next, if the number of regions contained in a contiguous area is not less than 4, the area is defined as a valid contiguous area, and the opposite is an invalid contiguous area. Invalid contiguous areas are considered as a kind of noisy data to be removed. Then, the pixel values in an image are sorted from smallest to largest and the pixel value located at 85% is taken as the initialization threshold. Regions with pixel values greater than the threshold are set to 1 and those less than it are set to 0. Finally, for each tile, a 16×16 pixels binary map is obtained. Experience shows that the best performance is achieved when the final threshold is the pixel value located at 85–90%, and the binary map after removing invalid contiguous areas represents the electricity detection result.

Model fusion: In order to detect settlements without electricity, the results of settlement detection and electricity detection need to be fused. The fusion strategy is shown in Fig. 4(a) and (b) are the results of settlement detection and electricity detection, respectively, and their corresponding regions with pixel values of 1 are (c) and (d). If there is an overlap (e) between (c) and (d), the final detection result (f) is obtained by modifying the 1 at the overlap position in (a) to 0.

D. Results and Discussion

Settlement model: Different networks [27], [28] are trained on the Sentinel-2 dataset. Table II shows the performance comparison of multiple models and operations on the validation set, and it can be seen that SENet154 performs the best, so it is still used in the test set. In addition, the data preprocessing method can significantly improve the settlement model performance.

Electricity model: Based on the performance of different models in the settlement detection, we continue to use SENet154 as the backbone architecture of the electricity model. Since there are only two images classified as having the presence of electricity in the validation set, which is much less than

TABLE II
PERFORMANCE COMPARISON ON THE VALIDATION SET

Settlement model	ResNet34	ResNet50	SENet50	PNASNet	SENet154
Recall	0.8224	0.8052	0.8927	0.8575	0.8549
Precision	0.8576	0.8745	0.8037	0.8275	0.8600
F1	0.8396	0.8384	0.8459	0.8422	0.8575
Operation	Single	+ Truncation		+ Augmentation	
F1	0.8575	0.8592		0.8678	
Operation	Integrated model		+ Electricity model		
F1	0.8735		0.8756		

TABLE III
PERFORMANCE OF DIFFERENT OPERATIONS ON THE TEST SET

Metric	F1
Settlement model	84.14%
+ Model integration	85.30%
+ Electricity model	88.64%
+ Correction	89.39%

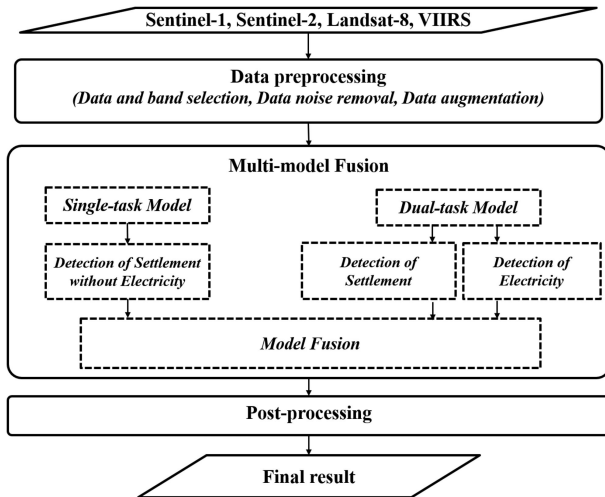


Fig. 5. Flowchart of the proposed method for DSE.

the number of images with electricity in the test set. However, Table II also shows the remarkable role of the electricity model.

Table III shows the improvement in F1 scores on the test set by adding various operations. When using only a single settlement model for prediction, F1 is 84.14%, and after integrating the results of five settlement models, F1 reaches 85.30%. Fusing the integrated results with those of the electricity model, F1 is 88.64%, and after postprocessing, F1 can reach 89.39%. It can be seen that the electricity detection is crucial and improves the F1 score by about 3%, while the post processing based on the expert priors also shows a favorable performance.

V. SECOND PLACE TEAM OF TRACK DSE

A. Proposed Framework

The flowchart of our method is shown in Fig. 5. The method mainly consists of three steps, which can be summarized as follows.

1) *Data Preprocessing*: The steps of data and band selection, data noise removal and data augmentation are implemented to process multimodal and multitemporal remote sensing data, so as to obtain a reliable and considerable training dataset.

2) *Multimodel Fusion*: Single-task model is built to detect settlements without electricity and dual-task model is built to detect settlements and electricity individually, and the strategy of integrating different models is applied for further improving the result.

3) *Postprocessing*: In order to further eliminate the interference of noise, the threshold segmentation methods are adopted for producing the final result.

The specific operations of each step and experimental results will be introduced in the following sections.

B. Data Preprocessing

The data preprocessing step is quite essential in the whole framework, for the reason that data selection lays the foundation for the subsequent results. In this section, three parts are included: data and band selection, data noise removal and data augmentation, which aim at removing degraded data and retaining valid data.

The first step is to select useful data and bands for the task of the detection of settlements without electricity. The optical imagery is used for the detection of settlements by providing abundant spectral information. Landsat-8 and Sentinel-2 are both optical data, while the spatial resolution of Sentinel-2 is higher than that of Landsat-8 and the cloud pollution of Landsat-8 data is more severe than that of Sentinel-2 data. Thus, Sentinel-2 data are selected for the detection of settlements. According to the commonly used building indices like NDBI and UI [29], classical bands for building detection including red, green, blue, near-infrared, SWIR-1 and SWIR-2 of the Sentinel-2 image are picked for the following steps. In addition, although SAR data can be effective for the settlement detection with structural information, the provided Sentinel-1 data with severe speckle noises will result in poor results, thus, the Sentinel-1 data are discarded. In order to detect the electricity condition, multitemporal VNP46A1 products are also chosen for the following parts.

The second preprocessing step aims to remove the data noise. Optical images are inevitably contaminated by clouds, thus, the cloud noise removal is of vital importance. A modified median filter is adopted in this part to composite cloud-free images from multitemporal data [30]. Furthermore, the VNP46A1 product data has light value variations at different times, and effective feature information should be extracted from multitemporal data to avoid noise interference. The accumulative, median and maximum values of multitemporal night time data are composited to generate new images. Moreover, a 2% linear gray scale stretching algorithm is then implemented on all selected and generated bands to eliminate extreme noisy points and enhance the image contrast.

The final preprocessing step is data augmentation, which can increase the amount of training data to avoid the overfitting of the model and improve the model generalization ability. In this

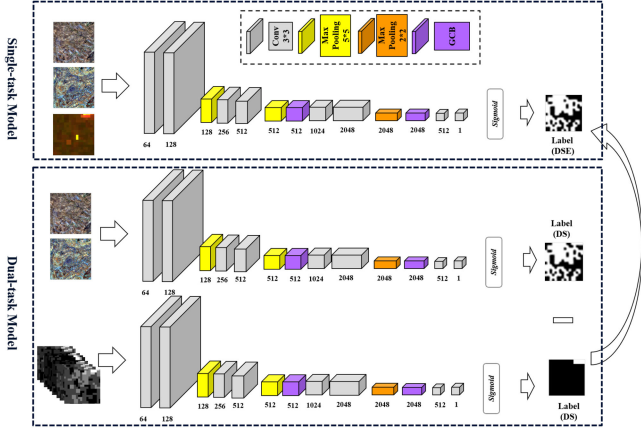


Fig. 6. Workflow of single-task and dual-task model.

part, the operation of rotation and flipping, and first rotation then flipping multiplies the amount of data eightfold, providing the foundation of robust network training [31].

The single task of DSE can be decomposed into two tasks: detection of settlement (DS) and detection of electricity (DE), so that a single-task and a dual-task model can be both formed. As shown in Fig. 6, in these two models, an end-to-end global context convolutional neural network (GC-CNN) is designed and applied with the same architecture. However, the input and output of two models are different. The single-task model has only one GC-CNN, the input of which is the cascaded data of the 800×800 Sentinel-2 imagery and night time data, and the output of which is a 16×16 segmentation map of the settlement without electricity. In the dual-task model, two GC-CNN are adopted with different input data: Sentinel-2 data and the original night time imagery. The output of two networks are the segmentation map of settlement and electricity. The final result can be obtained by subtracting the two results.

C. Multimodel Fusion

In the GC-CNN, three basic modules are utilized: convolution block, pooling block and global context (GC) block [32]. Among them, the convolution block is used for better extracting the deep features in the imagery, while the pooling block is applied to reduce the dimensionality of information. The GC block plays a vital role in extracting the global context information in the image and capture the long-distance dependence of the deep neural network. The specific operations of the global context block adopted in this paper are shown in Fig. 7. First, a global attention pooling operation is used for context modeling, and a bottleneck transformation is then applied to capture channel dependencies. Finally, by the operation of broadcast element addition, features can be further fused. In addition, due to the imbalance of label categories, the focal loss function is used in network training, which is described as follows:

$$\begin{aligned} \text{FL}(p, y) = & -\alpha y(1-p)^\gamma \log(p) \\ & - (1-\alpha)(1-y)p^\gamma \log(1-p). \end{aligned} \quad (2)$$

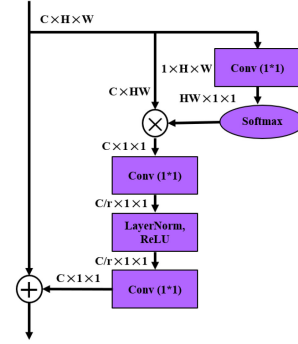


Fig. 7. Overview of global context block.

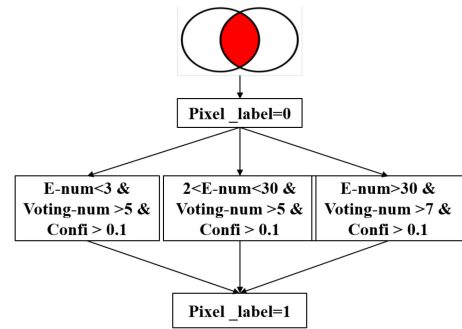


Fig. 8. Principle of model fusion.

In (2), p denotes the predicted result, and y represents the true label. α and γ are the weight parameters, which are always set as 0.6 and 2, respectively [33]. It is worth noting that a creative voting strategy is applied in the testing phase. Among the eight predicted results generated by different data augmentation methods, if more than two prediction results are labeled as 1, the pixel will be marked as positive. On the one hand, although the single-task model can directly produce the final result, the fusion of night data and optical data may be insufficient. On the other hand, the dual-task model makes full use of the characteristics of the data, but there may be error accumulation in the multitask prediction process. Therefore, an empirical fusion strategy is presented in Fig. 8. The first step of model fusion is to take the intersection of the two model results for producing a high-confidence label, but there are many omissions and errors. On the basis of these results, the confidence map of the network and the results of DE can be further processed by the classic threshold segmentation method, so that some high-confidence residential areas without electricity can be added to the intersection result. In Fig. 8, E-num represents the number of powered pixels in each tile, while Voting-num represents the number of positive results in the voting strategy (DS-GC-CNN). Confi is the confidence value of each pixel in DSE-GC-CNN.

D. Postprocessing

After the abovementioned steps, there are still some noise points in the multimodel fusion results, which can be further refined by threshold segmentation. First, we suppose that when the number of settlements without electricity in a tile is less

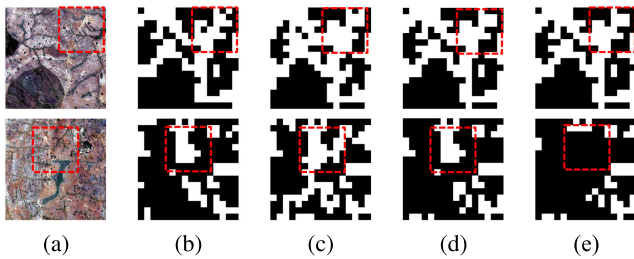


Fig. 9. Detection results of different methods and periods. (a) RGB image. (b) Result of single-task model. (c) Result of dual-task model. (d) Result of model fusion. (e) Result of postprocessing.

TABLE IV
PERFORMANCE ON THE TEST DATA IN THE CONTEST

Method	F1	Recall	Precision
Single-task Model	0.8623	0.8857	0.8401
Dual-task Model	0.8581	0.9014	0.8188
Model Fusion	0.8744	0.8717	0.8771
Post-processing	0.8806	0.8711	0.8896

than three, the entire tile has no residential areas. In addition, when each tile has a tile with E-num greater than 30, pixels with a maximum value between 18 and 30 and pixels with a total value greater than 30 are regarded as powered pixels and will be excluded.

E. Results and Analysis

The multitemporal and multimodal experimental data are provided by the organizer. The preprocessed training data is split with 80% used for model training and 20% used for model validation. The final performance of our proposed approach is evaluated on the test data of 19 tiles by F1 score, which is a comprehensive evaluation index that combines the precision and recall rate.

The visualization detection results of diverse models and periods are shown in Fig. 9, where the results of intermediate processes and methods including single-task model, dual-task model, model fusion, and postprocessing are compared. The corresponding quantitative results of the detection of settlements without electricity are displayed in Table IV. Several conclusions can be drawn by comparing visual results and quantitative results. First, in the both results of the single-task model and the dual-task model, the precision value is lower than the recall value, which indicates that the problem of overdetection exists in both two models. In addition, although the F1 score values of the two models are close, the visual detection results are inconsistent. The recall value of the dual-task model is higher than that of the single-task model, and the precision value of the single-task model is higher than that of the dual-task model. Besides, the F1 score value rises and the recall value and precision value is closer to each other after model fusion of the single-task model and dual-task model, demonstrating the integration of different models works. Finally, the postprocessing step makes an improvement in terms of F1 score and the precision value. Overall, the experiment results have illustrated

that our multimodel fusion framework effectively improves the final prediction results.

VI. CONCLUSION

Remote sensing and earth observation can and do play a key role for reaching the sustainable development goals defined by the UN. They are of direct relevance across multiple targets, goals, and indicators - among them target 7.1 to “ensure universal access to affordable, reliable, and modern energy services”. Remote sensing makes on of its strongest contributions in the monitoring of remote areas: A multitude of space-borne sensors with various spectral properties provide imagery with a high spatial and spectral resolution. This data can be leveraged to automatically detect settlements as well as indications regarding their access to electricity.

In this article, we summarize the Track DSE of the 2021 IEEE GRSS Data Fusion Contest, organized by the IEEE GRSS Image Analysis and Data Fusion Technical Committee, that was dedicated to exactly this real-world social challenge: To detect settlements without access to electricity. We describe the challenge to create well-performing machine learning models for the detection of settlements without electricity based on multisensor imagery if only noisy, low-resolution labels are available for training. To this aim, Track DSE provided Sentinel-1 polarimetric SAR, Sentinel-2, and Landsat 8 multispectral data, as well as VIIRS night time data. The images are geographically distributed over two different sites and split into 98 800 × 800 pixel large tiles that were further split into disjoint training, validation and test sets. Additionally, reference data was provided, which consisted of carefully manually annotated semantic maps that indicate the presence/absence of human settlements and access to electricity. The very different image content (e.g., optical versus SAR) as well as sensor characteristics (e.g., large differences in spatial resolution) as well as low resolution labels made the contest challenging yet realistic.

This is also reflected by the strategies the participants employed to address these challenges. The top ranking teams all heavily relied on data preprocessing such as normalization to handle the diversity of the input images. Data augmentation was used to increase the amount of training data, while ensemble learning, model fusion, and/or test-time data augmentation helped to mitigate the label noise caused by the low resolution of the reference data. Interestingly, at least within the context of this contest, the Sentinel-2 multispectral images were sufficient to detect settlements. While some participants decoupled settlement and electricity detection and fused the respective results, others addressed the two together. Both approaches yielded a similar performance. While deep neural networks were clearly the dominant machine learning method, other, more traditional approaches showed to be useful as well, including data normalization, simple thresholding, and Random Forests.

The four top ranked solutions of this track presented their methods at IGARSS 2021, while the two top ranking solutions are described in this article in more detail. As in previous years, the DFC2021 attracted global attention with participants well distributed over the world, different affiliations, and

career stages. This clearly illustrates the interest of the remote sensing and earth observation community to use the available tools and expertise to contribute to the social good. Furthermore, many of contest participants were students, which shows that the data fusion contest is introduced to early career scientists and used for educational purposes.

The data remains accessible after the DFC2021 on the globally-accessible data platform IEEE DataPort² to allow further research and contributions. The CodaLab evaluation server and its public leaderboard³ was reopened and made accessible from the contest website⁴. Thus, anyone can submit prediction results, obtain performance statistics, compare to other users and hopefully improve on the results presented in this article.

The work described in this article is a first step in the journey helping Solaraid in the detection of settlements without access to electricity at a large geographical scale. The effort made by the different contributors is an invaluable basis to support the collaboration involving Solaraid, Hewlett Packard Enterprise, and DSE. Additional work has already been started to automate the data preparation for a detection at the level of a whole country. Besides, certain teams have expressed their interest in pursuing further experiments to help scaling up the work to a continental level. Discussions are also ongoing to understand the optimal output layout of the algorithm for an easy adoption by the field and an actually operational service.

In addition to the societal impact, the DFC2021 Track DSE states a very interesting challenge from a scientific point of view. Questions like if and how multiple and diverse sensor data should be fused to improve predictions, how to process different spatial and spectral resolutions within one common framework, how to mitigate issues due to label noise or otherwise degraded reference data (e.g., lower resolutions), and how to deal with a limited amount of labeled data in general are far from being solved. Instead, all of these questions represent very active research directions where the future promises significant advancements. In this regard, the data of the DFC2021 provides a valuable benchmark dataset that can be used to evaluate all or only some of these aspects.

ACKNOWLEDGMENT

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

REFERENCES

- [1] N. Yokoya *et al.*, "2021 data fusion contest: Geospatial artificial intelligence for social good [technical committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 287–C3, Mar. 2021.
- [2] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.
- [3] F. Pacifici, F. Del Frate, W. J. Emery, P. Gamba, and J. Chanussot, "Urban mapping using coarse SAR and optical data: Outcome of the 2007 GRSS data fusion contest," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 331–335, Jul. 2008.
- [4] G. Licciardi *et al.*, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, Nov. 2009.
- [5] N. Longbotham *et al.*, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 331–342, Feb. 2012.
- [6] F. Pacifici and Q. Du, "Foreword to the special issue on optical multiangular data exploitation and outcome of the 2011 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 3–7, Feb. 2012.
- [7] C. Berger *et al.*, "Multi-modal and multi-temporal data fusion: Outcome of the 2012 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1324–1340, Jun. 2013.
- [8] C. Debes *et al.*, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.
- [9] W. Liao *et al.*, "Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2984–2996, Jun. 2015.
- [10] M. Campos-Taberner *et al.*, "Processing of extremely high resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS data fusion contest Part A: 2D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5547–5559, Dec. 2016.
- [11] A.-V. Vo *et al.*, "Processing of extremely high resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS data fusion contest Part B: 3D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5560–5575, Dec. 2016.
- [12] L. Mou *et al.*, "Multi-temporal very high resolution from space: Outcome of the 2016 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3435–3447, Aug. 2017.
- [13] 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.
- [14] 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.
- [15] B. Le Saux, N. Yokoya, R. Hänsch, M. Brown, and G. Hager, "2019 data fusion contest [technical committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 103–105, Mar. 2019.
- [16] 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.
- [17] 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, 2021.
- [18] 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, 2021.
- [19] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, "Sen2Cor for Sentinel-2" in *Image and Signal Processing for Remote Sensing XXIII*, L. Bruzzone, Ed., vol. 10427. Bellingham, WA, USA: SPIE, 2017, pp. 37–48.
- [20] M. O. Román *et al.*, "NASA's black Marble nighttime lights product suite," *Remote Sens. Environ.*, vol. 210, pp. 113–143, Jun. 2018.
- [21] Y. Ma *et al.*, "Multisource data fusion for the detection of settlements without electricity," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 1839–1842.
- [22] Y. Xia, Q. Huang, and H. Zhang, "A multi-model fusion of convolution neural network and random forest for detecting settlements without electricity," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 1843–1846.
- [23] T. D. Martino, M. Lenormand, and E. C. Koeniguer, "Multi-branch deep learning model for detection of settlements without electricity," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 1847–1850.
- [24] R. Feng *et al.*, "DO-UNet, DO-LinkNet: UNet, D-LinkNet with DO-Conv for the detection of settlements without electricity challenge," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 1851–1854.

²[Online]. Available: <https://iee-dataport.org/competitions/2021-ieee-grss-data-fusion-contest-track-dse>

³[Online]. Available: <https://competitions.codalab.org/competitions/22289>

⁴[Online]. Available: <https://www.grss-ieee.org/community/technical-committees/2021-ieee-grss-data-fusion-contest-track-dse/>

- [25] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
- [26] M. Schmitt, J. Prexl, P. Ebel, L. Liebel, and X. X. Zhu, "Weakly supervised semantic segmentation of satellite images for land cover mapping—challenges and opportunities," 2020, *arXiv:2002.08254*.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [28] C. Liu *et al.*, "Progressive neural architecture search," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 19–34.
- [29] H. Xu, "A new index for delineating built-up land features in satellite imagery," *Int. J. Remote Sens.*, vol. 29, no. 14, pp. 4269–4276, 2008.
- [30] H. Zhang, Q. Huang, H. Zhai, and L. Zhang, "Multi-temporal cloud detection based on robust PCA for optical remote sensing imagery," *Comput. Electron. Agriculture*, vol. 188, 2021, Art. no. 106342.
- [31] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [32] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 1971–1980.
- [33] T.-Y. Ross and G. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2980–2988.

Yanbiao Ma received the B.S. degree in intelligence science and technology in 2020 from Xidian University, Xi'an, China, where he is currently working toward the Ph.D. degree in computer science and technology with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education.

His research interests include deep metric learning, few-shot learning, and data augmentation.

Yuxin Li received the B.S. degree in information management and information system from the North University of China, Taiyuan, China, in 2020. She is currently working toward the M.S. degree in computer science and technology with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, China.

Her research interests include deep metric learning and multimodal retrieval.

Kexin Feng received the B.S. degree in communication engineering from Jilin University, Changchun, China, in 2020. She is currently working toward the M.Eng. degree in electronic and information engineering with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, China.

Her research interests include deep learning and object detection for remote sensing images.

Yu Xia received the B.S. degree in photogrammetry and remote sensing in 2018 from Wuhan University, Wuhan, China, where he is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing.

His research interests include cloud removal, image synthesis, and deep learning.

Qi Huang received the B.S. degree in photogrammetry and remote sensing in 2019 from Wuhan University, Wuhan, China, where she is currently working toward the M.S. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing.

Her research interests include cloud detection for optical remote sensing images and deep learning.

Hongyan Zhang (Senior Member, IEEE) received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

Dr. Zhang has been currently a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, since 2016. He is a Young Chang-Jiang Scholar appointed by the Ministry of Education of China. He is a Reviewer of more than 30 international academic journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and so on.

Colin Prieur received the Graduate degree in signal processing from the Grenoble Institute of Technology, Grenoble, France, in 2022.

In 2020 and 2021, he was a Research Engineer with Data Science Experts (Grenoble, France), leading several projects involving the development of cutting-edge deep learning algorithms for the processing and analysis of remote sensing data. He also spent a four months research stay with the Norwegian University of Science and Technology, Gjøvik, Norway, working on remote sensing of glaciers.

Giorgio Licciardi received the M.S. degree in telecommunication engineering and the Ph.D. degree in "geoinformation" from the Tor Vergata University, Rome, Italy, in 2005 and 2010, respectively.

In 2010, he was with the Laboratoire Grenoblois de l'Image, de la Parole, du Signal et de l'Automatique as a Postdoctoral Fellow. From 2014, he is a Research Engineer with the Grenoble Institute of Technology, and from 2019, he was the Grenoble Multidisciplinary Institute of Artificial Intelligence as Senior Researcher. His main research interests include hyperspectral image processing, including feature extraction techniques, spectral unmixing, super-resolution, and pansharpening.

Dr. Licciardi serves as a Referee for several scientific journals such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Hana Malha received the M.Sc. degree in telecommunication and network engineering from the National Institute of Applied Sciences (INSA Lyon), Bellignat, France, in 2012.

She was with the HPC&AI Competency Center, Grenoble, France, where she focused on High Performance Computing benchmarking and postsales activity. Since 2018, she has been working on the optimization of AI workloads, through the evaluation of a variety of accelerators and software optimization tools for both training and inference.

Jocelyn Chanussot (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree in electrical engineering from the Université de Savoie, Annecy, France, in 1998.

Since 1999, he has been with Grenoble INP, where he is currently a Professor of signal and image processing. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning, and artificial intelligence.

Prof. Chanussot has been a visiting scholar with Stanford University, KTH (Sweden), and NUS (Singapore). Since 2013, he is an Adjunct Professor with the University of Iceland. In 2015–2017, he was a visiting professor with the University of California, Los Angeles, USA. He holds the AXA chair in remote sensing and is an Adjunct Professor with the Chinese Academy of Sciences, Aerospace Information research Institute, Beijing, China. He is an honorary member of the Institut Universitaire de France and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters, since 2018).

Pedram Ghamisi (Senior Member, IEEE) received the M.Sc. (hons.) degree in remote sensing from the K. N. Toosi University of Technology, Tehran, Iran, in 2012, and the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2015.

He is the Head of the machine learning group, Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology (HZDR-HIF), Germany; Co-founder of VasoGnosis Inc., with two branches in the USA; Visiting Professor and Leader of AI4RS at the Institute of Advanced Research in Artificial Intelligence, Austria. He was the Vice-Chair of the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee (2019–2021).

Ronny Hänsch (Senior Member, IEEE) received the Diploma degree in computer science and the Ph.D. degree from the Technische Universität Berlin, Berlin, Germany, in 2007 and 2014, respectively.

His current research interest includes ensemble methods for image analysis.

Dr. Hänsch was Co-Chair (2017–2021) and is the current chair (2021–2023) of the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee and cochair of the International Society for Photogrammetry and Remote Sensing Working Group II/1 (Image Orientation). He serves as Associate Editor for the Geoscience and Remote Sensing Letters and as Editor in Chief for the GRSS eNewsletter.

Naoto Yokoya (Member, IEEE) received the M.Eng. and Ph.D. degrees in aerospace engineering from the University of Tokyo, Tokyo, Japan, in 2010 and 2013, respectively.

He is a Lecturer with the University of Tokyo and a unit leader with the RIKEN Center for Advanced Intelligence Project, Tokyo, where he leads the Geoinformatics Unit. He was the chair (2019–2021) of the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee. He is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.