

EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification

Patrick Helber¹, Benjamin Bischke², Andreas Dengel, and Damian Borth

Abstract—In this paper, we present a patch-based land use and land cover classification approach using Sentinel-2 satellite images. The Sentinel-2 satellite images are openly and freely accessible, and are provided in the earth observation program Copernicus. We present a novel dataset, based on these images that covers 13 spectral bands and is comprised of ten classes with a total of 27 000 labeled and geo-referenced images. Benchmarks are provided for this novel dataset with its spectral bands using state-of-the-art deep convolutional neural networks. **An overall classification accuracy of 98.57% was achieved with the proposed novel dataset.** The resulting classification system opens a gate toward a number of earth observation applications. We demonstrate how this classification system can be used for detecting land use and land cover changes, and how it can assist in improving geographical maps. The geo-referenced dataset EuroSAT is made publicly available at <https://github.com/phelber/eurosat>.

Index Terms—Dataset, deep convolutional neural network, deep learning, earth observation, land cover classification, land use classification, machine learning, remote sensing, satellite image classification, satellite images.

I. INTRODUCTION

WE ARE currently on the edge of public and continuous access to satellite image data for earth observation. Governmental programs such as ESA's Copernicus and NASA's Landsat are making significant efforts to make such data freely available for commercial and non-commercial purposes with the intention to fuel innovation and entrepreneurship. With access to such data, applications can be realized in the domains of agriculture, disaster recovery, climate change, urban development, and environmental monitoring [2], [3], [5], [36]. However, to fully utilize the data for such applications, satellite images must first be processed and transformed into structured semantics [35]. One such type of fundamental semantics is *land use and land*

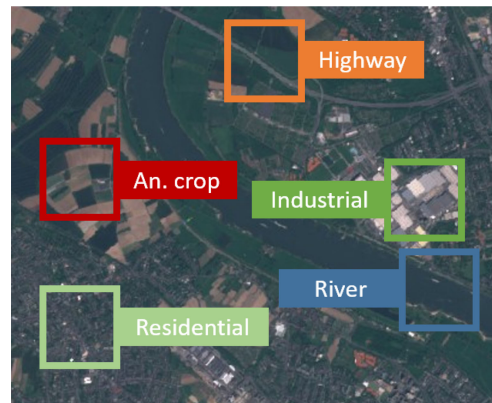


Fig. 1. Land use and land cover classification based on Sentinel-2 satellite images. Patches are extracted for class identification. This visualization highlights the classes: Annual Crop, River, Highway, Industrial Buildings, and Residential Buildings.

cover classification [1], [29]. **The aim of land use and land cover classification is to automatically provide labels describing the represented physical land type and usage** (e.g., residential, industrial, etc.). Fig. 1 shows an example of land use and land cover classification based on a Sentinel-2 satellite image. The illustration in Fig. 2 shows an overview of land use and land cover classification process using satellite images.

As is often observed in supervised machine learning, the performance of classification systems strongly depends on the availability of high-quality datasets with a suitable set of classes [21]. In particular, when considering the recent success of deep *convolutional neural networks (CNNs)* [12], having large quantities of training data available is crucial in order to train such a network. Unfortunately, current land use and land cover datasets are small-scale or rely on data sources that do not allow for the mentioned domain applications.

In this paper, we propose a novel satellite image dataset for the task of land use and land cover classification. The proposed EuroSAT dataset consists of 27 000 labeled images with ten different land use and land cover classes. **The proposed dataset is unique from previous datasets in which it is multi-spectral, covering 13 spectral bands in the visible, near infrared, and short wave infrared parts of the spectrum.** In addition, the proposed dataset is geo-referenced and based on openly and freely accessible earth observation data, and therefore, allowing for a unique range of applications. The labeled dataset EuroSAT is made

Manuscript received September 15, 2018; revised January 8, 2019 and March 21, 2019; accepted May 4, 2019. Date of publication June 13, 2019; date of current version July 30, 2019. This work was supported in part by the BMBF Project DeFuseNN under Grant 01IW17002 and in part by Nvidia. (Corresponding author: Patrick Helber.)

P. Helber, B. Bischke, and A. Dengel are with the Technische Universität Kaiserslautern, Kaiserslautern 67663, Germany, and also with the German Research Center for Artificial Intelligence, Kaiserslautern 67663, Germany (e-mail: Patrick.Helber@dfki.de; Benjamin.Bischke@dfki.de; Andreas.Dengel@dfki.de).

D. Borth is with the Institute for Computer Science, University of St. Gallen, St. Gallen 9000, Switzerland (e-mail: damian.borth@unisg.ch).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2019.2918242

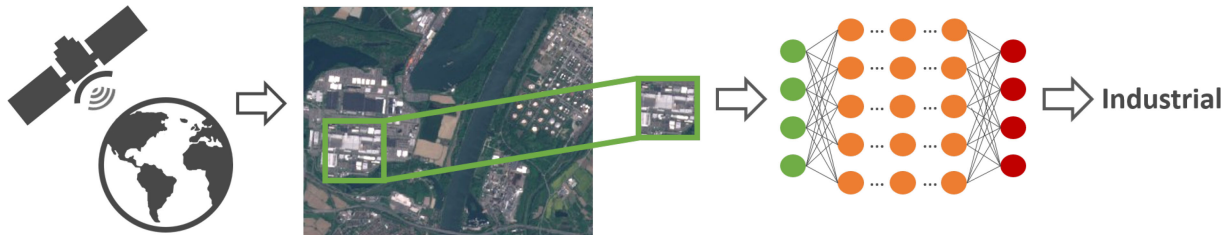


Fig. 2. This illustration shows an overview of the patch-based land use and land cover classification process using satellite images. A satellite scans the earth acquiring images. Patches are then extracted out of these images to be used for classification. The aim is to automatically provide labels describing the represented physical land type or land usage. To do so, an image patch is fed into a classifier (e.g., a neural network, as shown in this illustration) and the classifier outputs the class shown on the image patch.

publicly available at <https://github.com/phelber/eurosat>. Further, we provide a full benchmark demonstrating robust classification performance, which is the basis for developing applications in the previously mentioned domains. We outline how the classification model can be used for detecting land use or land cover changes and how it can assist in improving geographical maps.

We provide this paper in the context of the recently published EuroSAT dataset, which can be used similar to [18] as a basis for large-scale training of deep neural networks for the task of satellite image classification.

In this paper, we make the following *contributions*:

- 1) We introduce the first large-scale patch-based land use and land cover classification dataset based on Sentinel-2 satellite images. Every image in the dataset is labeled and geo-referenced. We release the RGB and the multi-spectral (MS) version of the dataset.
- 2) We provide benchmarks for the proposed EuroSAT dataset using CNNs.
- 3) We evaluate the performance of each spectral band of the Sentinel-2 satellite for the task of patch-based land use and land cover classification.

II. RELATED WORK

In this section, we review previous studies concerning land use and land cover classification. In this context, we present remotely sensed aerial and satellite image datasets. Furthermore, we review state-of-the-art image classification methods for land use and land cover classification.

A. Classification Datasets

The classification of remotely sensed images is a challenging task. The progress of classification in the remote sensing area has particularly been inhibited because of the lack of reliably labeled ground truth datasets. A popular and intensively studied [6], [19], [20], [27], [29] remotely sensed image classification dataset known as *UC Merced* (UCM) land use dataset was introduced by Yang *et al.* [29]. The dataset consists of 21 land use and land cover classes. Each class contains 100 images, each measuring 256×256 pixels with a spatial resolution of about 30 cm/pixel. All images are in the RGB color space and were extracted from the USGS National Map Urban Area Imagery

collection, i.e., the underlying images were acquired from an aircraft. Unfortunately, a dataset with 100 images per class is of small-scale. In an attempt to enhance the dataset situation, various works used commercial Google Earth images to manually create novel datasets, [22], [27], [28], [30] such as the two benchmark datasets PatternNet [37] and NWPU-RESISC45 [7]. The datasets are based on very-high-resolution images with a spatial resolution of up to 30 cm/pixel. Since the creation of a labeled dataset is extremely time-consuming, these datasets, thus, consist of only a few hundred images per class. One of the largest datasets is the *aerial image dataset* (AID) [27]. The AID dataset consists of 30 classes with 200–400 images per class. The 600×600 high-resolution images were also extracted from Google Earth imagery.

Compared to the EuroSAT dataset presented in this paper, the previously listed datasets rely on commercial very-high-resolution and preprocessed images. Using commercial and preprocessed very-high-resolution image data makes these datasets unsatisfying for real-world Sentinel-2 earth observation applications as proposed in this paper. Furthermore, while these datasets put a strong focus on strengthening the number of covered classes, the datasets suffer from a low number of images per class. With a spatial resolution of up to 30 cm/pixel and the possibility of identifying and distinguishing classes like churches, schools, etc., these previously defined datasets are difficult to compare to the dataset proposed in this paper.

A study more similar to this paper, provided by Penatti *et al.* [20], analyzed remotely sensed satellite images with a spatial resolution of 10 m/pixel to classify coffee crops. On the basis of these images, Penatti *et al.* [20] introduced the *Brazilian Coffee Scene (BCS) dataset*. The BCS dataset covers the two classes: coffee crop and non-coffee crop. Each class is comprised of 1423 images. The images consist of a red, green, and near-infrared band.

Similar to the currently proposed EuroSAT dataset, Basu *et al.* [1] introduced the *SAT-6* dataset relying on aerial images. This dataset was extracted from images with a spatial resolution of 1 m/pixel. The image patches were created using images from the National Agriculture Imagery Program (NAIP). The SAT-6 covers the six different classes: Barren Land, Trees, Grassland, Roads, Buildings, and Water Bodies. The proposed patches have a size of 28×28 pixels per image and consist of a red, green, blue, and a near-infrared band.

B. Land Use and Land Cover Classification

CNNs are a type of neural network [13], which through impressive results on image classification challenges [12], [21], [23] became well known as the state-of-the-art image classification method in computer vision and machine learning.

To classify remotely sensed images, various different feature extraction and classification methods (e.g., random forests) were evaluated on the basis of introduced datasets. Yang *et al.* evaluated Bag-of-Visual-Words (BoVW) and spatial extension approaches on the UCM dataset [29]. Basu *et al.* analyzed deep belief networks, basic CNNs, and stacked denoising autoencoders on the SAT-6 dataset [1]. Basu *et al.* also presented their own framework for the land cover classes introduced in the SAT-6 dataset. The framework extracts features from the input images, normalizes the extracted features, and uses the normalized features as an input to a deep belief network. Along with low-level color descriptors, Penatti *et al.* also evaluated deep CNNs on the UCM and BCS dataset [20]. In addition to deep CNNs, Castelluccio *et al.* intensively evaluated various machine learning methods (e.g., BoVWs, spatial pyramid match kernels) for the classification of the UCM and BCS dataset.

In the context of deep learning, the deep CNNs used in this paper have been trained from scratch or fine-tuned by using a pretrained network [6], [7], [16], [19], [31]. The networks were mainly pretrained on the ILSVRC-2012 image classification challenge [21] dataset. Even though these pretrained networks were trained on images from an entirely different domain, the features generalized well. Therefore, the pretrained networks proved to be suitable for the classification of remotely sensed images [17]. The presented work extensively evaluated all proposed machine learning methods and concluded that deep CNNs outperformed non-deep learning approaches on the considered datasets [6], [15], [17], [27].

III. DATASET ACQUISITION

Alongside NASA with its Landsat Mission, the European Space Agency (ESA) is making increased efforts to improve earth observation within its Copernicus program. Under this program, ESA operates a series of satellites known as Sentinels.

In this paper, we use multi-spectral image data, provided by the Sentinel-2A satellite, in order to address the challenge of land use and land cover classification. The Sentinel-2A is one of the two satellites in the two-satellite constellation consisting of the identical land monitoring satellites Sentinel-2A and Sentinel-2B. The satellites were successfully launched in June 2015 (Sentinel-2A) and March 2017 (Sentinel-2B). Both sun-synchronous satellites capture earth's global land surface with a multispectral imager (MSI) covering the 13 different spectral bands listed in Table I. The three bands B01, B09, and B10 are intended to be used in the correction of atmospheric effects (e.g., aerosols, cirrus, or water vapor). The remaining bands are primarily intended for identifying and monitoring land use and land cover classes. In addition to mainland coverage, large islands as well as inland and coastal waters are covered by these two satellites. Each satellite will deliver imagery for at least seven years with a spatial resolution of up to 10 m/pixel. Both satellites carry

TABLE I
ALL 13 BANDS COVERED BY SENTINEL-2'S MULTISPECTRAL IMAGER (MSI)

| Band | Spatial Resolution <i>m</i> | Central Wavelength <i>nm</i> |
|-------------------|--------------------------------|---------------------------------|
| B01 - Aerosols | 60 | 443 |
| B02 - Blue | 10 | 490 |
| B03 - Green | 10 | 560 |
| B04 - Red | 10 | 665 |
| B05 - Red edge 1 | 20 | 705 |
| B06 - Red edge 2 | 20 | 740 |
| B07 - Red edge 3 | 20 | 783 |
| B08 - NIR | 10 | 842 |
| B08A - Red edge 4 | 20 | 865 |
| B09 - Water vapor | 60 | 945 |
| B10 - Cirrus | 60 | 1375 |
| B11 - SWIR 1 | 20 | 1610 |
| B12 - SWIR 2 | 20 | 2190 |

The identification, the spatial resolution, and the central wavelength are listed for each spectral band.

fuel for up to 12 years of operation, which allows for an extension of the operation. The two-satellite constellation generates a coverage of almost the entire earth's land surface approximately every five days, i.e., the satellites capture each point in the covered area about every five days. This short repeat cycle, as well as the future availability of the Sentinel satellites, allows for continuous monitoring of the earth's land surface for the next 20–30 years. Most importantly, the data are openly and freely accessible and can be used for any application (commercial or non-commercial).

We are convinced that the large volume of satellite data in combination with powerful machine learning methods will influence future research. Therefore, one of our key research aims is to make this large amount of data accessible for machine learning based applications. To construct an image classification dataset, we performed the following two steps:

- 1) Satellite image acquisition: We gathered satellite images of European cities distributed in over 34 countries as shown in Fig. 5.
- 2) Dataset creation: On the basis of the obtained satellite images, we created a dataset of 27 000 geo-referenced and labeled image patches. The image patches measure 64×64 pixels and have been manually checked.

A. Satellite Image Acquisition

We have downloaded satellite images taken by the satellite Sentinel-2A via Amazon S3. We chose satellite images associated with the cities covered in the European Urban Atlas. The covered cities are distributed over the 34 European countries: Austria, Belarus, Belgium, Bulgaria, Cyprus, Czech Republic (Czechia), Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy/Holy See, Latvia, Lithuania, Luxembourg, Macedonia, Malta, Republic of Moldova, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, and United Kingdom.

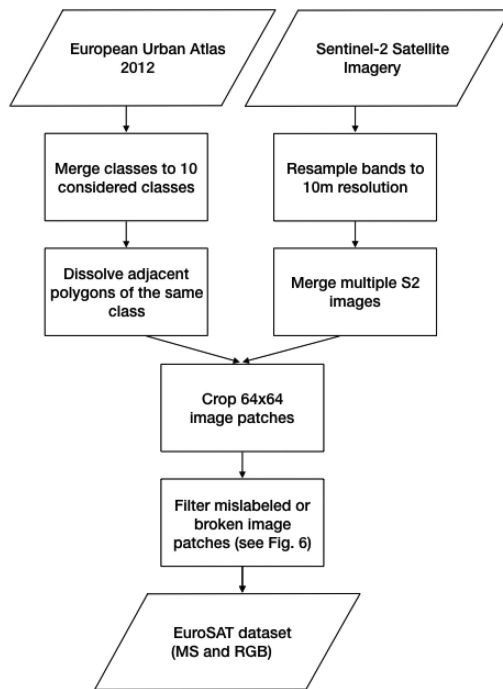


Fig. 3. Diagram illustrates the EuroSAT dataset creation process.

In order to improve the chance of attaining valuable image patches, satellite images with a low cloud level were selected. Along with the possibility to generate a cloud mask, ESA also provides a cloud level value for each satellite image; thus, allowing for quick selections of images with low percentages of clouds covering the land scene.

We aimed with the objective of covering as many countries as possible in the EuroSAT dataset in order to cover the high intra-class variance inherent to remotely sensed images. Furthermore, we have extracted images recorded throughout the year to attain a variance as high as possible within the covered land use and land cover classes. Within each class of the EuroSAT dataset, different land types of the class are represented, such as different types of forests in the Forest class or different types of industrial buildings in the Industrial Building class. Between the classes, there is a low positive correlation. The classes most common to each other are the two presented agricultural classes and the two classes representing residential and industrial buildings. The composition of the individual classes and their relationships are specified in the mapping guide of the European Urban Atlas [38]. An overview diagram of the dataset creation process is shown in Fig. 3.

The proposed EuroSAT dataset poses challenges for classification stemming from several sources of intra-class variation. For example, the class River (see Fig. 4) contains patches where the width of the stream or its curvature varies significantly. Similarly, the angle of the streets in the Highway class are randomly distributed, making low level features insufficient for classification. On the other hand, a limited inter-class variation can be observed between different classes, such as Pasture and Forest. Moreover, similar textures can be observed between Industrial Buildings and Residential Buildings as well as between the two

kinds of crops defined in the dataset. The mentioned properties make this classification task much more challenging as models cannot rely on simple low level features such as color histograms or edge detection methods. Similar properties are present in the other classes (also see the classes Sea & Lake and Herbaceous Vegetation in Fig. 4).

B. Dataset Creation

The Sentinel-2 satellite constellation provides approximately 1.6 TB of compressed images per day. Unfortunately, supervised machine learning is restricted by the lack of labeled ground truth data, even with this amount of data. There were two motivations for the generation of the benchmarking EuroSAT dataset: first, the objective of making this already open and free satellite data accessible to various earth observation applications, and second, the observation that existing benchmark datasets are not suitable for the intended applications with Sentinel-2 satellite images. The dataset consists of 10 different classes with 2000–3000 images per class. In total, the dataset has 27 000 images. The patches measure 64×64 pixels. We have chosen ten different land use and land cover classes based on the criteria that they are distinguishable at a resolution of 10 m/pixel and are frequently covered by the European Urban Atlas, thus are able to generate thousands of image patches. To differentiate between different agricultural land uses, the proposed dataset covers the classes: Annual Crop, Permanent Crop (e.g., fruit orchards, vineyards, or olive groves), and Pastures. The dataset also discriminates built-up areas by covering the classes: Highway, Residential Buildings, and Industrial Buildings. The Residential Buildings class was created using the urban fabric classes described in the European Urban Atlas. Different water bodies appear in the class River and in the class Sea & Lake. Furthermore, undeveloped environments, such as forest and herbaceous vegetation are included, each in their own class. An overview of the covered classes with four samples per class is shown in Fig. 4.

We manually checked all 27 000 images multiple times and corrected the ground truth by sorting out mislabeled images as well as images complicated by snow or ice. Examples of discarded images are shown in Fig. 6. The sample images are intended to show industrial buildings. Clearly, however, no industrial building is visible in the image. Note that the proposed dataset has not received atmospheric correction, which can result in images with a color cast. Extreme cases are visualized in Fig. 7. With the intention of the classifier also learning these cases, we did not filter the respective samples and let them flow into the dataset.

Along with the 13 covered spectral bands, the new dataset also has three further central innovations. First, the dataset is based on free satellite images, unlike Google Earth imagery, and does not rely on data sources that are not updated on a high-frequency basis (e.g., NAIP used in [1]). Instead, an open and free earth observation program whose satellites deliver images for the next decades is used allowing real-world earth observation applications. Second, the dataset uses a ten times lower spatial resolution than that of the benchmark dataset most similar to our research, and also distinguishes ten classes instead of six.

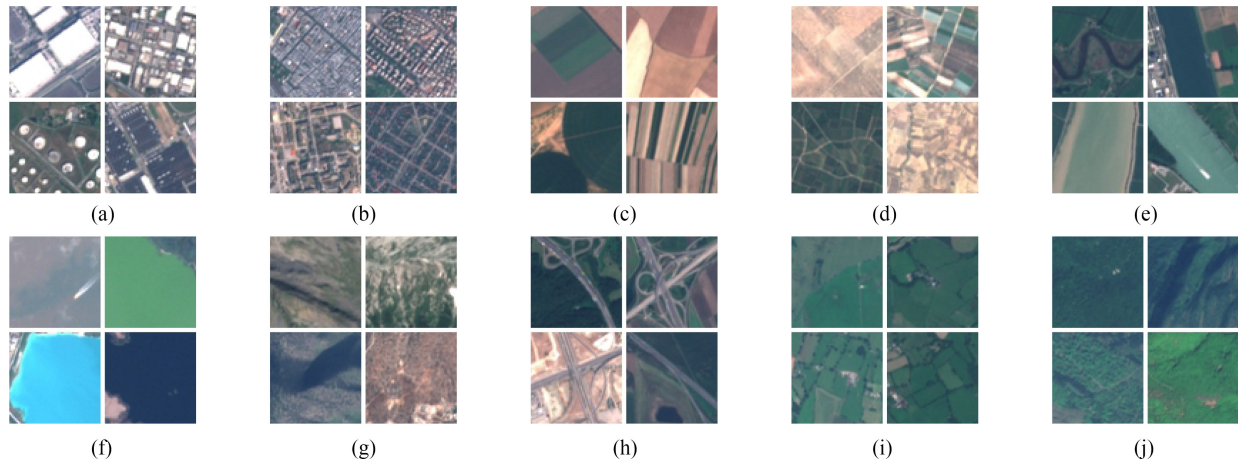


Fig. 4. This overview shows sample image patches of all ten classes covered in the proposed EuroSAT dataset. The images measure 64×64 pixels. Each class contains 2000–3000 images. In total, the dataset has 27 000 geo-referenced images. (a) Industrial buildings. (b) Residential buildings. (c) Annual crop. (d) Permanent crop. (e) River. (f) Sea and lake. (g) Herbaceous vegetation. (h) Highway. (i) Pasture. (j) Forest.

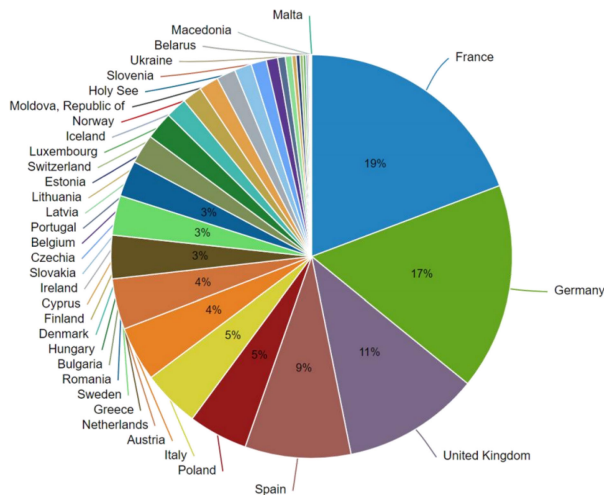


Fig. 5. EuroSAT dataset distribution. The geo-referenced images are distributed all across Europe. The distribution is influenced by the number of represented cities per country in the European Urban Atlas.

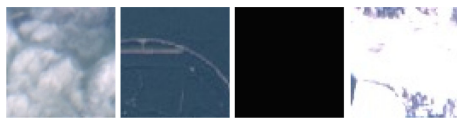


Fig. 6. Four examples of unideal image patches, which are intended to show industrial buildings. Clearly, no industrial building is recognizable due to clouds, mislabeling, dead pixels, or ice/snow.

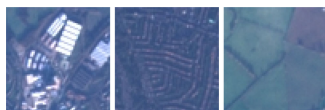


Fig. 7. Three image patches affected by color casts due to atmospheric effects.

For instance, we distinguish between built-up class, separating into a residential and an industrial class, and distinguish between different agricultural land uses. Third, we release the EuroSAT dataset in a geo-referenced version.

With the release of the geo-referenced EuroSAT dataset, we aim to make the large amount of Sentinel-2 satellite imagery accessible for machine learning. The effectiveness of these approaches was successfully demonstrated in [32]–[34].

IV. DATASET BENCHMARKING

As shown in previous work [6], [15], [17], [19], deep CNNs have been shown to outperform non-deep learning approaches in land use and land cover image classification. Accordingly, we use the state-of-the-art deep CNN models GoogleNet [25] and ResNet-50 [9], [10] for the classification of the introduced land use and land cover classes. The networks make use of the inception module [14], [24]–[26] and the residual unit [9], [10]. For the proposed EuroSAT dataset, we also evaluated the performance of the 13 spectral bands with respect to the classification task in which we evaluated the classification performance using single-band and band combination images.

A. Comparative Evaluation

Each dataset was divided into a training set and a test set (80/20 ratio, respectively). We ensured that the split was applied class-wise. While three (the red, green, and blue) bands have already been covered by almost all aerial and satellite image datasets, the proposed EuroSAT dataset consists of 13 spectral bands. For the comparative evaluation, we computed images in the RGB color space by combining the three bands: red (B04), green (B03), and blue (B02). All images have been rescaled to 224×224 pixels. The 16-bit Sentinel-2 satellite images were converted to 8-bit images by projecting values in the range 0–2750 onto values in the range 1–255 (with a no data value of 0).

TABLE II
CLASSIFICATION ACCURACY (%) OF DIFFERENT TRAINING-TEST SPLITS ON THE EUROSAT DATASET

| Method | 10/90 | 20/80 | 30/70 | 40/60 | 50/50 | 60/40 | 70/30 | 80/20 | 90/10 |
|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| BoVW (SVM, SIFT, k = 10) | 54.54 | 56.13 | 56.77 | 57.06 | 57.22 | 57.47 | 57.71 | 58.55 | 58.44 |
| BoVW (SVM, SIFT, k = 100) | 63.07 | 64.80 | 65.50 | 66.16 | 66.25 | 66.34 | 66.50 | 67.22 | 66.18 |
| BoVW (SVM, SIFT, k = 500) | 65.62 | 67.26 | 68.01 | 68.52 | 68.61 | 68.74 | 69.07 | 70.05 | 69.54 |
| CNN (three layers) | 79.65 | 87.24 | 88.75 | 90.74 | 92.48 | 92.62 | 92.34 | 92.61 | 93.62 |
| ResNet-50 | 87.36 | 90.00 | 92.35 | 93.70 | 94.41 | 95.26 | 96.30 | 95.43 | 96.37 |
| GoogLeNet | 87.42 | 90.97 | 92.17 | 93.26 | 94.85 | 95.54 | 96.69 | 96.60 | 96.17 |

TABLE III
HYPERPARAMETERS USED IN THE EXPERIMENTS

| Parameter | Value |
|-----------------------|--------------------------|
| Initial learning rate | 1e-3 |
| Epochs | 120 |
| Batch size | 16 |
| Decay | 1e-6 |
| Momentum (Nesterov) | 0.9 |
| Optimizer | SGD |
| Loss | Categorical Crossentropy |

For benchmarking, we evaluated the performance of the BoVW approach using SIFT features and a trained SVM. In addition, we trained a shallow CNN (three layers), a ResNet-50, and a GoogLeNet model on the training set. The shallow CNN consists of three layers each composed of a 3×3 convolutional layer (stride = 1, activation = relu) and a subsequent 4×4 max pooling layer (stride = 2). The layers are followed by a fully connected layer. Table III lists the used hyperparameters. Throughout the training of the CNN models, multiple data augmentation strategies have been used. The used strategies include horizontal image flipping, shearing with a range of 0.2, and random zooming with a range of 0.2. In the training process, we reduced the learning rate by a factor of ten when a plateau in the validation loss was detected (no improvement over five epochs).

We calculated the overall classification accuracy to evaluate the performance of the different models on the considered datasets. In Table II, we show how the approaches performed with different training-test splits for the EuroSAT RGB dataset. It can be seen that all CNN approaches outperformed the BoVW method and, overall, deep CNNs performed better than that of the shallow CNNs. Nevertheless, the shallow CNN classified the EuroSAT classes with a classification accuracy of up to 93.62%. We present how the different CNN approaches perform with different training-test splits for the UCM dataset in Table VI, for the AID dataset in Table VII, for the SAT-6 dataset in Table VIII, and for the BCS dataset in Table IX. It can be seen that the shallow CNN is sufficient for the datasets SAT-6 and BCS. Deep CNNs significantly outperform the shallow CNN, however, on more challenging datasets, such as UCM and AID. Further experiments on the presented datasets can be found in [6], [19], [22], [27].

Table IV lists the classification results achieved for the two best performing CNN models, GoogLeNet and ResNet-50. In this experiment, the GoogLeNet and ResNet-50 CNN models were pretrained on the ILSVRC-2012 image classification

TABLE IV
BENCHMARKED CLASSIFICATION ACCURACY (%) OF THE TWO BEST PERFORMING CLASSIFIERS GOOGLNET AND RESNET-50 WITH A 80/20 TRAINING-TEST SPLIT

| Method | UCM | AID | SAT-6 | BCS | EuroSAT |
|-----------|--------------|--------------|--------------|--------------|--------------|
| ResNet-50 | 96.42 | 94.38 | 99.56 | 93.57 | 98.57 |
| GoogLeNet | 97.32 | 93.99 | 98.29 | 92.70 | 98.18 |

Both CNNs have been pretrained on the image classification dataset ILSVRC-2012 [21].

TABLE V
CLASSIFICATION ACCURACY (%) OF A FINE-TUNED RESNET-50 CNN ON THE PROPOSED EUROSAT DATASET WITH THE THREE DIFFERENT BAND COMBINATIONS COLOR-INFRA-RED (CI), SHORTWAVE-INFRA-RED (SWIR), AND RGB AS INPUT

| Band Combination | Accuracy (ResNet-50) |
|------------------|----------------------|
| CI | 98.30 |
| RGB | 98.57 |
| SWIR | 97.05 |

dataset [21]. In all fine-tuning experiments, we first trained the last layer with a learning rate of 0.01. Afterward, we fine-tuned throughout the entire network with a low learning rate between 0.001 and 0.0001. With a fine-tuned network, we were able to achieve a classification accuracy of about 2% higher compared to randomly initialized versions of the networks that have been trained on the EuroSAT dataset with the same training-test split (see Table II).

The deep CNNs achieved state-of-the-art results on the UCM dataset and outperformed previous results on the other three presented datasets by about 2%–4% (AID, SAT-6, BCS) [6], [19], [22]. Table IV shows that the ResNet-50 architecture performed best on the introduced EuroSAT land use and land cover classes. In order to allow for the evaluation on the class level, Fig. 8 shows the confusion matrix of this best performing network. It is shown that the classifier sometimes confused the agricultural land classes as well as the classes highway and river.

B. Band Evaluation

In order to evaluate the performance of deep CNNs using single-band images as well as shortwave-infrared and color-infrared band combinations, we used the pretrained ResNet-50 with a fixed training-test split to compare the performance of the different spectral bands. For the single-band image evaluation, we used input comprised of images consisting of information gathered from a single spectral band on all three input channels. We analyzed all spectral bands, even bands not

TABLE VI
CLASSIFICATION ACCURACY (%) OF DIFFERENT TRAINING-TEST SPLITS ON THE UCM DATASET

| Method | 10/90 | 20/80 | 30/70 | 40/60 | 50/50 | 60/40 | 70/30 | 80/20 | 90/10 |
|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| CNN (three layers) | 36.45 | 49.58 | 64.62 | 69.60 | 70.76 | 76.78 | 76.83 | 81.43 | 82.38 |
| ResNet-50 | 40.63 | 50.89 | 65.51 | 72.51 | 75.61 | 84.40 | 84.33 | 84.76 | 90.00 |
| GoogleNet | 45.39 | 60.95 | 70.34 | 77.38 | 82.47 | 85.60 | 88.88 | 91.43 | 93.80 |

TABLE VII
CLASSIFICATION ACCURACY (%) OF DIFFERENT TRAINING-TEST SPLITS ON THE AID DATASET

| Method | 10/90 | 20/80 | 30/70 | 40/60 | 50/50 | 60/40 | 70/30 | 80/20 | 90/10 |
|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| CNN (three layers) | 50.31 | 60.61 | 65.91 | 70.21 | 69.36 | 73.85 | 74.85 | 76.35 | 78.70 |
| ResNet-50 | 47.87 | 60.22 | 61.14 | 62.08 | 67.42 | 79.48 | 73.26 | 77.95 | 83.70 |
| GoogleNet | 54.18 | 63.26 | 75.44 | 76.66 | 79.02 | 83.28 | 83.07 | 84.95 | 87.10 |

TABLE VIII
CLASSIFICATION ACCURACY (%) OF DIFFERENT TRAINING-TEST SPLITS ON THE SAT-6 DATASET

| Method | 10/90 | 20/80 | 30/70 | 40/60 | 50/50 | 60/40 | 70/30 | 80/20 | 90/10 |
|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| CNN (three layers) | 95.32 | 97.03 | 97.62 | 98.49 | 98.72 | 98.83 | 99.25 | 99.13 | 99.11 |
| ResNet-50 | 96.23 | 96.67 | 96.84 | 98.96 | 98.44 | 98.05 | 98.12 | 98.61 | 99.35 |
| GoogleNet | 96.75 | 97.01 | 97.15 | 98.12 | 98.68 | 98.95 | 99.29 | 99.21 | 99.18 |

TABLE IX
CLASSIFICATION ACCURACY (%) OF DIFFERENT TRAINING-TEST SPLITS ON THE BCS DATASET

| Method | 10/90 | 20/80 | 30/70 | 40/60 | 50/50 | 60/40 | 70/30 | 80/20 | 90/10 |
|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| CNN (three layers) | 83.08 | 85.23 | 86.09 | 86.67 | 87.41 | 86.45 | 88.31 | 88.54 | 89.33 |
| ResNet-50 | 82.12 | 85.02 | 87.88 | 89.51 | 89.46 | 88.62 | 91.32 | 90.10 | 89.23 |
| GoogleNet | 83.05 | 85.19 | 87.48 | 88.29 | 89.51 | 88.54 | 90.85 | 89.40 | 89.32 |



Fig. 8. Confusion matrix of a fine-tuned ResNet-50 CNN on the proposed EuroSAT dataset using satellite images in the RGB color space.

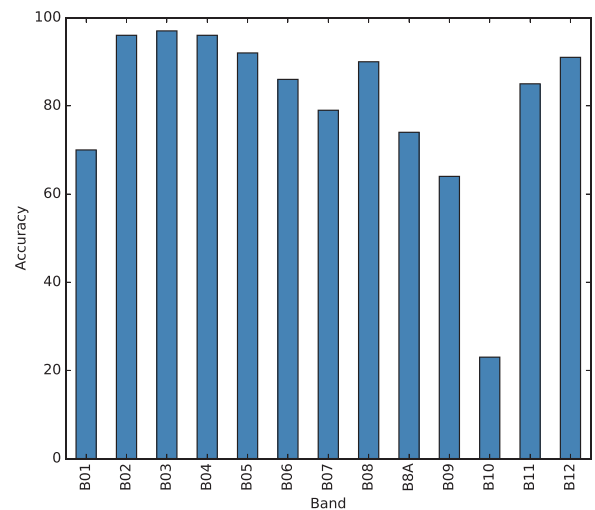


Fig. 9. Overall classification accuracy (%) of a fine-tuned ResNet-50 CNN on the given EuroSAT dataset using single-band images.

intended for land monitoring. Bands with lower spatial resolution have been upsampled to 10 m/pixel using cubic-spline interpolation [8]. Fig. 9 shows a comparison of each spectral band's performance. It is shown that the red, green, and blue bands outperformed all other bands. Interestingly, the bands red edge 1 (B05) and shortwave-infrared 2 (B12), with an original spatial resolution of merely 20 m/pixel, showed an impressive performance. These two bands even outperformed the near-infrared band (B08), which has a spatial resolution of 10 m/pixel.

In addition to the RGB band combination, we also analyzed the performance of the shortwave-infrared and color-infrared band combinations. Table V shows a comparison of the performance of these combinations. As shown, band combination images outperformed single-band images. Furthermore, images in the RGB color space performed best in the introduced land use and land cover classes. Note that the networks pretrained on the ILSVRC-2012 image classification dataset have initially not been trained on images other than RGB images.

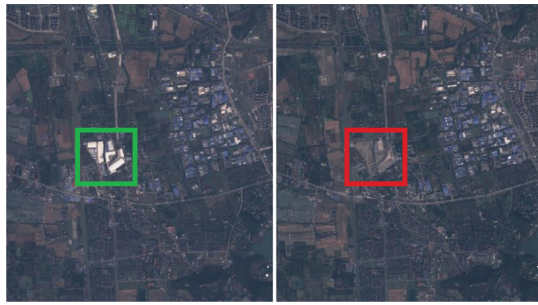


Fig. 10. Left image, acquired in the surroundings of Shanghai in December 2015, shows an area classified as industrial. The right image shows the same region in December 2016, revealing that the industrial buildings have been demolished.

V. APPLICATIONS

The openly and freely accessible satellite images allow for a broad range of possible applications. In this section, we demonstrate that the novel dataset published with this paper allows for real-world applications. The classification result with an overall accuracy of 98.57% paves the way for these applications. We show land use and land cover change detection applications as well as how the trained classifier can assist in keeping geographical maps up-to-date.

A. Land Use and Land Cover Change Detection

As the Sentinel-2 satellite constellation will scan the earth's land surface for the next 20–30 years on a repeat cycle of about every five days, a trained classifier can be used to continuously monitor land surfaces and detect changes in land use or land cover. To demonstrate land use and land cover change detection, we selected images from the same spatial region but from different points in time. Using the trained classifier, we analyzed 64×64 image regions. A change is considered to have occurred if the classifier delivers different classification results for different time points from patches taken from the same spatial 64×64 region. In the following, we show three examples of noted changes. In the first example shown in Fig. 10, the classification system recognized that the land cover had changed in the highlighted area. The left image, acquired in the surroundings of Shanghai, China, in December 2015, depicts an area classified as Industrial. The right image, depicting the same area in December 2016, reveals that the industrial buildings have been demolished. The second example is illustrated in Fig. 11. The left image, acquired in the surroundings of Dallas, USA, in August 2015, shows no dominant residential buildings in the highlighted area. The right image shows the same area in March 2017. The system has identified a change in the highlighted area, revealing that residential buildings have been constructed over this time period. The third example presented in Fig. 12 shows that the classification system detected deforestation near Villamontes, Bolivia. The left image was acquired in October 2015. The right image shows the same region in September 2016, revealing that a large area has been deforested. The presented examples are particularly of interest to urban area development, nature protection, and



Fig. 11. Left image, acquired in the surroundings of Dallas, USA, in August 2015, shows no dominant residential buildings in the highlighted area. The right image shows the same area in March 2017 with residential buildings having been built.

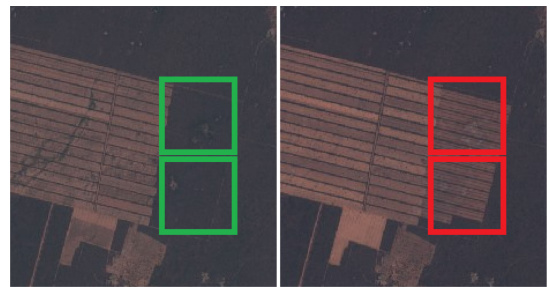


Fig. 12. Left image was acquired near Villamontes, Bolivia, in October 2015. The right image shows the same area in September 2016, revealing that a large area of land has been deforested.

sustainable development. For instance, deforestation is a main contributor to climate change, therefore, the detection of deforested land is of particular interest (e.g., to recognize illegal clearing of forests).

B. Assistance in Mapping

While a classification system trained with 64×64 image patches does not allow for finely graduated per-pixel segmentation, it can not only detect changes as shown in the previous examples, but it can also facilitate keeping maps up-to-date. This is extremely helpful with maps created in a crowdsourced manner like OpenStreetMap (OSM). A possible system can verify already tagged areas, identify mistagged areas, and include large area tagging. The proposed system is based on the trained CNN classifier, and provides classification results for each image patch created in a sliding windows-based manner.

As shown in Fig. 13, the industrial buildings seen in the left up-to-date satellite image are almost completely represented in the corresponding OSM mapping. The right up-to-date satellite image also shows industrial buildings. However, major aspects of the industrial buildings are not represented in the corresponding map. Due to the high temporal availability of Sentinel-2 satellite images in the future, this paper, together with the published dataset, can be used to build systems that assist in keeping maps updated. A detailed analysis of the respective land area can then be provided using high-resolution satellite images and an advanced segmentation approach [4], [11].



Fig. 13. Patch-based classification system can verify already tagged areas, identify mistagged areas, and do large area tagging as shown in the above images and maps. The left Sentinel-2 satellite image was acquired in Australia in March 2017. The right satellite image was acquired in the surroundings of Shanghai, China, in March 2017. The corresponding up-to-date OpenStreetMap (OSM) mapping images show that the industrial areas in the left satellite image are almost completely represented (colored gray). However, the industrial areas in the right satellite image are not properly represented.

VI. CONCLUSION

In this paper, we have addressed the challenge of land use and land cover classification. For this task, we presented a novel dataset based on remotely sensed satellite images. To obtain this dataset, we used openly and freely accessible Sentinel-2 satellite images provided in the earth observation program Copernicus. The proposed dataset consists of ten classes, covers 13 different spectral bands, and includes a total of 27 000 labeled and geo-referenced images. We provided benchmarks for this dataset with its spectral bands using state-of-the-art deep CNNs. For this novel dataset, we analyzed the performance of the 13 different spectral bands. As a result of this evaluation, the RGB band combination, with an overall classification accuracy of 98.57%, outperformed the shortwave-infrared and the color-infrared band combination and leads to a better classification accuracy than all single-band evaluations. Overall, the available free Sentinel-2 satellite images offer a broad range of possible applications. This paper is the first important step to make use of the large amount of available satellite data in machine learning, allowing for monitoring of earth's land surfaces on a large scale. The proposed dataset can be leveraged for multiple real-world earth observation applications. Possible applications are land use and land cover change detection and the improvement of geographical maps.

ACKNOWLEDGMENT

The authors would like to thank Nvidia for the support within the Nvidia AI Lab program.

REFERENCES

- [1] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "DeepSat: A learning framework for satellite imagery," in *Proc. 23rd ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2015, Art. no. 37.
- [2] B. Bischke, P. Bhardwaj, A. Gautam, P. Helber, D. Borth, and A. Dengel, "Detection of flooding events in social multimedia and satellite imagery using deep neural networks," in *Proc. MediaEval*, 2017.
- [3] B. Bischke, D. Borth, C. Schulze, and A. Dengel, "Contextual enrichment of remote-sensed events with social media streams," in *Proc. ACM Multimedia Conf.*, 2016, pp. 1077–1081.
- [4] B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel, "Multi-task learning for segmentation of buildings footprints with deep neural networks," *IEEE Int. Conf. Image Process.*, 2019, arXiv:1709.05932.
- [5] B. Bischke, P. Helber, C. Schulze, V. Srinivasan, and D. Borth, "The multimedia satellite task: Emergency response for flooding events," in *Proc. MediaEval*, 2017.
- [6] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," 2015, arXiv: 1508.00092.
- [7] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state-of-the-art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
- [8] C. De Boor, *A Practical Guide to Splines*, vol. 27. New York, NY, USA: Springer-Verlag, 1978.
- [9] 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.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 630–645.
- [11] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2016, pp. 680–688.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [13] Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
- [14] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, arXiv: 1312.4400.
- [15] F. P. Luus, B. P. Salmon, F. van den Bergh, and B. Maharaj, "Multiview deep learning for land-use classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2448–2452, Dec. 2015.
- [16] Z. Ma, Z. Wang, C. Liu, and X. Liu, "Satellite imagery classification based on deep convolution network," *World Acad. Sci. Eng. Technol. Int. J. Comput. Elect., Autom., Control Inf. Eng.*, vol. 10, no. 6, pp. 1113–1117, 2016.
- [17] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning earth observation classification using ImageNet pretrained networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 105–109, Jan. 2016.
- [18] K. Ni *et al.*, "Large-scale deep learning on the YFCC100M dataset, 2015, arXiv: 1502.03409.
- [19] K. Nogueira, O. A. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, 2017.

- [20] O. A. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 44–51.
- [21] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
- [22] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *Int. J. Remote Sens.*, vol. 33, no. 8, pp. 2395–2412, 2012.
- [23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv: 1409.1556.
- [24] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. Thirty-First AAAI Conf. Artif. Intell.*, 2017.
- [25] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
- [27] G.-S. Xia *et al.*, "Aid: A benchmark dataset for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, 2017.
- [28] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maître, "Structural high-resolution satellite image indexing," in *Proc. ISPRS TC VII Symp. 100 Years Int. Soc. Photogrammetry Remote Sens.*, vol. 38, 2010, pp. 298–303.
- [29] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.
- [30] L. Zhao, P. Tang, and L. Huo, "Feature significance-based multibag-of-visual-words model for remote sensing image scene classification," *J. Appl. Remote Sens.*, vol. 10, no. 3, 2016, Art. no. 035004.
- [31] K. Ahmad, K. Pogorelov, M. Riegler, N. Conci, and H. Pal, "CNN and gan based satellite and social media data fusion for disaster detection," in *Proc. MediaEval Workshop, Dublin, Ireland*, 2017.
- [32] G. Chen *et al.*, "Training small networks for scene classification of remote sensing images via knowledge distillation," *Remote Sens.*, vol. 10, no. 5, 2018, Art. no. 719.
- [33] S. Roy, E. Sangineto, N. Sebe, and B. Demir, "Semantic-fusion gans for semi-supervised satellite image classification," in *Proc. 25th IEEE Int. Conf. Image Process.*, 2018, Art. no. 684–688.
- [34] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Introducing EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 204–207.
- [35] L. Huang *et al.*, "Opensarship: A dataset dedicated to sentinel-1 ship interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 195–208, Jan. 2018.
- [36] M. Ponti, A. A. Chaves, F. R. Jorge, G. B. P. Costa, A. Colturato, and K. R. L. J. C. Branco, "Precision agriculture: Using low-cost systems to acquire low-altitude images," *IEEE Comput. Graph. Appl.*, vol. 36, no. 4, pp. 14–20, Jul./Aug. 2016.
- [37] W. Zhou, S. Newsam, C. Li, and Z. Shao, "PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS J. Photogrammetry Remote Sens.*, 2018.
- [38] European Commission. Mapping Guide for a European Urban Atlas. 2012. [Online]. Available: https://ec.europa.eu/regional_policy/sources/tender/pdf/2012066/annexe2.pdf.

Patrick Helber received the bachelor's and master's degrees in computer science (with a major in artificial intelligence and a minor in economics) from University of Kaiserslautern, in 2015 and 2017, respectively. He is currently working toward the Ph.D. degree in the Nvidia AI Laboratory at the German Research Center for Artificial Intelligence, Kaiserslautern, Germany.

He was a Visiting Ph.D. Researcher with the University of Oxford, where he worked in the Frontier Development Laboratory (an artificial intelligence research accelerator established by NASA and ESA). He is a co-organizer of the MediaEval—Multimedia Satellite Challenge. His main research interests include machine learning particularly focusing on AI research using earth observation data to address pressing global challenges such as fighting economic inequalities or supplying aid in natural disaster situations.

Benjamin Bischke received the bachelor's degree in computer science (with a major in artificial intelligence) from Baden-Württemberg Cooperative State University Mannheim, Mannheim, Germany, in 2012, and the master's degree in computer science (with a major in artificial intelligence) in 2016 from the University of Kaiserslautern, Kaiserslautern, Germany, where he is currently working toward the Ph.D. degree.

He was a Visiting Researcher with the University of Oxford, where he was with the Frontier Development Laboratory, an artificial intelligence research accelerator established by NASA and ESA. He is currently the Lead Organizer of the Multimedia Satellite Benchmark at MediaEval. His research interests include deep learning for the analysis of remote sensing and social media data to improve natural disaster management such as improved disaster mitigation or faster emergency response at the German Research Center for Artificial Intelligence, Kaiserslautern, Germany at the Nvidia AI Lab.

Andreas Dengel is currently the Site Head with the German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany, and the Scientific Director of the Smart Data & Knowledge Services Research Department, DFKI. In 1993, he became a Professor of the Computer Science Department, University of Kaiserslautern. Since 2009, he further holds a Professorship (kiyakuin) with the Department of Computer Science and Intelligent Systems, Graduate School of Engineering, Osaka Prefecture University, Sakai, Japan. He has authored 13 books and is the author of more than 350 peer-reviewed scientific publications, several of which received best-paper awards. He supervised more than 250 Ph.D., master's, and bachelor's theses. His main research interests include machine learning, pattern recognition, immersive quantified learning, data mining, and semantic technologies.

Prof. Dengel was the Program/Technical Chair of many international conferences, and is an Editorial Board Member of international journals and book series. Moreover, he is Founder, Initiator, and Mentor of many successful start-up companies, two of which received a Pioneer Spirit Award as well as the Cebit Innovation Award. For his contributions, he was honored by the prize Founding Promoter of the Year. Furthermore, he is a Fellow of the International Association for Pattern Recognition and the Chairman of the Flexible Factory Partner Alliance. He is also an Advisor for academic institutions, research programs as well as ministries, national and international. He is the recipient of the Alcatel/SEL Award (one of the most prestigious personal scientific awards in Germany,) on technical communication and was appointed as Distinguished Honorary Professor (tokubetu eiyo kyoju) at the Osaka Prefecture University (an honor only five researchers received within 135 years).

Damian Borth received the Ph.D. degree from the University of Kaiserslautern and DFKI, Kaiserslautern, Germany, in 2014.

He is currently a Professor of artificial intelligence and machine learning with the Institute for Computer Science, University of St. Gallen, St. Gallen, Switzerland. Previously, he was the Director of the Deep Learning Competence Center, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern and co-founded Sociovestix Laboratories, a social enterprise in the area of financial data science. During his Ph.D., he was also a Visiting Researcher with the Digital Video and Multimedia Laboratory, Columbia University. He did his postdoctoral research at UC Berkeley and the International Computer Science Institute (ICSI) in Berkeley.