

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

Patrick Helber^{1,2} Benjamin Bischke^{1,2} Andreas Dengel^{1,2} Damian Borth²

¹TU Kaiserslautern, Germany ²German Research Center for Artificial Intelligence (DFKI), Germany
 {Patrick.Helber, Benjamin.Bischke, Andreas.Dengel, Damian.Borth}@dfki.de

arXiv:1709.00029v2 [cs.CV] 1 Feb 2019

Abstract—In this paper, we address the challenge of land use and land cover classification using Sentinel-2 satellite images. The Sentinel-2 satellite images are openly and freely accessible provided in the Earth observation program Copernicus. We present a novel dataset based on Sentinel-2 satellite images covering 13 spectral bands and consisting out of 10 classes with in total 27,000 labeled and geo-referenced images. We provide benchmarks for this novel dataset with its spectral bands using state-of-the-art deep Convolutional Neural Network (CNNs). With the proposed novel dataset, we achieved an overall classification accuracy of 98.57%. The resulting classification system opens a gate towards 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—Remote Sensing, Earth Observation, Satellite Images, Satellite Image Classification, Land Use Classification, Land Cover Classification, Dataset, Machine Learning, Deep Learning, Deep Convolutional Neural Network

I. INTRODUCTION

We are currently at the edge of having public and continuous access to satellite image data for Earth observation. Governmental programs such as ESA's Copernicus and NASA's Landsat are taking significant efforts to make such data freely available for commercial and non-commercial purpose with the intention to fuel innovation and entrepreneurship. With access to such data, applications in the domains of agriculture, disaster recovery, climate change, urban development, or environmental monitoring can be realized [37], [2], [3], [5]. However, to fully utilize the data for the previously mentioned domains, first satellite images must be processed and transformed into structured semantics [35]. One type of such fundamental semantics is *Land Use and Land Cover Classification* [1], [29]. The aim of land use and land cover classification is to automatically provide labels describing the represented physical land type or how a land area is used (e.g., residential, industrial).

As often 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 (CNN)* [12], it is crucial to have large quantities of training data available to train such

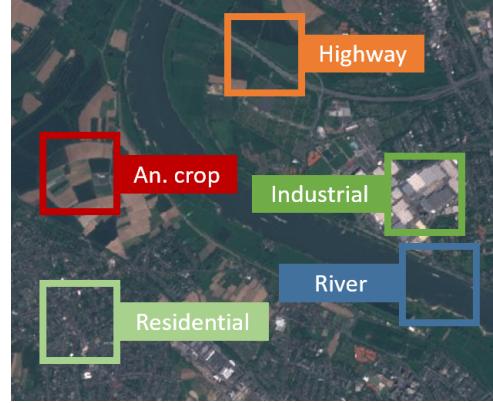


Fig. 1: Land use and land cover classification based on Sentinel-2 satellite images. Patches are extracted with the purpose to identify the shown class. This visualization highlights the classes annual crop, river, highway, industrial buildings and residential buildings.

a network. Unfortunately, current land use and land cover datasets are small-scale or rely on data sources which do not allow 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 10 different land use and land cover classes. A significant difference to previous datasets is that the presented satellite image dataset is multi-spectral covering 13 spectral bands in the visible, near infrared and short wave infrared part of the spectrum. In addition, the proposed dataset is georeferenced and based on openly and freely accessible Earth observation data allowing a unique range of applications. The labeled dataset EuroSAT is made publicly available at <https://github.com/phelber/eurosat>. Further, we provide a full benchmark demonstrating a robust classification performance which is the basis for developing applications for 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 work in the context of the recently published EuroSAT dataset, which can be used similar to [18] as a basis for a large-scale training of deep neural networks for

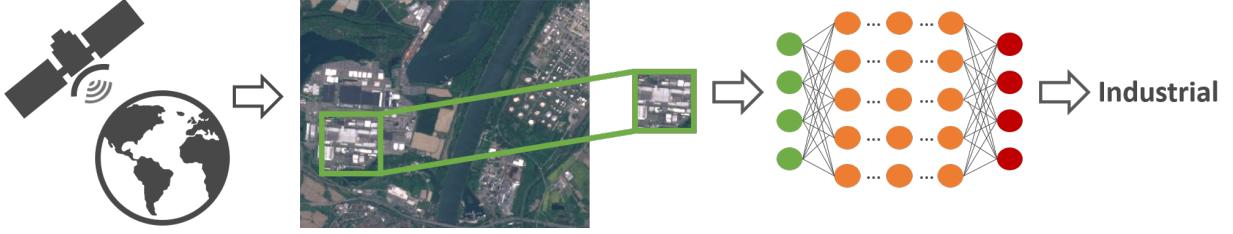


Fig. 2: This illustration shows an overview of the patch-based land use and land cover classification process using satellite images. A satellite scans the Earth to acquire images of it. Patches extracted out of these images are used for classification. The aim is to automatically provide labels describing the represented physical land type or how the land is used. For this purpose, an image patch is feed into a classifier, in this illustration a neural network, and the classifier outputs the class shown on the image patch.

the task of satellite image classification.

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

- 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.
- We provide benchmarks for the proposed EuroSAT dataset using Convolutional Neural Networks (CNNs).
- 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 in land use and land cover classification. In this context, we present remotely sensed aerial and satellite image datasets. Furthermore, we present 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 due to 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 has 100 images and the contained images measure 256x256 pixels with a spatial resolution of about 30 cm per 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 small-scale. Trying 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 [39] and NWPU-RESISC45 [36]. The datasets are based on very-high-resolution images with a spatial resolution of up to 30 cm

per pixel. Since the creation of a labeled dataset is extremely time-consuming, these datasets consist likewise of only a few hundred images per class. One of the largest datasets is the *Aerial Image Dataset (AID)*. AID consists of 30 classes with 200 to 400 images per class. The 600x600 high-resolution images were also extracted from Google Earth imagery.

Compared to the EuroSAT dataset presented in this work, the previously listed datasets rely on commercial very-high-resolution and preprocessed images. The fact of 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 work. 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. The fact of a spatial resolution of up to 30 cm per pixel, with the possibility to identify and distinguish classes like churches, schools etc., make the presented datasets difficult to compare with the dataset proposed in this work.

A study closer to our work, provided by Penatti et al. [20], analyzed remotely sensed satellite images with a spatial resolution of 10 meters per pixel to classify coffee crops. Based on these images, Penatti et al. [20] introduced the *Brazilian Coffee Scene (BCS) dataset*. The dataset covers the two classes coffee crop and non-coffee crop. Each class consists of 1,423 images. The images consist of a red, green and near-infrared band.

Similar to the proposed EuroSAT dataset, Basu et al. [1] introduced the *SAT-6* dataset relying on aerial images. This dataset has been extracted from images with a spatial resolution of 1 meter per pixel. The image patches are created using images from the National Agriculture Imagery Program (NAIP). SAT-6 covers the 6 different classes: barren land, trees, grassland, roads, buildings and water bodies. The proposed patches have a size of 28x28 pixels per image and consist of a red, green, blue and a near-infrared band.

B. Land Use and Land Cover Classification

Convolutional Neural Networks (CNNs) are a type of Neural Networks [13], which became with the impressive results on image classification challenges [12], [21], [23] the state-of-the-art image classification method in computer vision and machine learning.

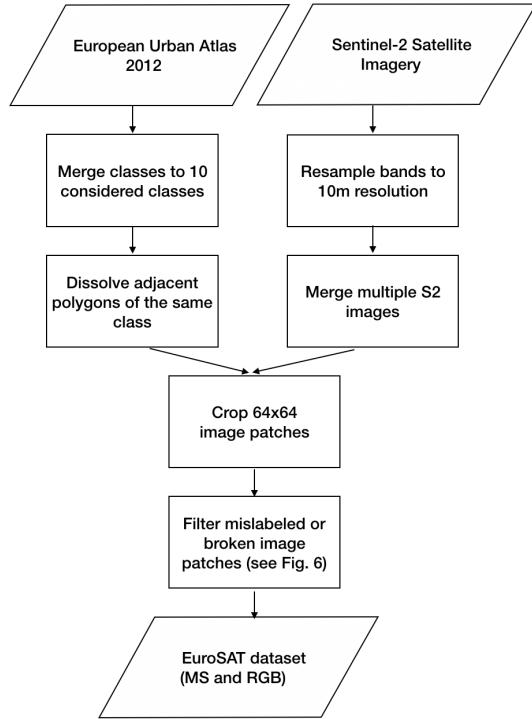


Fig. 3: The diagram illustrates the EuroSAT dataset creation process.

To classify remotely sensed images, various different feature extraction and classification methods (e.g., Random Forests) were evaluated on the 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 an 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 used the normalized features as input to a deep belief network. Besides 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., Bag-of-Visual-Words, spatial pyramid match kernels) for the classification of the UCM and BCS dataset.

In the context of deep learning, the used deep CNNs have been trained from scratch or fine-tuned by using a pretrained network [6], [19], [31], [36], [16]. The networks were mainly pretrained on the ILSVRC-2012 image classification challenge [21] dataset. Even though these pretrained networks were trained on images from a totally different domain, the features generalized well. Therefore, the pretrained networks proved to be suitable for the classification of remotely sensed images [17]. The presented works extensively evaluated all proposed machine learning methods and concluded that that deep CNNs outperform non-deep learning approaches on the considered datasets [6], [17], [15], [27].

III. DATASET ACQUISITION

Besides NASA with its Landsat Mission, the European Space Agency (ESA) steps up 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. Sentinel-2A is one satellite in the two-satellite constellation 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 the global Earth's land surface with a Multi-spectral Imager (MSI) covering the 13 different spectral bands listed in Table I. The three bands B01, B09 and B10 are intended to be used for the correction of atmospheric effects (e.g., aerosols, cirrus or water vapor). The remaining bands are primarily intended to identify and monitor land use and land cover classes. In addition to mainland, large islands as well as inland and coastal waters are covered by these two satellites. Each satellite will deliver imagery for at least 7 years with a spatial resolution of up to 10 meters per pixel. Both satellites carry 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 about 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 a continuous monitoring of the Earth's land surface for about the next 20 - 30 years. Most importantly, the data is openly and freely accessible and can be used for any application (commercial or non-commercial use).

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: Based on the obtained satellite images, we created a dataset of 27,000 georeferenced and labeled image patches. The image patches measure 64x64 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,

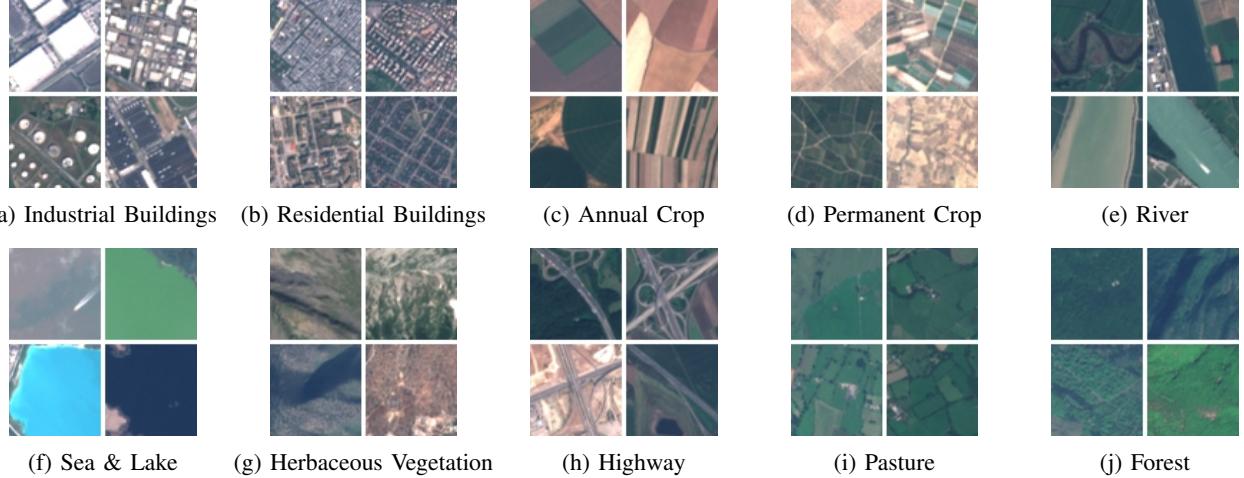


Fig. 4: This overview shows sample image patches of all 10 classes covered in the proposed EuroSAT dataset. The images measure 64x64 pixels. Each class contains 2,000 to 3,000 images. In total, the dataset has 27,000 geo-referenced images.

Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine and United Kingdom.

In order to improve the chance of getting valuable image patches, we selected satellite images with a low cloud level. Besides the possibility to generate a cloud mask, ESA provides a cloud level value for each satellite image allowing to quickly select images with a low percentage of clouds covering the land scene.

We aimed for the objective to cover 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 all over the year to get a variance as high as possible inherent in the covered land use and land cover classes. Within one class of the EuroSAT dataset, different land types of this 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 [40]. An overview diagram of the dataset creation process is shown in Fig. 3

B. Dataset Creation

The Sentinel-2 satellite constellation provides about 1.6 TB of compressed images per day. Unfortunately, supervised machine learning is restricted even with this amount of data by the lack of labeled ground truth data. The generation of the benchmarking EuroSAT dataset was motivated by the objective of making this open and free satellite data accessible to various Earth observation applications and 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 2,000 to 3,000 images per class. In total, the dataset has 27,000 images. The patches measure 64x64 pixels. We have chosen 10 different

TABLE I: All 13 bands covered by Sentinel-2's Multispectral Imager (MSI). The identification, the spatial resolution and the central wavelength is listed for each spectral band.

| Band | Spatial Resolution m | Central Wavelength nm |
|-------------------|-------------------------|--------------------------|
| 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 |

land use and land cover classes based on the principle that they showed to be visible at the resolution of 10 meters per pixel and are frequently enough covered by the European Urban Atlas 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. It therefore covers the classes highway, residential buildings and industrial buildings. The residential class is created using the urban fabric classes described in the European Urban Atlas. Different water bodies appear in the classes river and sea & lake. Furthermore, undeveloped environments such as forest and herbaceous vegetation are included. 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 full of snow or ice. Example images, which have been discarded, are shown in Fig. 6. The samples are intended to show industrial buildings. Clearly, no industrial

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 (two layers) | 75.88 | 79.84 | 81.29 | 83.04 | 84.48 | 85.77 | 87.24 | 87.96 | 88.66 |
| ResNet-50 | 75.06 | 88.53 | 93.75 | 94.01 | 94.45 | 95.26 | 95.32 | 96.43 | 96.37 |
| GoogleNet | 77.37 | 90.97 | 90.57 | 91.62 | 94.96 | 95.54 | 95.70 | 96.02 | 96.17 |

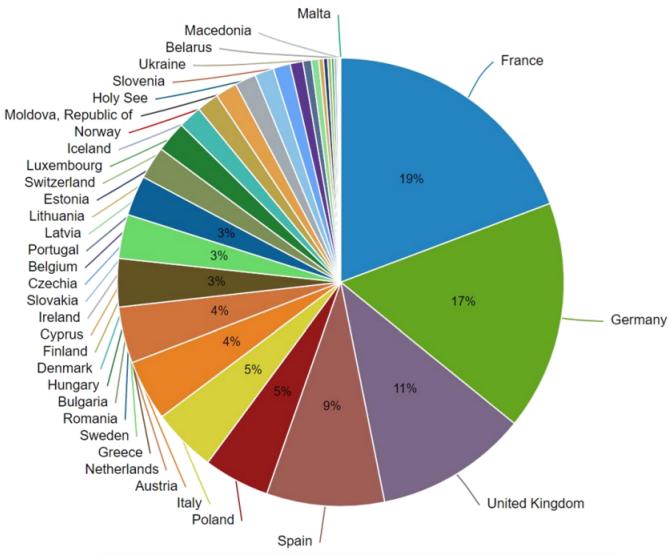


Fig. 5: EuroSAT dataset distribution. The georeferenced images are distributed all over Europe. The distribution is influenced by the number of represented cities per country in the European Urban Atlas.

building is visible. Please note, the proposed dataset has not received atmospheric correction. This can result in images with a color cast. Extreme cases are visualized in Fig. 7. With the intention to advocate the classifier to also learn these cases, we did not filter the respective samples and let them flow into the dataset.

Besides the 13 covered spectral bands, the new dataset has three further central innovations. Firstly, the dataset is not based on non-free satellite images like Google Earth imagery or relies on data sources which are not updated on a high-frequent basis (e.g., NAIP used in [1]). Instead, an open and free Earth observation program whose satellites deliver images for the next 20 - 30 years is used allowing real-world Earth observation applications. Secondly, the dataset uses a 10 times lower spatial resolution than the benchmark dataset closest to our research but at once distinguishes 10 classes instead of 6. For instance, we split up the built-up class into a residential and an industrial class or distinguish between different agricultural land uses. Thirdly, we release the EuroSAT dataset in a georeferenced version.

With the release of the geo-referenced EuroSAT we aim to make the large amount of Sentinel-2 satellite imagery accessible for machine learning approaches. Their effectiveness was successfully demonstrated in [32], [33], [34].

IV. DATASET BENCHMARKING

As shown in previous work [6], [15], [17], [19], deep CNNs have demonstrated 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 [25], [26], [24], [14] 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 this context, we evaluate the classification performance using single-band and band combination images.

TABLE III: Benchmarked classification accuracy (%) of the two best performing classifiers GoogLeNet and ResNet-50 with a 80/20 training-test split. Both CNNs have been pre-trained on the image classification dataset ILSVRC-2012 [21].

| 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 |

A. Comparative Evaluation

We respectively split each dataset in a training and a test set (80/20 ratio). We ensured that the split is applied class-wise. While the red, green and blue bands are 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 combining the bands red (B04), green (B03) and blue (B02). For benchmarking, we evaluated the performance of the Bag-of-Visual-Words (BoVW) approach using SIFT features and a trained SVM. In addition, we trained a shallow Convolutional Neural Network (CNN), a ResNet-50 and a GoogleNet model on the training set. 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 perform in case of different training-test splits for the EuroSAT RGB dataset.

It can be seen that all CNN approaches outperform the BoVW method and, overall, deep CNNs perform better than shallow CNNs. Nevertheless, the shallow CNN classifies the EuroSAT classes with a classification accuracy of up to 89.03%. Please note [6], [19], [22] for the benchmarking performance of the other datasets on different training-test splits.

Table III lists the achieved classification results for the two best performing CNN models GoogLeNet and ResNet-50. In



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



Fig. 7: Color cast due to atmospheric effects.

this experiment, the GoogleNet and ResNet-50 CNN models were pretrained on the ILSVRC-2012 image classification dataset [21]. In all fine-tuning experiments, we first trained the last layer with a learning rate of 0.01. Afterwards, we fine-tuned through the entire network with a low learning rate between 0,001 and 0,0001. With a finetuned network we achieve a classification accuracy of about 2% higher compared to randomly initialized versions of the networks which have been trained on the EuroSAT dataset with the same training-test split (see Table II).

The deep CNNs achieve state-of-the-art results on the UCM dataset and outperform previous results on the other three presented datasets by about 2-4% (AID, SAT-6, BCS) [6], [19], [22]. Table III shows that the ResNet-50 architecture performs best on the introduced EuroSAT land use and land cover classes. In order to allow an evaluation on the class level, Fig. 8 shows the confusion matrix of this best performing network. It is shown that the classifier sometimes confuses 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 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 images as input consisting of the information gathered from a single spectral band on all three input channels. We analyzed all spectral bands, even the bands not intended for land monitoring. Bands with a lower spatial resolution have been upsampled to 10 meters per pixel using cubic-spline interpolation [8]. Fig. 9 shows a comparison of the spectral band's performance. It is shown that the red, green and blue bands outperform all other bands. Interestingly, the bands red edge 1 (B05) and shortwave-infrared 2 (B12) with an original spatial resolution of merely 20 meters per pixel showed an impressive performance. The two bands even outperform the near-infrared band (B08) which has a spatial resolution of 10 meters per pixel.

In addition to the RGB band combination, we also analyzed the performance of the shortwave-infrared and color-infrared

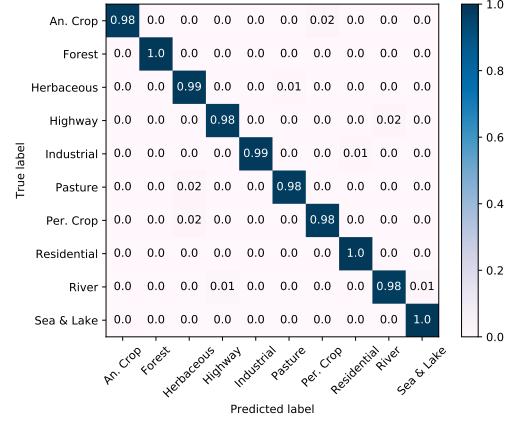


Fig. 8: Confusion matrix of a fine-tuned ResNet-50 CNN on the proposed EuroSAT dataset using satellite images in the RGB color space.

TABLE IV: Classification accuracy (%) of a fine-tuned ResNet-50 CNN on the proposed EuroSAT dataset with the three different band combinations color-infrared (CI), shortwave-infrared (SWIR) and RGB as input.

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

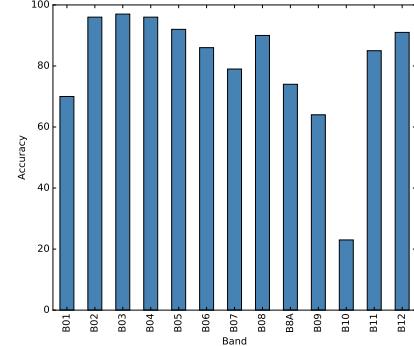


Fig. 9: Overall classification accuracy (%) of a fine-tuned ResNet-50 CNN on the given EuroSAT dataset using single-band images.

band combination. Table IV shows a comparison of the performance of these combinations. As shown, band combination images outperform single-band images. Furthermore, images in the RGB color space performed best on the introduced land use and land cover classes. Please note, networks pretrained on the ILSVRC-2012 image classification dataset have initially not been trained on images other than RGB images.

V. APPLICATIONS

The openly and freely accessible satellite images allow a broad range of possible applications. In this section, we demonstrate that the novel dataset published with this paper allows 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

Since the Sentinel-2 satellite constellation will scan the Earth's land surface for about the next 20 - 30 years on a repeat cycle of about five days, a trained classifier can be used for monitoring 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 64x64 image regions. A change has taken place if the classifier delivers different classification results for patches taken from the same spatial 64x64 region. In the following, we show three examples of spotted changes. In the first example shown in Fig. 10, the classification system recognized that the land cover has changed in the highlighted area. The left image was acquired in the surroundings of Shanghai, China in December 2015 showing an area classified as industrial. The right image shows the same area in December 2016 revealing that the industrial buildings have been demolished. The second example is illustrated in Fig. 11. The left image was acquired in the surroundings of Dallas, USA in August 2015 showing no dominant residential buildings in the highlighted area. The right image shows the same area in March 2017 revealing that residential buildings have been built up.



Fig. 11: The left image was acquired in the surroundings of Dallas, USA in August 2015 showing no dominant residential buildings in the highlighted area. The right image shows the same area in March 2017 showing that residential buildings have been built up.

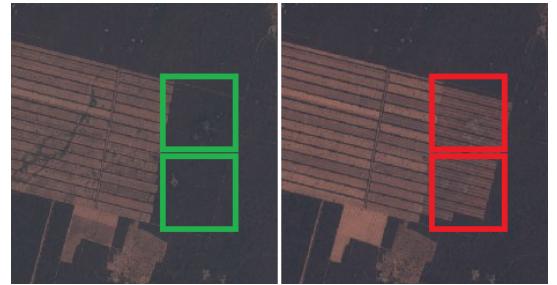


Fig. 12: The left image was acquired near Villamontes, Bolivia in October 2015. The right image shows the same area in September 2016 revealing that a large land area has been deforested.

B. Assistance in Mapping

While a classification system trained with 64x64 image patches does not allow a finely graduated per-pixel segmentation, it cannot only detect changes as shown in the previous examples, it can also facilitate keeping maps up-to-date. This is an extremely helpful assistance with maps created in a crowdsourced manner like OpenStreetMap (OSM). A possible system can verify already tagged areas, identify mistagged areas or bring large area tagging. The proposed system is based on the trained CNN classifier providing a classification result 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 covered in the corresponding OSM mapping. The right up-to-date satellite image also shows industrial buildings. However, a major part of the industrial buildings is not covered in the corresponding map. Due to the high temporal availability of Sentinel-2 satellite images in the future, this work together with the published dataset can be used to build systems which assist in keeping maps up-to-date. 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. 10: The left image was acquired in the surroundings of Shanghai in December 2015 showing an area classified as industrial. The right image shows the same region in December 2016 revealing that the industrial buildings have been demolished.



Fig. 13: A patch-based classification system can verify already tagged areas, identify mistagged areas or bring 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 covered (colored gray). However, the industrial areas in the right satellite image are not properly covered.

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 have used the openly and freely accessible Sentinel-2 satellite images provided in the Earth observation program Copernicus. The proposed dataset consists of 10 classes covering 13 different spectral bands with in total 27,000 labeled and geo-referenced images. We provided benchmarks for this dataset with its spectral bands using state-of-the-art deep Convolutional Neural Network (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 work is a first important step to make use of the large amount of available satellite data in machine learning allowing to monitor 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 or the improvement of geographical maps.

ACKNOWLEDGMENT

This work was partially funded by the BMBF project DeFuseNN (01IW17002). The authors 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 *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 37. ACM, 2015.
- [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 *MediaEval*, 2017.
- [3] B. Bischke, D. Borth, C. Schulze, and A. Dengel. Contextual Enrichment of Remote-Sensed Events with Social Media Streams. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 1077–1081. ACM, 2016.
- [4] B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel. Multi-Task Learning for Segmentation of Buildings Footprints with Deep Neural Networks. In *arXiv preprint arXiv:1709.05932*, 2017.
- [5] B. Bischke, P. Helber, C. Schulze, V. Srinivasan, and D. Borth. The Multimedia Satellite Task: Emergency Response for Flooding Events. In *MediaEval*, 2017.
- [6] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092*, 2015.
- [7] G. Cheng, J. Han, and X. Lu. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 2017.
- [8] C. De Boor. *A practical guide to splines*, volume 27. Springer-Verlag New York, 1978.
- [9] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [10] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645. Springer, 2016.
- [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 *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [13] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
- [14] M. Lin, Q. Chen, and S. Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.
- [15] F. P. Luus, B. P. Salmon, F. van den Bergh, and B. Maharaj. Multiview deep learning for land-use classification. *IEEE Geoscience and Remote Sensing Letters*, 12(12):2448–2452, 2015.
- [16] Z. Ma, Z. Wang, C. Liu, and X. Liu. Satellite imagery classification based on deep convolution network. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 10(6):1113–1117, 2016.
- [17] D. Marmanis, M. Datcu, T. Esch, and U. Stilla. Deep learning earth observation classification using imagenet pretrained networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1):105–109, 2016.
- [18] K. Ni, R. Pearce, K. Boakye, B. Van Essen, D. Borth, B. Chen, and E. Wang. Large-scale deep learning on the yfcc100m dataset. *arXiv preprint arXiv:1502.03409*, 2015.

- [19] K. Nogueira, O. A. Penatti, and J. A. dos Santos. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, 61:539–556, 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 44–51, 2015.
- [21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):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. *International journal of remote sensing*, 33(8):2395–2412, 2012.
- [23] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [24] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*, 2016.
- [25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.
- [26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.
- [27] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, and L. Zhang. Aid: A benchmark dataset for performance evaluation of aerial scene classification. *arXiv preprint arXiv:1608.05167*, 2016.
- [28] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maître. Structural high-resolution satellite image indexing. In *ISPRS TC VII Symposium-100 Years ISPRS*, volume 38, pages 298–303, 2010.
- [29] Y. Yang and S. Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 270–279. ACM, 2010.
- [30] L. Zhao, P. Tang, and L. Huo. Feature significance-based multibag-of-visual-words model for remote sensing image scene classification. *Journal of Applied Remote Sensing*, 10(3):035004–035004, 2016.
- [31] Kashif Ahmad, Konstantin Pogorelov, Michael Riegler, Nicola Conci, and H Pal. Cnn and gan based satellite and social media data fusion for disaster detection. In *Proc. of the MediaEval 2017 Workshop, Dublin, Ireland*, 2017.
- [32] Guanzhou Chen, Xiaodong Zhang, Xiaoliang Tan, Yufeng Cheng, Fan Dai, Kun Zhu, Yuanfu Gong, and Qing Wang. Training small networks for scene classification of remote sensing images via knowledge distillation. *Remote Sensing*, 10(5):719, 2018.
- [33] Subhankar Roy, Enver Sangineto, Nicu Sebe, and Begüm Demir. Semantic-fusion gans for semi-supervised satellite image classification. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 684–688. IEEE, 2018.
- [34] Patrick Helber, Benjamin Bischke, Andreas Dengel, Damian Borth. Introducing EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification In *Geoscience and Remote Sensing Symposium (IGARSS), 2018 IEEE International*. IEEE, 2018.
- [35] Lanqing Huang, Bin Liu, Boying Li, Weiwei Guo, Wenhao Yu, Zenghui Zhang, and Wenxian Yu. Opensarship: A dataset dedicated to sentinel-1 ship interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(1):195–208, 2018.
- [36] Gong Cheng, Junwei Han, and Xiaoqiang Lu. Remote sensing image scene classification: benchmark and state of the art. *Proceedings of the IEEE*, 105(10):1865–1883, 2017.
- [37] Moacir Ponti, Arthur A Chaves, Fábio R Jorge, Gabriel BP Costa, Adimara Colturato, and Kalinka RLJC Branco. Precision agriculture: Using low-cost systems to acquire low-altitude images. *IEEE computer graphics and applications*, 36(4):14–20, 2016.
- [38] Weixun Zhou, Shawn Newsam, Congmin Li, and Zhenfeng Shao. Patternnet: a benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018.
- [39] Weixun Zhou, Shawn Newsam, Congmin Li, and Zhenfeng Shao. Patternnet: a benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018.
- [40] European Commission. Mapping guide for a European urban atlas. https://ec.europa.eu/regional_policy/sources/tender/pdf/2012066/annexe2.pdf, 2012.