

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

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

ABSTRACT

In this paper, we address the challenge of land use and land cover classification using remote sensing satellite images. For this challenging task, we use the openly and freely accessible Sentinel-2 satellite images provided within the scope of the Earth observation program Copernicus. The key contributions are as follows. We present a novel dataset based on satellite images covering 13 different spectral bands and consisting of 10 classes with in total 27,000 labeled images. We evaluate state-of-the-art deep Convolutional Neural Network (CNNs) on this novel dataset with its different spectral bands. We also evaluate deep CNNs on existing remote sensing datasets and compare the obtained results. With the proposed novel dataset, we achieved an overall classification accuracy of 98.57%. The classification system resulting from the proposed research opens a gate towards a number of Earth observation applications. We demonstrate how the classification system can be used for detecting land use or land cover changes and how it can assist in improving geographical maps.

Keywords

Remote Sensing, Earth Observation, Satellite Images, Satellite Image Classification, Land Use Classification, Land Cover Classification, Dataset, Machine Learning, Deep Learning, Deep Convolutional Neural Network

1. 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 [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. 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, performance of classification systems strongly depends on the availability

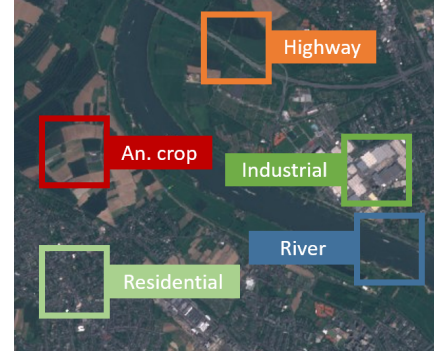


Figure 1: Land use and land cover classification using satellite images. Patches are extracted from a satellite image with the purpose to identify the showed land use or land cover class. This visualization highlights the classes annual crop, river, highway, industrial area and residential area.

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 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 a total of 10 different classes. A significant difference to previous datasets is that the presented satellite image dataset covers 13 spectral bands allowing to investigate multi-modal fusion approaches in the context of these bands. This is in particular an open challenge when considering deep neural networks for classification. In addition, the proposed dataset is based on openly and freely accessible Earth observation data allowing a unique range of applications. The labeled dataset EuroSAT is made publicly available¹. Further, we provide a full benchmark by using deep CNNs on the dataset demonstrating robust classification performance which can be taken as foundation to develop 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 pub-

¹<http://madm.dfki.de/downloads>



Figure 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 land use and land cover classification. The aim is to automatically provide labels describing the represented physical land type or how land is used. For this purpose, an image patch is feed into a classifier, in this case a neural network, and the classifier delivers the land use or land cover class represented on the image patch.

lished EuroSAT dataset, which can be used similar to [18] as a basis for large-scale training of deep learning networks for satellite images.

2. RELATED WORK

In this section, we review previous research in the area of land use and land cover classification. In this context, we present datasets consisting of airborne images as well as datasets consisting of satellite images. Furthermore, we present state-of-the-art methods in land use and land cover classification.

2.1 Classification Datasets

Remote sensing image classification 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 researched [6, 19, 20, 27, 29] remote sensing image classification dataset known as *UC Merced Land Use Dataset (UCM)* 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 very small-scale. Trying to enhance the dataset situation, various researchers used non-freely usable Google Earth² images to manually create novel datasets [22, 27, 28, 30]. The datasets use 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 extracted from Google Earth imagery. However, the fact of using preprocessed and non-free image data makes these datasets unsatisfying for real-world 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 very low number of images per class. The fact of a spatial resolution of up to 30 cm per pixel, with the possibility to identify

²<https://www.google.de/earth/>

and distinguish classes like churches, schools etc., make the presented datasets difficult to compare with the dataset proposed in this work.

Research closer to our work was provided by Penatti et al. [20] and Basu et al. [1]. Penatti et al. investigated remote sensing satellite images having a spatial resolution of 10 meters per pixel. Based on these images, [20] introduced the *Brazilian Coffee Scene dataset (BCS)*. 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. 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 by 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.

2.2 Land Use and Land Cover Classification

While CNNs are an established classification method [13], primarily with the impressive results on image classification challenges [12, 21, 23], deep CNNs became a common and popular image classification method. In remote sensing image classification, various different feature extraction and classification methods (e.g., Random Forest) were evaluated on the introduced datasets. Yang et al. evaluated bag-of-visual-words and spatial extension approaches on the UCM dataset [29]. Basu et al. investigated deep belief networks, basic CNNs and stacked denoising autoencoders on the SAT-6 dataset [1]. Basu et al. also represented an own framework for the land cover classes introduced in the SAT-6 dataset which extracts features from 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. evaluated deep CNNs on the UCM and BCS dataset [20]. In addition to deep CNNs, Castelluccio et al. intensively evaluated various methods (e.g., bag-of-visual-words, spatial pyramid match kernel) 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, 7, 16]. Mainly, the networks were pretrained on the dataset from the ILSVRC-2012 image classification challenge [21]. Thus, these pretrained networks were trained on images from a



Figure 3: These images visualize different band combinations of the spectral bands provided by ESA’s satellite Sentinel-2A. The left image shows a color-infrared image by combining the near-infrared (B08), red (B04) and green (B03) band. The middle image displays a combination of the shortwave-infrared 2 (B12), red edge 4 (B08A) and red band (B04). The right image is constructed using merely the near-infrared (B08) band.

totally different domain. However, the features generalized well and therefore these pretrained networks proved to be suitable for remote sensing image classification [17]. The examined works show that deep CNNs outperform all previous state-of-the-art approaches on the introduced datasets [6, 17, 15, 27].



Figure 4: This image shows a Sentinel-2 satellite image in the RGB color space generated by combining the red (B04), green (B03) and blue (B02) spectral band. The shown image gives an impression of the visible objects at a spatial resolution of 10 meters per pixel.

3. DATASET ACQUISITION

Besides NASA with its Landsat Mission³, the European Space Agency (ESA) steps up efforts to improve Earth observation within the scope of the Copernicus program⁴. Under this program, ESA operates a series of satellites known as Sentinels⁵. In order to address the challenge of land use and land cover classification, we use multispectral image data provided by the Sentinel-2A satellite.

Sentinel-2A is one out of a two-satellite constellation consisting of the identical land monitoring satellites Sentinel-2A and Sentinel-2B. The satellites were successfully launched in June 2015 and March 2017. Both sun-synchronous satellites

capture the global Earth’s land surface with a Multispectral Imager (MSI) covering 13 different spectral bands listed in Table 1. 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 vegetation. In addition to mainland, large islands as well as inland and coastal waters are covered by these two satellites. Each satellite will deliver optical 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 every five days, i.e. the satellites scan each point in the covered area every five days.⁶ This short repeat cycle as well as the future operation allows a continuous monitoring of the Earth’s land surface for about the next decade. Most importantly, the data is openly and freely accessible and can be used for any application (commercial or non-commercial use). Detailed information about the Sentinel-2 satellites is covered in the corresponding documentation.⁷

In order to show the spatial resolution of 10 meters per pixel, Figure 4 illustrates a sample scene originated from the combination of the red (B04), green (B03) and blue (B02) band. To emphasize different optical aspects, several spectral bands can be combined. In remote sensing, these combinations are used to make different aspects visible, which are poorly visible or cannot be seen using RGB color space images. One example is the combination of shortwave-infrared, near-infrared and green light to identify floods. An example of a color-infrared image, which results from the combination of the near-infrared (B08), red (B04) and green (B03) band, is shown in Figure 3. A shortwave-infrared image arisen from the combination of the shortwave-infrared 2 (B12), red edge 4 (B08A) and red (B04) band is also shown in Figure 3. Furthermore, Figure 3 shows an image consisting of merely the near-infrared (B08) band.

We are convinced that the vast data volume of these satellites in combination with powerful machine learning methods will influence future research. Therefore, one of our key research aims is to make this vast data source accessible for

³https://www.nasa.gov/mission_pages/landsat/main

⁴<http://www.copernicus.eu/>

⁵<https://sentinels.copernicus.eu>

⁶The one-satellite constellation has a repeat cycle of 10 days.

⁷https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook

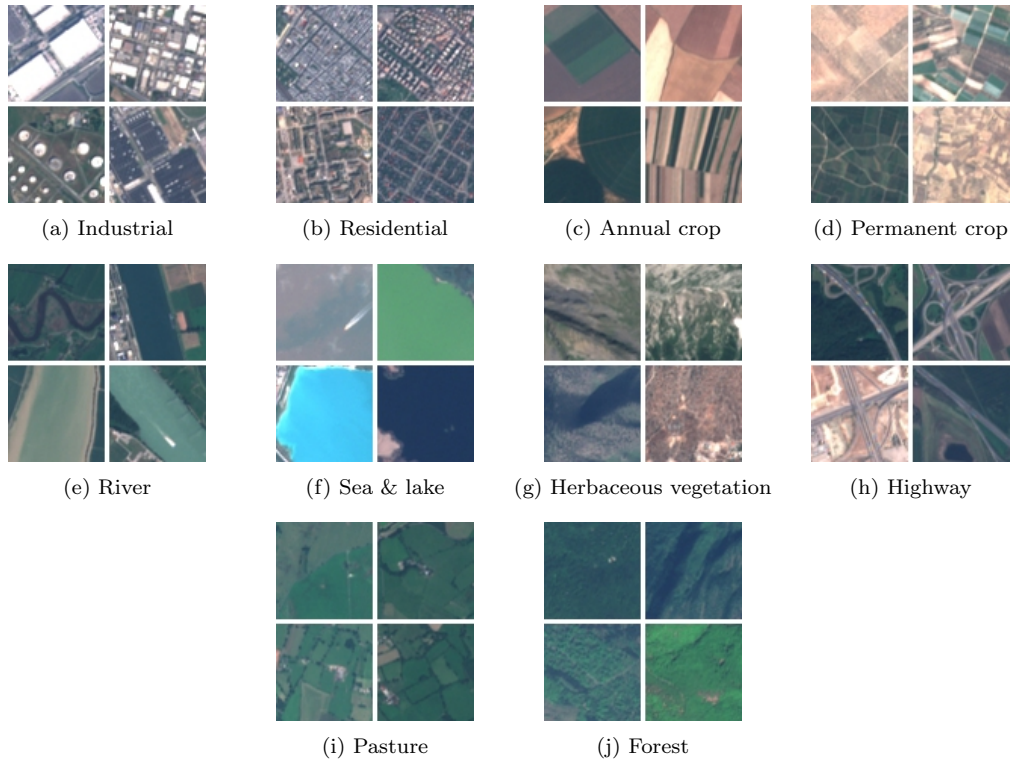


Figure 5: This overview shows sample images of all 10 classes covered in the proposed EuroSAT dataset. The image patches measure 64x64 pixels. Each class contains 2,000 to 3,000 image patches. In total, the dataset has 27,000 images.

machine learning applications. To construct an image classification dataset, we conducted the following steps:

1. Satellite Image Acquisition: We gathered satellite images of cities distributed in over 30 European countries.
2. Dataset Creation: Based on the obtained satellite images, we created a dataset of 27,000 labeled image patches. The image patches measure 64x64 pixels and have been manually checked.

3.1 Satellite Image Acquisition

We have downloaded satellite images recorded by the satellite Sentinel-2A via Amazon S3⁸. We chose satellite images associated with the cities covered in the European Urban Atlas⁹. The covered cities are distributed over the following 30 countries: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland 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 offering a quick solution to choose images with a low percentage of clouds covering the land scene.

We aimed for the objective to comprise as many countries as possible in order to cover the high intra-class variance

⁸<https://aws.amazon.com/de/public-datasets/sentinel-2/>

⁹<http://land.copernicus.eu/local/urban-atlas/urban-atlas-2012>

Table 1: All 13 bands covered by Sentinel-2’s Multispectral Imager (MSI). For each spectral band, its identification, the spatial resolution and the central wavelength are listed.

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

inherent to remote sensing image classes. Furthermore, we have extracted images recorded all over the year to get variance as high as possible inherent in the covered land use and land cover classes.

3.2 Dataset Creation

The amount of available satellite data is tremendous (the two-satellite constellation provides about 1.6 TB of com-

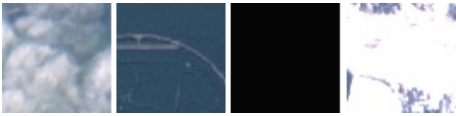


Figure 6: Four examples of bad image patches are shown. All four samples are intended to show industrial areas. Clearly, no industrial area is shown due to clouds, mislabeling, dead pixels or ice/snow.

pressed images per day). Unfortunately, even with this amount of data, supervised machine learning is constrained due to the lack of ground truth data. Motivated by the observation that existing benchmark datasets are not satisfying for the intended applications with Sentinel-2 satellite images and the objective to make this open and free satellite data accessible to various Earth observation applications, we generated a labeled dataset. 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 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 and industrial areas. Different water bodies appear in the classes river and sea & lake. Furthermore, undeveloped environments such as forest and herbaceous vegetation are comprised. An overview of the covered classes with four samples per class is illustrated in Figure 5.

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/ice. Example images which have been discarded are shown in Figure 6. The samples are intended to show industrial areas. Please note, the proposed dataset has not received atmospheric correction. This can result in images with a color cast. Extreme cases are visualized in Figure 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 two central innovations. Firstly, the dataset is not based on non-free satellite images like Google Earth imagery or rely on old remote sensing data which is not available on a high-frequency basis in future (e.g., NAIP used in [1]). Instead, an open and free Earth observation program whose satellites deliver images for the next 7 to 12 years is used allowing real-world Earth observation applications¹⁰. 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.

¹⁰The satellites Sentinel-2C and Sentinel-2D are already planned to continue the Earth observation mission afterwards.

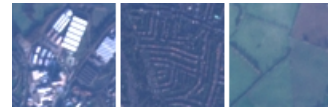


Figure 7: Color cast due to atmospheric effects.

Table 2: Classification accuracy (%) of fine-tuned GoogLeNet and ResNet-50 CNNs on the investigated datasets. Both CNNs have been pretrained on the image classification dataset ILSVRC-2012 [21].

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

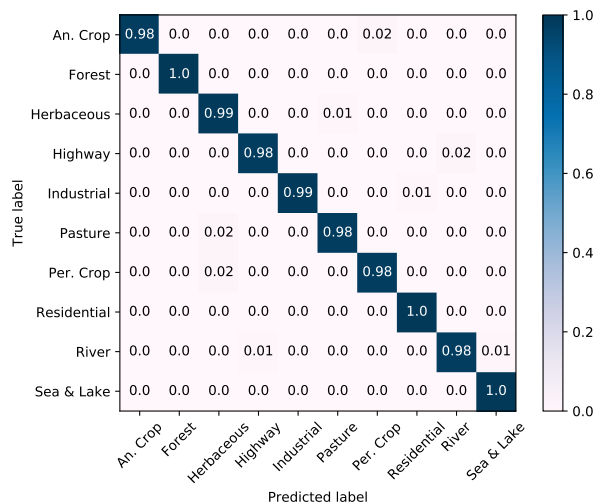


Figure 8: Confusion matrix of a fine-tuned ResNet-50 CNN on the proposed EuroSAT dataset using satellite images in the RGB color space.

4. DATASET BENCHMARKING

As shown in previous work [6, 15, 17, 19], deep CNNs have demonstrated to outperform all previous approaches on land use and land cover classification datasets. Accordingly, we use the state-of-the-art deep CNN models GoogleNet [25] and ResNet-50 [9, 10] for the classification of the introduced datasets. These networks evolved by innovations such as the inception module [25, 26, 24, 14] and the residual unit [9, 10].

For the proposed EuroSAT dataset, we also investigate the performance of the 13 spectral bands with respect to the classification task. In this context, the classification performance using single-band images as well as images based on common band combinations are evaluated.

4.1 Comparative Evaluation

In order to train and evaluate deep CNNs on the proposed novel and the introduced existing datasets, we respectively split the data for each dataset into a training and test set at the ratio of 80:20. We ensured that the split is applied class-

Table 3: Classification accuracy (%) of a fine-tuned ResNet-50 CNN on the proposed EuroSAT dataset with three different band combinations as input. A color-infrared (CI) image results from the combination of the near-infrared (B08), red (B04) and green (B03) band. An RGB image is obtained by combining the red (B04), green (B03) and blue (B02) band. A shortwave-infrared (SWIR) image is the result of the combination of the shortwave-infrared 2 (B12), red edge 4 (B08A) and red (B04) band.

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

wise. While the red, green and blue bands are covered by almost all airborne or satellite image classification 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 bands red (B04), green (B03) and blue (B02). For the respective dataset, we fine-tuned GoogLeNet and ResNet-50 CNN models, which were pretrained on the ILSVRC-2012 image classification dataset. [21]. In all evaluations, 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.

We computed the overall classification accuracy to evaluate the performance of the different CNN models on the investigated datasets. Table 2 lists the achieved classification results for the different CNN models. 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 2 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, Figure 8 shows the confusion matrix of this best performing network. Even if rarely, it is shown that the classifier sometimes confuses the agricultural land classes as well as the classes highway and river.

4.2 Band Evaluation

In order to investigate the performance of deep CNNs using single-band images as well the common 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. We investigated 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]. Figure 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.

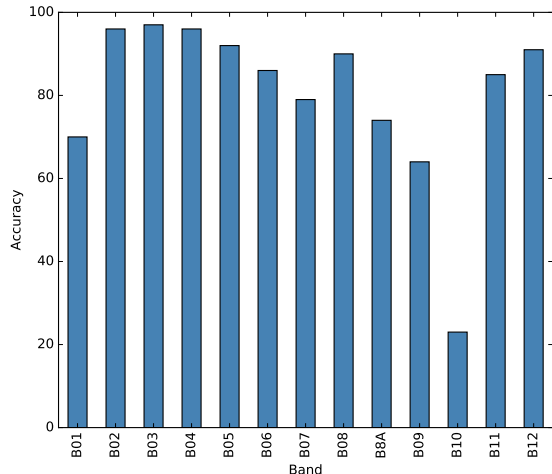


Figure 9: Overall classification accuracy (%) of a fine-tuned ResNet-50 CNN on the given EuroSAT dataset using single-band images.

In addition to the RGB band combination, we also investigated the performance of the shortwave-infrared and color-infrared band combinations. Table 3 shows a comparison of the performance of these combinations. As shown, image combinations 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.

5. APPLICATIONS

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

5.1 Land Use and Land Cover Change Detection

Since the two-satellite constellation will scan the Earth’s land surface for about the next decade on a repeat cycle of 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 investigated 64x64 image regions. A change has taken place if the classifier delivers different classification results on patches taken from the same spatial 64x64 region. In the following, we show three examples of spotted changes. In the first example shown in Figure 10, the classification system recognized that in the marked area the land has changed.

The left image was acquired in the surroundings of Shanghai, China in December 2015 showing an area classified as industrial. The right image was acquired in the same area in December 2016 showing that the industrial buildings have been demolished. The second example is illustrated in Figure 11. The left image was acquired in the surroundings of Dallas, USA in August 2015 showing no dominant residential area in the highlighted area. The right image was acquired in the same area in March 2017. The system has identified a change in the highlighted area showing a residential area has been constructed. The third example presented in Figure 12 shows that the system detected deforestation near Villamontes, Bolivia. The left image was acquired in October 2015. The right image was acquired in September 2016 showing that a large area has been deforested.

The shown examples find their usage in urban area development, nature protection or sustainable development. For instance, since deforestation is a main contributor to climate change, the detection of deforestation is particularly of interest (e.g., to notice illegal clearing of forest).

5.2 Assistance in Mapping

While a classification system using 64x64 patches does not allow a finely graduated segmentation or mapping, it cannot only detect changes as shown in the previous examples, it can also facilitate keeping maps up-to-date. This foundation is an extremely helpful assistance with maps created in a crowdsourced manner like OpenStreetMap (OSM). A possible system could verify already tagged areas, identify mistagged areas or bring large area tagging. As shown in Figure 13, the industrial areas 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 areas. However, a major part of the industrial areas is not covered in the corresponding map. Due to the high temporal availability of Sentinel-2 satellite images in the future, the proposed research 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].



Figure 10: Change detection application 1a: The left image was acquired in the surroundings of Shanghai in December 2015 showing an area classified as industrial. The right image was acquired in December 2016 showing that the industrial buildings have been demolished.



Figure 11: Change detection application 1b: The left image was acquired in the surroundings of Dallas, USA in August 2015 showing no dominant residential area in the highlighted area. The right image was acquired in the same area in March 2017. The system classified the area as residential showing that a residential area has been built up.

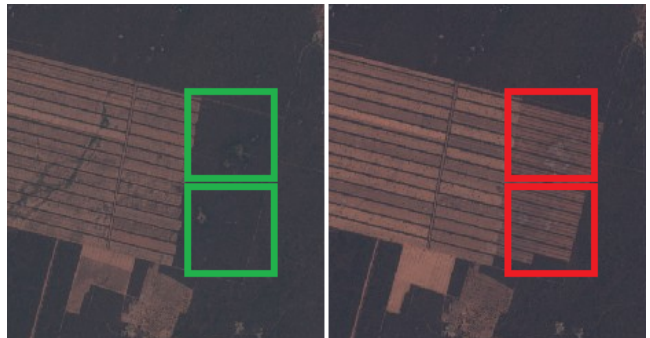


Figure 12: Change detection application 1c: The left image was acquired near Villamontes, Bolivia in October 2015. The right image was acquired in the same area in September 2016 showing that a large land area has been deforested.

6. CONCLUSIONS

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 remote sensing satellite images. To obtain this dataset, we have used the openly and freely accessible Sentinel-2 images provided within the scope of the Earth observation program Copernicus. The proposed dataset consists of 10 classes covering 13 different spectral bands with in total 27,000 labeled images. We evaluated state-of-the-art deep CNNs on this novel dataset. We also evaluated deep CNNs on existing remote sensing datasets and compared the obtained results. For the novel dataset, we investigated the performance based on different spectral bands. As a result of this evaluation, the RGB band combination outperformed all single-band images as well as the shortwave-infrared and the color-infrared band combination with an overall classification accuracy of 98.57%. For the existing datasets, we achieved state-of-the-art and outperforming results. Overall, the available free Sentinel-2 satellite images offer a broad choice of possible applications. The proposed research is a first important step to make use of the vast available satellite data in machine learning allowing classification and monitoring of Earth's land surfaces in future on a large scale. The proposed research can be

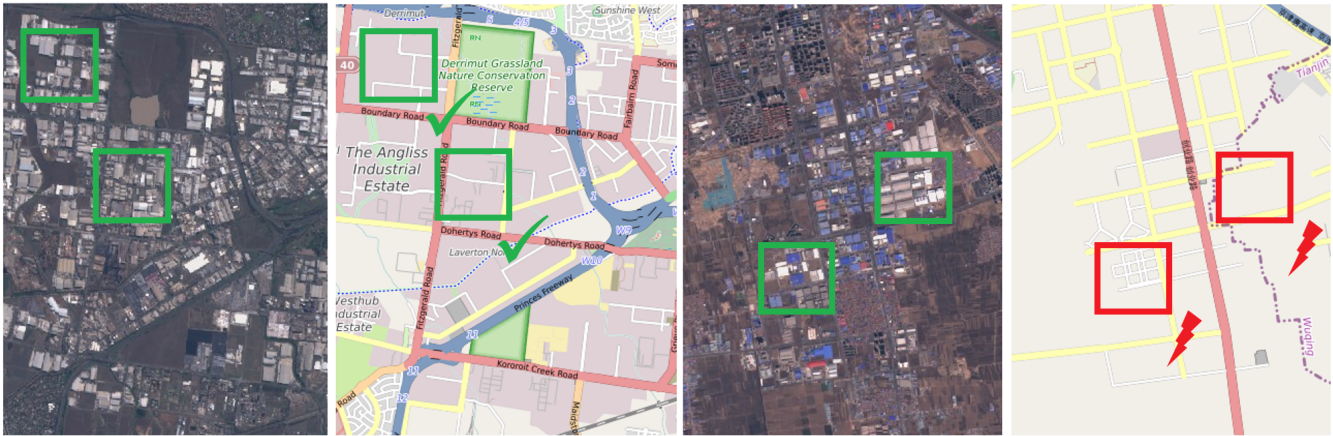


Figure 13: Application 2: A patch-based classification system can verify already tagged areas, identify mistagged areas or bring large area tagging as shown in the above images and maps. The left Sentinel-2 satellite image was acquired in Australia in March 2017. The right satellite image was acquired in the surroundings of Shanghai, China in March 2017. The corresponding up-to-date mapping images from OpenStreetMap 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.

leveraged for multiple real-world Earth observation applications. Possible applications in the area of land use and land cover change detection and improving geographical maps have been shown.

7. ACKNOWLEDGMENTS

This work was partially funded by the BMBF project Multimedia Opinion Mining (MOM: 01WI15002). The authors would like to thank NVIDIA for the support within the NVIDIA AI Lab program.

8. 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. To appear.
- [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 *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018. (Submitted).
- [5] B. Bischke, P. Helber, C. Schulze, V. Srinivasan, and D. Borth. The Multimedia Satellite Task: Emergency Response for Flooding Events. In *MediaEval*, 2017. To appear.
- [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, C. De Boor, E.-U. Mathématicien, C. De Boor, and 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. Jensen. 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.