

SEASONET: A SEASONAL SCENE CLASSIFICATION, SEGMENTATION AND RETRIEVAL DATASET FOR SATELLITE IMAGERY OVER GERMANY

Dominik Koßmann^{1,†}, Viktor Brack^{1,†}, Thorsten Wilhelm²

Pattern Recognition in Embedded Systems Group¹, Image Analysis Group²,
TU Dortmund University, Germany

ABSTRACT

This work presents SeasoNet, a new large-scale multi-label land cover and land use scene understanding dataset. It includes 1 759 830 images from Sentinel-2 tiles, with 12 spectral bands and patch sizes of up to 120 px × 120 px. Each image is annotated with large scale pixel level labels from the German land cover model LBM-DE2018 with land cover classes based on the CORINE Land Cover database (CLC) 2018 and a five times smaller minimum mapping unit (MMU) than the original CLC maps. We provide pixel synchronous examples from all four seasons, plus an additional snowy set. These properties make SeasoNet the currently most versatile and biggest remote sensing scene understanding dataset with possible applications ranging from scene classification over land cover mapping to content-based cross season image retrieval and self-supervised feature learning. We provide baseline results by evaluating state-of-the-art deep networks on the new dataset in scene classification and semantic segmentation scenarios.

Index Terms— land cover classification, mapping, retrieval, dataset, seasonal changes

1. INTRODUCTION

Automatic Earth monitoring and remote sensing applications produce a huge amount of unlabeled data every day. Fast analysis of this data is essential to land use and climate change monitoring as well as disaster prevention. An analysis includes solving various vision tasks to understand a satellite image scene to the fullest. Tasks range from land cover classification over image retrieval to semantically mapping each pixel. Further, all areas need to be agnostic to seasonal changes. Therefore, most approaches leverage deep learning architectures which need a huge amount of labeled data to train directly on the target remote sensing (RS) domain. Transfer learning scenarios between the RS domain and natural scene images with pre-trained features from ImageNet unfortunately fall short in performance [1]. While some large-scale RS benchmark datasets for scene classification (e.g., BigEarthNet [2] or SEN12MS [3]) exist, they do not



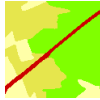



Sentinel-2	Multi-Label [2]	Pixel-Label [4]	This Work
	Broad-leaved forest, Non-irrigated arable land, Pastures		
	Coniferous forest, Discontinuous urban fabric, Non-irrigated arable land, Pastures		

Table 1: Overview of current land cover dataset resolutions and labels in comparison to our dataset. In contrast to image level labels [2] or pixel-level labels [4] this work adopts the land-cover and land-use maps of the German land cover model LBM-DE2018, which has a higher mapping resolution and therefore misses significantly less small objects.

offer pixel level labels or only provide small scale segmentation, see Table 1. This limits the potential of change detection and semantic land cover mapping. Furthermore, seasonal changes can introduce a significant domain shift. Current datasets include instances from multiple seasons only in different locations and without pixel labels. Thus, the domain shift can not be considered in image retrieval or segmentation scenarios.

In this work, we introduce SeasoNet¹, a new large-scale multi-spectral multi- and pixel-label remote sensing benchmark dataset. It consists of 1 759 830 Sentinel-2 image patches, annotated with multi-label land cover and land usage classes from the CORINE Land Cover database (CLC) 2018², covering the total area of Germany. To the best of our knowledge it is significantly larger than the currently biggest labeled multi-spectral archives (see Table 3). Further, we provide large scale pixel level labels based on these land cover classes. These annotations are adopted from a 5 hectare MMU version of the CLC database from the publicly available German land cover model LBM-DE2018³. It offers a 5 times better minimum object resolution than the original CLC database with 25 hectares. We offer individual sets for the four seasons and an additional snowy set. We believe

¹<https://doi.org/10.5281/zenodo.5850307>

²<https://land.copernicus.eu/pan-european/corine-land-cover/clc2018>

³<https://gdz.bkg.bund.de/index.php/default/catalog/product/view/id/1071/s/corine-land-cover-5-ha-stand-2018-clc5-2018/>

[†] Authors contributed equally






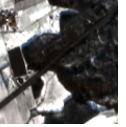





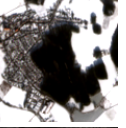






Image-level labels	Pixel-level labels	Spring (460 104)	Summer (481 456)	Fall (500 149)	Winter (218 663)	Snow (99 458)
Road and rail networks and associated land, Non-irrigated arable land, Pastures, Broad-leaved forest						
Discontinuous urban fabric, Industrial or commercial units, Non-irrigated arable land, Pastures, Coniferous forest, Mixed forest, Water bodies						
Discontinuous urban fabric, Road and rail networks and associated land, Construction sites, Pastures, Broad-leaved forest, Coniferous forest, Mixed forest						

Table 2: Example of the Sentinel-2 RGB channels in SeasoNet for four given seasons and the snow set.

this dataset serves well as a resource for the development and evaluation of new models in scene classification, mapping and retrieval. It may also serve as a basis for transfer-learning to other data in the RS domain or pre-training of features in a self-supervised setup. The dataset will be made publicly available alongside code for data preparation and evaluation.

2. CHALLENGES IN REMOTE SENSING SCENE UNDERSTANDING

One of the main challenges in remote sensing scene understanding is the huge amount of data necessary to learn domain agnostic features with modern deep architectures. For natural scene images supervised pre-training with ImageNet [5] is well established. Unfortunately, in a transfer learning setup the generated features do not always generalize well to new domains like remote sensing imagery. Furthermore, RS imagery often contains multi or hyper-spectral images, which do not match features learned from ImageNet RGB channels, thus making the domain gap more severe. There are currently multiple large-scale archives to learn in domain features (Table 3), but these still do not match the size of ImageNet [5] with its 1,2 million images.

Another issue are seasonal changes to vegetation and weather, which create a domain shift in the appearance. In current remote sensing archives, e.g., SEN12MS [3] or SeasonalContrast [1], multiple seasons are already included, but they are sampled over different locations and thus learning seasonal changes on an object level becomes difficult. Current works try to tackle seasonal changes by e.g., style transfer with generative adversarial networks [6]. But, it remains a challenge to content-based image retrieval [6] as well as change detection and classification tasks [1, 7].

For land cover mapping and change detection on pixel level not many large-scale datasets with a diverse set of classes and a non-local coverage exist so far, see Table 3. In addition, they can suffer from erroneous labels, since creating a grid over a given annotated land cover map to create image

patches can result in small areas at the edges of images [4]. These areas are impossible to segment on a pixel level or classify on an image level without the necessary context information. Thus, this problem affects scene classification and segmentation datasets alike. Moreover, the underlying mapping resolution might not fit the given satellite image resolution, making false pixel labels at the borders of regions likely.

3. THE SEASONET DATASET

Our dataset aims to tackle the above RS challenges. SeasoNet consists of 1 759 830 multi-spectral multi- and pixel-label image patches from the Sentinel-2 mission, covering the whole area of Germany. The dataset is constructed from 311 Sentinel-2 tiles covering Germany, acquired between April 2018 and February 2019. We use the same 12 spectral bands from Sentinel-2 as [2] with Level-2A Bottom-of-Atmosphere correction. Two sets of patches were created from two regular grids over the selected tiles. A singular grid consists of non overlapping connected patches. The two grids are shifted by half the patch size in both dimensions and thus overlap. By this process we were able to sample different large scale regions, since each grid avoids different small scale cut off regions at image borders. Each patch includes sizes of $120 \text{ px} \times 120 \text{ px}$, $60 \text{ px} \times 60 \text{ px}$, $20 \text{ px} \times 20 \text{ px}$ for 10 m, 20 m, 60 m Sentinel-2 bands, respectively. In total, we sample from 519 547 unique patch locations. The acquisition of images per patch location has been split by season into four sets plus an extra snowy set. Season date boundaries are based on their meteorological definitions, with an added gap of one month between them, ensuring that each image is representative for its season. All seasons except winter include only images with less than 1% snow and less than 5% clouds. For winter these thresholds were both set to 10%, because of the high confusion rate between frost, snow and clouds. The minimum snow amount of the snowy set is also 10%, aligning with the maximum threshold during winter. A fixed maxi-

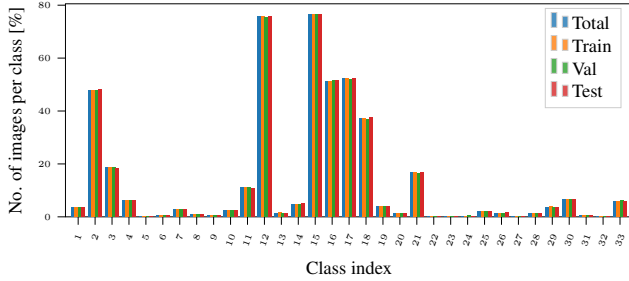


Fig. 1: Histogram of relative class distribution for the total number of 879 374 samples and the 70% training/20% test/10% validation split partitioning for grid 2 of SeasoNet. The class distribution over the individual splits is nearly identical and thus provides a good basis for evaluation.

imum value for clouds was not set on the snowy images, instead a classifier was trained to specifically find and remove cloudy images, with a focus on high recall of cloudy images. After preprocessing, there are 181 480 sample locations for which an image could be found in each season, thus a subset of $181\,480 \text{ locations} \cdot 4 \text{ seasons} = 725\,920$ dataset samples has full season coverage. For a total of 88% of all patch locations we found images for at least three seasons. The snowy set consists of 99 458 images, see Table 2. Thus, it is possible to track seasonal changes over the same patch location while providing visually varying and diverse scenes.

In Germany 35 of the 43 level-3 CLC classes are present. Two classes (*Glaciers and perpetual snow*, *Burnt Areas*) are additionally omitted due to an insufficient number of samples. As a result, 33 classes are included in SeasoNet. The number of samples per land cover class varies significantly over the dataset, thus making it a challenging but realistic RS scene understanding dataset (cf. Fig. 1). Additionally, we included the level of urbanization inside the metadata of each patch, provided by the GE250 region classification map⁴, enabling further dataset splits for e.g., retrieval tasks.

Our large-scale dataset helps in tackling the transfer learning domain gap and improving land cover classification tasks. It further provides a benchmark for understanding seasonal changes not only on an image but also on a mapping level of individual objects, thus supporting the evaluation and creation of new methods for style transfer and content-based retrieval methods. The high resolution land cover pixel annotations avoid problems with small cut off regions and the sampling over two grids enables more context information being represented inside the dataset, avoiding the difficulties described in [4]. Because of this, learning a segmentation of small regions like streets or railways on a huge dataset becomes possible. Even self-supervised feature learning methods relying on seasonal changes [1] now are possible on an object [8] and not only an image level.

⁴<http://gdz.bkg.bund.de/index.php/default/gebietseinheiten-1-250-000-ge250.html>

Dataset name	Image Type	Annotation Type	Number of Images
Agriculture-Vision [9]	Aerial Multispectral	Multi-Label + Pixel Label	94 986
MLRSNet [10]	Aerial/Sat RGB	Multi-Label	109 161
BigEarthNet [2]	Sat. Multispectral	Multi-Label	590 326
SEN12MS [3]	Sat. Multispectral	Multi-Label + Region Label	541 986
This Work	Sat. Multispectral	Multi-Label + Pixel-Label	1 759 830

Table 3: Examples of current remote sensing scene benchmarks.

4. EXPERIMENTAL RESULTS

For supervised learning tasks we provide two sets of roughly 880 000 images, one from each grid of SeasoNet, while prohibiting sample locations from the training set to appear in the test set. To ensure a machine learning benchmark suitable evaluation protocol, we provide an official split for the various possible remote sensing learning tasks. Since land cover data is inherently highly imbalanced, we focused on creating a set of training / validation / test splits with low variation between the class distribution over the total dataset versus that of the individual splits, see Fig. 1. In analogy to average precision thresholds, e.g. AP50, from instance segmentation tasks, we define three region size thresholds for our dataset, namely *easy*, *medium* and *hard*, with 300 px, 100 px and 0 px region size thresholds, respectively (Table 4). To control label noise introduced by small cutoff regions from neighboring patches, every region with a number of pixels below the chosen threshold is not considered during the training and evaluation process. This also controls the given multi-labels on an image level. We evaluate our dataset in land cover classification with DenseNet121 [11] and in segmentation with DeepLabV3 [12] using the proposed split over grid 2. All spectral bands are upsampled to the maximum patch resolution and used during training. For classification, we use a binary cross entropy loss and for segmentation the standard cross entropy loss. The results shown in Table 5 serve as a baseline. The resulting performance from current deep learning models is already producing decent accuracies for the usage in remote sensing applications. However, there is still room left for improvement on minority class and small region detection, deteriorating the average performance. Especially, the consideration of context from different regions and minority class agnostic objective functions or sampling, as in [13], are promising areas of future research.

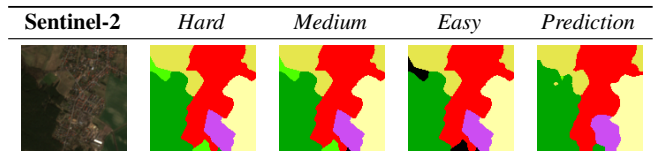


Table 4: Example of the region size threshold levels controlling the number of small regions flowing into the dataset’s pixel and image level ground-truth annotations. Black regions do not contribute to the patch annotations and evaluation anymore. Rightmost picture: Prediction from DeepLabV3 model, trained with the *hard* annotations.

	DenseNet121		DenseNet121 PT		DeepLabV3	DeepLabV3 PT
	$F1_{macro}$	$F1_{micro}$	$F1_{macro}$	$F1_{micro}$	mIoU	mIoU
<i>hard</i>	60.02	84.14	59.39	83.07	47.53	48.69
<i>medium</i>	61.48	85.06	60.42	83.46	47.78	48.95
<i>easy</i>	61.47	84.10	59.75	82.11	48.49	49.66
< 300 px	-	-	-	-	15.98	16.11
< 100 px	-	-	-	-	10.07	10.13

Table 5: Results on standard classification and segmentation tasks for the SeasoNet grid 2 test set. PT: Models pre-trained on ImageNet. $F1_{macro}$ averages over all classes equally, while $F1_{micro}$ is sample weighted and thus favors correct majority class predictions. Four different models were trained on the *hard* threshold image regions, thus including all small regions. Five different evaluations were carried out per model, three with the different levels of region thresholding and two specifically only on small regions below 300/100 px. The last two show the challenge of detecting cutoff regions.

5. CONCLUSION

We presented a new large-scale benchmark dataset for remote sensing scene understanding comprised of 1 759 830 Sentinel-2 image patches. It is annotated with large scale pixel level labels (5 ha MMU) and provides pixel synchronous examples from all four seasons, plus an additional snowy set. SeasoNet will provide the basis for future research of deep learning models in RS, specifically in the areas of scene classification, mapping and retrieval, and domain adaption in the context of seasons. Self-supervised learning approaches on a pixel level might now be possible and could be trained on this dataset. Experimental results show decent accuracies and the potential of SeasoNet. By providing an official evaluation protocol for various scene tasks we aim to foster the comparability of future research in this domain.

6. ACKNOWLEDGEMENT

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 269661170.

7. REFERENCES

- [1] O. Mañas, A. Lacoste, D. Vazquez X. Giro-i Nieto, and P. Rodriguez, “Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data,” Mar. 2021.
- [2] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, “Bigearthnet: A large-scale benchmark archive for remote sensing image understanding,” in *IGARSS 2019 - IEEE International Geoscience and Remote Sensing Symposium*. 2019, pp. 5901–5904, IEEE.
- [3] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, “Sen12ms – a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W7, pp. 153–160, 2019.
- [4] T. Wilhelm and D. Koßmann, “Land cover classification from a mapping perspective: Pixelwise supervision in the deep learning era,” in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 2496–2499.
- [5] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* Ieee, 2009, pp. 248–255.
- [6] Y. Li, J. Ma, and Y. Zhang, “Image retrieval from remote sensing big data: A survey,” *Information Fusion*, vol. 67, pp. 94–115, 2021.
- [7] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, “Urban change detection for multispectral earth observation using convolutional neural networks,” in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 2115–2118.
- [8] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, and L. Van Gool, “Unsupervised semantic segmentation by contrasting object mask proposals,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 10052–10062.
- [9] M. T. Chiu, X. Xu, Y. Wei, Z. Huang, A. G. Schwing, R. Brunner, H. Khachatrian, H. Karapetyan, I. Dozier, G. Rose, et al., “Agriculture-vision: A large aerial image database for agricultural pattern analysis,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 2828–2838.
- [10] X. Qi, P. Zhu, Y. Wang, L. Zhang, J. Peng, M. Wu, J. Chen, X. Zhao, N. Zang, and P. T. Mathiopoulos, “Mlrsnet: A multi-label high spatial resolution remote sensing dataset for semantic scene understanding,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 337–350, 2020.
- [11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 4700–4708.
- [12] L. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017.
- [13] D. Koßmann, T. Wilhelm, and G. A. Fink, “Towards tackling multi-label imbalances in remote sensing imagery,” in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, January 2020, pp. 5782–5789.