REMOTE SENSING IMAGE CLASSIFICATION WITH THE SEN12MS DATASET

M. Schmitt[a,b] , Y.-L. Wu[b]

[a]Department of Geoinformatics, Munich University of Applied Sciences, Munich, Germany - michael.schmitt@hm.edu
[b]Department of Aerospace and Geodesy, Technical University of Munich, Munich, Germany

KEY WORDS: deep learning, image classification, land cover, convolutional neural networks, data fusion, benchmarking

ABSTRACT:

Image classification is one of the main drivers of the rapid developments in deep learning with convolutional neural networks for computer vision. So is the analogous task of scene classification in remote sensing. However, in contrast to the computer vision community that has long been using well-established, large-scale standard datasets to train and benchmark high-capacity models, the remote sensing community still largely relies on relatively small and often application-dependend datasets, thus lacking comparability. With this paper, we present a classification-oriented conversion of the SEN12MS dataset. Using that, we provide results for several baseline models based on two standard CNN architectures and different input data configurations. Our results support the benchmarking of remote sensing image classification and provide insights to the benefit of multi-spectral data and multi-sensor data fusion over conventional RGB imagery.

1. INTRODUCTION

One of the most crucial preconditions for the development of machine learning models for the interpretation of remote sensing data is the availability of annotated datasets. While well-established *shallow* learning approaches were usually trained on small datasets, modern *deep* learning requires large-scale data to reach the desired generalization performance. In computer vision, the great success of deep learning was largely driven by the desire to solve the image classification problem, i.e. assigning one or more labels to a given photograph. For this purpose, many researchers have relied on the ImageNet database (Deng et al., 2009), which contains millions of annotated images. In remote sensing, the same task is often called *scene* classification, which similarly aims at assigning one or more labels to a remote sensing image, i.e., a scene. As (Cheng et al., 2020) summarizes, there has also been a lot of progress in this field in recent years, with a growing number of dedicated datasets (cf. Tab. 1). As can be seen from this non-complete selection, most datasets built for remote sensing image classification deal with high-resolution aerial imagery, usually providing three or four spectral channels (RGB, or RGB plus near-infrared). Only EuroSat and BigEarthNet provide spaceborne multi-spectral imagery, with So2Sat LCZ42 being the only existing scene classification dataset covering the other large data modality – synthetic aperture radar (SAR) data[1]. Combining all points, i.e. dataset size, availability of more than just a single sensor modality, and versatility, it becomes obvious that most existing datasets lack the power to train generic, region-agnostic models exploiting multi-sensor information.

With this paper, we present the conversion of the SEN12MS dataset to the image classification purpose as well as a couple of baseline models including their evaluation. Since SEN12MS is – in terms of spatial coverage and sensor modalities – significantly larger than all other available datasets, and sampled in a more versatile manner, this will enhance the possibility to benchmark future model developments in a transparent way, and to pre-train remote sensing-specific models that can later be fine-tuned to individual problems and user needs.

2. SEN12MS FOR IMAGE CLASSIFICATION

In this section, the SEN12MS dataset in its new image classification variant is described. All resources, most notably labels or pre-trained baseline models, can be downloaded from `https://github.com/schmitt-muc/SEN12MS` in an open access manner. The goal of both the repository and this paper is to support the establishment of standardized benchmarks for better comparability in the field.

2.1 The original SEN12MS Dataset

The SEN12MS dataset (Schmitt et al., 2019) was published in 2019 and contains 180,662 so-called patches, which are distributed across the world and all seasons. For each of those patches, the dataset provides, at a pixel sampling of 10m and a size of 256×256 pixels,

- a Sentinel-1 SAR image with two polarimetric channels (VV, VH)

- a Sentinel-2 optical image with 13 multi-spectral channels

- four different land cover maps following different classification schemes.

The SEN12MS dataset was designed having the following key features in mind:

- Its main distinction from other deep learning-oriented datasets was (and is) its focus on multi-sensor data fusion. Instead of containing only optical imagery, SEN12MS provides both SAR and multi-spectral optical data to cover the most relevant modalities in remote sensing.

- Instead of being sampled over a single study area or a geographical region of limited extent (e.g. individual countries or continents) the data of SEN12MS is sampled from all

---

[1]After this paper was accepted for publication at ISPRS Congress 2021, it came to the authors' attention that BigEarthNet in the meantime was extended by BigEarthNet-S1, a collection of Sentinel-1 images corresponding to the Sentinel-2 images contained in the original dataset. Thus, now another multi-modal scene classification dataset exists.

Table 1: Non-exhaustive list of datasets for remote sensing image classification

Dataset	Data Type	Number of Images	Image Size	Reference
UC-Merced	Aerial RGB	2,100	256×256	(Yang and Newsam, 2010)
WHU-RS19	Aerial RGB	1,005	600×600	(Xia et al., 2010)
SIRI-WHU	Aerial RGB	2,400	200×200	(Zhao et al., 2016)
NWPU-RESISC45	Aerial RGB	31,500	256×256	(Cheng et al., 2017)
AID	Aerial RGB	10,000	600×600	(Xia et al., 2017)
PatternNet	Aerial RGB	30,400	256×256	(Zhou et al., 2018)
EuroSat	Satellite multi-spectral	$27,000$	64×64	(Helber et al., 2019)
BigEarthNet	Satellite multi-spectral	590,326	120×120	(Sumbul et al., 2019)
So2Sat LCZ42	Satellite multi-sensor	400,673	32×32	(Zhu et al., 2020)
SEN12MS	*Satellite multi-sensor*	*180,662*	256×256	*(Schmitt et al., 2019)*

inhabited continents. This makes the dataset unique with regard to the possibility to train generalizing models that are comparably agnostic with respect to target scenes.

- In contrast to its predecessor, the SEN1-2 dataset, all images contained in SEN12MS come as geotiffs, i.e. they include geolocalization information. On the one hand, this information can be used as an additional input feature (c.f. Uber's CoordConv solution (Liu et al., 2018)). On the other hand, it can be used to pair the SEN12MS data with external geodata.

- By providing dense – albeit coarse – land cover labels for each patch, semantic segmentation for land cover classification was intended to be one of the main application areas of the dataset.

Since its publication, SEN12MS has been used in many studies on deep learning applied to multi-sensor remote sensing imagery. Examples include image-to-image translation (Abady et al., 2020, Yuan et al., 2020) and land cover mapping with focuses put on weakly supervised learning (Schmitt et al., 2020, Yu et al., 2020), model generalization (Hu et al., 2020), and meta-learning (Rußwurm et al., 2020). This shows the dataset's potential for both application-oriented as well as methodical research.

2.2 Creation of Scene Labels from Dense Labels

The original SEN12MS dataset contains four different schemes of MODIS-derived land cover labels. From those schemes, the IGBP scheme (see, e.g., (Sulla-Menashe et al., 2019)) was chosen as background for the conversion into a classification dataset. This was done because the IGBP scheme features rather generic classes, including both natural and urban environments with a moderate level of semantic granularity. The other, LCCS-based, classification schemes, in contrast are less generic and over-focus on different topics of interest, e.g. land use or surface hydrology. As already proposed by (Yokoya et al., 2020), the 17 original IGBP classes were converted to the simplified IGBP scheme (cf. Table 2) to ensure comparability to other land cover schemes such as FROM-GLC10 (Gong et al., 2019), and to mitigate the class imbalance of SEN12MS to some extent.

For the generation of single-label scene annotations, simply the land cover class corresponding to the mode of the pixel-based land cover distribution in that scene was used. For the generation of multi-label scene annotations, the histogram of land cover appearances within a scene was converted to a probability distribution. Then, to remove visually underrepresented classes, only classes with a probability larger than 10% were kept. An illustration of some example images including their single-label and multi-label scene annotations is provided in Fig. 1.

2.3 Dataset Statistics

The class distribution of the different label representations is summarized in Fig. 2. It can be seen that the dataset is fairly imbalanced, with the classes *Savanna*, *Croplands*, and *Grassland* being very frequent and classes such as *Wetlands*, *Barren*, and *Water* being comparably underrepresented. The class *Snow/Ice* can basically be considered as non-existing in SEN12MS. Due to the generally uniform spatial sampling of the SEN12MS data, this imbalanced class distribution is a representation of the natural imbalance of the real world, but should be considered when the data is used for the training and evaluation of machine learning models.

While (Sulla-Menashe et al., 2019) provides a rough estimate of the accuracy of the original, global IGBP land cover map at 500 m GSD – namely 67% –, there is no accuracy assessment of the upsampled IGBP maps provided as dense annotations in the original SEN12MS dataset. So, to provide a better intuition of the accuracy of the single-label and multi-label scene annotations presented with this paper, we have randomly selected 600 patches for human annotation. This human annotation was conducted by visual inspection of the corresponding high-resolution aerial imagery in Google Earth. While this form of evaluation suffers from a certain subjectivity and the difficulty of distinguishing some land cover classes visually, it still serves the purpose to gain a better feeling for the label quality, if the numbers provided in Tab. 3 are taken with a grain of salt. In this evaluation, a single human label was given to each patch, and Top-1 accuracy refers to the case in which the MODIS-derived single-label scene annotation matches this human label, whereas Top-3 accuracy refers to the case in which one of the top-3 multi-label scene annotations matches this human label. As can be seen, the average accuracy is somewhere around 80%, which is significantly better than the original accuracy of the global IGBP land cover map. This is, of course, caused by the simplification of the IGBP scheme, as well as the reduction to just few scene labels instead of coarse-resolution pixel-wise annotations.

3. BASELINE MODELS FOR SINGLE-LABEL AND MULTI-LABEL SCENE CLASSIFICATION

To provide the community with both baseline models and results, we have selected two well-established convolutional neural network (CNN) architectures for image classification: ResNet and DenseNet. The architectures and the necessary adaptations and settings are shortly described in the following.

3.1 ResNet

The ResNet architecture (He et al., 2016) was designed to mitigate the problem of vanishing gradients, which tended to appear for very deep CNNs before. This is realized by the introduction

Table 2: The simplified IGBP land cover classification scheme.

IGBP Class Number	IGBP Class Name	Simplified Class Number	Simplified Class Name	Description	Color
1 2 3 4 5	Evergreen Needleleaf Forest Evergreen Broadleaf Forest Deciduous Needleleaf Forest Deciduous Broadleaf Forest Mixed Forest	1	Forest	Lands covered by woody vegetation at $> 60\%$ and height exceeding 2 m	009900
6 7	Closed Shrublands Open Shrublands	2	Shrubland	Lands with shrub canopy cover $> 10\%$ and < 2 m tall	c6b044
8 9	Woody Savannas Savanna	3	Savanna	Lands with understory systems, and with forest cover between 10% and 60% and > 2 m tall	fbff13
10	Grasslands	4	Grassland	Herbaceous lands with $< 10\%$ trees/shrubs	b6ff05
11	Permanent Wetlands	5	Wetlands	Lands with a permanent mixture of water and herbaceous or woody vegetation	27ff87
12 14	Croplands Cropland / Natural Vegetation Mosaics	6	Croplands	Lands covered with temporary crops followed by harvest and a bare soil period	c24f44
13	Urban and Built-up Lands	7	Urban/Built-up	Land covered by buildings and other man-made structures	a5a5a5
15	Permanent Snow and Ice	8	Snow/Ice	Lands under snow/ice cover throughout the year	69fff8
16	Barren	9	Barren	Lands with exposed soil, sand, rocks	f9ffa4
17	Water Bodies	10	Water	Oceans, seas, lakes, reservoirs, and rivers	1c0dff

Barren   **Croplands**   **Urban/Built-Up**   **Croplands**   **Urban/Built-Up**   **Savanna**
Forest   Savanna
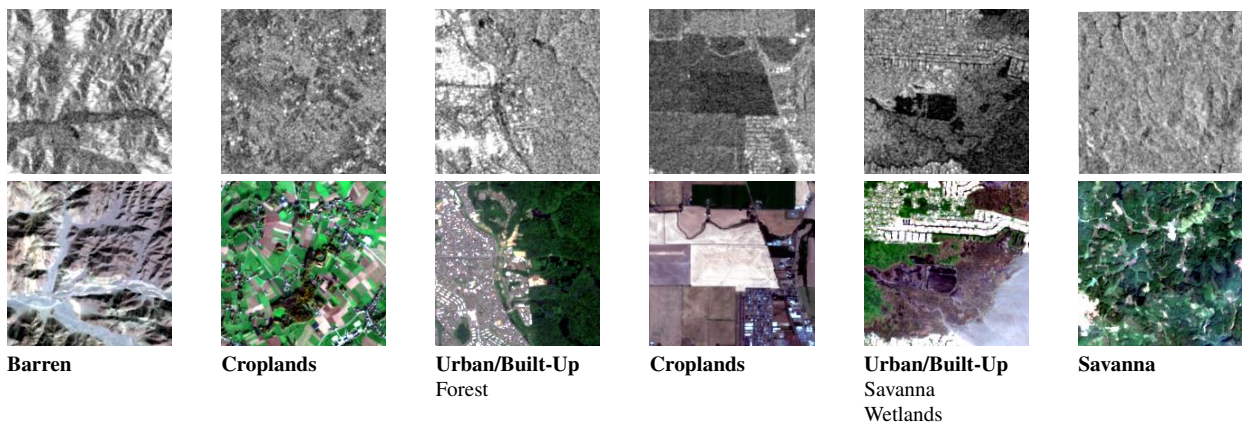   Wetlands

Figure 1: Some randomly selected samples from the SEN12MS dataset, including simplified IGBP land cover annotations on scene level. First row: Sentinel-1 (VV backscatter). Second row: Sentinel-2 (RGB). The main class of the scene is set in bold, additional classes in the multi-label case are set in regular font.

of so-called shortcut connections, i.e. instead of learning a direct mapping from input to output layer, the shortcut connection skips one or more layers by passing the original input through the network without modifications. Then, the network learns residual mappings with respect to this input. In this work, we used a ResNet50, i.e. a variant with a depth of 50 layers.

3.2 DenseNet

The DenseNet architecture (Huang et al., 2017) is based on the finding that CNNs can be deeper, more accurate and efficient to train if they contain shorter connections between layers close to the input and layers close to the output. Thus, DenseNets directly connect each layer to every other layer in order to ensure

maximum information flow between layers in the network. Different to ResNets, the features are not combined through summation; instead, they are concatenated. In this work, we employ a DenseNet121 with a depth of 121 layers.

3.3 Training Details

Both models were trained using binary cross entropy with logit loss for multi-label classification. To keep everything simple, optimization was performed with an Adam optimizer, a learning rate of 0.001, a decay rate of 10^{-5}, and a batch size of 64.

In order to evaluate the usefulness of different input data configurations, separate models were trained for the following cases:
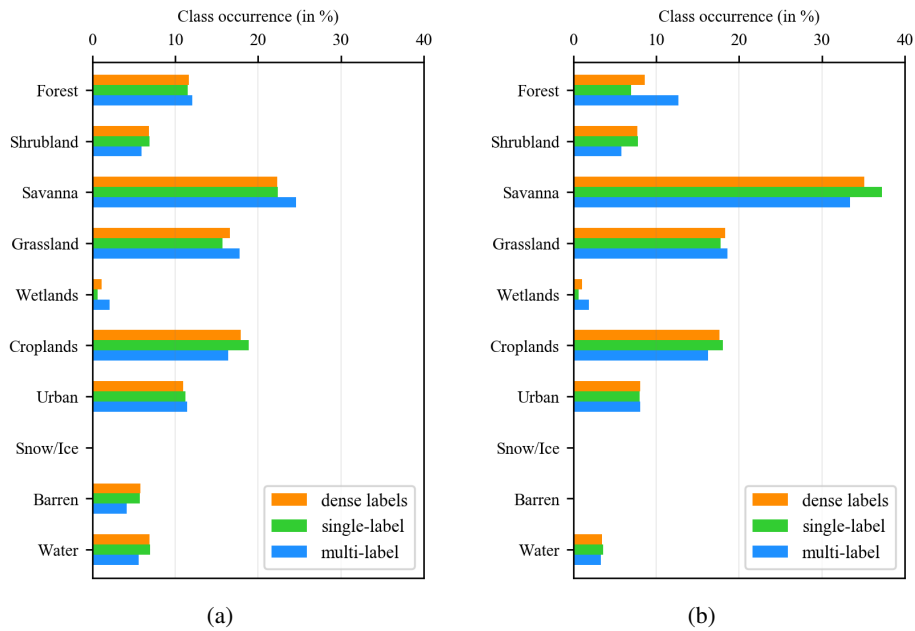
Figure 2: Class occurrences in the different formats of the SEN12MS dataset. (a) Training split, (b) hold-out testing split. Note that the number of *Barren* samples in the dense and the multi-label variant of the test set is actually not zero, but very small in comparison to the other classes.

Table 3: Validation of the Scene Labels with Respect to Human Annotation

Class	Number	Top-1 Acc.	Top-3 Acc.
Forest	66	80.3%	97.0%
Shrubland	26	46.2%	73.1%
Savanna	106	88.7%	95.3%
Grassland	64	68.8%	73.4%
Wetlands	3	66.7%	66.7%
Croplands	162	73.5%	84.6%
Urban/Built-Up	90	80.0%	93.3%
Snow/Ice	-	-	-
Barren	39	92.3%	94.9%
Water	44	97.7%	97.7%
Average	67	77.1%	86.2%

- S2_RGB: Sentinel-2 RGB data only (as a computer vision-like baseline)

- S2_MS: all 10 surface-related spectral channels of Sentinel-2

- S1+S2: Sentinel-1 dual-polarimetric data plus the 10 surface-related channels of Sentinel-2

Each model was trained from scratch on the official SEN12MS training split with early stopping based on a validation set randomly selected from the training set.

4. BENCHMARK RESULTS

Figure 3 illustrates multi-label prediction results for three example patches. Table 4 contains a summary of accuracy metrics for the multi-label classification results. The F1 score is the harmonic mean of the precision and recall metrics, i.e.

$$F_1 = 2\frac{p \cdot r}{p + r}, \qquad (1)$$

where the precision p is the fraction of correct predictions per all predictions of a certain class, and the recall r is the fraction of correct predictions per all appearances of a class in the reference annotations.

From the results, different insights can be drawn:

- There is generally not a huge difference between the two baseline CNN architectures. This is particularly interesting, because both models are of significantly different depths. Thus, this provides a hint towards the hypothesis that the achievable predictive power is more limited by the training data than by the model capacity.

- Overall, the weakest performances are achieved for those models that take only optical RGB imagery as input. However, a measurable improvement is observed when multi-spectral data is used, with even more improvement provided by the data fusion-based model that exploits both optical and SAR data. This suggests that spectral diversity is of high importance in remote sensing-based land cover mapping and confirms once more that there is a significant difference between the analysis of conventional photographs and remote sensing imagery.

- While SAR-optical data fusion is helpful on average, for some classes, e.g. *Shrubland*, *Wetlands*, and *Croplands* it does not seem to be very helpful. This indicates that observation-level fusion based on a simple channel concatenation is not enough and more sophisticated fusion strategies, e.g. with sensor-dependent CNN streams, are needed.

- Due to the reduction of dense labels to scene labels, the Barren class is underrepresented – there are only 29 patches carrying a Barren label in the multi-label test dataset. This renders the results for the Barren class insecure. It is still interesting to note that for both CNN architectures, data fusion provides the best result for this class, while multi-spectral imagery yields the worst result – even worse than RGB only.
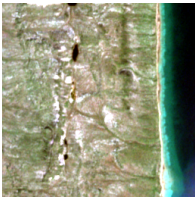
	ResNet50	DenseNet121	*Reference*
	Savanna Wetlands Water	Savanna Wetlands Water	*Savanna* *Wetlands* *Water*
	Savanna Urban/Built-up	Savanna Croplands	*Savanna*
	Urban/Built-up	Urban/Built-up	*Urban/Built-up*

Figure 3: Multi-label predictions for three example patches using the S1+S2 input data configuration. The second example is particularly noteworthy: While the reference denotes the patch as pure *Savanna*, the models add finer information by recognizing urban or cropland structures, respectively.

Table 4: F1 scores for the different models in the multi-label case.

		ResNet50			DenseNet121		
Class	Number	S2_RGB	S2_MS	S1+S2	S2_RGB	S2_MS	S1+S2
Forest	3529	60.9	69.3	70.6	65.4	68.9	70.6
Shrubland	1611	56.0	50.2	50.2	58.3	43.2	45.0
Savanna	9276	79.3	78.5	79.9	76.7	80.2	85.4
Grassland	5171	50.5	58.4	64.5	50.9	57.4	60.0
Wetlands	523	11.0	62.6	55.8	26.5	53.7	55.9
Croplands	4511	63.4	66.9	64.3	66.3	59.1	63.2
Urban/Built-Up	2235	82.8	82.6	84.1	84.1	84.9	83.1
Snow/Ice	-	-	-	-	-	-	-
Barren	29	1.3	0.0	11.5	8.0	0.6	8.7
Water	919	70.4	93.0	93.7	69.8	93.8	93.7
Average F1 Score		52.9	62.4	63.9	56.2	60.2	62.8
Overall F1 Score		66.5	69.9	71.4	66.9	68.9	72.0

- As already discussed in (Schmitt et al., 2020), the IGBP-based *Savanna* label can be problematic: The MODIS-derived reference considers the scene of the second example in Fig. 3 as *Savanna*, although it visually seems to rather be a mixture of built-up structures and croplands. Interestingly, both baseline models are able to identify those classes (i.e. *Urban / Built-Up* for ResNet50 and *Croplands* for DenseNet121) – albeit not both at the same time, and without removing the *Savanna* class.

All in all, the achieved accuracies are of the same order of magnitude as the accuracies reported by (Sumbul et al., 2019) for the BigEarthNet dataset. While SEN12MS and BigEarthNet are not directly comparable (as SEN12MS is more versatile and contains Sentinel-1 SAR imagery, but also uses a simpler class scheme), this indicates the usability of SEN12MS for the training and evaluation of scene classification models. Besides, the fact that achievable accuracies on both datasets are similar, this suggests a certain saturation for off-the-shelf image classification models for remote sensing scene classification. Investigating this further would be an interesting future research direction.

5. SUMMARY & CONCLUSION

With this paper, we have presented the SEN12MS dataset repurposed for remote sensing image classification. To achieve this, the original land cover annotations, which are provided at a resolution of 500 m and a pixel spacing of 10 m, are converted to both single-label and multi-label annotations. Based on a randomized validation of 600 patches by a human expert, an average accuracy of about 80% was estimated. Using the dataset and two standard CNN image classification architectures, we have trained and evaluated several baseline models, which can serve as a baseline for future developments and provide insight to the benefit of using multi-sensor and multi-spectral data over plain RGB imagery.

ACKNOWLEDGMENTS

REFERENCES

Abady, L., Barni, M., Garzelli, A. and Tondi, B., 2020. GAN generation of synthetic multispectral satellite images. In: Proc. SPIE 11533, p. 115330L.

Cheng, G., Han, J. and Lu, X., 2017. Remote sensing image scene classification: Benchmark and state of the art. Proc. IEEE 105(10), pp. 1865–1883.

Cheng, G., Xie, X., Han, J., Guo, L. and Xia, G. S., 2020. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. IEEE J. Sel. Top. Appl. Earth. Observ. Remote Sens. 13, pp. 3735–3756.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: Proc. CVPR, pp. 248–255.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., Liu, Q. and Song, L., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Science Bulletin 64(6), pp. 370–373.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In: Proc. CVPR, pp. 770–778.

Helber, P., Bischke, B., Dengel, A. and Borth, D., 2019. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12(7), pp. 2217–2226.

Hu, L., Robinson, C. and Dilkina, B., 2020. Model generalization in deep learning applications for land cover mapping. arXiv:2008.10351.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q., 2017. Densely connected convolutional networks. In: Proc. CVPR, pp. 2261–2269.

Liu, R., Lehman, J., Molino, P., Such, F. P., Frank, E., Sergeev, A. and Yosinski, J., 2018. An intriguing failing of convolutional neural networks and the CoordConv solution. arXiv:1807.03247.

Rußwurm, M., Wang, S., Körner, M. and Lobell, D., 2020. Meta-learning for few-shot land cover classification. In: Proc. CVPRW.

Schmitt, M., Hughes, L., Qiu, C. and Zhu, X., 2019. SEN12MS – a curated dataset of georeferenced multi-spectral Sentinel-1/2 imagery for deep learning and data fusion. In: ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. IV-2/W7, p. 153–160.

Schmitt, M., Prexl, J., Ebel, P., Liebel, L. and Zhu, X., 2020. Weakly supervised semantic segmentation of satellite images for land cover mapping – challenges and opportunities. In: ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. V-3-2020, pp. 795–802.

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P. and Friedl, M. A., 2019. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. Remote Sens. Env. 222, pp. 183–194.

Sumbul, G., Charfuelan, M., Demir, B. and Markl, V., 2019. BigEarthNet: A large-scale benchmark archive for remote sensing image understanding. In: Proc. IEEE IGARSS, pp. 5901–5904.

Xia, G., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L. and Lu, X., 2017. AID: A benchmark data set for performance evaluation of aerial scene classification. IEEE Trans. Geosci. Remote Sens. 55(7), pp. 3965–3981.

Xia, G.-S., Yang, W., Delon, J., Gousseau, Y., Sun, H. and Maître, H., 2010. Structural high-resolution satellite image indexing. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. 38-7A, pp. 298–303.

Yang, Y. and Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. In: Proc. 18th ACM SIGSPATIAL, pp. 270–279.

Yokoya, N., Ghamisi, P., Hänsch, R. and Schmitt, M., 2020. 2020 IEEE GRSS Data Fusion Contest: Global land cover mapping with weak supervision. IEEE Geosci. Remote Sens. Mag. 8(1), pp. 154–157.

Yu, Q., Liu, W. and Li, J., 2020. Spatial resolution enhancement of land cover mapping using deep convolutional nets. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. XLIII-B1-2020, pp. 85–89.

Yuan, Y., Tian, J. and Reinartz, P., 2020. Generating artificial near infrared spectral band from RGB image using conditional generative adversarial network. In: ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. V-3-2020, pp. 279–285.

Zhao, B., Zhong, Y., Xia, G. and Zhang, L., 2016. Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 54(4), pp. 2108–2123.

Zhou, W., Newsam, S., Li, C. and Shao, Z., 2018. Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval. ISPRS J. Photogramm. Remote Sens. 145, pp. 197–209.

Zhu, X. X., Hu, J., Qiu, C., Shi, Y., Kang, J., Mou, L., Bagheri, H., Haberle, M., Hua, Y., Huang, R., Hughes, L., Li, H., Sun, Y., Zhang, G., Han, S., Schmitt, M. and Wang, Y., 2020. So2Sat LCZ42: A benchmark data set for the classification of global local climate zones. IEEE Geosci. Remote Sens. Mag. 8(3), pp. 76–89.