Explicit Haze & Cloud Removal for Global Land Cover Classification

Ziqi Gu[1,2], Patrick Ebel[1], Qiangqiang Yuan[2], Michael Schmitt[3,4], Xiao Xiang Zhu[1,3]
[1]Technical University of Munich (TUM), [2]Wuhan University,
[3]German Aerospace Center (DLR), [4]Bundeswehr University Munich
{ziqi.gu, patrick.ebel, xiaoxiang.zhu}@tum.de,
qqyuan@sgg.whu.edu.cn, michael.schmitt@unibw.de

Abstract

Haze and clouds in Earth's atmosphere obstruct a seamless monitoring of our planet via optical satellites. Prior work shows that models can learn to adapt and perform remote sensing downstream tasks even in the presence of such sensor noise. So what are the auxiliary benefits of incorporating an explicit cloud removal task, and what is its relation to other tasks in the remote sensing pipeline? We address these questions and show that explicit cloud removal makes models for land cover classification furthermore robust to haze and clouds. Finally, we explore the relation to a self-supervised pre-text task (including abundant cloudy data) and demonstrate how to further ease the need for costly annotations on the land cover classification task.

1. Introduction

On average, over half of Earth is shrouded by haze and clouds [9]—impeding the capabilities of spaceborne sensors to continuously monitor our planet. While established benchmarks in remote sensing are carefully curated and cleared of any artifacts and noise [7, 10, 15, 17, 19, 20], there exist models that have specifically been investigated for their resilience to cloud coverage, as may be encountered in practical use cases: Notably, [5, 14] perform crop type classification and learn to ignore cloudy time points irrelevant to the target task.

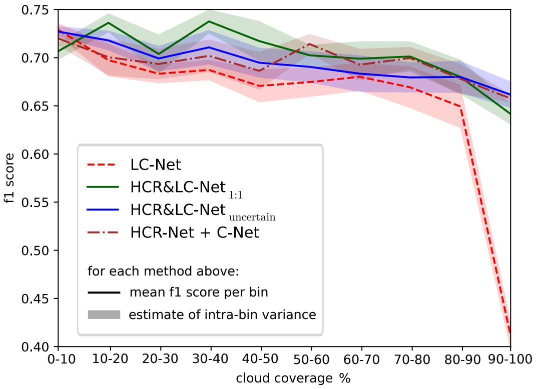

Figure 1. F1 Score of top models in Table 2 on LCC task as function of cloudy pixel %. In contrast to the models not trained on cloud removal, the ones that are perform robustly even in the presence of heavy cloud coverage.

Moreover, the model of [12] performs semantic segmentation and, as a side effect, learns to reconstruct cloud covered information. Together, the works of [5, 12, 14] raise the question of what the benefits may be of including an explicit haze & cloud removal (HCR) task [6, 11]. Herein, we investigate this question by analysing the benefits of HCR with respect to a global land cover classification (LCC) target task [16, 17, 19, 20]. We show that explicit cloud removal makes models for land cover classification more robust to haze and clouds. Specifically, we consider a multi-task setup, where a multi-task network with one specific branch for cloud removal and another one for land cover clas-

Architecture	# Parameters
LC-Net	103,851,050
HCR&LC-Net	106,899,319
HCR-Net + C-Net	107,473,335

Table 1. Comparison of learnable parameters per model.

sification is created. Finally, we explore the relation to a self-supervised pre-text task (including abundant cloudy data) and show how to further reduce the need for costly annotations in the land cover classification context.

2. Data

This work uses Sentinel-1 (S1) radar data and Sentinel-2 (S2) optical data from ESA's Copernicus mission. Combining S1 & S2 has shown beneficial for e.g. semantic segmentation, change detection, LCC, and HCR tasks [3,4,11,15–17]. We take S1 and cloud-free S2 data from the SEN12MS data set [15] with the LCC labels of [16] for patch-wise land cover classification. We take co-registered cloudy S2 data from the associated SEN12MS-CR data set [2]. The geo-spatial coverage of images contained in SEN12MS-CR is a subset of the geo-spatial coverage of images contained in the original SEN12MS dataset, so we focus on that subset for which all data modalities are available. The train and test splits are the intersections of those defined in [2, 15] and of sizes 109,549 and 12,666, respectively. The cloud coverage is at 58 (± 37) % and 66 (± 37) % per split, as estimated by pixel-wise cloud masks m computed via s2cloudless [21]. In sum, each sample is a triplet $(S1, S2_{clear}, S2_{cloudy})$ of $(256\ px)^2$ patches with masks m and associated multi-class target labels t as in [16].

3. Methods

As a backbone architecture for our model, U-Net [13] is chosen for it's dual purpose of extracting features related to the *what* and *where* of task-related information, being equally valuable for image reconstruction and classification. Furthermore, it is close to the architecture of [12]. For the training of the global LCC task, we use the cross-entropy loss. For the HCR task, a pixel-wise CARL image reconstruction loss as in [11] is utilized.

We train all models for 10 epochs with batch size 64 via ADAM with learning rate 10^{-3} as well as weight decay 10^{-5} and observed convergences within this schedule. For each model, its best checkpoint as assessed on a validation split (a fixed 10% random sub-set of the training split) is chosen for subsequent testing. The Multi-task training of LCC and HCR is done via (a) naive 1:1 task weighting and (b) the uncertainty weighting of [8]. These models are denoted as *HCR&LC-Net*$_{1:1}$ and *HCR&LC-Net*$_{uncertain}$. Importantly, the baseline *LC-Net* only trained for LCC (without an explicit cloud removal task) is readily trained on the cloud-covered S2 data in order to learn implicitly ignoring task-irrelevant cloudy pixels. This is to follow the approaches of [5, 12, 14] and to get competitive baselines. Moreover, while this network misses layers dedicated to image reconstruction, we controlled for parametric comparability across models by spending additional learnable parameters on the LCC branch of the baseline. As another baseline, we also consider a sequential ensemble of an HCR net followed by a down-stream LCC net, and call this baseline *HCR-Net + C-Net*. The parametric complexities of all models are given in Table 1.

As a final experiment, to explore how the reliance on costly global land cover annotations for the LCC target task can be reduced, we investigate a self-supervised pretext task in the setting of fewer supervised training data points. Our pretext task is to predict the geo-spatial relationship between the concatenated S1 and S2 input bands. While there's a 100% spatial overlap for the co-registered S1 and S2 data of section 2 as used in the supervised setting, here we randomly pair 1 of 4 neighboring S1 patches to a given S2 input with 50% overlap among another. Specifically, the network is tasked to classify the paired $(S1,\ S2_{cloudy})$ input's spatial relation as quantified via a cross entropy loss. That is, the network classifies whether the S1 patch is north, west, south or east of its paired S2 patch. This serves to pre-train the weights of the classification branch (depicted in red and green

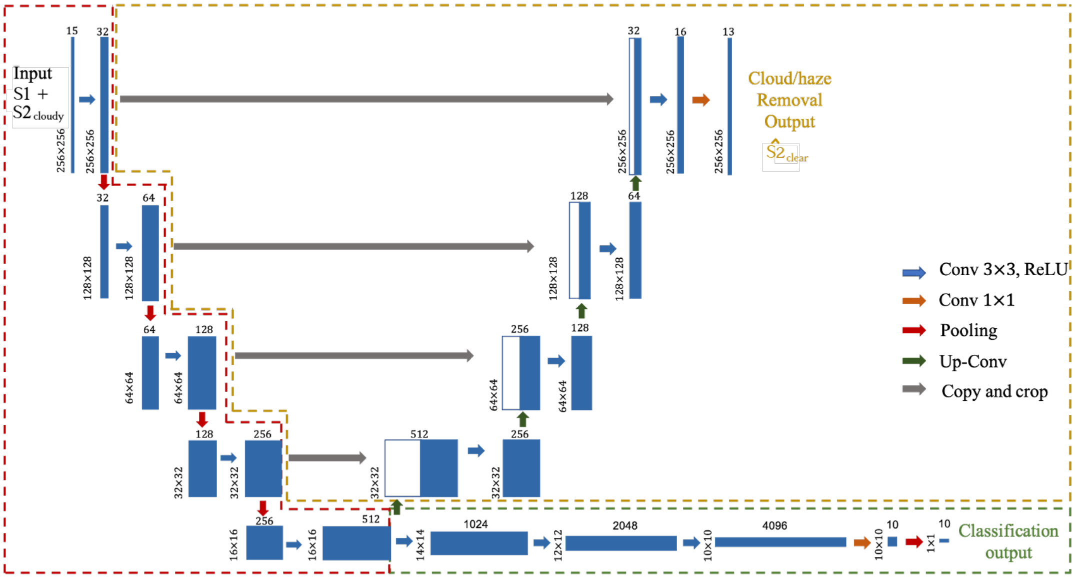

Figure 2. Architecture of HCR&LC-Net. The red sub-net contains shared layers for both tasks. The yellow and green parts are for HCR and LCC tasks, respectively. The joint red and yellow parts correspond to the classical U-Net [13].

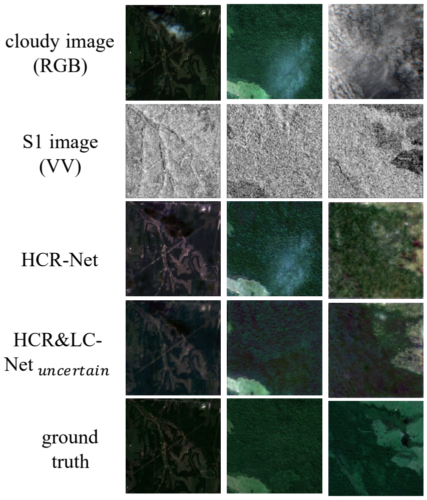

Figure 3. Three exemplary cloud removal results. Rows: Cloudy image, S1 image, HCR-Net prediction, HCR&LC-Net$_{uncertain}$ prediction, ground truth. Our multi-task network can successfully remove most clouds.

in Fig. 2) without any reliance on costly LCC labels. Including pre-training, optimization is done in 3 steps: 1) The classification part of our multi-task network (without the cloud removal branch) is pre-trained for 20 epochs under the pretext task. 2) We freeze the classification weights learned in the first step and supervisedly train our multi-task network for the layers not included before (i.e. the cloud removal layers) for 5 epochs. 3) Finally, we unfreeze these weights and train all layers of the multi-task network together for another 10 epochs as described earlier to obtain the final model.

4. Results

The models of section 3 are evaluated in terms of their goodness on the LCC task (Precision, Recall, F1 score) and on image reconstruction (PSNR, SSIM [18]), as shown in Table 2 with (second) best results highlighted in bold (/italic). On average, all networks or ensembles involving HCR outperform those without on the LCC task. Moreover, our proposed multi-task network performs best on LCC and is a close second on the HCR task, only behind the model specialising on cloud removal. Exemplary HCR predictions are depicted in Fig. 3.

Table 2. Quantitative results for LCC and HCR tasks. (Second) best results highlighted in bold (/italic).

	Land Cover Classification			Cloud/Haze Removal	
	Precision	Recall	F1	PSNR	SSIM
LC-Net	0.6759	0.6650	0.6704	\	\
HCR-Net	\	\	\	**27.9342**	**0.8975**
HCR&LC-Net$_{1\&1}$	0.7042	0.6572	0.6799	27.4370	0.8844
HCR&LC-Net$_{uncertain}$	0.6697	0.7004	**0.6847**	*27.8973*	*0.8941*
HCR-Net + C-Net	0.6831	0.6823	*0.6827*	\	\

To further analyze LCC performances, Fig.1 evaluates each model's F1 score as a function of cloud coverage. Mean scores are evaluated in bins of 10 % steps, standard deviations are estimated by sub-sampling each bin into 10 sub-groups to compute within-bin variances. Fig.1 shows that models including an explicit cloud removal task (at no additional parametric cost) are more robust to dense and heavy coverage than nets learning implicitly to ignore cloudy pixels irrelevant to the LCC task. In summary, the results show that adding an explicit HCR task into a LCC pipeline further improves the final classification results, especially when dealing with high cloud coverage.

Finally, we investigate the benefits of the proposed self-supervised pretext task. Fig. 4 shows performances as a function of annotated training data set size (in steps of 25%). While the model without self-supervision suffers severely from the lack of labeled data, the self-supervised model remains relatively robust. Mind the gaps widening further with fewer data and note that with only 25% data, we can already achieve around 95% of the final F1 score when utilizing the proposed pretext task. Finally, with an F1 score of 0.703 the self-supervised model trained on all data performs best on the LCC task, out of all benchmarked models. These findings underline the effectiveness of the proposed self-supervised pretext task for scenarios where costly annotated geospatial data is rare.

5. Conclusion

This work demonstrates the benefits of an explicit image reconstruction task for cloud removal and its benefits on a common global LCC remote sensing downstream task. While preceding work demonstrated that neural networks can learn ignoring data points noisy (i.e. cloud-covered) or unrelated to the target task [5, 12, 14], we show that incorporating an explicit cloud removal task can make the model even more robust at no additional parametric costs. Changes to networks are kept minimal, and more advanced architectures will be addressed in the future. Rather, we focused on parameter-neutral adjustments of tasks and investigating their interactions. Finally, a self-supervised pretext task was proposed to further improve our results and extend global LCC to the common scenario of few annotated data. We plan to extend our analysis to related tasks and other data sets [1].

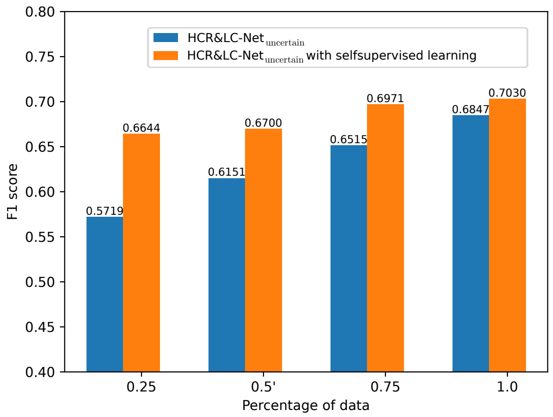

Figure 4. F1 score on LCC task for 25, 50, 75 and 100 % of labeled LCC training data. With versus without self-supervised pre-training. With self-supervision, the model is more robust to lack of annotated data.

References

[1] Miriam Cha, Kuan Wei Huang, Morgan Schmidt, Gregory Angelides, Mark Hamilton, Sam Goldberg, Armando Cabrera, Phillip Isola, Taylor Perron, Bill Freeman, et al. MultiEarth 2022–Multimodal Learning for Earth and Environment Workshop and Challenge. *arXiv preprint arXiv:2204.07649*, 2022. 4

[2] Patrick Ebel, Andrea Meraner, Michael Schmitt, and Xiao Xiang Zhu. Multisensor Data Fusion for Cloud Removal in Global and All-season Sentinel-2 Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):5866–5878, 2020. 2

[3] Patrick Ebel, Sudipan Saha, and Xiao Xiang Zhu. Fusing multi-modal data for supervised change detection. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:243–249, 2021. 2

[4] Patrick Ebel, Yajin Xu, Michael Schmitt, and Xiao Xiang Zhu. SEN12MS-CR-TS: A Remote-Sensing Data Set for Multimodal Multitemporal Cloud Removal. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022. 2

[5] Vivien Sainte Fare Garnot, Loic Landrieu, Sebastien Giordano, and Nesrine Chehata. Satellite Image Time Series Classification with Pixel-set Encoders and Temporal Self-attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12325–12334, 2020. 1, 2, 4

[6] Ziqi Gu, Zongqian Zhan, Qiangqiang Yuan, and Li Yan. Single Remote Sensing Image Dehazing Using A Prior-based Dense Attentive Network. *Remote Sensing*, 11(24):3008, 2019. 1

[7] Pascal Kaiser, Jan Dirk Wegner, Aurélien Lucchi, Martin Jaggi, Thomas Hofmann, and Konrad Schindler. Learning Aerial Image Segmentation from Online Maps. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6054–6068, 2017. 1

[8] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7482–7491, 2018. 2

[9] Michael D. King, Steven Platnick, W. Paul Menzel, Steven A. Ackerman, and Paul A. Hubanks. Spatial and Temporal Distribution of Clouds Observed by MODIS Onboard the Terra and Aqua Satellites. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7):3826–3852, Jul 2013. 1

[10] Lukas Kondmann, Aysim Toker, Marc Rußwurm, Andrés Camero, Devis Peressuti, Grega Milcinski, Pierre-Philippe Mathieu, Nicolas Longépé, Timothy Davis, Giovanni Marchisio, et al. DENETHOR: The DynamicEarthNET Dataset for Harmonized, Inter-Operable, Analysis-Ready, Daily Crop Monitoring from Space. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021. 1

[11] Andrea Meraner, Patrick Ebel, Xiao Xiang Zhu, and Michael Schmitt. Cloud Removal in Sentinel-2 Imagery Using A Deep Residual Neural Network and SAR-optical Data Fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:333–346, 2020. 1, 2

[12] Muhammad Usman Rafique, Hunter Blanton, and Nathan Jacobs. Weakly Supervised Fusion of Multiple Overhead Images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1479–1486. IEEE, 2019. 1, 2, 4

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 234–241. Springer, 2015. 2, 3

[14] Marc Rußwurm and Marco Körner. Temporal Vegetation Modelling Using Long Short-term Memory Networks for Crop Identification From Medium-resolution Multi-spectral Satellite Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 11–19, 2017. 1, 2, 4

[15] Michael Schmitt, Lloyd Haydn Hughes, Chunping Qiu, and Xiao Xiang Zhu. SEN12MS – A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume IV-2/W7, pages 153–160, 2019. 1, 2

[16] M. Schmitt and Y. L. Wu. Remote Sensing Image Classification with the SEN12MS Dataset. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 52:101–106, June 2021. 1, 2

[17] Gencer Sumbul, Marcela Charfuelan, Begüm Demir, and Volker Markl. BigEarthNet: A

Large-scale Benchmark Archive for Remote Sensing Image Understanding. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 5901–5904. IEEE, 2019. 1, 2

[18] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, Apr 2004. 3

[19] Yi Yang and Shawn 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, 2010. 1

[20] Xiao Xiang Zhu, Jingliang Hu, Chunping Qiu, Yilei Shi, Jian Kang, Lichao Mou, Hossein Bagheri, Matthias Haberle, Yuansheng Hua, Rong Huang, et al. So2Sat LCZ42: A Benchmark Data Set for the Classification of Global Local Climate Zones [Software and Data Sets]. *IEEE Geoscience and Remote Sensing Magazine*, 8(3):76–89, 2020. 1

[21] Anze Zupanc. Improving Cloud Detection with Machine Learning. `Sentinel-Hub.` https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13, 2017. Accessed: 2022-04-15. 2