Impact of Cloud Removal on Urban Classification Tasks

Matthias Aarnoutse
m.f.a.aarnoutse@umail.leidenuniv.nl
Leiden University
Leiden, The Netherlands

Niels Witte
n.witte@umail.leidenuniv.nl
Leiden University
Leiden, The Netherlands

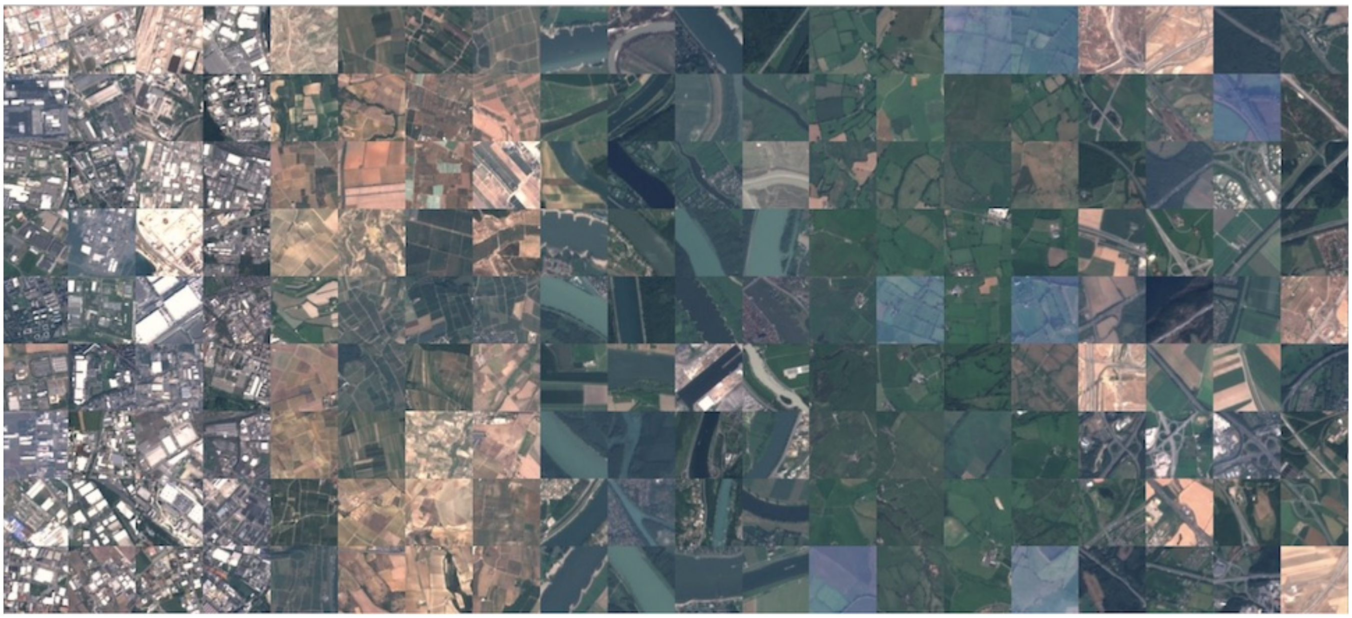


Figure 1. EuroSAT: Land Use and Land Cover Classification with Sentinel-2 [3]

Abstract

In this paper we explore the impact of cloud removal on the downstream task of land cover type classification. We use the DSen2-CR [7] network as our cloud removal algorithm. We use the data set SEN12MS-CR [7] containing pristine and cloudy Sentinel-2 imagery including Sentinel-1 SAR data on our cloud removal task. The resulting images are classified using a ResNet-50 classifier trained on the EuroSAT [3] data set containing 10 land cover types. We show that a significant improvement in performance can be gained when using Dsen2-CR as a pre-processing step on cloudy images for the land cover classification task.

Urban Computing Fall'21, January 2022, Leiden, Netherlands

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/1122445.1122456

Keywords: cloud removal, deep learning, land cover type classification

ACM Reference Format:
Matthias Aarnoutse and Niels Witte. 2021. Impact of Cloud Removal on Urban Classification Tasks. In *Proceedings of ACM Conference (Urban Computing Fall'21).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

1 Introduction

The Earth's surface is by 67% covered in clouds on average as shown by an analysis of over 12 years of observations made by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument of the satellites Terra and Aqua. If we only consider land surfaces, this percentage is reduced to 55%. Still, that is a lot of clouds on average. It makes sense to use cloud removal to obtain 'clean' data from satellite imagery, but how useful are these images?

In this paper we use a downstream task to evaluate the impact of cloud removal on a land cover classification task. We use the cloud removal pipeline from [Meraner et al.], which removes clouds from Sentinel-2 imagery using Sentinel-1 SAR data with a deep residual neural network called DSen-2CR. In the remainder of this paper we will be answering the following question; *what impact does cloud removal have on the accuracy for the land cover classification task in comparison to the pristine cloudy images?*

1.1 Related work

In *Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion* [7] the authors propose an extension to the DSen-2 [6] Deep Neural Network (DNN) architecture in order to more effectively remove clouds from satellite imagery. The authors claim that this is the first paper using non-generated clouds as their data set, which introduces a more accurate representation of real-world scenarios.

Missions. The DSen-2 model is trained on the optical data from the Copernicus Sentinel-2 mission. The mission provides data for tasks such as risk management, land cover and environmental monitoring. The Copernicus Sentinel-2 mission is a constellation of two identical satellites in the same orbit. As the mission aims to provide data for environmental monitoring it has a high revisit time of 2-3 days at mid-latitudes.

The Copernicus Sentinel-1 mission also provides imagery of coastal zones and landmasses at regular intervals. Sentinel-1 is equipped with a Synthetic Aperture Radar (SAR) which acquires data in the C-band, this refers to a portion of the electromagnetic spectrum in the 4GHz to 8GHz frequency range. This means the SAR can 'see' through the clouds as it uses microwave signals that bounces off the earth back into the sensor. You could think of this as a heightmap of the Earth.

DSen2-CR. DSen2-CR, the model proposed by [Meraner et al.] is based on Deep Sentinel-2 (DSen-2) [5] ResNet [2]. Cloud removal can be seen as an image reconstruction task, where the missing spatial and spectral information (behind the clouds) has to be restored in the image. With this goal in mind the authors introduce a novel loss function $\mathcal{L}_{\text{CARL}}$ which is a combination of the MAE and a *Cloud Adaptive Part* which ensures that as much of the original image not covered in clouds is retained.

The authors test their contributions by performing a series of experiments where they introduce their loss function and compare it to a more the more general `MAE` loss function. They also perform an experiment where they test the addition of the Sentinel-1 data and show that adding this data also shows significant improvements to the predictions.

2 Data

In order to find if cloud removal is viable, we will run two experiments. In the first experiment we remove clouds in order to generate cloudless images. In the next experiment we predict the land cover type based on these generated images.

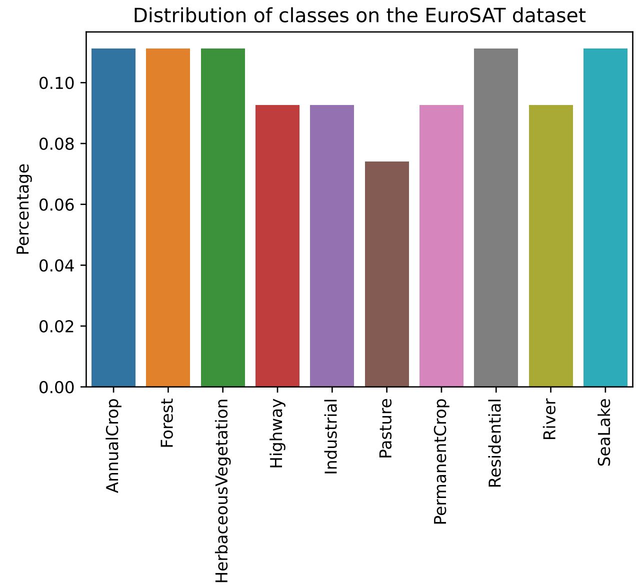


Figure 2. EuroSAT class distribution

For the cloud removal task we make use of the SEN12MS-CR dataset, these then serve as the input for the land cover classification task as well. To train the network of this latter task we use EuroSAT.

SEN12MS-CR. The SEN12MS-CR data set [7] is an evolution on the SEN12MS data set and is used to train the DSen2-CR model. This new data set contains triplets of cloudy, cloudless Sentinel-2 images and the corresponding Sentinel-1 image. The data set contains 169 non-overlapping regions of interest (ROI). The ROIs have an average size of approximately 5200×4000px. This corresponds to 52×40km of ground coverage. The ROIs have been adapted to 256×256px patches with a 128px stride, which means that there is a 50% overlap in the data set.

EuroSAT. The EuroSAT dataset [3] is based on Sentinel-2 satellite images covering 13 spectral bands and consisting of 10 classes with 27000 labelled and geo-referenced samples. The 10 generic land cover types are; *Highway*, *Industrial*, *Residential*, *Pasture*, *Forest*, *Herbaceous Vegetation*, *Sea Lake*, *River*, *Permanent Crop*, and *Annual Crop*. The distribution of these classes within the data are shown in Figure 2. Previous research has shown that an accuracy of 98.57% can be achieved using a ResNet-50 model [3].

3 Experiments

In order to measure the performance of the newly generated cloudless images using DSen2-CR we measure the performance using a downstream task. As one of the purposes of the Sentinel-2 mission is land cover monitoring, we chose land cover type classification as our downstream task.

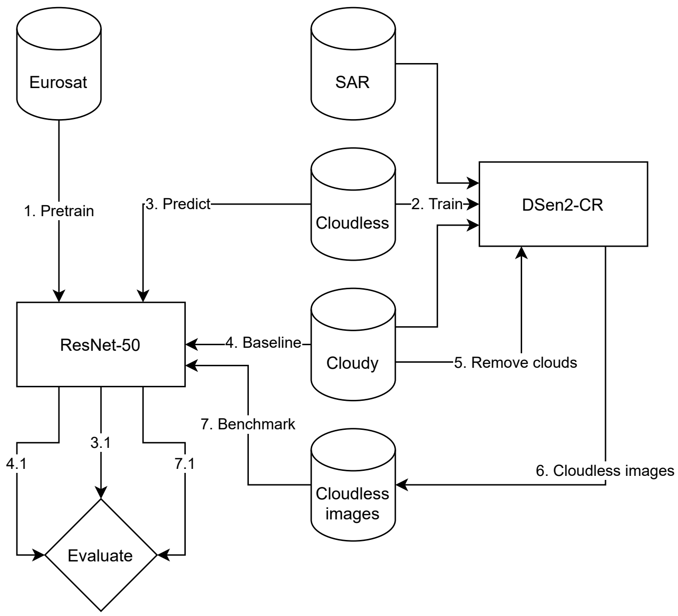


Figure 3. Visual representation of our experiment pipeline, the numbers indicate the order in which the steps happen. 4.1, 3.1, 7.1 represent the results from the steps 3, 4, 7 respectively. The results are all saved and evaluated.

A visual aid to our experiment pipeline can be found in Figure 3.

We first train a ResNet-50 classifier comparable to the one in [3] on the labelled EuroSAT dataset for our land cover type classification task. However, instead of using only a subset of bands such as RGB, we use the full spectrum of bands, this is done so the output of DSen2-CR can directly be used as input for the downstream task. It is however proven that ResNet-50 performs better when given just the RGB bands instead [3].

3.1 Setting a baseline

In the first experiment we use the pristine images from the Sentinel-2 data set used in the DSen-2CR paper [7] on our land cover classification task. The predicted labels will be our 'ground truth' for the next experiment with the generated cloudless images. We also predict the land cover type for all of the cloudy images to set a baseline. We then calculate the accuracy based on this ground truth for the generated cloudless images and the cloudy images. If the generated cloudless images score worse than the cloudy images, then there is no reason to apply cloud removal at all.

3.2 Prediction on the generated cloudless images

We then run these cloudy images through the DSen-2CR model which results in a set of generated cloudless images. These images are then classified using the same ResNet-50 classifier from the previous experiment which results in our final predictions. Comparing these new predictions to the earlier predicted 'ground truth' allows us to compute the accuracy on the ResNet-50 classifier, and more importantly, it will show us the impact of the cloud removal models as preprocessing steps on the accuracy.

4 Results

Setting a baseline. Our first course of action is to train the ResNet-50 model which will later be used to generate our 'ground truth' dataset. The classifier was implemented for us in the Tensorflow library [1]. We only configure this network to have an extra 10 node output layer to correspond with our classes. The top input layer is also configured to handle our 13 input channels compared to the 3 that ResNet-5 is usually trained with. The weights were initialised using a uniform random initializer. During training we use the Adam optimizer [4] using a custom learning schedule: the learning rate starts at 0.001 and is divided by 5 at epochs 50, 100, 200 and 250. This schedule is based on the work done by [Smith et al.]. Figure 5 presents the accuracy over 200 epochs for the ResNet-50 classifier. We achieve a maximum accuracy of 92.9%.

The distribution of the predicted classes on our 'ground truth' data set are shown in Figure 6a.

Predicting the land cover type of the cloudy images in the SEN12MS-CR dataset results in an accuracy of 39.87%.

Prediction on the generated cloudless images. An example of generating a cloudless image using the DSen2-CR network is shown in Figure 4. The DSen2-CR network receives Figure 4a and Figure 4b as their input and targets Figure 4c.

Using these generated cloudless images on the previously trained ResNet-50 classifier results in an accuracy of 75.65%. The results are shown in Table 1.

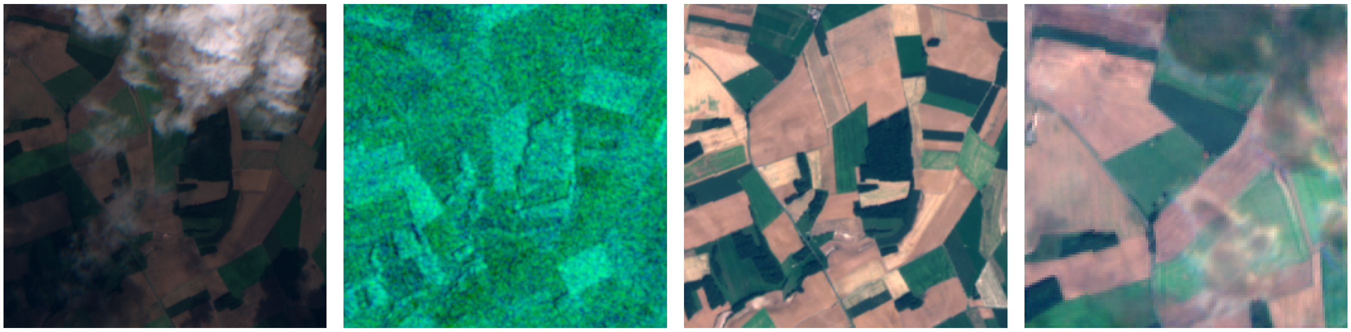
Baseline (Cloudy)	Cloudless
39.87%	75.65%

Table 1. Accuracy on the land cover type classification task

5 Discussion

Although we could not reproduce the performance of the the ResNet-50 network on the EuroSAT dataset shown in [3], we managed to achieve an accuracy of 92.9% on the validation set. We think that it is fairly interesting that we did not manage to reproduce the results as at first we used the exact parameters used in [3].

The prediction on the cloudless images leads to an accuracy of 75.6%, which is significantly lower than the accuracy on the EuroSAT dataset. We do still think that the images



(a) Cloudy **(b)** SAR **(c)** Pristine **(d)** Cloudless prediction

Figure 4. Results

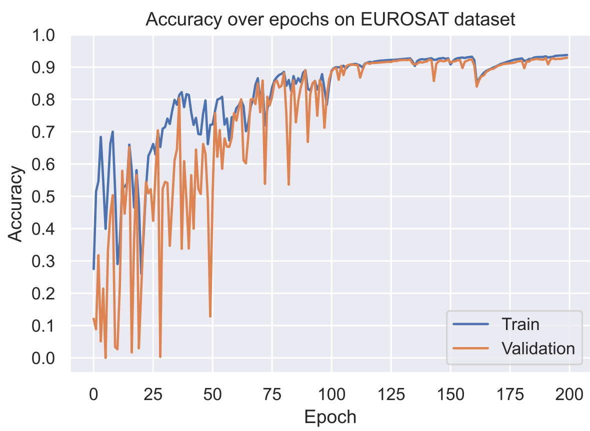
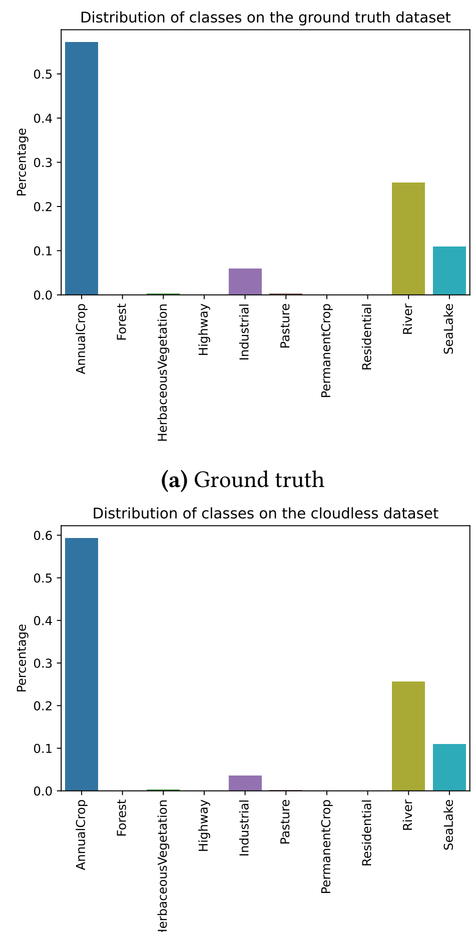


Figure 5. Accuracy over epochs for our ResNet-50 classifier trained for a total of 200 epochs



(a) Ground truth

(b) Cloudless

Figure 6. Class distributions

are viable to use for a downstream land cover classification task, as it is significantly more accurate than the baseline which only classified 39.87% of the samples correctly.

As shown in Figure 6a the predicted class distribution of our 'ground truth' is vastly different than the original class distribution in the EuroSAT dataset (Figure 2). This is something we cannot confirm due to the nature of our experimental setup. We have to assume this is correct. However, if we look at the distribution of the classes in the generated cloudless dataset we see a similar distribution. While this does not confirm that the ground truth is correct, it shows that we get consistent results. It could be that the distribution in the SEN12MS-CR dataset is skewed in regards to land cover types. Another hypothesis, which might be more likely, is that 'AnnualCrop' is predicted more often when it should have been 'HerbaceousVegetation' or 'Pasture' for example. Figure 7 confirms this hypothesis, showing that the classes 'Highway' and 'Pasture' are mislabelled as 'AnnualCrop'. Another interesting piece of information that this matrix presents is that predictions for the 'SeaLake' class seem to cover itself, as well as the entirety of the 'Residential' class.

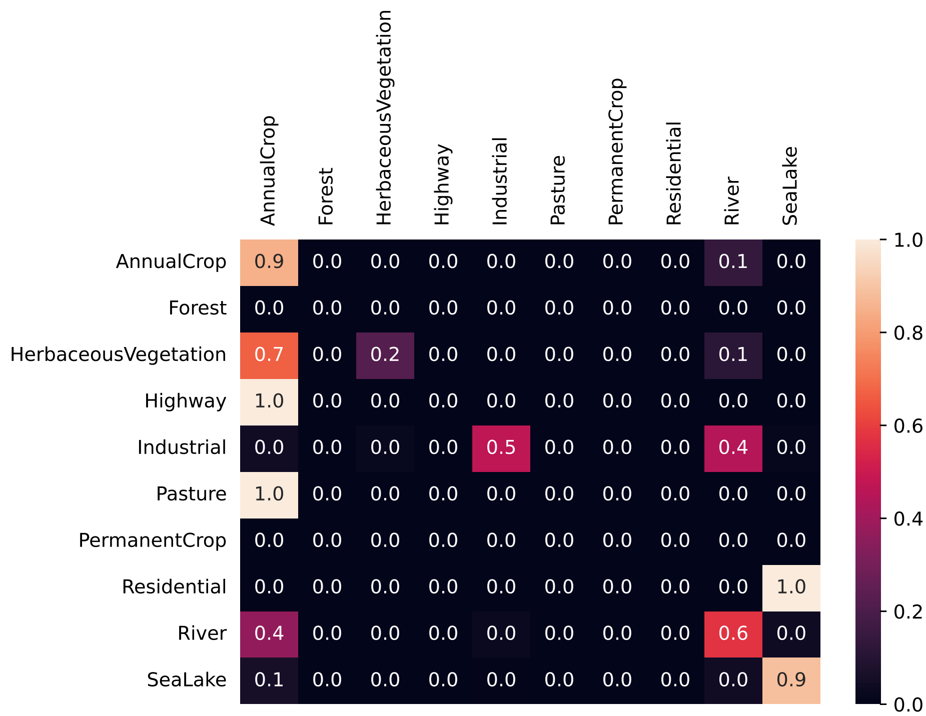


Figure 7. Confusion matrix for the predictions on the cloudless dataset compared to the ground truth dataset.

6 Conclusion

In this paper we show the viability of cloud removal to improve performance on downstream tasks, such as land cover classification, using the DSen2-CR model proposed by [Meraner et al.]. We do this by training a ResNet-50 model [2] which has proven its performance for this task on the EuroSAT dataset [3]. We then use this trained model to label an existing dataset used by the authors of the DSen2-CR paper [7]. We then use the same classifier to predict labels for cloudy images and for images which have their clouds removed. We show a significant increase in performance when applying DSen2-CR as a preprocessing step to cloudy images.

The idea seemed great at first, it is however difficult to confirm how realistic the results are. Compared to our ground truth the generated images perform well and could be used to 'create' information when clouds are blocking vision. There are still question marks surrounding our ground truth. Every result has to be taken with a grain of salt here.

In future work we would like to compare other cloud removal techniques using this downstream task. In fact our original plan included a comparison with *pix2pix*. Due to some issues we had with getting a working DSen2-CR model, we had to abandon this idea.

The DSen2-CR code from the authors' GitHub was fairly dated. This brought a lot of challenges with it trying to get it running. In the end we managed to update the code to TensorFlow 2.0 to get it working. However, with this basis it might be worth to compare it with other techniques.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/ Software available from tensorflow.org.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs.CV]

[3] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. 2019. EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. arXiv:1709.00029 [cs.CV]

[4] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG]

[5] Charis Lanaras, José Bioucas-Dias, Silvano Galliani, Emmanuel Baltsavias, and Konrad Schindler. 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing* 146 (Dec 2018), 305–319. https://doi.org/10.1016/j.isprsjprs.2018.09.018

[6] Charis Lanaras, José M. Bioucas-Dias, Silvano Galliani, Emmanuel Baltsavias, and Konrad Schindler. 2018. Super-Resolution of Sentinel-2 Images: Learning a Globally Applicable Deep Neural Network. *CoRR* abs/1803.04271 (2018). arXiv:1803.04271 http://arxiv.org/abs/1803.04271

[7] Andrea Meraner, Patrick Ebel, Xiao Xiang Zhu, and Michael Schmitt. 2020. 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 (2020), 333–346. https://doi.org/10.1016/j.isprsjprs.2020.05.013

[8] Samuel L. Smith, Pieter-Jan Kindermans, Chris Ying, and Quoc V. Le. 2018. Don't Decay the Learning Rate, Increase the Batch Size. arXiv:1711.00489 [cs.LG]