MULTI-SENSOR TIME SERIES CLOUD REMOVAL FUSING OPTICAL AND SAR SATELLITE INFORMATION

Patrick Ebel[1], *Yajin Xu*[1], *Michael Schmitt*[2,3], *Xiao Xiang Zhu*[1,2]

[1]Data Science in Earth Observation(SiPEO), Technical University of Munich (TUM), Munich, Germany
[2]Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany
[3]Chair of Earth Observation, Bundeswehr University Munich, Neubiberg, Germany

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9883238

ABSTRACT

On average, about half of all optical satellite data observing Earth is covered by haze or clouds. These atmospheric disturbances hinder the ongoing observation of our planet and prevent the seamless application of established remote sensing methods. Accordingly, to allow for an ongoing monitoring of Earth, approaches to reconstruct optical spaceborne observations are required. This work introduces a new data set, SEN12MS-CR-TS, for the purpose of multi-sensor time series cloud removal. SEN12MS-CR-TS consists of co-registered radar and optical satellite data, featuring a sequence of bi-weekly observations throughout an entire year. Finally, we demonstrate the usability of our novel data set by developing a new multi-sensor time-series cloud removal architecture. We are positive that our curated data set as well as the proposed model will advance future research in satellite image reconstruction and benefit the expanding adaptation of global and all-weather remote sensing applications.

Index Terms— synthetic aperture radar, optical imagery, image reconstruction, time series, data fusion

1. INTRODUCTION

The majority of Earth's surface is covered by clouds [1], impacting the aim of missions like ESA's Copernicus to reliably provide noise-free observations at a high frequency, a prerequisite for applications such as change detection [2]. Subsequently, there exists a need for techniques that allow established remote sensing applications to remain operation even in the presence of haze and clouds. With the aim of removing such noise from optical satellite observations, cloud removal has become an increasingly active domain of research. Preceding methods have followed a multi-sensor data fusion approach [3, 4], reconstructing the cloud-covered optical imagery with the aide of complementary synthetic aperture radar (SAR) data, which is practically unaffected by haze or cloud coverage. While SAR allows reconstructing terrain and coarse shapes, there is a considerable domain gap that incorporating historical optical data may allow to bridge. In this line, multi-temporal cloud removal approaches integrate spectral information across time [5, 3]. The contribution of our work is in combining multi-modal with multi-temporal cloud removal approaches, and in providing a data set for training and benchmarking such methodology. Our new data set, SEN12MS-CR-TS, contains globally sampled time series of co-registered synthetic aperture radar (SAR) Sentinel-1 (S1) and optical Sentinel-2 (S2) observations. It builds on the mono-temporal SEN12MS-CR benchmark [6] and complements it with time series observations. As a proof of concept, we design a neural network architecture that makes use of multi-temporal SAR and optical data. The proposed model is trained and evaluated on SEN12MS-CR-TS to demonstrate the opportunities of leveraging multi-modal multi-temporal information for cloud removal purposes in remote sensing applications.

2. RELATED WORK

The data set described in this work builds on the preceding mono-temporal SEN12MS-CR [6] data set. Both utilize the same geospatial definitions of ROI and share compatible train and test split definitions. In fact, the ROI and splits of SEN12MS-CR-TS are a subset of those defined for SEN12MS-CR and the two data sets are compatible with one another. An overview of both data sets' ROI is given in Fig. 1. We encourage using both together, as exemplified in this work, to combine the benefits of multi-temporal information processing with an extensive geospatial coverage. In terms of its network architecture, our proposed model builds on the preceding works of [4, 5]. While those proceeding approaches are either exclusively multi-modal or solely multi-temporal, our network makes use of both multi-sensor and time-series information. Finally, our developed method differs from the work of [7], which introduced a sequence-to-sequence cloud removal technique trained directly on but only applicable to the target data. In comparison, our method performs sequence-to-point cloud removal and, once trained, can be generically applied to any ROI on Earth. This makes our model generally applicable to any ROI according to the needs of remote sensing practitioners.

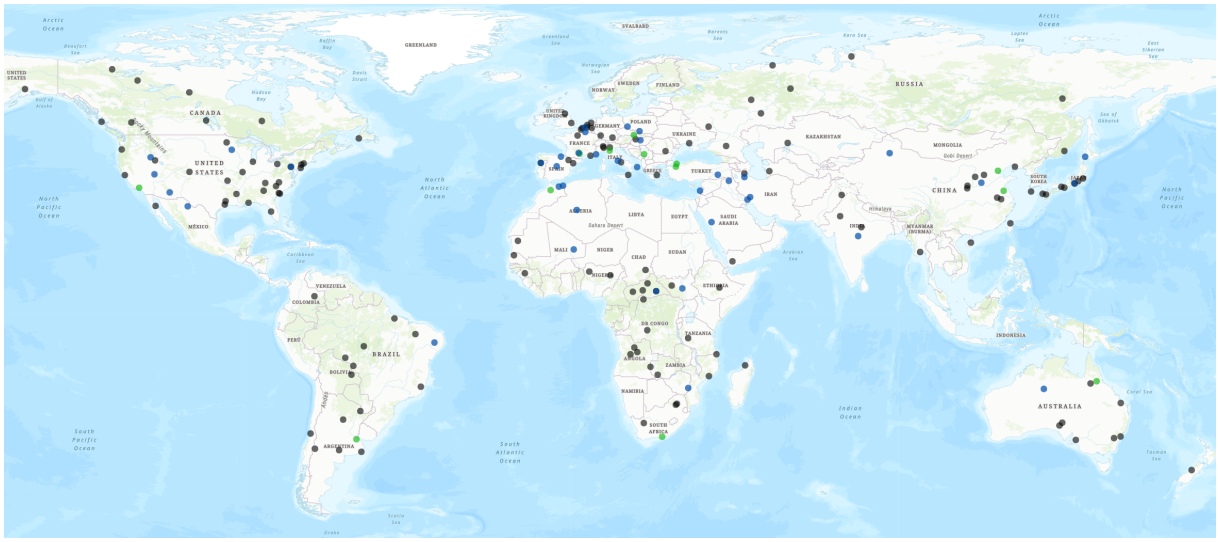

Fig. 1: Geospatial distribution of the ROI in SEN12MS-CR-TS. Train split ROI are colored blue, test split ROI are colored green. The ROI specific to SEN12MS-CR [6], compatible to our ROI and splits, are shown in gray color. Pins are plotted semi-transparently and with overlay so close-by dots can be discerned easier.

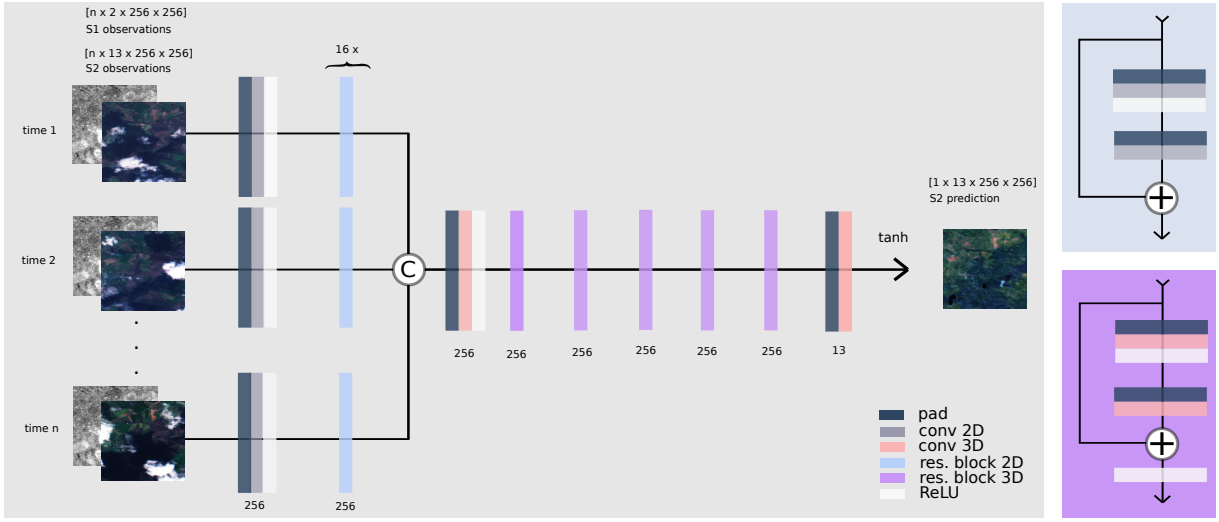

Fig. 2: A schema of the proposed multi-sensor time-series cloud removal architecture. The model is based on the Branched ResNet generator architecture of [5] and contains n Siamese ResNet branches [4] performing mono-temporal cloud removal on n input samples, separately. The resulting features are stacked in the temporal dimension and 3D convolutions are applied to integrate information across time. The output of the network is a single cloud-free image prediction.

3. DATA

To curate SEN12MS-CR-TS [1] we collect co-registered as well as paired paired S1 (ground range detected) and S2 (top-of-atmosphere reflectance) time series data of ESA's Copernicus mission. Each time series consists of $N = 30$ images over a given ROI at subsequent points in time, sampled throughout the full year of 2018. The data set consists of 53 geospatially separate ROI distributed across the entire globe, forming a subset of the ROI of SEN12MS-CR [6]. 40 ROI are defined as training and validation data, whereas 13 hold-out ROI are reserved for testing. The geospatial locations of all ROI included in our data set are illustrated in Fig. 1. All data is collected via Google Earth Engine and in the World Geodetic System 1984 (WGS84) coordinate reference system. For each ROI, the full-scene observations are collected within one sensor pass to minimize mosaicing effects. Subsequently, the images are sliced into non-overlapping patches of dimensions

[1] data available at: `https://patrickTUM.github.io/cloud_removal`

model	NRMSE (all)	NRMSE (cloudy)	NRMSE (clear)	PSNR	SSIM	SAM
least cloudy	0.079	0.082	**0.031**	—	0.815	0.213
mosaicing	0.062	0.064	0.036	**31.68**	0.811	0.250
ResNet	0.060	0.062	0.040	26.04	0.810	0.212
STGAN	0.057	0.059	0.050	25.42	0.818	0.219
ours	**0.051**	**0.052**	0.040	26.68	**0.836**	**0.186**

Table 1: Quantitative evaluation of the proposed model with baseline approaches in terms of NRSME, PSNR, SSIM and the SAM metric. Our model performs best in the majority of metrics, demonstrating that a deep neural network approach yields further improvements over simple solutions to the multi-modal time-series cloud removal problem.

% cloud coverage	NRMSE (all)	NRMSE (cloudy)	NRMSE (clear)	PSNR	SSIM	SAM
0-10 %	**0.041**	**0.046**	**0.041**	**28.59**	**0.870**	**0.143**
10-20 %	0.044	**0.046**	0.043	27.69	0.848	0.166
20-30 %	0.046	0.047	0.044	27.25	0.841	0.169
30-40 %	0.048	0.050	0.045	26.77	0.830	0.169
40-50 %	0.047	0.048	0.045	26.86	0.830	0.167
50-60 %	0.049	0.494	0.048	26.55	0.825	0.185
60-70 %	0.052	0.052	0.043	26.10	0.817	0.184
70-80 %	0.049	0.050	0.044	26.59	0.816	0.179
80-90 %	0.050	0.050	0.044	26.54	0.820	0.175
90-100 %	0.063	0.063	—	24.79	0.786	0.222

Table 2: Performance of our cloud removal model, depending on the extent of cloud coverage. All n=3 input samples are drawn to contain a specified percentage of cloudy pixels. The analysis highlights that the quality of the reconstructed images depends on the percentage of cloud coverage.

$[256 \times 256]\ px^2$. Finally, in the context of our experiments, all samples utilized by our model have their values clamped and rescaled. The two S1 bands are clamped to ranges [-25; 0], [-32.5; 0] and normalized to values in [0; 2]. The S2 bands are clipped to [0; 10000] and mapped into the range [0; 5], in line with the pre-processing pipeline proposed in [4].

4. METHODOLOGY

To highlight the benefits of the data set detailed in section 3, we design a multi-modal multi-temporal deep neural network architecture for cloud removal. The network is based on the Branched ResNet generator architecture of [5] and consists of n Siamese ResNet branches [4]. The mono-temporal ResNet branches are integrated into a multi-temporal representation by means of feature map stacking and 3D convolutions. The prediction of the network is a single time point cloud-free image prediction of the same spatial and spectral dimensions as any given S2 input sample. The architecture of the proposed model is conceptualized in Fig. 2.

5. EXPERIMENTS AND RESULTS

The model detailed in section 4 is trained and evaluated on SEN12MS-CR-TS. For this purpose, we initially pre-trained a ResNet as specified in section 4 and [4] on the training split of the mono-temporal SEN12MS-CR data set [6]. Subsequently, the multi-modal multi-temporal model is trained on the train split of SEN12MS-CR-TS. The model is trained on $n = 3$ input time points of co-registered S1 & S2 tuples and learns to predict a cloud-free reference S2 sample. Training is done by means of ADAM optimization and a combination of $\mathcal{L}_1$ and perceptual losses for 10 epochs at a batch size of 1. For the perceptual loss, we pre-trained and utilized a VGG-16 network as in [7]. Hyperparameters are chosen as in [5].

At test time, predictions are evaluated in terms of normalized root mean square error (NRMSE), peak signal-to-noise-ratio (PSNR), structural similarity (SSIM) [8] and the spectral angle mapper (SAM) [9] metric. Additionally, we differentiate the pixel-wise NRMSE into errors evaluated over all pixels (all), only over pixels cloudy in all input samples (cloudy) and solely over pixels clear in all input samples (clear). To test our proposed model, we evaluate it in two experiments: First, in a benchmark against concurrent approaches. Second, as a function of varying cloud coverage.

5.1. Experiment I: Benchmarking

To compare our proposed model with existing approaches of cloud removal, related techniques are benchmarked: "least cloudy" denotes the strategy of just outputting the least cloudy input image and "mosaicing" refers to a temporal integration of cloud-free pixels across all input time points. ResNet is the residual architecture detailed in section 4 and STGAN denotes the "Branched ResNet generator (IR)" baseline of [5]. The results in Table 1 show that the deep learning based cloud removal approaches tend to clearly outperform the simply heuristics on most metrics. The multi-temporal model provides benefits over the mono-temporal one, and the multi-sensor time-series network performs best.

5.2. Experiment II: Effects of cloud coverage

In a second experiment, we consider the proposed multi-sensor time-series cloud removal network and how its performance varies as a function of cloud coverage. For this, all input time samples exhibit a cloud coverage within the specified range. Manipulating cloud coverage this way, Table 2 shows that the model performs best at a minimum percentage of cloud coverage. While the decrease in performance is not monotonous, there is a clear trend of cloud coverage negatively affecting image reconstruction.

6. CONCLUSION

This study introduced SEN12MS-CR-TS, a multi-sensor time-series data set for training and testing methodology for the purpose of cloud removal in optical satellite images. Our data set is unique in providing a benchmark for multi-sensor time-series satellite image reconstruction, sampled from a global distribution of ROI. To demonstrate the

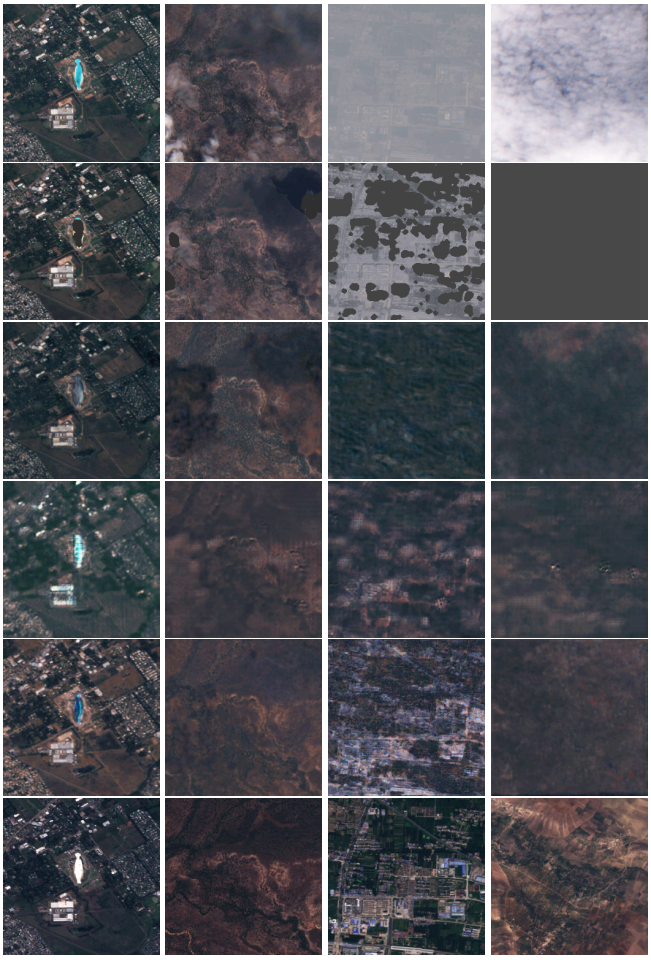

Fig. 3: Quantitative analysis of exemplary predictions of all methods reported in Table 1 and cloud-free reference samples Columns: Four test split samples. The illustrated samples represent cases that are cloud-free, partly-cloudy, cloud-covered with no visibility except for a single time point and cloud-covered with no visibility in any time point. Rows: Predictions of least cloudy, mosaicing, ResNet, STGAN, ours, and the cloud-free target sample. The outcomes show that deep networks outperform the simple heuristics and that multi-sensor time-series data may benefit image reconstruction.

benefits of SEN12MS-CR-TS, we designed a multi-modal multi-temporal deep neural network architecture that was trained and subsequently evaluated on the novel data set. The model outperformed simple baselines based on heuristics as well as preceding networks utilizing only mono-temporal or single-sensor information, respectively. In a final experiment, we probed the effects that the extent of cloud coverage has on the quality of the reconstructed image.

We are positive that our curated data set and the proposed model will advance further research in satellite image reconstruction and benefit the expanding application of subsequent remote sensing methodology.

7. REFERENCES

[1] 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*, vol. 51, no. 7, pp. 3826–3852, Jul 2013.

[2] 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*, vol. 43, pp. 243–249, 2021.

[3] Patrick Ebel, Michael Schmitt, and Xiao Xiang Zhu, "Cloud removal in unpaired Sentinel-2 imagery using cycle-consistent GAN and SAR-optical data fusion," *IGARSS 2020 IEEE International Geoscience and Remote Sensing Symposium, 2020*, 2020.

[4] 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*, 2020.

[5] Vishnu Sarukkai, Anirudh Jain, Burak Uzkent, and Stefano Ermon, "Cloud removal from satellite images using spatiotemporal generator networks," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 1796–1805.

[6] 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*, 2020.

[7] Patrick Ebel, Michael Schmitt, and Xiao Xiang Zhu, "Internal learning for sequence-to-sequence cloud removal via synthetic aperture radar prior information," in *IGARSS 2021-2021 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2021, p. In press.

[8] 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*, vol. 13, no. 4, pp. 600–612, 2004.

[9] Fred A Kruse, AB Lefkoff, JW Boardman, KB Heidebrecht, AT Shapiro, PJ Barloon, and AFH Goetz, "The spectral image processing system (sips)-interactive visualization and analysis of imaging spectrometer data," in *AIP Conference Proceedings*. American Institute of Physics, 1993, vol. 283, pp. 192–201.