IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10282605

ONBOARD CLOUD DETECTION AND ATMOSPHERIC CORRECTION WITH DEEP LEARNING EMULATORS

Gonzalo Mateo-García [1,2,], Cesar Aybar [1,2,*], Giacomo Acciarini [1,3,*], Vít Růžička [1,4], Gabriele Meoni [5], Nicolas Longépé [5], Luis Gómez-Chova [2]*

[1] Trillium Technologies, London, United Kingdom
[2] University of Valencia, Valencia, Spain
[3] University of Surrey, Guildford, United Kingdom
[4] University of Oxford, Oxford, United Kingdom
[5] European Space Agency Φ-Lab, Frascati, Italy

ABSTRACT

This paper introduces DTACSNet, a Convolutional Neural Network (CNN) model specifically developed for efficient onboard atmospheric correction and cloud detection in optical Earth observation satellites. The model is developed with Sentinel-2 data. Through a comparative analysis with the operational Sen2Cor processor, DTACSNet demonstrates a significantly better performance in cloud scene classification (F2 score of 0.89 for DTACSNet compared to 0.51 for Sen2Cor v2.8) and a surface reflectance estimation with average absolute error below 2% in reflectance units. Moreover, we tested DTACSNet on hardware-constrained systems similar to recent deployed missions and show that DTACSNet is 11 times faster than Sen2Cor with a significantly lower memory consumption footprint. These preliminary results highlight the potential of DTACSNet to provide enhanced efficiency, autonomy, and responsiveness in onboard data processing for Earth observation satellite missions.

***Index Terms*—** Sentinel-2, Sen2Cor, onboard processing, atmospheric correction, cloud detection, deep learning, CNN

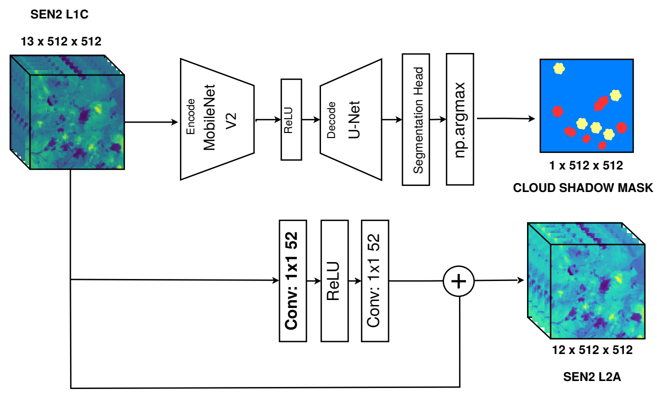

Fig. 1. DTACSNet architecture: the top branch shows the cloud scene classification (SC) network and the bottom one the atmospheric correction (AC) network. Architectures were chosen based on best trade-off between accuracy and inference speed.

1. INTRODUCTION

Onboard data processing in Earth observation satellites offers numerous benefits such as discarding cloud-contaminated imagery in real-time [1] or selectively identifying and prioritizing relevant targets to process, such as algae blooms, oil spills, floating debris, crop damage, fires, flood events, or methane leaks [2]. Furthermore, onboard processing allows for data optimization, reducing the size of products that need to be downlinked to Earth [3] and enabling the prioritization of specific tiles for download [4]. These capabilities could enhance the overall efficiency and effectiveness of Earth observation missions, facilitating timely decision-making and resource allocation.

Nevertheless, processing data onboard requires the integration and calibration of raw data before it is used by most applications. While this is normal practice on the ground, preparation of data to create Analysis Ready Data (ARD) products is not straightforward since the processes running on the ground are less constrained than those running onboard. For instance, onboard hardware has lower memory and processing capability, and the access to ancillary data of onboard processes is very restrictive (if any).

One ubiquitous process to create Analysis Ready Data (ARD) products for optical sensors is atmospheric correction. Atmospheric correction is a sophisticated procedure

This work has been funded by ESA Cognitive Cloud Computing in Space initiative (project number D-TACS I-2022-00380). G.M.-G. and L.G.-C. have been supported by the Spanish Ministry of Science and Innovation (project PID2019-109026RB-I00 funded by MCIN/AEI/10.13039/501100011033) and the European Social Fund.
* Authors with equal contribution.

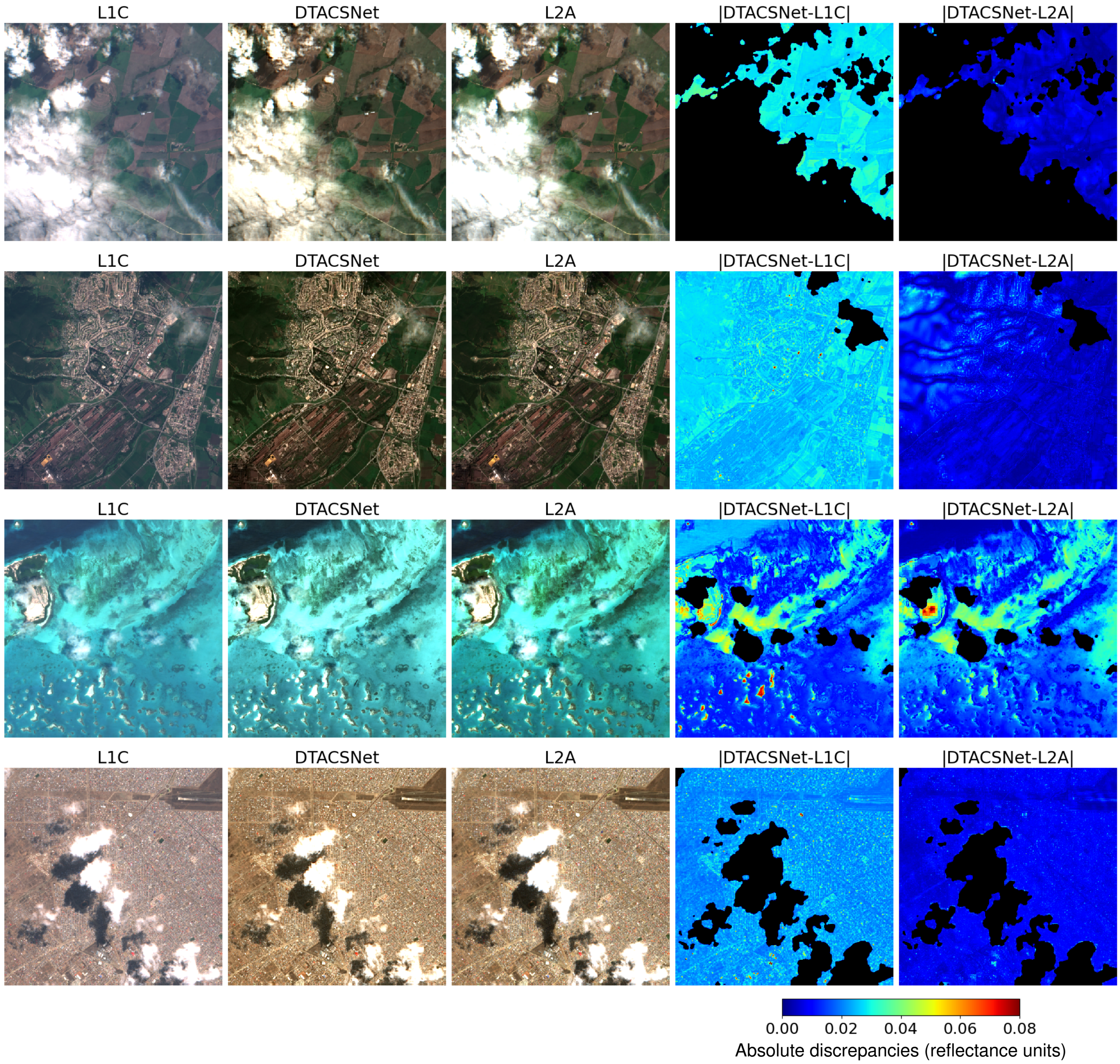

Fig. 2. Samples from the test set of CloudSEN12. First column: Sentinel-2 TOA image (RGB channels). Second column: image corrected with DTACSNet AC network. Third column: Sentinel-2 image corrected with Sen2Cor. Fourth column: absolute differences between L1C TOA reflectance and DTACSNet output across all bands. Fifth column: absolute differences between Sen2Cor and DTACSNet outputs across all bands. In the last two columns, clouds and shadows are masked out (black pixels) using the output of the DTACSNet scene classification network.

involving two core steps. a) Cloud scene classification (SC): identifying cloud contaminated pixels where the the signal from the surface cannot be recovered (thick clouds). b) Atmospheric correction (AC): removing the perturbations introduced by the atmosphere in the observed at-sensor radiance, i.e. conversion from Top-of-Atmosphere (TOA) reflectance to Bottom-of-Atmosphere (BOA) surface reflectance. These perturbations are caused by absorption and scattering of atmospheric constituents (thin clouds, aerosols, water vapor, ozone ...) and by occlusions (cloud and terrain shadows). In this work, we have developed an atmospheric correction processor that can be run onboard with tight requirements of memory and processing capabilities. For developing this processor, that we call DTACSNet, we have taken Sentinel-2 mission as a reference, taking advantage of its publicly available TOA and BOA data (level 1C and level 2A products, respectively) [5].

We compare DTACSNet with the operational Sentinel-2 atmospheric correction processor: the Sen2Cor [5] software, which is publicly available at the ESA web page[1]. Sen2Cor produces accurate surface reflectance when compared with ground truth data according to the recent ESA-NASA Atmospheric Correction Intercomparison Exercise (ACIX) [6, 7]. However, its cloud detection is significantly worse than other approaches [8, 9]. When we look at critical variables for onboard processing (memory consumption and processing time), we have found that the current implementation of Sen2Cor (v2.11) is too demanding to be run onboard.

2. METHODOLOGY

The proposed DTACSNet model is based on convolutional neural network (CNN) architectures that perform both the semantic segmentation of Sentinel-2 Level-1C TOA images into cloud classes and its atmospheric correction (Fig. 1). DTACSNet is trained with data from the recently published CloudSEN12 dataset [9], and it is intended to work on remote embedded systems with limited hardware resources. The cloud detection branch is trained using the high-quality manually generated cloud mask included in CloudSEN12 as ground truth reference. On the other hand, the atmospheric correction branch is trained using the Sentinel-2 level 2A surface reflectance product generated by the Sen2Cor processor as reference. Therefore, the atmospheric correction branch should be considered as a deep learning emulator of the Sen2Cor processor [10], since it is not possible to have access to the actual surface reflectance values.

3. RESULTS

The developed DTACSNet model is validated in a large dataset of independent geographic locations from CloudSEN12 dataset (not used for training the models) and over a year-long time series over the same locations used in the ACIX experiment [6]. On the one hand, we found that our cloud scene classification model is significantly more accurate than Sen2Cor (F2 score of 0.89 for DTACSNet and 0.51 for Sen2Cor v2.8). On the other hand, the atmospheric correction model has reflectance discrepancies in the same order of magnitude as the errors of Sen2Cor in the ACIX exercise (around 1.8%). Figure 2 shows some representative samples over the test dataset. Additionally, we show that the atmospheric correction models work also with less spectral bands, mimicking other existing and prospective multispectral medium-size satellite missions. Finally, we have compared DTACSNet and Sen2Cor on different hardware configurations: an standard workstation with 30 Gb of RAM, 8 CPUs and a NVIDIA T4 GPU and a UNIBAP SpaceCloud flatsat with a low power CPU processor, 1.74Gb of RAM an a Myriad-X vision processing unit. The flasat configuration is very similar to the deployed at D-Orbit Wild Ride mission [11]. In the workstation, we found improvements in running time between $\times 4$ to $\times 11$. In the flatsat, Sen2Cor was not able to run at 10m resolution due to the low RAM memory available. At 20m resolution the running time of Sen2Cor was over an hour whereas DTACSNet took less than 7 minutes (running on a full Sentinel-2 product of 5490x5490 pixels). These results show that DTACSNet can be used operationally in an onboard setting.

[1]Sen2Cor processor for Sentinel-2 level 2A product generation: https://step.esa.int/main/snap-supported-plugins/sen2cor/

4. CONCLUSIONS

In this work, we presented some preliminary results of DTACSNet: an efficient deep learning model for atmospheric correction and cloud detection in Sentinel-2 imagery. We show that the model is able to produce accurate cloud detection and surface reflectance estimation in computationally constrained scenarios. In order to foster research in this field, we release some of our pre-trained models for research purposes at https://github.com/spaceml-org/DTACSNet.

5. REFERENCES

[1] Gianluca Giuffrida, Luca Fanucci, Gabriele Meoni, Matej Batič, Léonie Buckley, Aubrey Dunne, Chris Van Dijk, Marco Esposito, John Hefele, Nathan Vercruyssen, Gianluca Furano, Massimiliano Pastena, and Josef Aschbacher, "The ϕ-Sat-1 mission: the first onboard deep neural network demonstrator for satellite earth observation," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–1, 2021, Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

[2] Vít Růžička, Gonzalo Mateo-Garcia, Luis Gómez-Chova, Anna Vaughan, Luis Guanter, and Andrew Markham, "STARCOP: Semantic Segmentation of Methane Plumes with Hyperspectral Machine Learning Models," *Submitted to Scientific Reports*, May 2023.

[3] Gonzalo Mateo-Garcia, Joshua Veitch-Michaelis, Lewis Smith, Silviu Vlad Oprea, Guy Schumann, Yarin Gal, Atılım Güneş Baydin, and Dietmar Backes, "Towards global flood mapping onboard low cost satellites with machine learning," *Scientific Reports*, vol. 11, no. 1, pp. 7249, Mar. 2021, Number: 1 Publisher: Nature Publishing Group.

[4] Vít Růžička, Anna Vaughan, Daniele De Martini, James Fulton, Valentina Salvatelli, Chris Bridges, Gonzalo Mateo-Garcia, and Valentina Zantedeschi, "RaVÆn:

unsupervised change detection of extreme events using ML on-board satellites," *Scientific Reports*, vol. 12, no. 1, pp. 16939, Oct. 2022, Number: 1 Publisher: Nature Publishing Group.

[5] Jérôme Louis, Vincent Debaecker, Bringfried Pflug, Magdalena Main-Knorn, Jakub Bieniarz, Uwe Mueller-Wilm, Enrico Cadau, and Ferran Gascon, "Sentinel-2 Sen2Cor: L2A processor for users," in *Proceedings of ESA Living Planet Symposium 2016*. Spacebooks Online, 2016, pp. 1–8.

[6] Georgia Doxani, Eric Vermote, Jean-Claude Roger, Ferran Gascon, Stefan Adriaensen, David Frantz, Olivier Hagolle, André Hollstein, Grit Kirches, Fuqin Li, Jérôme Louis, Antoine Mangin, Nima Pahlevan, Bringfried Pflug, and Quinten Vanhellemont, "Atmospheric Correction Inter-Comparison Exercise," *Remote Sensing*, vol. 10, no. 2, pp. 352, Feb. 2018, Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[7] Nima Pahlevan, Antoine Mangin, Sundarabalan V. Balasubramanian, Brandon Smith, Krista Alikas, Kohei Arai, Claudio Barbosa, Simon Bélanger, Caren Binding, Mariano Bresciani, Claudia Giardino, Daniela Gurlin, Yongzhen Fan, Tristan Harmel, Peter Hunter, Joji Ishikaza, Susanne Kratzer, Moritz K. Lehmann, Martin Ligi, Ronghua Ma, François-Régis Martin-Lauzer, Leif Olmanson, Natascha Oppelt, Yanqun Pan, Steef Peters, Nathalie Reynaud, Lino A. Sander de Carvalho, Stefan Simis, Evangelos Spyrakos, François Steinmetz, Kerstin Stelzer, Sindy Sterckx, Thierry Tormos, Andrew Tyler, Quinten Vanhellemont, and Mark Warren, "ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters," *Remote Sensing of Environment*, vol. 258, pp. 112366, June 2021.

[8] Sergii Skakun, Jan Wevers, Carsten Brockmann, Georgia Doxani, Matej Aleksandrov, Matej Batič, David Frantz, Ferran Gascon, Luis Gómez-Chova, Olivier Hagolle, Dan López-Puigdollers, Jérôme Louis, Matic Lubej, Gonzalo Mateo-García, Julien Osman, Devis Peressutti, Bringfried Pflug, Jernej Puc, Rudolf Richter, Jean-Claude Roger, Pat Scaramuzza, Eric Vermote, Nejc Vesel, Anže Zupanc, and Lojze Žust, "Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2," *Remote Sensing of Environment*, vol. 274, pp. 112990, June 2022.

[9] Cesar Aybar, Luis Ysuhuaylas, Jhomira Loja, Karen Gonzales, Fernando Herrera, Lesly Bautista, Roy Yali, Angie Flores, Lissette Diaz, Nicole Cuenca, Wendy Espinoza, Fernando Prudencio, Valeria Llactayo, David Montero, Martin Sudmanns, Dirk Tiede, Gonzalo Mateo-García, and Luis Gómez-Chova, "CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2," *Scientific Data*, vol. 9, no. 1, pp. 782, Dec. 2022, Number: 1 Publisher: Nature Publishing Group.

[10] L. Gómez-Chova, G. Mateo-García, V. Laparra, E. Portales-Julià, and F. Niro, "Synergistic exploitation of PV-CC in combination with Sentinel-2 towards a machine learning based atmospheric correction," in *ESA Living Planet Symposium*, May 2022.

[11] Gonzalo Mateo-García, Josh Veitch-Michaelis, Cormac Purcell, Nicolas Longepe, Simon Reid, Alice Anlind, Fredrik Bruhn, James Parr, and Pierre Philippe Mathieu, "In-orbit demonstration of a re-trainable Machine Learning Payload for processing optical imagery," *Scientific Reports (Accepted)*, Aug. 2022.