This is a preprint of the paper published in IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, DOI: 10.1109/IGARSS53475.2024.10642339

A DEEP LEARNING APPROACH FOR IMAGERY MASKING OF SPECTRAL SENSORS

*Efrain Padilla-Zepeda [1,2] * †, Kevin Alonso [3], Raquel De Los Reyes [1], Deni Torres-Roman [2], Avi Putri Pertiwi [4]*

[1] German Aerospace Center (DLR), Earth Observation Center,
Photogrammetry and Image Analysis Department, Oberpfaffenhofen, Germany
[2] Center for Research and Advanced Studies of the National Polytechnic Institute (Cinvestav),
Telecommunications Group, Zapopan, Mexico
[3] Starion Group c/o European Space Agency (ESA), Frascati, Italy
[4] German Aerospace Center (DLR), Earth Observation Center,
Photogrammetry and Image Analysis Department, Berlin, Germany

ABSTRACT

Currently, some of the implemented atmospheric correction processors for remote sensing spectral sensors, use masking algorithms based on thresholding of spectral indices with sensor Top-Of-Atmosphere (TOA) reflectance. This concept allows the use of a limited amount of spectral bands, which is optimal for multi-spectral sensors ($\sim$ 10-20 bands), but for the case of the high spectral dimensionality of hyper-spectral sensors, spectral thresholding underutilizes the number of available bands. Given this limitation, we propose a masking algorithm which performs spatial and spectral feature extraction based on a 2D convolutional neural network, fitting the model with the available classification maps from the Python-based Atmospheric COrrection (PACO) processor. The training samples are selected based on their uncertainty to belong to a given class, The validation is performed using two independent human expert labelled datasets. The resulting classification maps show an improvement from the original ones of PACO.

Index Terms— classification, masking algorithm, spectral imagery, atmospheric correction, pixel filtering

1. INTRODUCTION

Masking algorithms generate classification maps of multi/hyper-spectral remotely sensed Earth's surface scenes. This information is produced during the processing chain to be internally used during the correction of the atmospheric effects, and afterwards delivered as part of the resulting Level-2A (L2A) data products. The atmospheric correction applied to each pixel depends on the class they belong to. For example, correction is not possible for completely opaque pixels by clouds. Atmospheric effects like semi-transparent clouds, cloud shadows, or haze, might be considered internally by the algorithms in different ways. Therefore, the atmospheric correction performance is directly related to the performance of the masking algorithms.

The Python-based Atmospheric COrrection (PACO) software [1], is an atmospheric correction processor developed at the German Aerospace Center (DLR) based on the well-established ATCOR processor [2]. PACO is able to process the Top-Of-Atmosphere (TOA) data from the multi-spectral sensors Sentinel-2 [3] and Landsat-8/9 [4][5][6], as well as the hyper-spectral sensors; DESIS [7] and EnMAP [8], to retrieve the Bottom-Of-Atmosphere (BOA) spectra.

In this work, we propose a Deep Learning (DL) approach for imagery masking, training a model having as an input the masking products of PACO. As a first learning stage, we show results only with data from the multi-spectral sensor Sentinel-2. However, the development plan of this work aims to implement this method for hyper-spectral sensors in the near future. The content of this paper presents the concept in section 2. A description of the data used for training and testing in section 3. In section 4, the DL model used and the pixel selection strategy for training. Subsequently, the results are presented in section 5, validating the classification maps with two different independent datasets. Finally, in section 6, the learned conclusions and future directions of the work are discussed.

2. CONCEPT

The pre-classification, based on empirically defined thresholds for spectral indices, is limited by the number of thresholds a human expert can define. In the case of hyperspectral imagery, it is rather challenging to consider the complete spectral information available with this approach. DL approaches can extract the spatial and spectral features that characterize each class and generate a classification map from it, but for supervised learning, a reference is necessary to train

*Supported by Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT) grant number 789304.

†Research stay supported the German Federal Ministry for Research and Technology (BMBF), DAAD-fellowship number 57645446.

the model. In order to fully exploit the performance of the DL models, the reference dataset used for the training must be more precise than the model sensitivity. This is not so easily achievable, and reference dataset masked by human experts are often quite manpower demanding. To overcome this limitation we use the thresholding based PACO initial image segmentation results along their confusion to train our DL models. Thus, the training datasets are constructed using only the lower uncertainty PACO classification maps. With this approach, the DL model can fit the easy cases of classification, often easily masked using multi-spectral bands, but finding better boundaries for difficult pixels that might barely skip the thresholding criteria. Using an analogy, this is equivalent to finding the edge thresholds automatically and taking into consideration complex spatial and spectral features, utilizing all the spectral information available.

3. DATA DESCRIPTION

The datasets used for this analysis are L1C scenes of Sentinel-2. This mission consists of a constellation of two sun-synchronous satellites, each vehicle carrying a multi-spectral instrument of 13 spectral bands with different spatial resolutions at the ground: four bands at 10m, six at 20m, and three at 60m. The sensor captures the visible, Near Infra-Red (NIR), and Short Wave Infra-Red (SWIR) segments of the electromagnetic spectrum. The bands of 10m and 60m were re-scaled to 20m and represented in TOA reflectances, obtained from the PACO software [1]. We use the tensor notation from [9] to represent spectral data as a 3rd-order tensor $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, where I_1 and I_2 correspond to the spatial domain and I_3 to the spectral domain. In the case of spectral data from Sentinel-2, the tensor dimensions are $\mathcal{S} \in \mathbb{R}^{5490 \times 5490 \times 13}$. Additionally, we include the Digital Elevation Model (DEM) [10] obtained from the Copernicus mission of the European Space Agency (ESA), represented as a matrix $\mathbf{D} \in \mathbb{Z}^{I_1 \times I_2}$ containing the height in Meters-Above-Sea-level (MASL) of each pixel, stacked as a new layer of the spectral cube, resulting in a new tensor $\mathcal{X} \in \mathbb{R}^{5490 \times 5490 \times 14}$. Given that this masking algorithm can classify every possible location of the earth's surface, the normalization has to be fixed, dividing the spectral data and the DEM by pre-defined values, which are empirically set to correspond to approximate theoretical maximums. To establish an upper limit for Top of Atmosphere (TOA) reflectance, we consider the scenario in which a pixel exhibits a reflectance of more than 1, indicating reflections from other pixels or light sources, empirically setted observing maximums of 1.5. For the normalization of the DEM, we take into account the highest point on Earth above the global mean sea level: Mount Everest, with an elevation of 8,850 meters. Conversely, negative MASL heights are adjusted to a minimum of 0 meters.

3.1. Training Dataset

We have limited out the dataset to 40 scenes to achieve faster training times during the study phase. Ideally the training set should contain a representation of each of the labels we want to mask, which should represent the Earth's surface and the atmospheric phenomena present in the masking labels. For the latest criteria we have used Sentinel-2 scenes overpassing AERONET validation sites [11], which are set to characterize the atmospheric conditions all around the globe, including also a diversity in ground surface.

The training reference labels are derived from PACO pre-classification, formatted in a binary vector of 21 classes, where each pixel is simultaneously classified into one or more classes. These vectors form a 3rd-order tensor $\mathcal{Y} \in \{0,1\}^{5490 \times 5490 \times 21}$. We take advantage of this format since it allows us to map for any set of classes targeted for classification. In this manner, we are able to fit the requirements of a wide range of applications. The mapping rules should be defined by a human expert, based on the information which defines every target class. The source dataset classes are described in [1].

3.2. Testing Dataset

We test the model with two independent datasets labelled by human experts with different criteria. The first testing dataset (TD1) includes 13 scenes, labelled for a reduced version of the set of classes for Sen2Cor [12]. The annotated pixels are uniformly distributed regions of the scenes. The class count of each class is shown in Table 1.

Table 1. Class count and color code for TD1.

Color	Class	Count	% of total
	Clouds_shadows	95063	13.50 %
	Clear	250647	35.59 %
	Water	132114	18.76 %
	Cloud_high_probability	109586	15.56 %
	Thin_cirrus	85792	12.18 %
	Snow	30965	4.40 %

The second testing dataset (TD2) was already tested in [13], comparing the performance of Fmask [14], ATCOR, and Sen2Cor. This dataset includes 20 scenes, and the annotated pixels were selected taking into consideration challenging classification cases. More information about this dataset in [13]. The class count of each class is shown in Table 2.

Given the bias difference of the datasets, we are able to evaluate different aspects of the resulting classification maps.

Table 2. Class count and color code for TD2.

Color	Class	Count	% of total
	Clear	872209	83.86 %
	Semi-transparent	19183	1.84 %
	Clouds	40641	3.91 %
	Cloud_shadows	70373	6.77 %
	Water	26310	2.53 %
	Snow	8781	0.84 %
	Topographic_shadows	2584	0.25 %

4. METHOD

The first step is to map the multi-label PACO masks cube in order to obtain a multi-class map. This process depends on the set of classes from the application. In the case of this work, the target classes are defined by the testing sets. As an example, TD1 was labelled according to the defined classes of Sen2Cor, but the map could match any other class definitions if there is a direct relation or combination of classes from the PACO masks cube.

4.1. Pixel Selection

The multi-label classification of PACO gives valuable information when an atmospheric effect is detected. For instance, a cloud shadow could be projected over land, water, or snow. Multi-class classification usually only informs where a cloud shadow is, losing the information of the surface where the shadow is projected. Given that PACO masking products classify every pixel with one or more classes, sometimes there is confusion, given that a pixel could be classified as two or more contradictory classes, e.g., cloud and snow, land and water, etc. We consider these cases as high-uncertainty pixels. To prevent the inclusion of misclassified pixels in the training set, we exclude those exhibiting a higher number of simultaneously active masks. To guarantee adequate representation for each class, we establish a maximum percentage threshold for the number of pixels to be discarded. We test the model performance by training the model filtering pixels with high uncertainty, and for comparison, by training without discarding any pixel.

4.2. Deep Learning Model

PACO's masking algorithms identify patterns in the spectral domain based on the spectral signature of each material on the Earth's surface. However, information in the spatial domain is also present due to the high spatial correlation in remote sensing imagery [15]. Therefore, the DL model examined in this paper is a 2DCNN [16], classifying pixel-by-pixel using information from neighboring pixels. In contrast to other popular DL models, such as those based on U-NeT [17], which require a fixed-size complete classification map for training, the 2DCNN is trained with only a few labelled pixels and images of any size, enabling a broad range of applications. While there are models like 3DCNN [18] that simultaneously extract spatial and spectral features, increasing computational complexity, we defer testing them for the future as the objective in this paper is to evaluate the general idea.

The architecture of the 2DCNN is based on the model presented in [19] for hyper-spectral imagery, but we have reduced it for multi-spectral imagery. A description of the tested architecture in Table 3. The input size of the model is a patch of 11×11 pixels, with the pixel of interest at the center, allowing to get information from neighbor pixels, even if they are not in the training set.

Table 3. 2DCNN architecture.

Main Layer	Activation function	Down sampling
Linear input (11 x 11 x 14)	-	-
Conv2D($50 \times 5 \times 5$)	ReLU	-
Conv2D($100 \times 5 \times 5$)	ReLU	POOL(2×2)
Dense(100)	ReLU	-
Dense(n_class)	Soft-max	-

4.3. Training and Testing

Given the large number of samples available for training and high spatial correlation, training with 10% is enough to get representation of all the pixels present in the image. The samples are randomly selected with a stratified strategy, keeping the same proportion of samples for each class. The remaining 90% is used in each epoch for validation, saving the model with higher accuracy. The model is trained using the cross-entropy loss function through the Adam optimizer for 3 epochs. The training set size and the number of epochs were empirically searched analyzing the loss values of each epoch. The testing results are reported using three wide-used classification metrics in the state-of-the-art [20, 21]: Overall Accuracy (OA)[0:1], Cohen's Kappa correlation coefficient (Kappa) [-1;1], and Matthew's Correlation Coefficient (MCC)[-1;1], where Kappa and MCC penalize the unbalanced dataset cases.

5. RESULTS

In Tables 4 and 5, we show the classification metrics for the tests with TD1 and TD2, respectively. For both testing datasets, we compare the results obtained by PACO masking and the predictions from our model, the first without discarding pixels (2DCNN) and the second filtering pixels with high uncertainty (2DCNN-PF).

Table 4. Classification metrics results with TD1.

Metric	PACO	2DCNN	2DCNN-PF
OA	0.7604	0.7673	0.7711
Kappa	0.6849	0.6929	0.6979
MCC	0.6906	0.7004	0.7055

Table 5. Classification metrics results with TD2.

Metric	PACO	2DCNN	2DCNN-PF
OA	0.5816	0.7294	0.7290
Kappa	0.2461	0.3796	0.3547
MCC	0.3111	0.4258	0.3882

The best obtained results testing with TD1 shows that the proposed method of 2DCNN-PF is able to fit the source classification maps from PACO with a slight improvement, from 0.6906 to 0.7055 of MCC, or 1.07% more correctly classified pixels in terms of OA. On the other hand, TD2's best results shows a significant improvement, from 0.3111 to 0.4258 of MCC for 2DCNN, or 14.78% more correctly classified pixels in terms of OA. Given that TD2 is biased towards outlier pixels [13], these are challenging edge cases for thresholding masking algorithms, in the other hand, the complex feature extraction of the 2DCNN is better able to define these boundaries; therefore, the performance enhancement is more noticeable than with TD1. Another aspect to look are the generated classification maps. In Figures 1, 2, and 3 we show three examples comparing PACO's and 2DCNN's classification maps with the correspondent quick-look for reference.

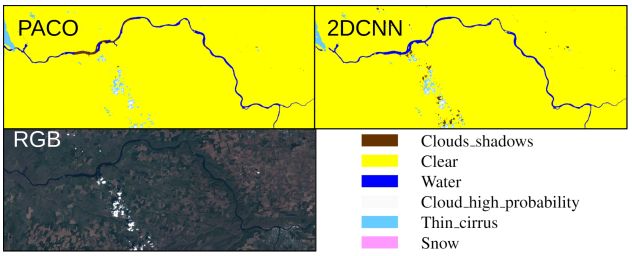

Fig. 1. TD1-Rimrock, USA. T11TMM, 2022-04-25. 2DCNN is able to correctly classify the majority of river's water pixels.

6. CONCLUSIONS

In this work, we provide a deep learning approach for improving masking algorithms in the context of the atmospheric corrector PACO software. Two independently labelled datasets (TD1 and TD2) were tested, showing improvement in both cases. TD2, which was labelled to target challenging classification cases, exhibited the greatest improvement, with 14.78% more correctly classified pixels. The generalization of the model is limited by the representation of each possible sub-class in the training set. Therefore, future work includes creating a large training dataset composed by low-uncertainty pixels, incorporating more testing datasets to evaluate different characteristics of the model predictions, and extending this method for hyperspectral sensors.

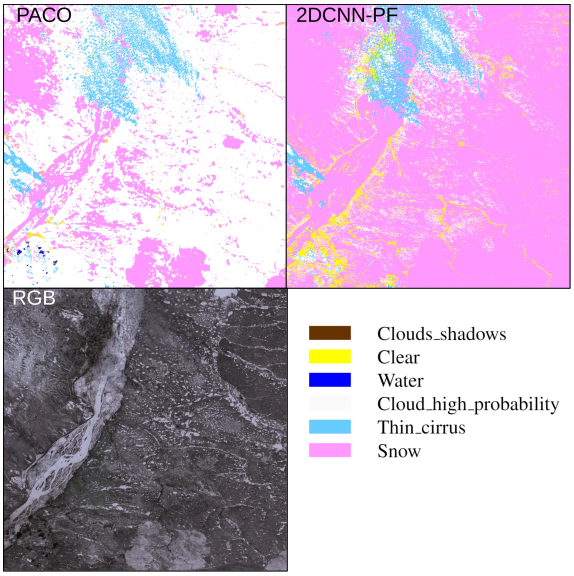

Fig. 2. TD1-Yakutsk, Russia. T52VEP 2023-05-01. The majority of the ice/snow pixels were misclassified by PACO as cloud high probability, improved it by 2DCNN-PF classifying as snow.

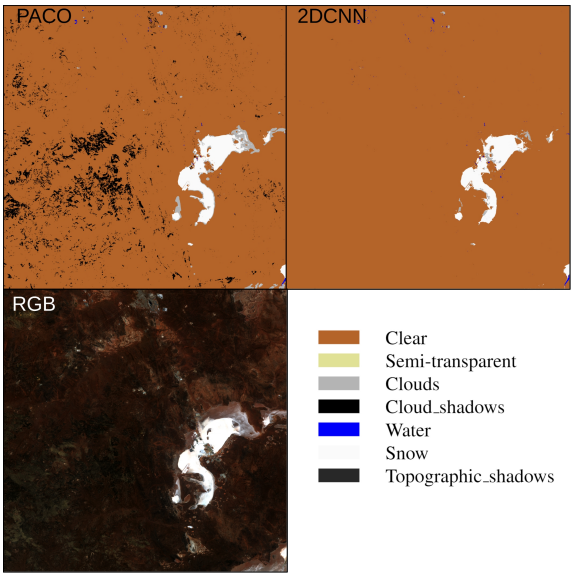

Fig. 3. TD2-Lake Lefroy, Australia. T51JUF 2018-08-19. Lake Lefroy is dry, misclassified as snow by both predictions. PACO predict cloud shadows and clouds not present in the scene, 2DCNN performs better in this case.

7. REFERENCES

[1] Raquel de Los Reyes, Maximilian Langheinrich, Peter Schwind, Rudolf Richter, Bringfried Pflug, Martin Bachmann, Rupert Müller, Emiliano Carmona, Viktoria Zekoll, and Peter Reinartz, "PACO: Python-Based Atmospheric COrrection," *Sensors (Basel, Switzerland)*, vol. 20, no. 5, pp. 1428, 3 2020.

[2] Rudolf Richter, "Correction of satellite imagery over mountainous terrain," *Applied Optics, Vol. 37, Issue 18, pp. 4004-4015*, vol. 37, no. 18, pp. 4004–4015, 6 1998.

[3] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, 5 2012.

[4] Julia Barsi, Kenton Lee, Geir Kvaran, Brian Markham, and Jeffrey Pedelty, "The spectral response of the landsat-8 operational land imager," *Remote Sensing*, vol. 6, pp. 10232–10251, 10 2014.

[5] Eric Vermote, Chris Justice, Martin Claverie, and Belen Franch, "Preliminary analysis of the performance of the landsat 8/oli land surface reflectance product," *Remote Sensing of Environment*, vol. 185, pp. 46 – 56, 2016, Landsat 8 Science Results.

[6] Jeffrey G. Masek, Michael A. Wulder, Brian Markham, Joel McCorkel, Christopher J. Crawford, James Storey, and Del T. Jenstrom, "Landsat 9: Empowering open science and applications through continuity," *Remote Sensing of Environment*, vol. 248, pp. 111968, 2020.

[7] Kevin Alonso, Martin Bachmann, Kara Burch, Emiliano Carmona, Daniele Cerra, Raquel de los Reyes, Daniele Dietrich, Uta Heiden, Andreas Hölderlin, Jack Ickes, Uwe Knodt, David Krutz, Heath Lester, Rupert Müller, Mary Pagnutti, Peter Reinartz, Rudolf Richter, Robert Ryan, Ilse Sebastian, and Mirco Tegler, "Data Products, Quality and Validation of the DLR Earth Sensing Imaging Spectrometer (DESIS)," *Sensors*, vol. 19, no. 20, 2019.

[8] Luis Guanter, Hermann Kaufmann, Karl Segl, Saskia Foerster, Christian Rogass, Sabine Chabrillat, Theres Kuester, André Hollstein, Godela Rossner, Christian Chlebek, Christoph Straif, Sebastian Fischer, Stefanie Schrader, Tobias Storch, Uta Heiden, Andreas Mueller, Martin Bachmann, Helmut Mühle, Rupert Müller, Martin Habermeyer, Andreas Ohndorf, Joachim Hill, Henning Buddenbaum, Patrick Hostert, Sebastian der Linden, Pedro J Leitão, Andreas Rabe, Roland Doerffer, Hajo Krasemann, Hongyan Xi, Wolfram Mauser, Tobias Hank, Matthias Locherer, Michael Rast, Karl Staenz, and Bernhard Sang, "The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation," *Remote Sensing*, vol. 7, no. 7, pp. 8830–8857, 2015.

[9] Tamara G. Kolda and Brett W. Bader, "Tensor decompositions and applications," 2009.

[10] Luca Cenci, Marco Galli, Carla Santella, Valentina Boccia, and Clément Albinet, "Analyzing the impact of the different instances of the Copernicus DEM dataset on the orthorectification of VHR optical data," in *Proceedings of 2022 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022, pp. 6001–6004.

[11] B.N. Holben, T.F. Eck, I. Slutsker, D. Tanré, J.P. Buis, A. Setzer, E. Vermote, J.A. Reagan, Y.J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov, "Aeronet—a federated instrument network and data archive for aerosol characterization," *Remote Sensing of Environment*, vol. 66, no. 1, pp. 1 – 16, 1998.

[12] ESA, "Sentinel-2 Level-2A Algorithm Theoretical Basis Document (ATBD) - S2-PDGS-MPC-ATBD-L2A 2.10," 11 2021.

[13] Viktoria Zekoll, Magdalena Main-Knorn, Kevin Alonso, Jerome Louis, David Frantz, Rudolf Richter, and Bringfried Pflug, "Comparison of Masking Algorithms for Sentinel-2 Imagery," *Remote Sensing*, vol. 13, no. 1, 2021.

[14] Zhe Zhu, Shixiong Wang, and Curtis E. Woodcock, "Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images," *Remote Sensing of Environment*, vol. 159, pp. 269–277, 2015.

[15] Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi, "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 10 2016.

[16] Peiyuan Jia, Miao Zhang, Wenbo Yu, Fei Shen, and Yi Shen, "Convolutional neural network based classification for hyperspectral data," *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2016-November, pp. 5075–5078, 11 2016.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lecture Notes in Computer Science*

(including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351, pp. 234–241, 2015.

[18] Ying Li, Haokui Zhang, and Qiang Shen, "Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network," *Remote Sensing 2017, Vol. 9, Page 67*, vol. 9, no. 1, pp. 67, 1 2017.

[19] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," 12 2019.

[20] Amalia Luque, Alejandro Carrasco, Alejandro Martín, and Ana de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 7 2019.

[21] Margherita Grandini, Enrico Bagli, and Giorgio Visani, "Metrics for Multi-Class Classification: an Overview," *arXiv e-prints*, 8 2020.