FILLING CLOUD GAPS IN SATELLITE AOD RETRIEVALS USING AN LSTM CNN-AUTOENCODER MODEL

Jacob Daniels[1], Colleen P. Bailey[1], Lu Liang[2]*

[1]University of North Texas, Department of Electrical Engineering
[2]University of North Texas, Department of Geography and the Environment

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9884482

ABSTRACT

Satellite imagery enables spatially-temporally continuous monitoring and understanding of global environmental factors for a range of applications. This data, however, often suffers from gaps in retrieval from sensor malfunction or atmospheric interference, particularly dense clouds that obscure parts or all of an area. For dynamic datasets such as the MCD19A2 Aerosol Optical Depth (AOD) dataset, gap filling is especially challenging. The difficulty lies in the often large, continuous blocks of cloudy pixels with missing data that limit the ability of spatial filling and the daily fluctuation in features such as AOD that incur high difficulty in gap filling from temporal trends. In this study, we propose a spatio-temporal long short-term memory (LSTM) convolutional autoencoder method that effectively reconstructs missing data resulting from thick cloud interference for MODIS AOD data. The proposed method outperforms previous methods of reconstructing data lost to thick cloud interference in AOD retrievals with a generalized network achieving a weighted average PSNR, SSIM, and R^2 of 47.2, 0.992, and 0.941, respectively, between original, cloud-free days and those same days masked with simulated thick cloud interference without the need for additional covariates.

Index Terms— remote sensing, data reconstruction, AOD, machine learning, cloud filling

1. INTRODUCTION

Satellite imagery is an efficient tool for retrieving a large number of environmental, biological, and ecological variables at many different scales with temporal continuity [1], [2]; however, these benefits come with the challenge of collecting surface or atmospheric variables through the interference of atmospheric contamination. Gaps in satellite retrieval occur due to multiple factors including sensor malfunction, sensor limitations, and the presence of thick clouds [3]. Since many techniques including classification, environmental modeling, and trend analysis rely on complete imagery to accurately arrive at their respective solutions, filling these data gaps remains an important topic for remote sensing applications.

Many techniques have been proposed to address the missing data found in remote sensing products including spatial, temporal, and spectral methods [4], [5]. The solutions that integrate multiple methods of gap filling such as the spatial-spectral-temporal model proposed in [4] have shown significant promise in approximating the missing data in satellite imagery. While these methods perform well for some applications like gap filling evenly-spaced dead pixel lines found in Landsat's ETM+ data after its scan line corrector (SLC) failed, they are less performant for large cloud removal. Unlike dead lines, clouds can often remove large swaths of information from a region resulting in no neighboring details to use in the filling process.

Prior methods were also optimized to work under select product conditions. Spectral-based models assume there is at least one other accompanying band correlated with that of the missing data and that these correlated bands do not likewise have missing data. Similarly, the visible spectrums of Landsat imagery frequently used in prior studies also assume relatively consistent information between observations over relatively large time frames, a feature non-characteristic of all satellite imagery.

In this paper, we propose a spatio-temporal long short-term memory (LSTM) convolutional autoencoder that works for dynamic, single-band datasets. This study uses the NASA MODIS Level 2 MCD19A2 data product and is conducted over MODIS tile h05v09 crossing a geographically heterogeneous area across portions of Texas, New Mexico, and Colorado, among other states in the United States, to create a generalized model capable of reconstructing satellite imagery over a diverse set of locations. As a product of various algorithms, MCD19A2 is a single-band imagery that measures the geophysical parameter of aerosols [6]. This particular L2 product combines the data from the MODIS Terra and Aqua satellites and uses the Multi-angle Implementation of Atmospheric Correction (MAIAC) algorithm to derive Land Aerosol Optical Depth (AOD) [7], a parameter that fluctuates daily [5] causing gap-filling techniques relying on long-term consistency to be ineffective. We further investigate the model performance at varying levels of cloud cover to analyze how the performance changes as less information is available to the model.

*This research was supported by NSF grant BCS-2117433.

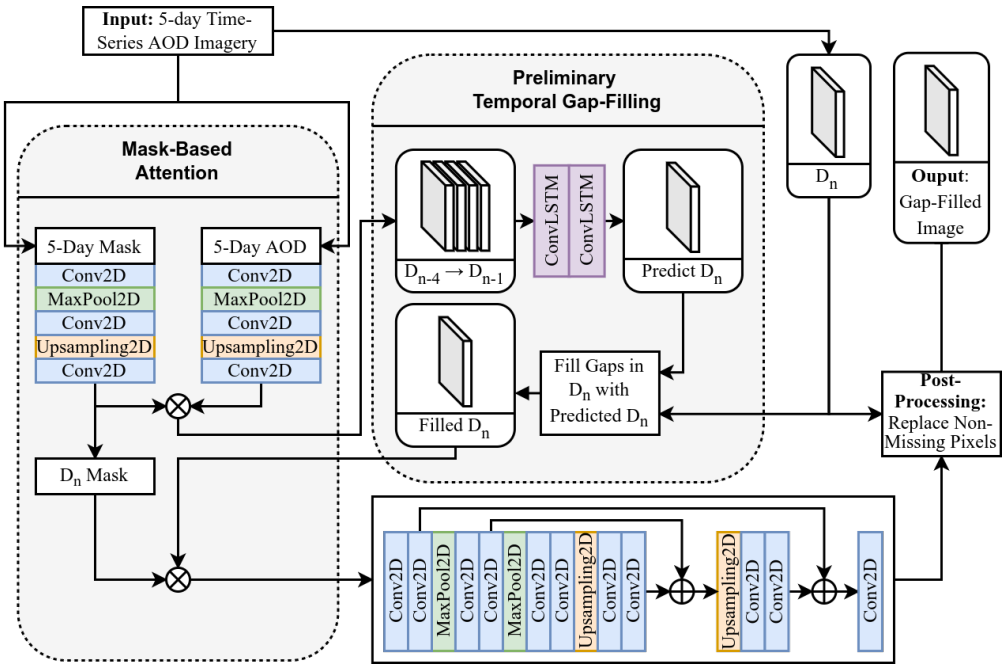

Fig. 1. Gap filling system architecture. D_n is the satellite image on day *n*.

2. METHODOLOGY

2.1. Data Preparation

Before training the model to fill the missing cloud gaps, the raw AOD imagery was filtered and masked to provide an optimal training set. The imagery first must be filtered using the provided quality assurance (QA) masks to remove pixels corresponding to those with cloud interference; for this step, any pixels with clouds detected by the MAIAC cloud detection algorithms are labeled as missing. Those labeled as possibly suffering from cloud interference are retained as this filter is known to erroneously erase non-degraded data over urban areas [7]. Following QA masking, any pixels over bodies of water not removed prior to the publication of the MCD19A2 dataset were removed and flagged as water body pixels as AOD over water is often unreliable or missing [7].

2.2. Dataset Creation

For each satellite AOD image, the quantity of cloudy pixels was computed for later use in evaluating the model at varying data availability. Using this statistic for each image, two groups were formed: those with no missing information and those with cloud interference present, obscuring AOD retrieval. Those images that are complete and clear of clouds were utilized in training and testing the model as the target images. From the incomplete images, the empty pixel positions are extracted to create cloud masks for simulating cloudy conditions over the clear inputs for training and testing; since these cloud masks are extracted from actual satellite retrievals, they are more representative and accurate than artificially simulated cloud blobs.

After labeling, grouping, and extracting cloud masks from the images based on the quantity of missing data, the dataset used for training and testing the model is created. This dataset uses each of the complete images as the base for the training samples, and, by stacking the previous four days of imagery with the current day's complete AOD retrieval, a time series of AOD retrievals was created. The span of previous imagery was determined experimentally to optimize performance. After each time series is constructed, the current day's AOD map is corrupted using the previously extracted cloud masks by zeroing the corresponding cloud pixel positions in the complete image to simulate cloudy conditions.

2.3. Model Architecture and Implementation

The architecture of the proposed spatial-temporal LSTM convolutional autoencoder model is shown in Fig. 1, with the major components being the mask-based attention, LSTM gap prediction, and the convolutional autoencoder. The network input is a time series of five days of satellite AOD imagery, with the final day in the series being the current day, D_n, to be filled. From this time series, the 5-day masks are extracted, and these and the raw 5-day AOD are fed through identical autoencoders. The novel mask autoencoder structure extracts a dynamic attention map based on the location of corrupted pixels while the AOD autoencoder mirrors this functionality to match the new attention map, and these are multiplied to focus the model on the non-corrupted data.

The attention-based representation of the previous four days' images, D_{n-4}-D_{n-1}, are then fed into the convolutional LSTM layers. These are similar to normal LSTM layers, but the input transformations and recurrent transformations are convolutional allowing further feature extraction characteristic of convolutional layers. The first layer returns the full sequence of the output while the second convolutional LSTM layer returns only the final predicted output for D_n. Using this predicted AOD, the missing-data pixels of the raw corrupted D_n image are filled with the corresponding values in the predicted AOD image, completing the preliminary temporal gap-filling procedure.

Before the convolutional autoencoder, the filled D_n AOD is multiplied once again by the mask-based attention map to reinforce the attention toward the known pixels and give less weight to those previously predicted. The convolutional autoencoder, comprised of the encoder and decoder systems, takes this weighted AOD image and is trained to approximate the corresponding non-corrupted image. The encoder attempts to downsample the image into an optimized feature set, keeping the features most representative of the target image. It is comprised of all layers prior to the first upsampling layer with the max pooling layers downsampling the data and the convolutional layers being used for feature extraction.

The decoder performs the inverse operations of the encoder, upsampling the image representation at the bottleneck, or the output of the encoder. The upsampling layers mirror the downsampling of the max pool layers while the convolutional layers learn to extract relevant information from the compressed feature set. During the decoding of the compressed feature set, residual information from the encoding layers are passed through to the decoder to provide fine-resolution information about the original image. Finally, this is followed by a 2D convolution layer with a single output filter to transform the previous layer's representation into the dimensions of the input image. The final post-processing step replaces non-corrupted pixels from the original D_n image in the convolutional autoencoder's output to create the final reconstructed AOD.

3. RESULTS AND DISCUSSION

Similar studies have attempted to fill AOD satellite imagery but achieved non-continuous reconstruction with varying completeness levels of 67.7% [5] and 90% [8], respectively. Studies achieving complete reconstruction, however, can be seen in Table 1 and are more comparable to this work. We use R^2 here for comparison as this metric measures error with respect to the variance around the mean of each dataset's samples and, therefore, measures a model's ability to explain the variance in the target variable (AOD, in this case); this eliminates problems that arise from cross-dataset comparison with non-relative error metrics like RMSE or MAE. As can be seen in this table, the model proposed here outperforms

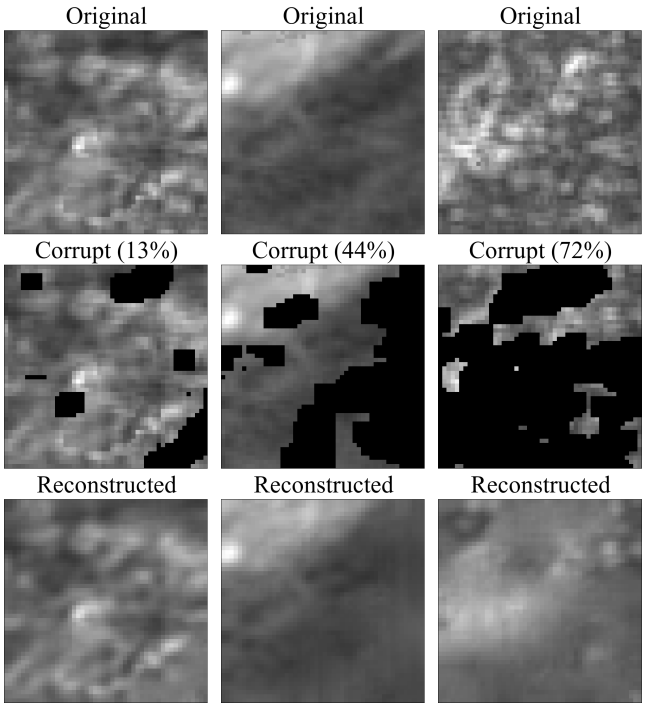

Fig. 2. Gap filled AOD at varying cloud covers. Percentage is amount of corrupted data.

Model	Timescale	R^2	Ref
Full Residual Network	Weekly	0.94	[10]
Non-Full Residual Network	Weekly	0.86	[10]
Non-Linear GAM	Weekly	0.81	[10]
Multiple Imputation	Daily	0.77	[9]
LSTM CNN-Autoencoder	Daily	0.94	*

Table 1. Performance comparison across similar studies. *=This Study

the similar daily model seen in [9] with a 22% increase in R^2. The other models detailed in Table 1 reconstruct corrupted images on a weekly timescale as opposed to the daily timescale of our model; since there is larger variation on the daily timescale, we expect poorer performance from daily models. Even with the increased complexity of creating a daily model, when comparing our performance to that reported in [10], our model outperforms both the Non-Full Residual Network and Non-Linear GAM and performs comparably to the Full Residual Network. Other than the ability to reconstruct daily images, one major advantage of our model is the lack of required covariates. All three networks described in [10] require 13 covariates in addition to AOD. The model with the highest performance of the three, the Full Residual Network, also requires that all AOD pixels contain $\geq$60% valid retrievals during the study month to perform reconstruction and consequently allows only intermittent re-

Corruption Level	PSNR	SSIM	RMSE	R^2
0.01 - 0.1	67.22	1.000	0.002	0.999
0.11 - 0.2	59.45	1.000	0.005	0.997
0.21 - 0.3	57.36	0.999	0.007	0.994
0.31 - 0.4	55.47	0.999	0.008	0.989
0.41 - 0.5	54.00	0.999	0.010	0.987
0.51 - 0.6	51.79	0.992	0.013	0.980
0.61 - 0.7	50.12	0.992	0.016	0.972
0.71 - 0.8	48.53	0.998	0.019	0.957
0.81 - 0.9	43.98	0.992	0.032	0.905
0.91 - 1.0	39.57	0.907	0.053	0.555

Table 2. Results of gap filling method. The table is organized by proportion of corruption.

construction. In contrast, this work requires only the retrieved AOD and, therefore, operates continuously, regardless of the presence of other retrievals.

The results of the trained model with varying levels of corruption are shown in Table 2 as well as samples of test images, their cloud-masked counterparts, and the reconstructed output from our model in Fig. 2 for varying levels of cloud cover. The proposed model is able to achieve a weighted average peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and R^2 of 47.2, 0.992, and 0.941, respectively. As can be seen in both the table and the reconstruction samples, as the level of cloud interference increases, the performance degrades as expected due to a higher reliability on prior days' retrievals, but the trends in areas with higher and lower levels of AOD are still reliably reconstructed with a maximum RMSE of 0.053. The sharp drop in performance when the corruption nears 100% is likely due to the need of the model to rely heavily on the regularly fluctuating temporal AOD data to fill cloud gaps.

4. CONCLUSION

As satellite imagery continues to be an integral component of fields including biological conservation, urban planning, disaster mitigation, and many others, the need for techniques that fill information gaps in retrieval with high accuracy persists. In this paper, we propose a spatio-temporal LSTM convolutional autoencoder that is able to restore data in remotely sensed images resulting from dense cloud interference with an average PSNR, SSIM, and R^2 of 47.2, 0.992, and 0.941, respectively, using the ability of LSTM-based models to learn temporal characteristics and convolutional autoencoders' aptitude for optimizing relevant spatial feature extraction and suppressing noise. The technique is able to outperform past methods for this task, allowing higher quality reproduction of lost data in single-band satellite products with high levels of temporal inconsistency.

5. REFERENCES

[1] Harini Nagendra and Madhav Gadgil, "Satellite imagery as a tool for monitoring species diversity: an assessment," *Journal of Applied Ecology*, vol. 36, no. 3, pp. 388–397, 1999.

[2] Aleem Khaliq, Vittorio Mazzia, and Marcello Chiaberge, "Refining satellite imagery by using uav imagery for vineyard environment: A cnn based approach," in *2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, 2019, pp. 25–29.

[3] Huanfeng Shen, Xinghua Li, Qing Cheng, Chao Zeng, Gang Yang, Huifang Li, and Liangpei Zhang, "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 61–85, 2015.

[4] Qiang Zhang, Qiangqiang Yuan, Chao Zeng, Xinghua Li, and Yancong Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274–4288, 2018.

[5] Jing Yang and Maogui Hu, "Filling the missing data gaps of daily modis aod using spatiotemporal interpolation," *Science of The Total Environment*, vol. 633, pp. 677–683, 2018.

[6] "Data processing levels," https://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/, Dec 2021.

[7] Alexei Lyapustin and Yujie Wang, *MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC) Data User's Guide*, NASA, Jun 2018.

[8] Yufeng Chi, Zhifeng Wu, Kuo Liao, and Yin Ren, "Handling missing data in large-scale modis aod products using a two-step model," *Remote Sensing*, vol. 12, no. 22, 2020.

[9] Qingyang Xiao, Yujie Wang, Howard H. Chang, Xia Meng, Guannan Geng, Alexei Lyapustin, and Yang Liu, "Full-coverage high-resolution daily pm2.5 estimation using maiac aod in the yangtze river delta of china," *Remote Sensing of Environment*, vol. 199, pp. 437–446, 2017.

[10] Lianfa Li, Meredith Franklin, Mariam Girguis, Frederick Lurmann, Jun Wu, Nathan Pavlovic, Carrie Breton, Frank Gilliland, and Rima Habre, "Spatiotemporal imputation of maiac aod using deep learning with downscaling," *Remote Sensing of Environment*, vol. 237, pp. 111584, 2020.