

Deforestation Area Segmentation in Satellite Image from Multimodal Remote Sensing Data

Dongoo Lee,¹ Yeonju Choi,¹ and SungTae Moon²

¹Satellite Application Division, Korea Aerospace Research Institute, Address, Daejeon, South Korea

Email: ldg810@kari.re.kr, choiyj@kari.re.kr

²Department of Intelligent System Robotics, Chungbuk National University, Cheongju, South Korea

Email: stmoon@cbnu.ac.kr

Deforestation of the Amazon rainforest is approaching the worst in history. To protect against deforestation, it is necessary to accurately estimate the deforestation area. However, it is difficult to analyze large areas without direct human access. In addition, even if deforestation is estimated using satellite images, the presence of extensive cloud cover during the rainy season makes it challenging to obtain a clear view of the ground surface. In this paper, we propose a segmentation method based on deep learning and post-processing to predict the deforestation status in the Amazon rainforest area. To train and predict the deforestation area, we utilize a multi-modal satellite imagery dataset, including Sentinel-1, Sentinel-2, and Landsat 8. The proposed approach achieves the highest performance in the official CVPR MultiEarth Workshop 2023 challenge.

Introduction: The Amazon is the most important and priceless resource on Earth. It plays a key role in reducing the negative consequences of climate change by actively absorbing greenhouse gases and producing oxygen, acting as a crucial regulator. However, the problem of Amazon deforestation is becoming increasingly serious. Unregulated deforestation has negative effects, such as the eradication of ecosystems, loss of biodiversity, soil erosion, and accelerated global warming. A daily average of 2,300 hectares of forest were destroyed in 2020 [1] and as a result, reducing the rate of Amazonian deforestation through the creation of protected zones has gained international attention. Effective forest management, including strengthened forest monitoring, is essential to addressing this urgent issue[2].

Recently, as the types of available satellite images have been diversified, the spatial resolution of satellites is improving, and deep learning is showing remarkable performance in the image analysis area [3, 4]. Especially, the multi-sensor data fusion strategy combining SAR (Synthetic Aperture Radar) and optical sensors such as Landsat has improved the accuracy of forest mapping [5]. The MultiEarth 2023 is the second CVPR workshop utilizing multi-satellites imagery to help monitor and analyze the health of these Earth ecosystems [6, 7]. The MultiEarth 2023 Challenge's objective is to combine optical and SAR imagery to carry out continuous assessments of forest monitoring at any time and in all weather. This research is the extending work of last year's deforestation challenge study [8]

The main contributions of this research are summarized as follows.

- Post-processing with effective cloud removal to accurately predict deforestation with multi-modal dataset is proposed.
- Time-series analysis method with adjacent month data is proposed.

The remainder of this paper is organized as follows. Section 2 introduced the contents and generation method of the dataset in detail and Section 2 described the proposed methodology for deforestation detection and in Section 2, prediction results are presented. The conclusion is presented in Section 2.

Dataset: The area of interest in this study is the area bounded by (latitude (LAT) :-3.33° ~ -4.39°, longitude (LON) :-54.48° ~ 55.2°), which is one of the areas where deforestation occurs very frequently in the Amazon. The region comprises a portion of dense tropical Amazon rainforest in Para, Brazil, containing thousands of species of broad-leaved evergreen trees. Historically, this has been one of the areas with the highest tree loss rate in the Amazon region, with pastures in the region nearly

Satellite	Time Range	Bands	Resolution
Landsat 8	2013-2021	9	30 m
Sentinel-1	2014-2021	2	10 m
Sentinel-2	2018-2021	12	10 m
Labeling	2016-2021	1	10 m

Table 1. Satellites and data specification.

doubling between 1997 and 2007.

We used three datasets obtained from Sentinel-1 (Sen1), Sentinel-2 (Sen2) Land Landsat 8 (Land8). The three types of satellites are equipped with different sensors, and have different temporal/spatial resolutions and acquisition cycles. Specific specifications for the utilization dataset are described in Section 1.

The final size of all dataset images is 256 x 256, and in the labeling image, the deforestation pixels are set to 1 and the forest or background pixels are set to 0.

Methodology: The deforestation estimation challenge of MultiEarth is to determine whether a region is deforested or not. To solve the problem, we adopted Masked-attention Mask Transformer (Mask2Former)[9], a new architecture capable of addressing any image segmentation task as depicted in Figure 1. With the Mask2Former deep neural network, we focused on data pre-processing and post-processing to improve performance. The overall procedures for deforestation estimation method is as follows.

Pre-processing: To use diverse data augmentation libraries, which are mostly support only 3-channel image, we only select RGB bands of Sen2 and Land8 imagery data. To make 3-channel image, a mock band filled with 0 is inserted to Sen1 imagery data. The pre-processing procedures of training dataset are summarized as follows.

1. RGB bands for Sen2 and Land8 are selected and VV/VH bands are selected for Sen1.
2. Generate a training input image with selected bands. a mock band is inserted to Sen1 to make it 3 channel image. The shape of the processed images are (3, 256, 256) for all types of satellite.
3. Remove lowest 2% and highest 2% data in the image and normalized to [0,1].

Method	ResNet-50	Swin-L
Landsat 8	79.445	80.84
Sentinel-1	82.13	79.42
Sentinel-2	78.145	77.41

Table 2. Pixel accuracy comparison according to backbone network

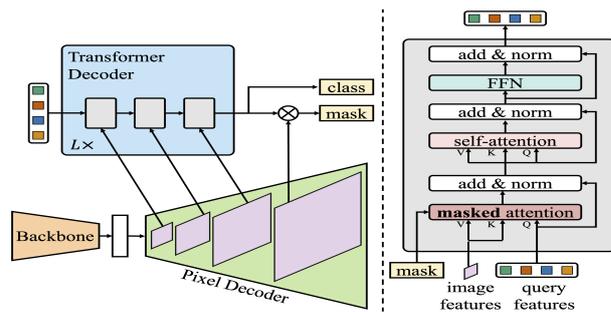


Fig 1 Mask2Former network architecture [9].

Training Network: To detect the deforested area in this challenge, we tried to adopt several networks using MMSegmentation [10] which is an open source semantic segmentation toolbox based on PyTorch. Through

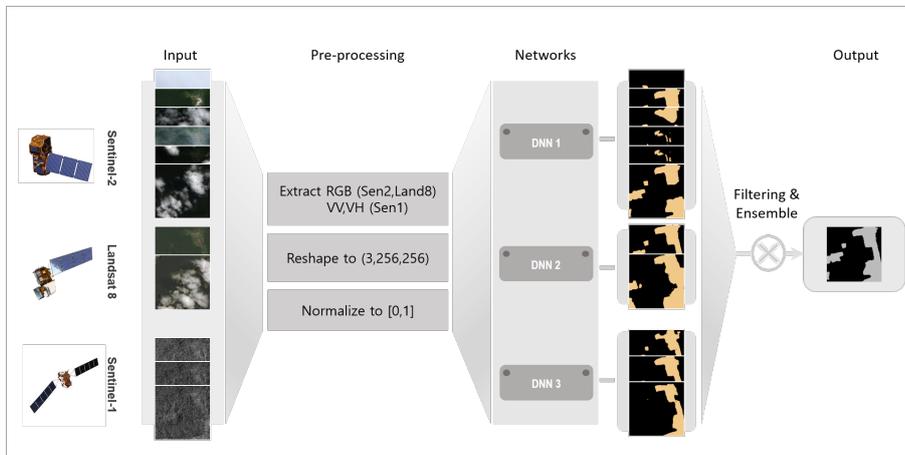


Fig 2 Overall learning and prediction strategy for the proposed method

multiple experiments, we have discovered that Mask2Former [9] is the most suitable network for this challenge. In addition, there were differences in pixel accuracy depending on the backbone network of the mask2former network, as shown in Table 2. Therefore, we selected the best backbone network for each satellite.

In this challenge, there are two detection classes. The deforested area, which is the target, is labeled as class 1, while the forested/other area is labeled as class 0. To compensate for the class imbalance, we adopted the combination of Binary Cross-Entropy loss[11] and the Dice loss[12]. In addition, the model training is performed with AdamW optimization, and the learning rate is adjusted by checking for validation loss every two epochs.

Post-processing: The deforestation estimation test queries consist of satellite images captured at specific times and locations, expressed in latitude and longitude coordinates. However, there are many cases where there are no satellite images available at the given query, and in some cases, there are only one or two images available as shown in Figure 3. In addition, test images data of the same region at different times are supported. Therefore, to address the absence of image data or insufficient information at the given query, we leveraged time-series data that incorporated images captured at the same location during similar periods. For this post processing, the deforestation estimation output of image from the adjacent month for a specific query are generated using different network according to Sen1, Sen2, and Land8. To enhance the importance of the images captured in the current month, higher weights were assigned compared to the preceding and succeeding months.

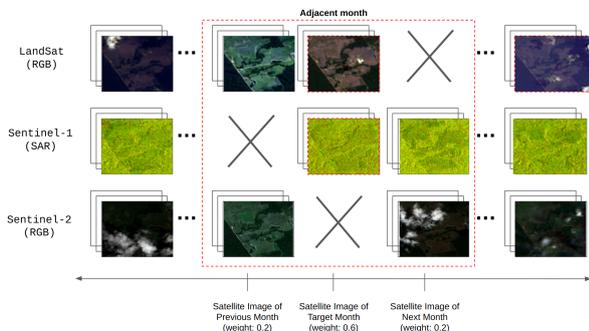


Fig 3 Post processing for the adjacent month.

On the other hands, the clouds in the image can lead to misleading results when estimating deforestation areas. To enhance the deforestation estimation performance, we exclude images that contain a significant amount of clouds. Especially, it was observed that the accuracy of the estimation decreased for Land8 and Sen2 images with a significant presence of clouds, as shown in Figure 4. This was attributed to the selection of RGB bands unlike Sen1. Therefore, for Land8 and Sen2, which are affected by clouds, images with a significant cloud presence

were removed. The removal criterion was determined by a simple rule: images in which all RGB values exceeded 160 were classified as clouds, with the threshold set at over 50% of the total image pixel. In case of Sen1 images, no cloud removal was performed because it were acquired using SAR.

Despite learning from the same network, some of the prediction results were completely black (no deforestation), and some present plausible deforestation results. Therefore, two-step filtering was applied to the output result to ensure the overall performance as shown in Figure 5. The clear outliers were filtered without three-sigma range ($\mu \pm 3\sigma$) at the first filtration step. Second filtration used one-sigma range ($\mu \pm \sigma$) using predicted deforestation percentage in the images. Here, μ is the mean and σ is the variation of deforestation ratio for the predicted output.

Using the filtered output images from the previous step, all images were averaged to create a single binary image. In this image, each pixel was assigned a value of 1 if the probability of deforestation in that pixel exceeded a specific threshold (we set the threshold at 40%). Conversely, if the probability was below the threshold, the pixel was assigned as 0.

Finally, the noise in binary image is removed with the morphology opening operation \ominus as described in Equation 1 that is dilation operation of erosion result.

$$I \circ M = (I \ominus M) \oplus M, \quad (1)$$

where, I, M are the original image and a structuring element and \ominus, \oplus are the erosion and dilation operation respectively. The erosion operation eliminates small objects and dilation restores the size and shape of the remaining objects in image.

Results: The test set comprises 1,000 queries spanning from August 2016 to August 2021, covering 135 regions. The amount of image information provided for each query was inconsistent. Therefore, we focus on the post-processing performance to improve the pixel accuracy. To evaluate the post-processing performance, each method is assessed from a evaluation website. Especially, With the cloud removal post processing, the pixel accuracy is increased up to 90.546. In addition, when there are only a few satellite images available at the given query time, considering time-series data increases pixel accuracy.

With cloud removal and time adjacent image data in the post-processing procedure as described in Section 2, final detection performance has been improved from the initial results. As a result, the proposed method finally achieves 91.13 pixel accuracy, 0.88 F1-score and 0.81 IoU for the test set as represented in Table 3.

Conclusion: In this study, we suggested the deforestation estimation model based on multi-modal dataset(Sentinel-1, Sentinel-2, and Landsat 8). A variety of augmentation libraries and the most recent deep neural networks are adopted using the MMsegmentation framework. The RGB bands of satellite data are selected for the application of data augmentation libraries. For each set of satellite data, the Mask2Former approach adopts the better of two alternative backbones. The final step involves

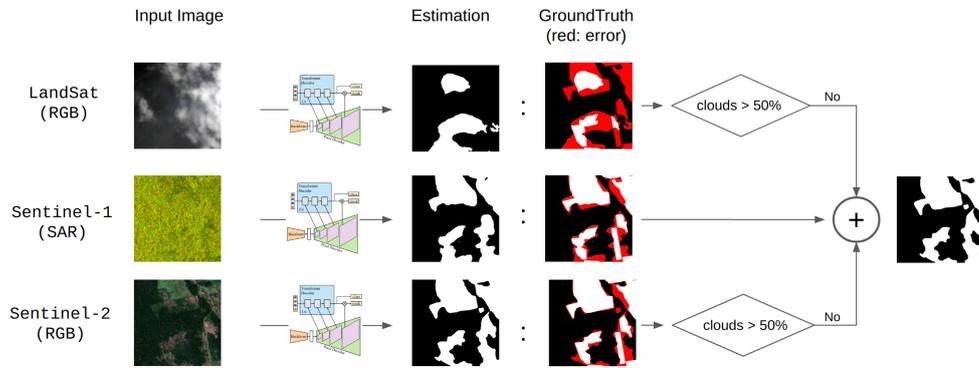


Fig 4 Post processing for cloud removal.

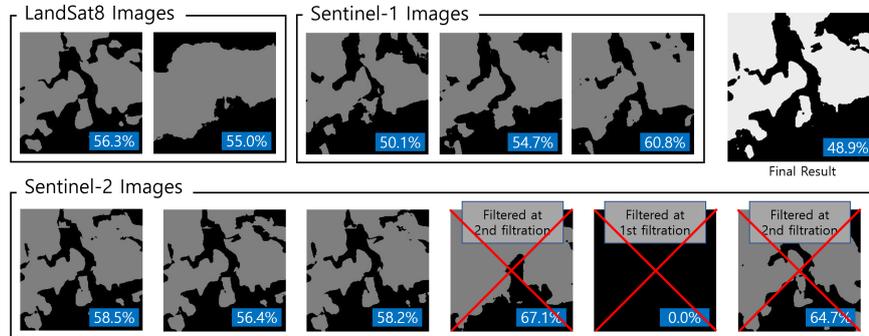


Fig 5 Post processing for outlier filter

Method	Pixel Accuracy	F1 Score	IoU
Mask2Former (original)	89.762	0.865	0.784
Mask2Former (with cloud removal)	90.536	0.879	0.801
Mask2Former (with cloud removal and adjacent month)	91.139	0.888	0.814

Table 3. Final statistics for test set evaluation.

post-processing, which includes removing cloud-covered photos, filtering outlier prediction findings, averaging them with data from the preceding and following months, and performing denoising operations. The suggested method outperforms in all evaluation criteria that exceed the state-of-the-art model, including pixel accuracy, F1, and IoU.

Acknowledgments: This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2022-00166347, No. 2023M3C1C1A0109841621)

© 2023 The Authors. *Electronics Letters* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Received: 10 January 2023 Accepted: 4 March 2023
doi: 10.1049/ell2.10001

References

- Annual report on deforestation in brazil. https://s3.amazonaws.com/alerta.mapbiomas.org/rad2020/RAD2021_-_MapBiomasaAlerta_EN.pdf (2020)
- Banskota, A., et al.: Forest monitoring using landsat time series data: A review. *Canadian Journal of Remote Sensing* 40(5), 362–384 (2014)
- de Andrade, R.B., Mota, G.L.A., da Costa, G.A.O.P.: Deforestation detection in the amazon using deepplabv3+ semantic segmentation model variants. *Remote Sensing* 14(19), 4694 (2022)
- John, D., Zhang, C.: An attention-based u-net for detecting deforestation within satellite sensor imagery. *International Journal of Applied Earth Observation and Geoinformation* 107, 102685 (2022)
- Walker, W.S., et al.: Large-area classification and mapping of forest and land cover in the brazilian amazon: A comparative analysis of alos/palsar and landsat data sources. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3(4), 594–604 (2010)
- Cha, M., et al.: Multiearth 2022 – multimodal learning for earth and environment workshop and challenge (2022)
- Cha, M., et al.: Multiearth 2023 – multimodal learning for earth and environment workshop and challenge (2023)
- Lee, D., Choi, Y.: Multiearth 2022 deforestation challenge – forest-gump (2022)
- Cheng, B., et al.: Masked-attention mask transformer for universal image segmentation. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1280–1289. (2022)
- Contributors, M.: MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. <https://github.com/open-mmlab/mms Segmentation> (2020)
- Yi-de, M., Qing, L., Zhi-Bai, Q.: Automated image segmentation using improved pcnn model based on cross-entropy. In: Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004., pp. 743–746. IEEE (2004)
- Sudre, C.H., et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 240–248. Springer (2017).
- FG Assis, L.F., et al.: Terrabrasilis: a spatial data analytics infrastructure for large-scale thematic mapping. *ISPRS International Journal of Geo-Information* 8(11), 513 (2019)