Cloud Publications
International Journal of Advanced Remote Sensing and GIS
2024, Volume 12, Issue 1, pp. 3531-3544
ISSN 2320 – 0243

Research Article

Enhancing Semantic Segmentation of Cloud Images Captured with Horizon-Oriented Cameras

Allan Cerentini[1*], Bruno Juncklaus Martins[1], Juliana Marian Arrais[1], Sylvio Luiz Mantelli Neto[2], Gilberto Perello Ricci Neto[1], Aldo von Wangenheim[1]

[1]PPGCC - Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.
[2]FOTOVOLTAICA-UFSC, INPE Brazilian National Institute for Space Research, São José dos Campos, São Paulo, Brazil.

Correspondence should be addressed to Allan Cerentini, *allan.c@posgrad.ufsc.br*

Publication Date: 3 June 2024

Abstract The segmentation of sky cloud images is a complex task essential for applications like weather analysis. Compared to all-sky imagers, horizon-oriented cameras provide a more detailed view of clouds near the horizon. In our study, we evaluated three semantic segmentation models: HRNet48, PPLite, and SegFormerB3, utilizing a variety of loss functions on a novel dataset of horizon cloud images. Throughout our experiments, we consistently observed segmentation leakage issues. To address this, we introduced machine learning-based post-processing methods, including random forest and xgboost, that leverage region-specific features to refine the segmentation. Our results showed notable improvements, with the Cumuliform class dice score increasing from 0.552 to 0.583, and Stratiform class accuracy improving from 0.49 to 0.511 when applying xgboost on SegFormerB3's output. The study revealed the relative contributions of the loss functions and post-processing steps.
Keywords *Remote Sensing; Segmentation; Sky Clouds; Deep Learning*

1. Introduction

Classifying clouds is crucial in meteorology, aviation, and environmental science for understanding weather patterns, precipitation, air quality, and flight safety. Visual observation is subjective, so artificial neural networks (ANNs) are used to automate cloud type classification from sky images, providing more accurate and consistent results (Veremev, 2021). At airports, accurately determining cloud ceiling height is vital for safe takeoffs and landings. For example, at Bombay airport, low cloud ceilings linked to wind shear impact visibility and operations (Kumar and Patkar, 2022). Also, studying cloud cover properties aids weather prediction, precipitation quantification, and analyzing air mass movement over oceans.

Cloud classification is a nuanced task traditionally undertaken by organizations like the World Meteorological Organization (WMO), which categorizes clouds by shape, clustering, and base height. The WMO Cloud Atlas[1] further subdivides clouds into specific groups. In addition to traditional classifications, clouds can be characterized by their albedo or reflective properties, which set them apart

[1] https://cloudatlas.wmo.int/en/cloud-classification-summary.html

from other outdoor objects. Due to their higher reflectivity in the visible spectrum, clouds present unique detection challenges, often constrained by camera scale limitations (Mantelli et al., 2020). Regular objects usually reflect local radiation, failing to capture the unique albedo attributes of clouds and the surrounding landscape in sunlight. Thus, brightness alone is inadequate for cloud distinction. Some modern computer vision (CV) techniques go beyond conventional approaches by utilizing cross-classification, dividing clouds into five physical forms:

- Stratiform (Cirrostratus, Altostratus, Stratus and Nimbostratus)
- Cirriform (Cirri)
- Stratocumulus (Cirrocumulus, Altocumulus and Stratocumulus)
- Cumuliform (Cumulus)
- Cumulonimbus (cumulonimbus)

These classifications consider opacity, structure, and formation processes, aligning with the methodologies proposed by (Song et al., 2020). The complexity of clouds extends to their segmentation and classification in ground-based images, a particularly arduous task (Fabel et al., 2022). Clouds, being amorphous structures, lack clear boundaries and sizes that are challenging to identify and track. Many regions within different cloud types appear similar, requiring broader context for accurate classification, and clouds can fluidly transition from one type to another without distinct boundaries (Ye et al., 2022). This complexity has led many researchers to reduce cloud segmentation and classification to a binary "cloud/not-cloud" problem (Ye et al., 2022). Others have focused on estimating cloud impact on specific applications by grouping clouds into a few superclasses that are easier to distinguish (Fabel et al., 2022). In the context of CV, clouds can be segmented, and their individual pathways tracked and predicted. Various segmentation techniques have been employed to identify and classify clouds, primarily focusing on attributes such as shape, texture, color similarity, brightness, and contour continuity within an image (Juncklaus Martins et al., 2022a).

In this context, semantic segmentation (SS) using deep learning (DL) techniques have been the most promising. However, there are situations where SS cloud segmentation can yield unusable results, as depicted in Figure 1, where several clouds have been both under- or over-segmented and misclassified. These errors in cloud segmentation and classification may arise from the inherent complexity of cloud formations, leading to a phenomenon called segment leakage (Wangenheim et al., 2007), variations in atmospheric conditions, or limitations in the algorithms being employed. Such results pose challenges for meteorological analysis and forecasting, underscoring the need for continual refinement and adaptation of the segmentation techniques employed.

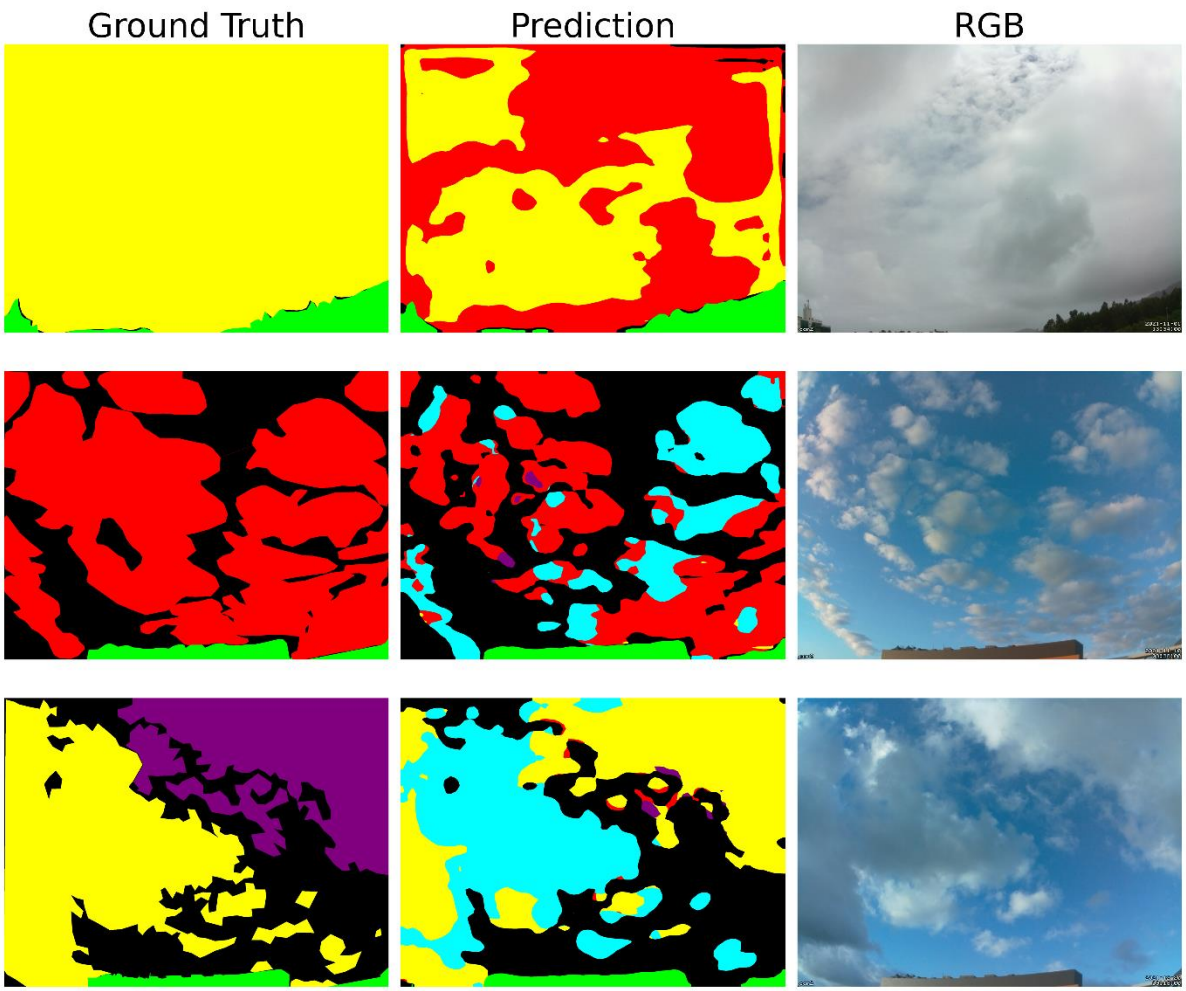

Figure 1: Examples of unreliable results obtained from traditional DL-based SS cloud segmentation. Where each color represents a different class.

2. Objectives

In this work, our research focuses on a few key objectives aimed at advancing the state of the art of segmentation of cloud images captured by horizon-aimed cameras, specifically to:

- Employ multiple segmentation models to evaluate their performance and applicability on our unique horizon cloud dataset.
- Investigate the influence of different loss functions on the segmentation process, seeking insights into how they affect segmentation quality and characteristics.
- Apply and develop innovative post-processing techniques to enhance the final segmentation result, working to refine the accuracy and robustness of the models.

Through these objectives, we aim to make valuable contributions to the field of cloud image segmentation, enabling more precise and efficient analysis of atmospheric phenomena captured by horizon-oriented imaging systems.

3. Methodology

3.1 Dataset

In this work we used the open source Clouds-1500 dataset (Arrais, 2023), that is an extension of the Clouds-1000 dataset (Juncklaus Martins et al., 2022b, p. 100), comprising in 1500 sky images captured between March 2021 and January 2023 using ground-based cameras at the Federal University of Santa Catarina and the Photovoltaic Energy Laboratory in Brazil. The images were manually annotated by a team of computer scientists, meteorologists, and an experienced sky observer using the Supervisely platform.

The dataset employs a practical cloud height-based classification system, categorizing clouds into four groups: Cirriforms, Cumuliforms, Stratiforms, and Stratocumuliforms, along with a category for background objects. This classification aims to enhance nowcasting in the solar energy sector by predicting solar radiation absorption by clouds covering solar energy facilities. To ensure dataset quality, a subset of images was inspected for annotation consistency. The dataset was then split into training and validation sets, and a semantic segmentation convolutional neural network was used to identify the 100 lowest-scoring images, which were manually reviewed and corrected by a meteorologist.

The images were captured using motionEye version 0.41 and Motion version 4.2.2, with a frame rate of 1 per minute between 08:00 and 22:00 GMT. The captured images have a resolution of 2592 x 1944 and are stored locally before being uploaded to Google Drive (Arrais, 2023).

The dataset distribution, as presented in

Table 1, reveals significant variations in the number of images and area percentage across different classes. The Object class consists of the largest number of images, totaling 1376, and covers 17.02% of the dataset's area. The Stratocumuliform class follows, with 1095 images and 35.64% of the dataset's area, marking the highest percentage coverage among cloud types. Stratiform is represented by 453 images, accounting for 11.01% of the area, while Cirriform and Cumuliform are the least prevalent classes, with 382 and 251 images, respectively, and corresponding to 4.81% and 3.58% of the dataset's area. This distribution highlights the prominence of certain classes and provides insight into the diversity and characteristics of the dataset, especially considering the regional climatic conditions that influence the formation of specific cloud types. Given the humid climate of the region, Cumulonimbus clouds are rarely formed, typically occurring in dryer regions, and in consequence, few instances of this cloud type are found in our dataset.

Table 1: Shows the distribution of classes in the clouds 1500 dataset. Where the object class represents pixels that do not belong to clouds and the area represents the number of pixels in a given class in relation to the total annotated.

Class Type		Images (quantity)	Area in dataset (%)
Object		1376	17.02
Stratocumuliform		1095	35.64
Stratiform		453	11.01
Cirriform		382	4.81
Cumuliform		251	3.58

3.2 Semantic Segmentation Models and Loss Functions

First, we needed to train and compare various semantic segmentation models to attain efficient segmentation, which is pivotal for the subsequent post-processing step. In our study, we selected two classic models, HRNet (Wang et al., 2020) and SegFormer (Zhu et al., 2021), which have demonstrated significant success across various semantic segmentation challenges, including benchmark datasets like Cityscapes. HRNet maintains high-resolution representations through the network without down-sampling, thus allowing better localization and dense prediction. SegFormer, on the other hand, combines the strengths of transformers and convolutional layers to address the scale variation problem, making it particularly appealing for our specific task of segmenting clouds from the sky. In addition to the models, we also chose to experiment with a newer, lightweight, and efficient model named PP-LiteSeg (Luo et al., 2022). Despite its compact design, PP-LiteSeg has been shown to be capable of achieving performances on par with much larger models. Its architectural efficiency emanates from a unique integration of pyramid pooling and Lite modules, making it an attractive choice for applications demanding reduced computational resources. By comparing these three models, our aim was to discern whether the newer, more lightweight PP-LiteSeg could stand alongside or even surpass the established HRNet and SegFormer in the specific context of cloud segmentation, thereby leading to a choice that balances both accuracy and efficiency.

For the semantic segmentation loss functions, we initially employed the cross-entropy loss, acknowledged for its effectiveness in classification tasks (Goodfellow et al., 2016). However, this alone could not deal with the class imbalance prevalent in our dataset. In this context, particularly when dealing with intricate patterns like segmenting clouds from the sky, employing a combination of Dice and Focal loss has proven to be remarkably effective. The Dice loss focuses on enhancing the spatial continuity and shape properties of the segmented regions, making it crucial for achieving a more realistic configuration of cloud formations (Milletari et al., 2016). Focal loss addresses the class imbalance problem by dynamically scaling the contribution of each instance to the loss based on its classification accuracy, thus emphasizing the learning from the under-represented class (Lin et al., 2017). Together, the combination of Dice and Focal loss harmonizes the requirements of maintaining spatial continuity with class balance, resulting in a more nuanced and accurate segmentation that would be difficult to achieve with either loss function alone. We also tested with a more recent type of loss, referred to in the literature as Semantic Connectivity-aware Loss (SCL) (Chu et al., 2021). This loss function is specifically designed to improve the quality of segmentation results by considering the connectivity perspective. While the paper focuses on portrait segmentation in the context of video conferencing, the principles behind this loss function could be applied to SS of clouds and sky. In the segmentation of cloud and sky, connectivity is a crucial aspect. Clouds often form complex, interconnected structures, and the sky itself is a continuous entity. Traditional loss functions might not adequately capture these connections, leading to fragmented or inconsistent segmentation. The SCL could address this issue by emphasizing the relationships between different regions of the cloud and sky. By ensuring that connected regions are segmented consistently, this loss function could lead to more accurate and coherent segmentation of cloud formations and sky areas.

The neural networks were trained using a dataset consisting of 1500 images, which were divided into two subsets: 375 images for validation and 1125 images for testing. To ensure consistency and compatibility with the network architecture, all images were resized to a uniform resolution of 648x486 pixels. During the training process, data augmentation techniques were employed to enhance the robustness and generalization capabilities of the models. These augmentation algorithms were applied dynamically, subjecting the images to various transformations. One such transformation was a vertical flip operation, which randomly flipped the images along the vertical axis.

Additionally, the brightness, contrast, and saturation of each image were adjusted by a random value within the range of +15% to -15%. This augmentation strategy helped to simulate different lighting conditions and variations that the models might encounter in real-world scenarios.

The training of each network was conducted for a total of 15,000 iterations, with a batch size of 2. The loss function used during this phase was CrossEntropy. To monitor the progress and performance of the networks, validation was performed after every 500 iterations. This allowed for the identification of the best-performing model, which was then selected as the basis for further fine-tuning.

The fine-tuning process was applied to the best model obtained from the initial training phase. This step aimed to optimize the model's performance by adapting it to specific loss functions. Two loss functions were considered for fine-tuning: Dice+Focal loss and Semantic Connectivity Loss. Each loss function was applied separately for 10,000 iterations, allowing the model to learn and adapt to the specific characteristics and objectives of each loss function. After the fine-tuning process, the best weights obtained from each loss function were retained and used as the foundation for our subsequent study.

3.3 Post-Processing Methodology

In our methodology for correcting cloud segmentations, we employed a post-processing technique that relies on previously segmented images. The process consists of two main steps: training and prediction, an overview can be seen on **Error! Reference source not found.**. During the training phase, we r emoved the object class from both the ground truth and the prediction of an image. We then used OpenCV (Bradski, 2000) to extract all the connected components of the prediction. For each component, we extracted information regarding the number of pixels for each class, total area, height, and width. Using the predicted component mask within the ground truth mask, we identified the predominant class in the ground truth image. The extracted features were used as input for classic machine learning models such as random forest (Breiman, 2001) and XGBoost (Chen and Guestrin, 2016), with the target being the class extracted from the ground truth. This procedure was carried out for each component of every image in the dataset. During the prediction phase, we followed a similar process. We extracted the characteristics of the components of the prediction and employed the trained models to determine the predominant class for each component. Finally, all components were relabeled with their predicted class. This approach allowed us to refine cloud segmentations by training machine learning models with features derived from each image's components, resulting in a more accurate representation of the clouds. By leveraging regional information such as pixel counts, area, height, and width, the models could learn to correct segmentation errors and leakages that the initial semantic segmentation model may have produced. For comparative purposes, we also employed a simpler "voting" method. In this method, for each connected component, we counted the number of pixels associated with each class and then labeled the entire component with the majority class. Our post-processing methodology aimed to enhance the segmentation results by incorporating additional spatial and contextual information that the semantic segmentation models might have overlooked. By training models specifically on the connected components, we could capture more nuanced patterns and correct errors at a finer granularity. This approach has the potential to significantly improve the accuracy and consistency of cloud segmentation, especially in challenging scenarios where clouds exhibit complex structures and boundaries.

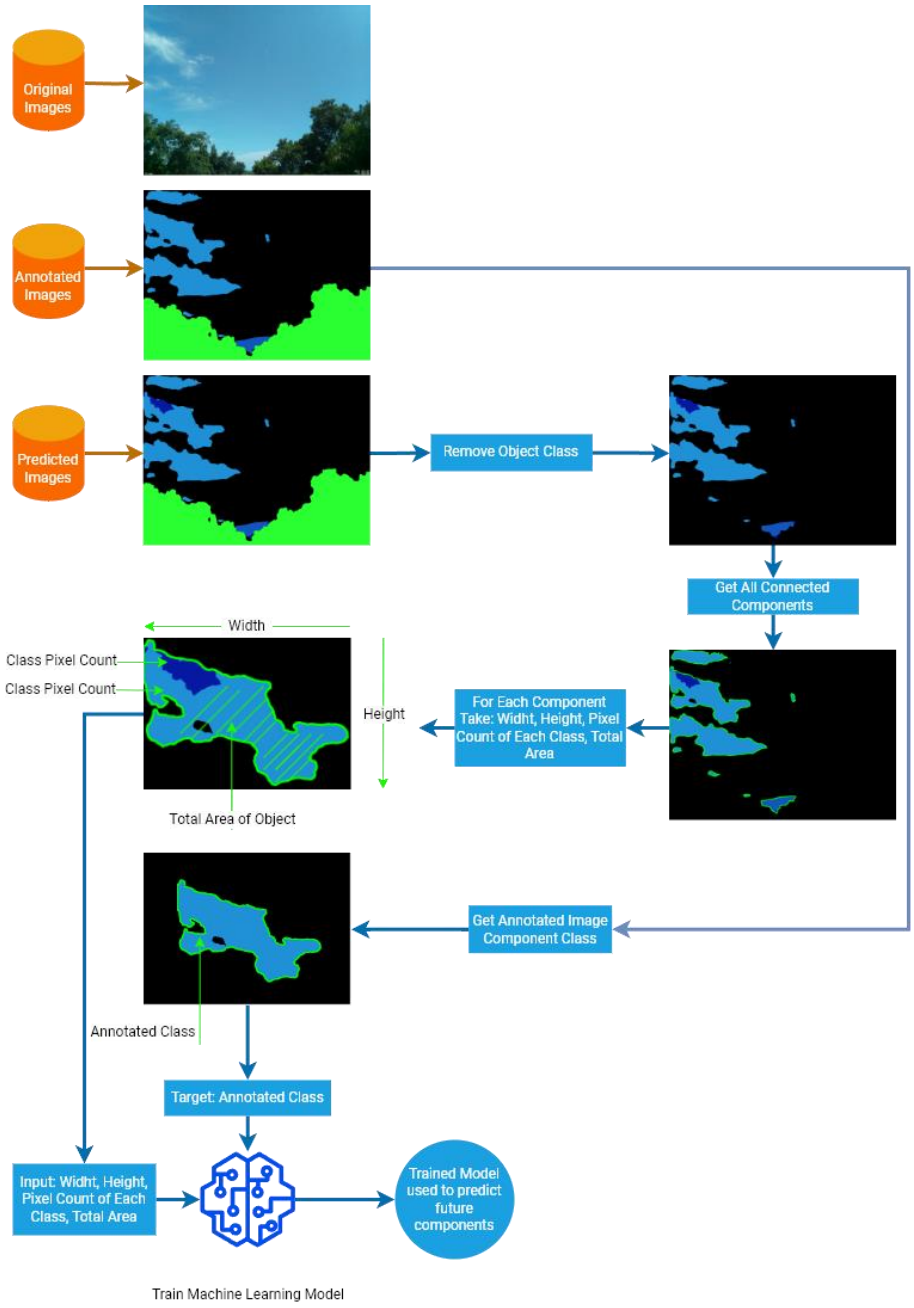

Figure 2: Presents an overview of the methodology applied. The methodology consists of using the image that has already been predicted and correcting it using information from each segmented cloud together with the predicted class.

4. Results and Discussions

Our first experiment aimed to evaluate the performance of three models – HRNet48, PP-LiteSeg, and SegFormerB3 – using three loss functions: Cross Entropy (CE), DiceFocal (DF), and Semantic Connectivity-Aware Loss (SCL). The models were assessed across six classes using the Dice coefficient as the evaluation metric.

Table *2* presents the results of these experiments.

SegFormerB3 consistently outperformed the other models across all loss functions and most classes, particularly in the 'Sky' and 'Cirriform' classes. These findings align with recent research highlighting the

effectiveness of transformer-based architectures in various computer vision tasks, including semantic segmentation (Dosovitskiy et al., 2020).

Conversely, PP-LiteSeg underperformed, especially in the 'Stratiform' and 'Cirriform' classes, potentially due to its lightweight architecture, which may lack the capacity to capture complex patterns in certain classes despite offering computational efficiency benefits (Li et al., 2020).

Comparing the effects of different loss functions revealed that the differences in Dice scores across the three loss functions for all models were marginal. The DiceFocal loss slightly enhanced the Dice scores for the 'Cumuliform' class in all models compared to the other loss functions, possibly due to its balanced approach that combines the advantages of Dice loss and Focal loss, providing more robust performance for difficult-to-segment classes (Lin et al., 2017; Milletari et al., 2016). However, these differences were relatively minor, suggesting that the choice of loss function may not have a major impact on the overall performance of these models. This observation aligns with previous research indicating that while the choice of loss function can influence model performance, it is often secondary to other factors such as model architecture and training regime (Rahman and Wang, 2016).

Table 2: Comparison of dice coefficients for different cloud types across various model architectures and loss functions. Each entry in the table represents the average dice coefficient over the test dataset for a given cloud type, model, and loss function.

Model+Loss	Sky	Object	Cirriform	Cumuliform	Stratiform	Stratocumuli
hrnet48-ce	0.876	0.972	0.539	0.430	0.526	0.799
hrnet48-df	0.875	0.975	0.579	0.455	0.528	0.803
hrnet48-scl	0.876	0.976	0.585	0.452	0.498	0.793
pplite-ce	0.878	0.976	0.488	0.490	0.420	0.768
pplite-df	0.877	0.976	0.499	0.479	0.430	0.763
pplite-scl	0.878	0.975	0.502	0.472	0.424	0.768
segformerb3-ce	0.897	0.977	0.619	0.552	0.490	0.816
segformerb3-df	0.897	0.979	0.643	0.520	0.466	0.793
segformerb3-scl	0.894	0.978	0.622	0.486	0.423	0.792

Figure 3 demonstrates that SegFormerB3 provided more consistent performance, showing fewer leakage issues in the segmentation of these images, with HRNet following closely behind. On the other hand, PP-LiteSeg exhibited several segmentation leaks. This figure highlights a class with less than 5% representation within the dataset. By considering the results from both the images and the data in Table 2, it can be inferred that PP-LiteSeg may not be well-suited for detecting classes with low representation in the dataset.

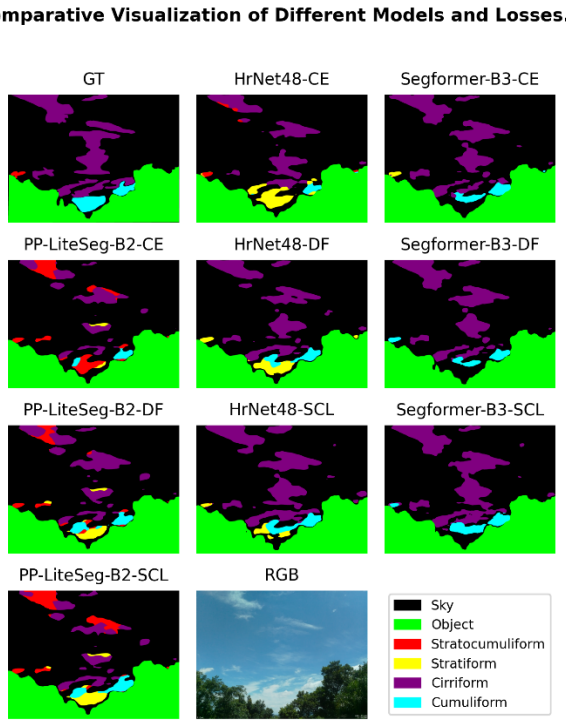

Figure 3: An example cirriform and cumuliform cloud image segmented by all networks with all loss functions. We can see that SegFormerB3 achieved a significantly better result than PP-LiteSeg in all cases and slightly outperformed HRNet.

However, when examining images that appear more frequently in the dataset, such as the stratocumuliform clouds in Figure 4, PP-LiteSeg's results are significantly better compared to the previous figure. In this case, PP-LiteSeg even has fewer leaks than SegFormerB3. Moreover, by referring to Table 2.

Table 2, it is evident that the differences in the predictions for this class are smaller across models. This implies that if the dataset is made more balanced, the performance gap between the segmentation models could be further reduced. It is important to note that while the Dice coefficient offers a valuable measure of model performance, it is not the sole factor to consider when assessing the effectiveness of a model. Other aspects such as computational efficiency, ease of implementation, and adaptability to different tasks should also be taken into account (Rousson et al., 2008). Due to its lightweight architecture, PP-LiteSeg was trained at least twice as fast, utilizing about 3GB of GPU VRAM. This can be beneficial if the objective is to deploy the model on embedded hardware for real-time prediction.

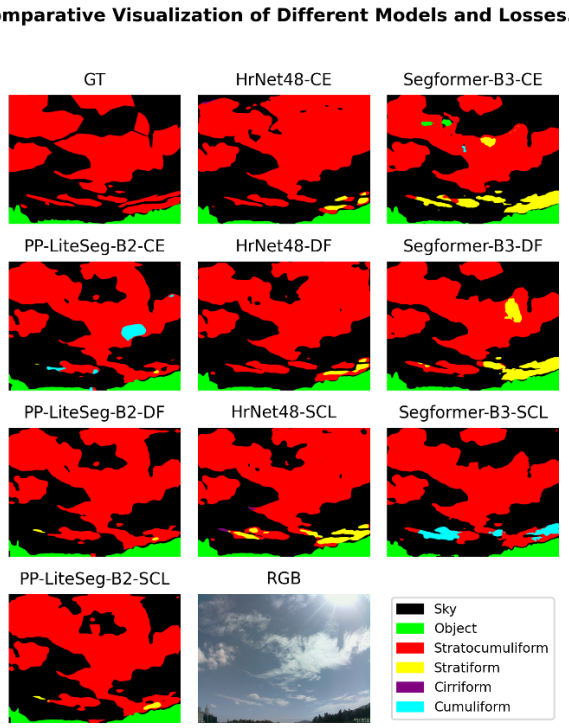

Figure 4: An example image with stratocumuliform clouds segmented by all networks with all loss functions. In this case, PP-LiteSeg and HRNet produced results with less leakage compared to SegFormerB3.

4.1 Post-Processing Results

Table 3: Comparison of dice coefficients for different cloud types before and after applying post-processing methods. Each entry in the table represents the average dice coefficient over the test dataset for a given cloud type, model, and post-processing method.

Experiment	Sky	Object	Cirriform	Cumuliform	Stratiform	Stratocumuli
pplite-scl	0.878	0.975	0.502	0.472	0.424	0.768
pplite-rf	0.878	0.975	0.413	0.468	0.453	0.778
pplite-vt	0.878	0.975	0.412	0.527	0.418	0.777
pplite-xgb	0.878	0.975	0.458	0.452	0.447	0.786
segformerb3-CE	0.897	0.977	**0.578**	0.552	0.49	**0.812**
segformerb3-rf	0.897	0.977	0.542	0.556	0.475	0.801
segformerb3-xgb	0.892	0.974	0.572	**0.583**	**0.511**	0.810
segformerb3-vt	0.897	0.977	0.521	0.572	0.483	0.808

In our comparative analysis of post-processing methods, we selected two segmentation models based on their performance in prior experiments: PP-LiteSeg with SCL and SegFormerB3 with Cross-Entropy (CE) loss. The motivation behind this selection was to evaluate the extremes, choosing the model with the overall best metric (SegFormerB3-CE) and the one with the worst (PP-LiteSeg-SCL).

Table 3 presents the results of applying various post-processing methods to the outputs of these models.

During our evaluations, it became evident that SegFormerB3 with CE loss often outperformed PP-LiteSeg with SCL, especially for the Cirriform cloud type. Applying the 'voting' leakage correction to SegFormerB3 did not drastically surpass the baseline in most categories, suggesting that simple majority-based corrections may not adequately address the complexities inherent in segmentation tasks. However, SegFormerB3 with this correction still outperformed PP-LiteSeg, implying inherent advantages of the SegFormerB3 model for specific cloud types.

We further investigated the use of machine learning methods for fixing segment leakage issues, focusing on random forest and XGBoost. SegFormerB3, when coupled with XGBoost, demonstrated notable improvements for the Cumuliform and Stratiform categories, with Dice coefficients increasing from 0.552 to 0.583 and from 0.490 to 0.511, respectively. This underscores the value of leveraging region-specific features in leakage correction. In contrast, the random forest method presented varied outcomes, excelling with Stratocumuli but falling behind with Cirriform clouds, emphasizing the importance of meticulous feature selection and optimization when applying machine learning corrections. (Martins et al., 2023)

An interesting observation was that even when the post-processing methods improved the scores for some classes, they sometimes lowered the overall image classification score, especially for Cirriform clouds. This drop can be attributed to our algorithm's reliance on OpenCV's connectivity feature. In certain prediction scenarios, the network identifies clouds as interconnected, which is common with the Cirriform class since they appear frequently and often look fragmented in images. When one cloud is identified and it touches another, our algorithm ends up changing the classification of that large cluster of clouds into a single segmentation, whereas ideally, they should be multiple distinct segments. As a result, the algorithm labels them under a single class. We are actively exploring ways to address this issue, such as experimenting with erosion and dilation techniques to better differentiate cloud classifications.

Comparing our main experiments, PP-LiteSeg-SCL and SegFormerB3-CE, with the leakage-corrected models highlights the importance of choosing the right model combination, loss function, and correction technique. Figure 5 shows that our algorithm enhances the visual consistency of image segmentation, which aids cloud ceiling observations. The steady improvements we observed with XGBoost demonstrate its effectiveness and provide valuable insights for future work on fixing segmentation leakage.

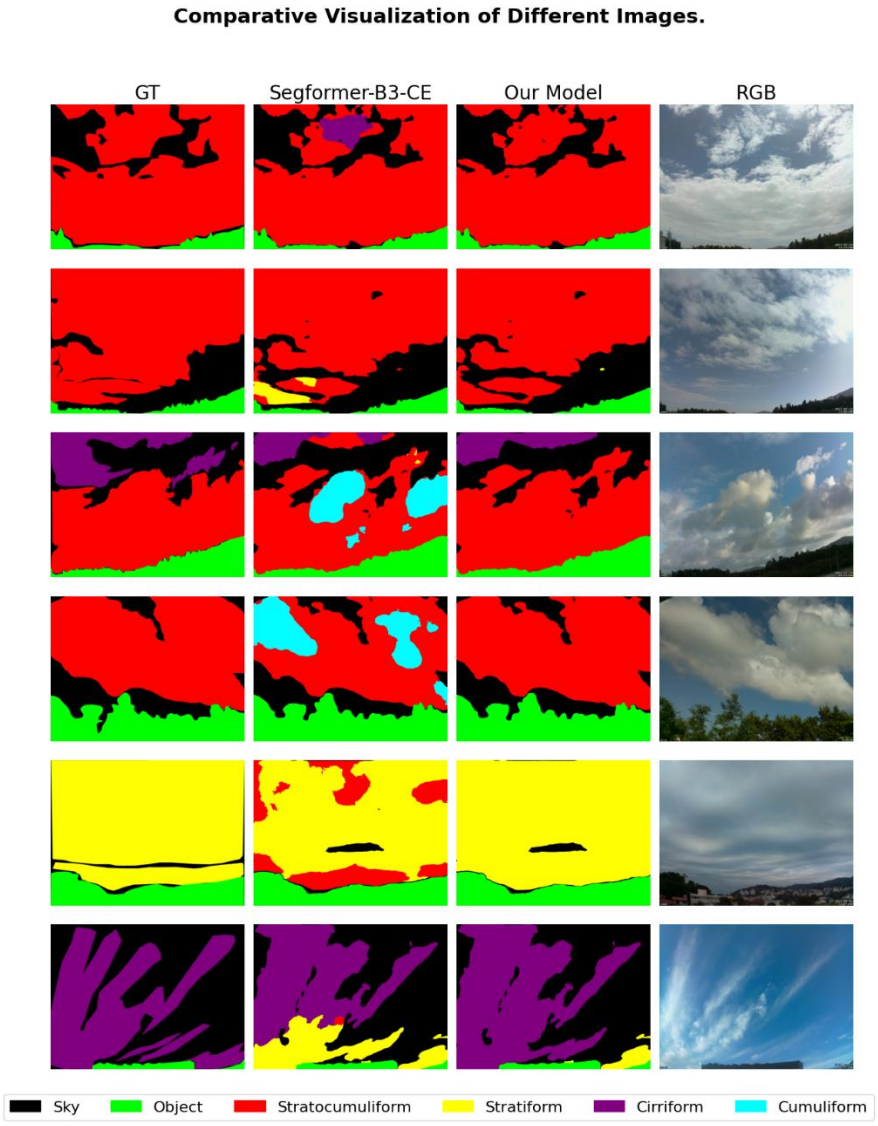

Figure 5: Comparison between ground truth, segformer B3 with CrossEntropy prediction and the result using our algorithm with xgboost

Conclusions

We evaluated the performance of SegFormerB3 with CE loss under three scenarios: 1) without any post-processing, 2) with only voting-based correction, and 3) with XGBoost-based correction. The results showed that while voting-based correction provided some improvements, the XGBoost-based approach consistently outperformed the others, especially for the Cumuliform and Stratiform classes. This confirms that using machine learning models to learn from region-specific features is a more effective strategy for fixing segmentation leaks compared to simple majority voting.

However, our current approach has limitations. The reliance on connected components can sometimes lead to over-grouping of fragmented clouds, as observed with the Cirriform class. This suggests that more advanced techniques for separating touching or overlapping clouds could be beneficial. Additionally, the computational overhead introduced by the post-processing steps needs to be carefully considered for real-time applications. While the improved accuracy is valuable, it comes at the cost of increased computation time. Future work could explore more efficient post-processing methods or ways to integrate the leakage correction into the main segmentation model itself.

Despite these challenges, our results demonstrate the potential of combining semantic segmentation with machine learning-based post-processing for improved cloud image segmentation. By leveraging the strengths of models like SegFormerB3 and XGBoost, we can obtain more accurate and consistent results, even for complex cloud scenes. This has important implications for downstream applications such as solar energy forecasting, where precise cloud segmentation is crucial for predicting irradiance levels and optimizing energy generation.

Furthermore, the ability to segment and classify clouds with higher accuracy can contribute to climate research and weather forecasting. By providing more detailed and reliable data on cloud distributions and their evolution over time, our approach can help validate and improve climate models, leading to better predictions of weather patterns and long-term climate trends.

References

Arrais, J.M., 2023. Clouds-1500. https://doi.org/10.17632/2KHCHJBGZR.2

Bradski, G., 2000. *The OpenCV Library*. Dr. Dobb's Journal of Software Tools.

Breiman, L., 2001. *Random Forests*. Machine Learning. 45, pp.5–32.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp.785–794.

Chu, L., Liu, Y., Wu, Z., Tang, S., Chen, G., Hao, Y., Peng, J., Yu, Z., Chen, Z., Lai, B., Xiong, H., 2021. PP-HumanSeg: Connectivity-Aware Portrait Segmentation with a Large-Scale Teleconferencing Video Dataset. https://doi.org/10.48550/ARXIV.2112.07146

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., others, 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, in: Proceedings of the International Conference on Learning Representations (ICLR).

Fabel, Y., Nouri, B., Wilbert, S., Blum, N., Triebel, R., Hasenbalg, M., Kuhn, P., Zarzalejo, L.F., Pitz-Paal, R., 2022. *Applying self-supervised learning for semantic cloud segmentation of all-sky images.* Atmospheric Measurement Techniques. 15, pp.797–809. https://doi.org/10.5194/amt-15-797-2022

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning, Adaptive computation and machine learning. MIT Press.

Juncklaus Martins, B., Cerentini, A., Neto, S.M., von Wangenheim, A., 2022a. Systematic Review of Nowcasting Approaches for Solar Energy Production based upon Ground-Based Cloud Imaging. Solar Energy Advances.

Juncklaus Martins, B., Polli, M., Cerentini, A., Mantelli, S., Chaves, T., Moreira Branco, N., von Wangenheim, A., Arrais, J., 2022b. Clouds-1000. https://doi.org/10.17632/4pw8vfsnpx.1

Kumar, S., Patkar, P., 2022. Low level wind shear over Bomba airport. MAUSAM.

Li, H., Wang, S., Zuo, W., Zhang, L., 2020. PPLite: Efficient Convolutional Neural Networks with Dynamic Pointwise Filters. arXiv preprint arXiv:2003.11506.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal Loss for Dense Object Detection, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988.

Luo, W., Zhao, H., Li, L., Wang, C., 2022. PP-LiteSeg: Lightweight Model for Real-Time Semantic Segmentation. arXiv preprint arXiv:2201.00239.

Mantelli, S.L., von Wangenheim, A., Pereira, E.B., Sobieranki, A.C., 2020. Hierarchical color similarity metrics for step-wise application on sky monitoring surface cameras. Earth and Space Science Open Archive 25. https://doi.org/10.1002/essoar.10503135.1

Martins, B.J., Arrais, J.M., Cerentini, A., Wangenheim, A. von, Neto, G.P.R., Mantelli, S., 2023. *Segmentation and Classification of Individual Clouds in Images Captured with Horizon-Aimed Cameras for Nowcasting of Solar Irradiance Absorption*. American Journal of Climate Change. 12, pp.628–654. https://doi.org/10.4236/ajcc.2023.124027

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, in: 3D Vision (3DV), 2016 Fourth International Conference On. IEEE.

Rahman, S., Wang, Y., 2016. Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation, in: International Symposium on Visual Computing. Springer, pp.234–244.

Rousson, M., Lenglet, C., Deriche, R., 2008. *The Dice Metric Considerations*. Insight Journal 2008, pp.1–10.

Song, Q., Cui, Z., Liu, P., 2020. *An Efficient Solution for Semantic Segmentation of Three Ground-based Cloud Datasets*. Earth and Space Science. 7, e2019EA001040. https://doi.org/10.1029/2019EA001040

Veremev, N., 2021. The use of artificial neural networks in the problem of classifying cloud types in wide-angle images of the visible hemisphere of the sky.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., others, 2020. Deep High-Resolution Representation Learning for Visual Recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Wangenheim, A. v, Bertoldi, R.F., Abdala, D.D., Richter, M.M., 2007. *Color image segmentation guided by a color gradient network*. Pattern Recognition Letters. 28, pp.1795–1803. https://doi.org/10.1016/j.patrec.2007.05.009

Ye, L., Wang, Y., Cao, Z., Yang, Z., Min, H., 2022. *A Self Training Mechanism with Scanty and Incompletely Annotated Samples for Learning-Based Cloud Detection in Whole Sky Images*. Earth and Space Science. 9. https://doi.org/10.1029/2022ea002220

Zhu, E., Zheng, K., Gao, K., Zhang, J., Yang, Z., Wang, Z., 2021. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. arXiv preprint arXiv:2105.15203.