EFFECT OF DATA QUALITY ON WATER BODY SEGMENTATION WITH DEEPLABV3+ ALGORITHM

Anirudh Edpuganti[1], Pillalamarri Akshaya[1], Jangala Gouthami[1], Sajith Variyar V.V[1] *, Sowmya V[1], Ramesh Sivanpillai[2]

[1]Amrita School of Artificial Intelligence, Coimbatore, Amrita Vishwa Vidyapeetham, India. (aniedpuganti@gmail.com, ORCID: 0009-0007-4926-214X; akshayapillalamarri310321@gmail.com, ORCID: 0009-0000-2157-9988; jangalagouthami2003@gmail.com, ORCID: 0009-0001-7900-755X; vv_sajithvariyar@cb.amrita.edu, ORCID: 0000-0003-3944-8155; v_sowmya@cb.amrita.edu, ORCID: 0000-0003-3745-6944) [2]Wyoming GIS Center, School of Computing, University of Wyoming, Laramie, WY, 82071, USA. (sivan@uwyo.edu, ORCID: 0000-0003-3547-9464)

KEY WORDS: Data Quality, Image Segmentation, Deep Learning, DeepLab, Satellite Images, Water Body.

ABSTRACT:

Training Deep Learning (DL) algorithms for segmenting features require hundreds to thousands of input data and corresponding labels. Generating thousands of input images and labels requires considerable resources and time. Hence, it is common practice to use opensource imagery data and labels available online. Most of these open-source data have little or no metadata describing their quality or suitability making it problematic for training or evaluating DL models. This study evaluated the effect of data quality on training DeepLabV3+, using Sentinel 2 A/B RGB images and labels obtained from Kaggle. We generated subsets of 256 x 256 pixels, and 10% of these images (802) were set aside for testing. First, we trained and validated the DeepLabV3+ model with the remaining images. Second, we removed images with incorrect labels and trained another DeepLabV3+ network. Finally, we trained the third DeepLabV3+ network after removing images with turbid water or with floating vegetation. All three trained models were evaluated with test images and then we calculated accuracy metrics. As the quality of the input images improved, accuracy of the predicted masks generated from the first model increased from 92.8% to 94.3% in the second model. The third model's accuracy was 96.4%, demonstrating the network's ability to better learn and predict water bodies when the input data had fewer class variations. Based on the results we recommend assessing the quality of open-source data for incorrect labels and variations in the target class prior to training DeepLabV3+ or any other DL network.

1. INTRODUCTION

Image segmentation, particularly semantic image segmentation, is a crucial task in computer vision. This process involves labelling each pixel in an image with a class, thereby providing a detailed understanding of the image at a granular level. Semantic image segmentation has broad applications ranging from autonomous driving to remote sensing (Subramanian et al., 2022) (Ramiya et al., 2016), medical image analysis, and robotics (Badrinarayanan et al., 2017) (Long et al., 2015) (Zhao et al., 2017).

One of the state-of-the-art architectures that have significantly impacted the field of semantic image segmentation is DeepLabV3+ (Chen et al., 2018) (Sunandini et al., 2023). This architecture combines the strengths of atrous convolutions, spatial pyramid pooling modules, and an encoder-decoder structure, thereby enhancing boundary detection and dealing with objects of different scales effectively (Yang et al., 2018) (George et al., 2023).

However, the performance of DeepLabV3+ like any other deep learning model depends on the quality of the training data. The more accurate and diverse the training data, the better the model's performance in segmenting images (George et al., 2023). This relationship is particularly noticeable in tasks such as segmenting water bodies from satellite images, which form the focus of this study (Harika et al., 2022).

Our objective is to delve deeper into this dependency on data quality. We investigate the influence of the quality of data while training the DeepLabV3+ algorithm for segmenting water bodies from satellite images. We scrutinize scenarios where there is a mismatch in the mask of the training data or when the training data consists of mixed water quality, such as clear and turbid water or water with floating vegetation. It is essential to understand whether excluding such cases would improve or degrade the model's performance (Jean et al., 2019) (Harika et al., 2022).

There is a general belief that cleaner and more consistent training data leads to better performance in machine learning models. However, when dealing with real-world scenarios, such as satellite image segmentation, data inconsistencies are often the norm rather than the exception. Hence, it is important to understand the behaviour of models like DeepLabV3+ under these circumstances, as they reflect realistic conditions that these models would encounter in actual deployments (Volpi and Tuia, 2016).

In the following sections, we present our findings, shedding light on the influence of data quality on the performance of DeepLabV3+ in the task of segmenting water bodies from satellite images.

* Corresponding author

2. MATERIALS AND METHODS

2.1 Dataset

This study utilizes an open-source Kaggle dataset that contains satellite images and masks captured by the Sentinel-2A and Sentinel2B satellite. The images and the mask in the dataset have three bands of image: red (R), green (G), and blue (B), allowing to capture both the spectral and spatial information essential for analysis. Figure 1 represents the RGB image and the corresponding water class and no water class mask.

Figure 1. Sample RGB (left) image and the corresponding mask with water class shown in white pixels (right).

2.2 Data Processing

Dataset consists of images and masks of varying sizes, to maintain uniformity and avoid bias in training the model, various sizes were resized to 256x256 pixels. This resizing ensures consistency in the data, allowing for efficient processing and analysis. The masks are recoded, where background pixels were labelled as 0, representing areas without water, while water pixels were labelled as 1, indicating the presence of water.

The dataset is split into three sets after being resized and recoded as training, testing, and validation, in a ratio of 80:10:10. The training set, includes 80% of the data, which is used to train the model. The testing set, includes 10% of the data, is used to evaluate how well the model performs on new, unseen images.

The validation set, also includes 10% of the data, is used to measure model's performance after each iteration and is helpful in adjusting the hyperparameters based on the feedback from the validation set. Splitting the dataset into separate sets allows for thorough model development and ensures that the model is tested on unbiased data to assess its ability to classify water effectively.

2.3 Data Refining

In the process of data refining, the objective was to enhance the performance of the model. To achieve this, three separate experiments are conducted, where each experiment involves training a new instance of DeeplabV3+.

These experiments aim to refine the data by incorporating various techniques such as removing images with mismatched masks and floating vegetation images. Table 1 contains information about the number of images used in the experiments.

	Training	Testing	Validation
Experiment - 1	6422	802	802
Experiment - 2	4841	593	611
Experiment - 3	2916	364	353

Table 1. Image count for each experiment after refining

2.3.1 Experiment 1: This dataset included all images with correct and incorrect masks and those with mixed water quality. Only the pre-processing described earlier as resizing to 256x256 was performed in this experiment and no refinement of data was made.

2.3.2 Experiment 2: From the original dataset, images with incorrect masks were removed to ensure that network learns from correct masks. After removing incorrect masks, the model with accurate training data will be helpful for giving better segmentation results. In Figure 2, it can be inferred that there is land in image, but NDWI (Normalized Difference Water Index) predicted land as water in mask.

Figure 2. Sample RGB (left) image and the incorrect mask (right) where land was identified as water.

2.3.3 Experiment 3: For the final experiment, images containing mixed water quality such as turbid water images and floating vegetation images were removed. This also resulted in addressing class imbalance. The dataset originally contained fewer samples for turbid and floating vegetation compared to images depicting good quality water. By removing these fewer samples, the dataset was refined to enhance the model's performance. Figure 3 and Figure 4 turbid water image and floating vegetation image with respective masks.

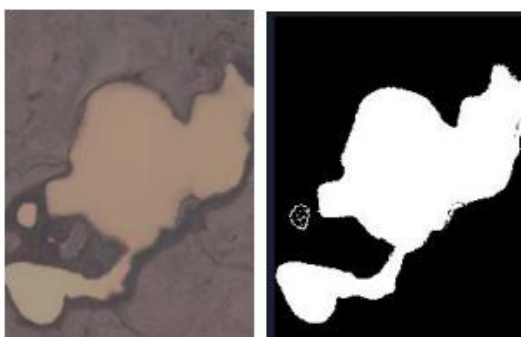

Figure 3. Sample RGB (left) image and corresponding mask (right) of a turbid water body.

Figure 4. Sample RGB (left) image and corresponding mask (right) of water body with floating vegetation.

2.4 Selection of DeepLabV3+ Architecture for Water Body Segmentation

DeepLabv3+ has distinguished itself as a potent tool for semantic image segmentation, a technique that assigns every pixel of an image to a specific class label, facilitating intricate analysis of images. The model's application ranges from augmenting the perception capability of autonomous vehicles to improving the precision of medical diagnoses. The choice of DeepLabv3+ for this study lies in its exceptional architecture that accommodates various scales and contexts within images and its proficiency in capturing details regardless of scale, a characteristic that aligns with the properties of our dataset.

The architecture of DeepLabv3+ incorporates an advanced Atrous Spatial Pyramid Pooling (ASPP) module and an encoder-decoder structure. This combination aids in maximizing the contextual information derived from the images while ensuring that fine details are not lost. The foundation of this architecture is the Exception model, a high-performing convolutional neural network used to extract preliminary features from input images.

The ASPP module, the core of the DeepLabv3+ architecture, utilizes filters at diverse scales concurrently. This technique enables the model to capture contextual information from different sized areas in the image, an indispensable feature for our dataset, given the varying sizes of water bodies. Once this multi-scale feature extraction is complete, these features become inputs for the decoder.

The decoder's role in the DeepLabv3+ architecture is to refine the segmentation results using the extracted features. It up samples these feature maps to a higher resolution and combines them with earlier-stage feature maps from the network. This blend of high-level contextual information and detailed spatial information allows the decoder to generate precise and granular segmentation maps.

The output from DeepLabv3+ is a categorically labelled image, noted for its crisp object boundaries and comprehensive contextual understanding. The incorporation of state-of-the-art techniques such as depth-wise separable convolutions enhances its computational efficiency. As a result of this, and by leveraging the capabilities of the ASPP module and the encoder-decoder structure, DeepLabv3+ serves as an ideal choice for our study, providing a perfect blend of efficiency and accuracy for the task of water body segmentation.

2.5 Accuracy and Loss plots

Accuracy and loss plots provide valuable insights into the learning process of a model during training. As training progresses, the model adjusts its parameters to improve accuracy and reduce loss. With more epochs, accuracy improves as the model learns the data's patterns. Ideally, validation accuracy converges with training accuracy, indicating effective generalization. Loss decreases as the model's predictions align better with true values. Eventually, the reduction in loss flattens as the model captures most relevant patterns and further adjustments have diminishing returns.

2.6 Metrics

Relying solely on accuracy as the sole evaluation metric may lead to an incomplete assessment of model performance, particularly in scenarios involving imbalanced datasets. While accuracy considers both classes, precision, recall, and F1 score focus specifically on the target class.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision quantifies the proportion of correctly predicted positive instances relative to the total predicted positives, whereas recall measures the proportion of true positive (TP) instances identified by the model.

$$\text{Precision}=\frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad (3)$$

F1 score harmonizes precision and recall into a single metric, providing a balanced assessment that accounts for both false positives (FP) and false negatives (FN).

$$\text{F1-Score}=\frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

By incorporating precision, recall, and F1 score alongside accuracy, a more comprehensive and nuanced understanding of the model's performance within the target class can be attained, enabling more informed decision-making.

3. RESULTS

3.1 Experiment 1:

The accuracy plot obtained when DeepLabV3+ was trained with all images (n = 6422), shows a gradual increase in the accuracy around 90% (Figure 5). However, the loss values shown in Figure 6 remain relatively high, indicating that the model faced some challenges in reducing its errors. The metrics obtained for this experiment are summarized in Table 2.

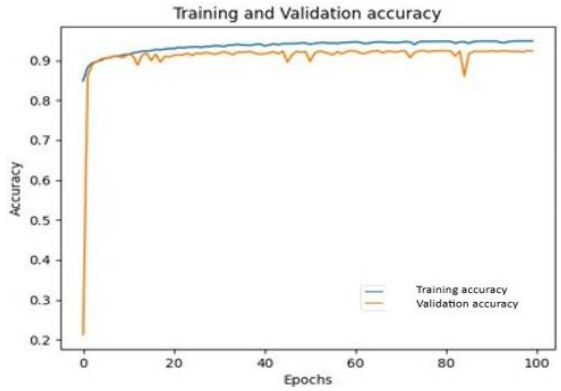

Figure 5. Accuracy plot when DeepLabV3+ network was trained with all RGB images (n = 6422).

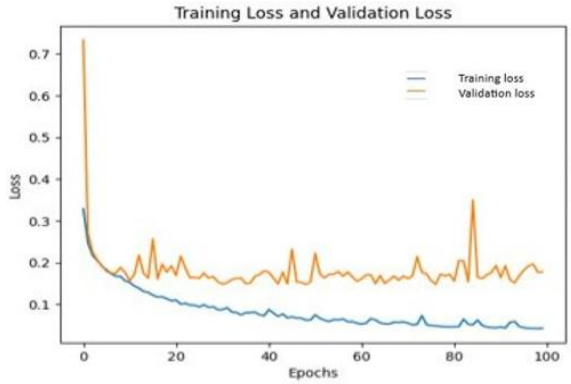

Figure 6. Loss plot when DeepLabV3+ network was trained with all RGB images (n = 6422).

3.2 Experiment 2:

The accuracy plot obtained when DeepLabV3+ was trained after removing RGB images with incorrect masks (n = 1581), shows an improvement in the model accuracy (Figure 7). The accuracy value steadily improved and reached approximately 95%.

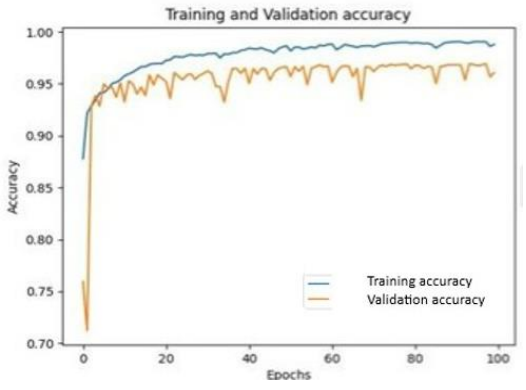

Figure 7. Accuracy plot when DeepLabV3+ network was trained with RGB images with correct masks (n = 4841).

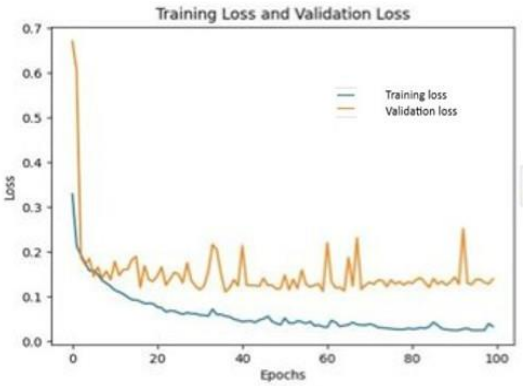

Figure 8. Loss plot when DeepLabV3+ network was trained with RGB images with correct masks (n = 4841).

However, the loss values remain relatively higher (Figure 8), indicating that the model continues to encounter challenges in minimizing its errors. In terms of evaluation metrics, a marginal change is observed in the results.

3.3 Experiment 3:

When DeepLabV3+ was trained with images of clear water bodies and correct masks, the accuracy and loss plots demonstrate remarkable improvement. Despite using fewer images, the overall accuracy surpassed 95% (Figure 9). This demonstrated that the model's ability to generalize improved when there was less variation in the training data.

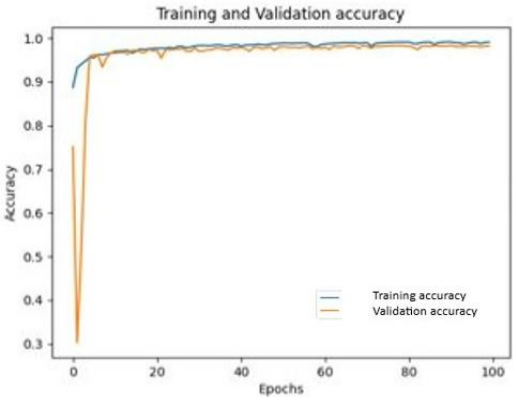

Figure 9. Accuracy plot when DeepLabV3+ was trained with RGB images of clear water bodies (n = 2916).

Moreover, the convergence of accuracy and loss closely matches the training accuracy, showcasing the model's robustness. This indicates enhanced learning capabilities compared to the previous experiments, evident in the gradual decrease in loss over time (Figure 10).

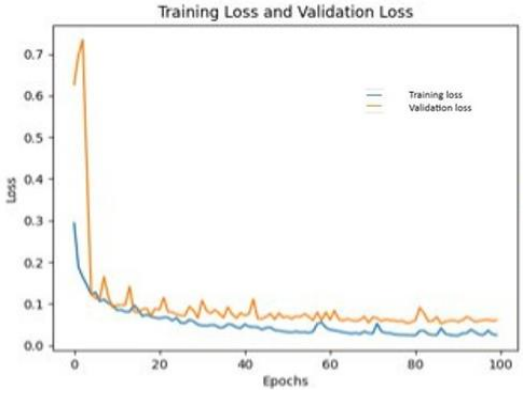

Figure 10. Loss plot when DeepLabV3+ was trained with RGB images of clear water bodies (n = 2916).

Precision score increased to 0.9792 from 0.4827 when RGB images of turbid water and floating vegetation were removed (Table 2). Recall and F1 scores also increased in the 3rd experiment. Metrics obtained for all three experiments are summarized in Table 2.

	Precision	Recall	F1-Score	Accuracy
Experiment - 1	0.4692	0.3362	0.361	0.928
Experiment - 2	0.4827	0.3555	0.3900	0.943
Experiment - 3	0.9792	0.7511	0.8205	0.963

Table 2. Evaluation results for experiments

The enhanced performance of the model underscores the importance of dataset cleaning and its influence on training deep learning models for image segmentation tasks.

4. CONCLUSION

In this study, we investigated the influence of data refinement on model predictions. Our findings underscore the importance of data quality in improving model performance. Prior to refining the data, the model achieved a precision of 0.4692. However, after implementing data refinement techniques, including the removal of mismatched masks and mixed water quality images, notable improvements in precision were observed. The precision increased to 0.4827 after removing mismatched masks and further improved to an impressive 0.9792 after excluding mixed water quality images. These results highlight the significance of addressing data quality issues to enhance model accuracy and reliability. By prioritizing data refinement, researchers and practitioners can optimize model performance and minimize potential errors. Overall, this study underscores the crucial relationship between data quality and model predictions, emphasizing the need for meticulous data refinement to achieve more accurate and reliable results.

5. REFERENCES

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).

George, B., Sajith, V. V., Sowmya, V., & Sivanpillai, R. (2023, March). Performance Improvement of Water Body Segmentation by DeeplabV3+ Using Two-Dimensional Variational Mode Decomposition. In 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 603-608). IEEE.

Harika, A., Sivanpillai, R., Sajith Variyar, V. V., & Sowmya, V. (2022). Extracting Water Bodies in Rgb Images Using Deeplabv3+ Algorithm. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 46, 97-101.

Jean, N., Wang, S., Samar, A., Azzari, G., Lobell, D., & Ermon, S. (2019, July). Tile2vec: Unsupervised representation learning for spatially distributed data. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 3967-3974).

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

Ramiya, A.M., Nidamanuri, R.R. and Krishnan, R., 2016. Object-oriented semantic labelling of spectral–spatial LiDAR point cloud for urban land cover classification and buildings detection. Geocarto international, 31(2), pp.121-139.

Subramanian, J., Thangavel, S.K. and Caianiello, P., 2022. Segmentation of Streets and Buildings Using U-Net from Satellite Image. In Disruptive Technologies for Big Data and Cloud Applications: Proceedings of ICBDCC 2021 (pp. 859-868). Singapore: Springer Nature Singapore.

Sunandini, G., Sivanpillai, R., Sowmya, V., & Variyar, V. S. (2023, March). Significance of Atrous Spatial Pyramid Pooling (ASPP) in Deeplabv3+ for Water Body Segmentation. In 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 744-749). IEEE.

Volpi, M., & Tuia, D. (2016). Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, 55(2), 881-893.

Yang, G., Li, G., Pan, T., Kong, Y., Wu, J., Shu, H., ... & Zhu, X. (2018, August). Automatic segmentation of kidney and renal tumor in ct images based on 3d fully convolutional neural network with pyramid pooling module. In 2018 24th International Conference on Pattern Recognition (ICPR) (pp. 3790-3795). IEEE.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2881-2890).