Evaluation of Deep Learning Transfer Techniques for Mangrove Segmentation with Images of the Sentinel-2A

Amanda Cristina Fraga de Albuquerque*, Helyane Bronoski Borges†
Federal University of Technology of Paraná, Ponta Grossa-PR, 84017-220, Brazil
Email: *amandaalbuquerque@alunos.utfpr.edu.br,†helyane@utfpr.edu.br

***Abstract*—Fine-tuning techniques allow the use of weights from pre-trained networks in other models across different contexts, potentially improving training performance as it generally requires fewer computational resources and less data. Fine-tuning has become more widespread in the natural domain (RGB) with the availability of pre-trained model weights from the ImageNet database. However, pre-trained models in the same domain are not readily available for the remote sensing domain, such as in mangrove identification. Both nationally and in the state of Paraná, there are few studies employing deep learning for mangrove segmentation. Developing models using deep learning transfer can help establish automated monitoring systems. Thus, this study evaluated fine-tuning techniques for mangrove segmentation in Paraná using the U-Net model with pre-trained encoders in the same domain, remote sensing, and the natural domain. The dataset for training the U-Net was generated using bands from the Sentinel-2A satellite and annotations from the MapBiomas project maps. The fine-tuned networks discussed in this study accurately identified mangroves in Paraná, all achieving accuracies above 95.1% and F-scores greater than 92.6%.**

I. Introduction

Mangroves are coastal ecosystems that serve as a transition zone between terrestrial and marine environments, influenced by tidal regimes and associated with a characteristic vegetation known as mangrove [1]. In Paraná, due to the proximity of urban centers and port activities, mangroves have been and continue to be degraded by deforestation, pollution from effluents and waste disposal [2].

Both nationally and within Paraná, various studies have used satellite images to monitor mangroves. However, more systematic and continuous monitoring has only recently been established through the MapBiomas project [3]. Using 30m resolution Landsat images, the project has identified mangroves across Brazil from 1985 to 2022 [4]. In its beta stage, the project identified mangroves from 2016 to 2022 using higher resolution (10m) Sentinel-2A images.

Few works have been developed with deep learning applied to mangrove segmentation, with more common machine learning methods in various countries being Support Vector Machine (SVM) and Random Forest (RF) [5]. However, deep learning models using Convolutional Neural Networks (CNNs) have proven to be more accurate in segmenting multispectral satellite images compared to traditional machine learning models [6]. Among convolutional architectures, one commonly used for mangrove segmentation is U-Net, specifically developed for image segmentation. It is suitable for land use classification due to its ability to maintain spatial relationships [7].

Since convolutional models generally require a significant amount of annotated data, it is necessary to explore approaches that overcome this limitation, such as deep learning transfer techniques. In deep learning transfer, pre-trained models on different but well-established datasets are used to reuse the weights of the initial layers in subsequent training with the available data [8].

In deep learning applications within the domain of natural images using RGB (Red, Green, Blue) channels, it is common practice to leverage publicly available weights from pre-trained models on the ImageNet dataset. However, pre-trained models are not publicly available in research involving multispectral data, such as satellite images. For this reason, some mangrove segmentation studies are limited to transferring knowledge only from the RGB domain to the multispectral domain [9]–[11], or from the multispectral domain using your own data [12]–[14].

Considering the potential advantages of fine-tuning with datasets in the remote sensing domain, such as better performance and lower demand for resources, this study evaluated fine-tuning for U-Net with pre-training of the encoder using a benchmark remote sensing dataset, EuroSAT [15]. Furthermore, aiming to develop mangrove identification techniques that facilitate automatic and continuous monitoring, U-Net segmentation models were trained with different fine-tuning approaches. For training the U-Net, a Mangrove dataset was created using Sentinel-2A data and annotations from MapBiomas classes. The models were evaluated for mangroves in Paraná, but the same fine-tuning methodologies have the potential to be extended to other regions in Brazil.

II. Materials and Methods

This study evaluated fine-tuning techniques for the U-Net model in mangrove segmentation. The procedures depicted in Fig. 1 were necessary to conduct the experiments. The Mangrove dataset, consisting of Sentinel-2A satellite images, was generated to train the U-Net. Additionally, to define the experiments, it was necessary to configure the types of encoder weight adjustments for the U-Net.

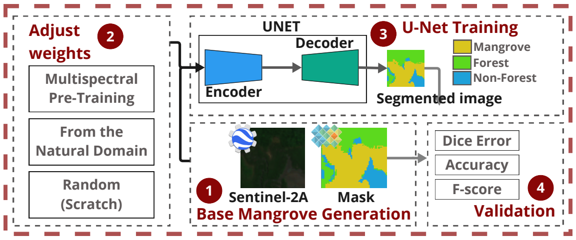

Fig. 1. Flowchart with the main procedures of this work.

A. Mangrove Base Generation

During the segmentation model training, two inputs are provided: a multispectral image from Sentinel-2A and its corresponding ground truth mask with segmentation classes. The Sentinel-2A images are publicly available resources from the Google Earth Engine (GEE) platform [16]. This study extracted images at the Level-2A (orthoimage Bottom-Of-Atmosphere - BOA), including reflectance correction. Since bands 1, 9, and 10 of Sentinel-2A are mainly used for atmospheric correction and have low relevance to mangrove classification [13], these bands were not considered.

In this study, image patches were extracted from the GEE platform covering the coastal regions of Paraná where mangroves are present, corresponding to the year of 2022. For the same image patches from Sentinel-2A in the area of interest, masks with MapBiomas classes [4] were also obtained from the GEE platform. MapBiomas land use maps are produced annually, and this study used 2022 data based on Sentinel-2A images. Considering that the objective of this study is mangrove identification, the MapBiomas classes were adapted and grouped into only three categories, as represented in Fig. 1: Forest; Mangrove; Non-Forest.

Patches measuring 64x64 were generated for both the cutouts of the Sentinel images and the respective masks. To balance the class distribution in the patches, only those containing at least 2% of the mangrove class were selected. The generated Mangrove dataset contains 2,030 pairs of multispectral images and their corresponding masks, and can be accessed via this link[1].

B. U-Net Segmentation Model Architecture

The U-Net network used in this study for mangrove segmentation is a convolutional model with an encoder-decoder architecture [17]. The encoder, through convolution and pooling operations, extracts features from the images and reduces dimensionality. Meanwhile, the decoder restores the image dimensions using upsampling while amplifying relevant low-level features.

As the encoder architecture of the U-Net, ResNet-50, a residual network, was used in this study. This structure's main difference from other convolutional networks is the presence of skip connections (shortcuts) between layers [18].

[1]Mangrove Paraná dataset: https://github.com/Amanda-Cristina/mangrove-segmentation

C. Definition of Experiments with and without Fine-Tuning for U-Net

Among the fine-tuning techniques for U-Net, the main focus of this study was to evaluate transfer learning between multispectral domains (MS→MS) using a pre-trained encoder with a benchmark dataset in remote sensing. Thus, it was necessary to pre-train the ResNet-50 encoder with the EuroSat dataset, a procedure detailed in Section II-D.

As indicated in Table I, transfer between different domains (RGB→MS) was also evaluated. For the transfer from the natural image domain (RGB), we utilized the weights of the encoder model pre-trained with the ImageNet dataset. These weights were obtained directly from the U-Net training framework presented in Section II-E.

For each domain transfer variation, a comparison was made between two fine-tuning techniques: the first involves keeping all weights of the pre-trained encoder unchanged during U-Net training, referred to as the "Freeze" method, and the second consists of retraining all weights, known as the "UnFreeze" method. Within the five experiments, Table I, training the U-Net without fine-tuning was also considered, where the encoder weights were randomly initialized (Scratch).

TABLE I
EXPERIMENT SETTINGS WITH AND WITHOUT FINE-TUNING FOR U-NET IN SEGMENTATION.

	EuroSAT UnFreeze	EuroSAT Freeze	ImageNet UnFreeze	ImageNet Freeze	Scratch
Encoder weights	Pre-training EuroSAT		Pre-training ImageNet		Random (Scratch)
Origin domain	Multispectral		RGB		———
Adjust weights	UnFreeze	Freeze	UnFreeze	Freeze	UnFreeze

D. U-Net Encoder Pre-Training

In the experiments involving fine-tuning in the multispectral domain, the ResNet-50 encoder was pre-trained using the EuroSAT dataset. The EuroSAT dataset is a reference for land use and land cover classification released in 2019 [15], accessible via Zenodo [19]. It comprises multispectral images from Sentinel-2A, with 10 of the 13 bands being used for training the ResNet, as indicated in Section II-A. In total, there are 27,000 annotated images across ten land use categories. For pre-training ResNet, the EuroSAT dataset was partitioned into training, validation, and test sets, with the following proportions, respectively: 60%, 20%, and 20%.

The ResNet architecture used in this study is described in [18], and its training was implemented based on PyTorch [20]. The training configurations were commonly used for EuroSAT classification with ResNet, in which cross entropy was used as the error function, and the Adam optimizer for the learning rate [21]. The initial learning rate was set to 0.0001, and the ReduceLROnPlateau algorithm was used to decrease this value during training. Data augmentation techniques such as horizontal flip, vertical flip, rotation, and resizing were applied on the training dataset. The model was trained for 100 epochs with a batch size of 50, within the limits in related works [21].

The F-score [12] metric was used for model validation, and both accuracy [12] and F-score metrics were considered during the testing stage. The weights of the trained and validated model, with performance levels from the literature, were saved and used in the transfer learning experiments for U-Net. They can be accessed via this link[2].

E. Training Segmentation Experiments with U-Net

For training the U-Net, the Mangrove dataset was partitioned into training, validation, and test sets with proportions of 60%, 20%, and 20%, respectively. The training was implemented using PyTorch [20] and Segmentation Models PyTorch [22]. Standard parameters from the literature were followed, including using the Dice loss function for weight adjustment during training [23]. The initial learning rate was set to 0.001, with the Adam optimizer and adjustment using the ReduceLROnPlateau algorithm. Data augmentation techniques such as horizontal flip, vertical flip, rotation, and resizing were also applied on the training dataset.

Each U-Net experiment was trained for 200 epochs with a batch size of 32, within the limits in related works [23], using an NVIDIA RTX 8000 graphics card with 48GB capacity and 32GB of RAM. The F-score [12] metric was used for model validation, and accuracy [12] was also considered during the testing stage. In the test, in addition to the average metrics considering all the three classes (Forest; Mangrove; Non-Forest), the metrics for each class were also calculated. Additionally, the training durations of each experiment were evaluated.

III. Results

In this section, the results of this study are presented. The first part, Section III-A, presents the results of the pre-training of the U-Net encoder in the multispectral domain. The last part, Section III-B, contains the results of the experiments with and without fine-tuning for U-Net in mangrove segmentation.

A. Pre-training the U-Net Encoder in the Multispectral Domain

In the pre-training of the ResNet-50 model using EuroSAT data and with 100 epochs, the average time for each epoch was 15.5 seconds, with a total duration of 25.8 minutes. The training stabilized between the 20th and 30th epochs, maintaining the lowest error (0.034) and the highest F-score (98.9%) after these epochs. In validation, the performance was slightly lower, with an F-score of 97.0% and an error of 0.097. In testing, both accuracy and F-score were 97.2%.

B. Experiments with and without Fine-Tuning for U-Net in Mangrove Segmentation

As indicated in the graphics (see Fig. 2), the training of all experiments for the U-Net model converged appropriately, stabilizing after a few epochs. The experiments were trained for 200 epochs, but all stabilized before the 30th epoch, with

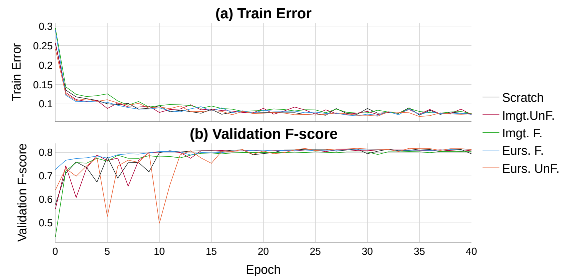

Fig. 2. Results of experiments with U-Net for mangrove segmentation: (a) Training error curves; (b) F-score curves in validation.

those using the "Freeze" technique converging more quickly, around the 10th epoch. From the training and validation metrics in Table II, it can be observed that all models in the training phase finished with an error of less than 0.075 and achieved accuracy greater than 95.4% in validation.

TABLE II
Results of experiments with and without fine-tuning segmentation in the training and validation phases.

	Error		F-Score		Accuracy		Epoch
	Train	Validation	Train	Validation	Train	Validation	(s)
Scratch	0.075	0.067	79.8%	**81.5%**	95.3%	**95.7%**	1.99
Imgt. F.	0.073	0.070	79.9%	80.6%	95.3%	95.4%	**1.96**
Imgt. UnF.	0.075	0.069	79.5%	81.3%	95.3%	95.6%	2.16
Eurst. F.	0.069	0.066	80.3%	81.4%	95.6%	**95.7%**	**1.97**
Eurst. UnF.	0.068	0.064	81.0%	**82.0%**	95.6%	**95.9%**	2.00

The "Freeze" experiments had shorter training times than the "UnFreeze" technique, but they exhibited lower F-score and accuracy. While the model with the highest F-score for ImageNet ("ImageNet UnFreeze") achieved an F-score of 81.3%, for the models using EuroSAT, the "EuroSAT UnFreeze" model obtained an F-score of 82.0% in validation. Thus, the EuroSAT model performed better and had a shorter training time when compared to these two experiments.

The error and accuracy results for the test, Table III, were similar to those found in training and validation. The F-score values in the test phase are higher than in the other two phases, because in this case it was not calculated as the average of the results for each image, but rather using the values from the confusion matrix considering all images.

TABLE III
Results of experiments with and without fine-tuning segmentation in the test phase.

	Error Overall	Accuracy Overall	F-score Overall	F-score Mangrove	F-score Forest	F-score Non-Forest
Scratch	0.069	95.6%	93.3%	94.3%	93.5%	92.1%
Imgt. F.	0.075	95.1%	92.6%	93.6%	93.1%	91.2%
Imgt. UnF.	0.071	95.4%	93.1%	94.1%	93.3%	91.8%
Eurst. F.	0.067	95.7%	93.5%	94.5%	93.6%	92.3%
Eurst. UnF.	0.065	95.8%	**93.6%**	**94.7%**	93.7%	92.5%

The fine-tuning techniques using ImageNet had the lowest overall F-scores in the test, even lower than training without fine-tuning ("Scratch"). Results with EuroSAT fine-tuning

[2]ResNet-50 model trained with Eurosat database: https://github.com/Amanda-Cristina/mangrove-segmentation

showed slightly better performance than training without fine-tuning (F-score = 93.3%), where the highest overall F-score of 93.6% was achieved with the "EuroSAT UnFreeze" experiment. The test F-score for mangrove identification was higher than the overall metric in multispectral experiments, with a value 94.7% for the "EuroSAT UnFreeze" experiment.

IV. Conclusion

This study evaluated fine-tuning techniques on the U-Net model for mangrove segmentation with the ResNet-50 encoder. The experiments considered the following variations: fine-tuning in the same domain (multispectral) versus fine-tuning across different domains (RGB to multispectral); fine-tuning with retraining of all encoder layers versus fine-tuning without retraining the encoder. All fine-tuning experiments showed accuracy exceeding 95.1%, comparable to similar works, and F-score greater than 92.6%.

The fine-tuning without retraining the initial layers converged more rapidly during training; however, they exhibited lower accuracy than the corresponding models with encoder weight retraining. Transfer learning with encoder weight retraining was more stable in the same domain. Even though the multispectral models were adapted to weights from ImageNet, natural domain, the fine-tuning results were inferior to the fine-tuning in the same domain and also to the non-fine-tuned multispectral experiment.

The model that best identified mangroves was the one fine-tuned in the same domain (multispectral) with encoder retraining, named "EuroSAT UnFreeze". For this result, both the "Forest" and "Non-Forest" classes had lower recognition compared to the "Mangrove" class. These results could be improved in future work with a more balanced class distribution in the dataset and the use of spectral indices as input data. Nevertheless, this model is already suitable for mangrove identification with an F-score of 94.7%.

Acknowledgment

The authors thank the Federal Technological University of Paraná, in particular the Information Technology Management Coordination for providing the infrastructure.

References

[1] ICMBio, "Atlas dos manguezais do brasil," 2018, last accessed: Feb. 2024. [Online]. Available: https://www.gov.br/icmbio/pt-br/centrais-de-conteudo/atlas-dos-manguezais-do-brasil-pdf

[2] LAGEAMB, "Saúde dos manguezais de paranaguá: um olhar para os bosques antropizados." 2021, last accessed: Feb. 2024. [Online]. Available: https://lageamb.ufpr.br

[3] C. Diniz, L. Cortinhas, G. Nerino, J. Rodrigues, L. Sadeck, M. Adami, and P. W. M. Souza-Filho, "Brazilian mangrove status: Three decades of satellite data analysis," *Remote Sensing*, vol. 11, no. 7, p. 808, 2019.

[4] C. M. Souza, J. Z. Shimbo, M. R. Rosa, L. L. Parente, A. A. Alencar, B. F. T. Rudorff, H. Hasenack, M. Matsumoto, L. G. Ferreira, P. W. M. Souza-Filho, S. W. de Oliveira, W. F. Rocha, A. V. Fonseca, C. B. Marques, C. G. Diniz, D. Costa, D. Monteiro, E. R. Rosa, E. Vélez-Martin, E. J. Weber, F. E. B. Lenti, F. F. Paternost, F. G. C. Pareyn, J. V. Siqueira, J. L. Viera, L. C. F. Neto, M. M. Saraiva, M. H. Sales, M. P. G. Salgado, R. Vasconcelos, S. Galano, V. V. Mesquita, and T. Azevedo, "Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine," *Remote Sensing*, vol. 12, no. 17, 2020.

[5] K. Maurya, S. Mahajan, and N. Chaube, "Remote sensing techniques: Mapping and monitoring of mangrove ecosystem—a review," *Complex & Intelligent Systems*, vol. 7, no. 6, pp. 2797–2818, 2021.

[6] Q. Li, F. K. K. Wong, and T. Fung, "Mapping multi-layered mangroves from multispectral, hyperspectral, and lidar data," *Remote Sensing of Environment*, vol. 258, p. 112403, 2021.

[7] T. L. Giang, K. B. Dang, Q. T. Le, V. G. Nguyen, S. S. Tong, and V.-M. Pham, "U-net convolutional networks for mining land cover classification based on high-resolution uav imagery," *IEEE Access*, vol. 8, pp. 186 257–186 273, 2020.

[8] K. Nogueira, O. A. Penatti, and J. A. Dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539–556, 2017.

[9] D. Lomeo and M. Singh, "Cloud-based monitoring and evaluation of the spatial-temporal distribution of southeast asia's mangroves using deep learning," *Remote Sensing*, vol. 14, no. 10, p. 2291, 2022.

[10] M. Guo, Z. Yu, Y. Xu, Y. Huang, and C. Li, "Me-net: a deep convolutional neural network for extracting mangrove using sentinel-2a data," *Remote Sensing*, vol. 13, no. 7, p. 1292, 2021.

[11] J. A. Arias-Aguilar, E. López-Jimenez, O. D. Ramírez-Cárdenas, J. C. Herrera-Lozada, and N. Hevia-Montiel, "Mangrove semantic segmentation on aerial images," *IEEE Latin America Transactions*, vol. 22, no. 5, pp. 379–386, 2024.

[12] T. V. Tran, R. Reef, X. Zhu, and A. Gunn, "Characterising the distribution of mangroves along the southern coast of vietnam using multi-spectral indices and a deep learning model," *Science of The Total Environment*, p. 171367, 2024.

[13] W. S. Maung and J. Sasaki, "Assessing the natural recovery of mangroves after human disturbance using neural network classification and sentinel-2 imagery in wunbaik mangrove forest, myanmar," *Remote Sensing*, vol. 13, no. 1, p. 52, 2020.

[14] Y. Li, B. Fu, X. Sun, D. Fan, Y. Wang, H. He, E. Gao, W. He, and Y. Yao, "Comparison of different transfer learning methods for classification of mangrove communities using mccunet and uav multispectral images," *Remote Sensing*, vol. 14, no. 21, p. 5533, 2022.

[15] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, 2019.

[16] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, 2017.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proceedings*, International Conference Medical image computing and computer-assisted intervention–MICCAI. Munich, Germany: Springer, 2015, pp. 234–241.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings*, IEEE conference on computer vision and pattern recognition. IEEE, 2016, pp. 770–778.

[19] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification," Mar. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7711810

[20] J. Ansel, E. Yang, H. He, N. Gimelshein, A. Jain, M. Voznesensky, B. Bao, P. Bell, D. Berard, E. Burovski *et al.*, "Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation," in *Proceedings*, vol. 2. ACM International Conference on Architectural Support for Programming Languages and Operating Systems., 2024, pp. 929–947.

[21] H. Dastour and Q. K. Hassan, "A comparison of deep transfer learning methods for land use and land cover classification," *Sustainability*, vol. 15, no. 10, p. 7854, 2023.

[22] P. Iakubovskii, "Segmentation models pytorch," 2019, last accessed: Feb. 2024. [Online]. Available: https://github.com/qubvel/segmentation_models.pytorch

[23] Y. Dong, K. Yu, and W. Hu, "Gc-unet: an improved unet model for mangrove segmentation using landsat8," in *Proceedings*, International Conference on Big Data Engineering. IEEE, 2021, pp. 58–63.