

An Attention-Based U-Net for Detecting Deforestation Within Satellite Sensor Imagery*

David John^a, Ce Zhang^{b,c,*}

^a*School of Computing and Communications, Lancaster University, Bailrigg, Lancaster, United Kingdom, LA1 4WA*

^b*Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster, United Kingdom, LA1 4YQ*

^c*UK Centre for Ecology & Hydrology, Library Avenue, Lancaster University, Bailrigg, Lancaster, United Kingdom, LA1 4AP*

Abstract

In this paper, we implement and analyse an Attention U-Net deep network for semantic segmentation using Sentinel-2 satellite sensor imagery, for the purpose of detecting deforestation within two forest biomes in South America, the Amazon Rainforest and the Atlantic Forest. The performance of the Attention U-Net is compared with U-Net, Residual U-Net, ResNet50-SegNet and FCN32-VGG16 across three different datasets (three-band Amazon, four-band Amazon and Atlantic Forest). Results indicate that the Attention U-Net provides the best deforestation masks when tested on each dataset, achieving average pixel-wise F1-scores of 0.9550, 0.9769 and 0.9461 for each dataset, respectively. Mask reproductions from each classifier were also analysed, showing that compared to the ground reference the Attention U-Net could more accurately detect non-forest polygons than U-Net and overall it provides more accurate land cover polygons than each of the other methods despite its reduced complexity and training time, thus being the first application of an Attention U-Net to a land cover segmentation problem. This paper concludes with a brief discussion on the ability of the attention mechanism to offset

*This work was supported by an industrial placement at the Centre of Excellence in Environmental Data Science (CEEDS).

*Corresponding author at: Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster, United Kingdom, LA1 4YQ

Email address: `c.zhang@lancaster.ac.uk` (Ce Zhang)

the reduced complexity of the Attention U-Net, as well as ideas for further research into optimising the architecture and applying attention mechanisms into other architectures for deforestation detection. Our code is available at <https://github.com/davej23/attention-mechanism-unet>.

Keywords: Attention Mechanism, Attention U-Net, Deep Learning, Deforestation Mapping, Sentinel-2

1. Introduction

The Amazon Rainforest represents around 40% of the remaining tropical forests on Earth (Hubbell et al., 2008), and provides refuge for 10% of the world’s species (WWF, 2020). Therefore, the enormous carbon sequestering capability of the Amazon Rainforest is pivotal to the regulation of the continental, and global climate, since it is estimated to store 76 billion tonnes of carbon in the form of 390 billion trees (Müller, 2020). However, the region has seen large-scale deforestation for agriculture, raw materials, and for land to build housing due to rapid development of South America (Garcia-Ayllon, 2016).

This destruction poses an existential threat to the Amazon Rainforest and threatens to further worsen the effects of global warming. It is estimated that the Amazon’s ability to act as a carbon sink will disappear in 2035 (Hubau et al., 2020), and is already showing signs of being close to this (Harris et al., 2021), thus resulting in extreme weather such as drought and forest fires locally and globally.

In 2020 alone, an average of 2309.5 hectares of forest per day was destroyed (MapBiomass, 2020); roughly equating to an area the size of Ottawa, the capital city of Canada, which covers 6287 square kilometers (Statistics Canada, 2011), every month. As a result, it has become a global priority to minimise the rate of deforestation of the Amazon by designating protected areas, campaigning against companies which produce products in illegally-cleared areas of the forest, as well as by regular monitoring (Tollefson, 2015). The latter has long been a problem as on-the-ground monitoring is infeasible due to the sheer amount of

surface area the Amazon Rainforest covers (Gong et al., 1994). This paper looks to further the effort towards remotely-sensed monitoring for detecting deforestation within the Amazon region primarily but also for use in other forest biomes, through the use of artificial intelligence (AI) in the form of an Attention U-Net deep neural network (Oktay et al., 2018).

1.1. Architecture Fundamentals

The Attention U-Net is based upon the U-Net architecture (Ronneberger et al., 2015), which itself is a specific type of fully convolutional network (FCN); a family of neural networks characterised by an encoder-decoder, or contraction and expansion, structure. These are designed for semantic segmentation, also known as pixel-wise classification.

U-Net builds upon the standard FCN architecture by introducing skip connections, meaning that blocks of layers within the contraction phase can pass their output directly to blocks within the expansion phase, which greatly improves the ability to extract high-level features within training images. Previously, the U-Net has been applied to the task of semantic segmentation of the Amazon Rainforest using Sentinel-2 satellite imagery with high success (Bragagnolo et al., 2021), and the aim of this paper is to explore the incorporation of an attention mechanism into U-Net to improve upon this benchmark.

An attention mechanism aims to replicate the human ability to direct focus, or to concentrate on, specific stimuli. In the domain of neural networks, this involves learning which parts of the input to focus on during the process of training and prediction. Attention mechanisms are prominently used within the field of natural language processing (NLP), where they focus on sections of an input corpus, which is useful within tasks such as sentiment analysis (Galassi et al., 2020).

The Attention U-Net is created by adding an attention gate to the skip connection used within U-Net. Rather than concatenating each upscaled layer in the expansion phase with the appropriate contraction-phase layer, the upscaled layer is concatenated with the output from the attention mechanism, a function

of the pre-upscaled layer and the aforementioned contraction-phase layer.

1.2. Previous Work

Machine learning-based forest cover change monitoring of the Amazon has been ongoing for almost a decade (Souza et al., 2013), with deep learning (DL) methods being the current state-of-the-art. This has been demonstrated within comparisons of cutting-edge methods such as U-Net, ResUNet (Diakogiannis et al., 2020) and SharpMask (Pinheiro et al., 2016), using Landsat imagery of the Amazon Rainforest, versus less sophisticated methods such as the multi-layer perceptron (MLP) and random forests (de Bem et al., 2020).

Previous segmentation work using U-Net, and involving Sentinel-2 satellite data, has also been carried out, such as detecting change within Ukrainian forests (Isaienkov et al., 2021), as well as mapping irrigation systems (Graf, 2020). Other examples include the use of a spatio-temporal FCN for land cover segmentation of Slovenia (Zupanc et al., 2019). Spatio-temporal approaches use a short-term collection of images to map changes over time, thus learning how to recognise deforested areas. This type of approach has been used with early and late fusion spatio-temporal U-Nets and have been shown to provide marginal improvements upon U-Net (Maretto et al., 2021) at mapping deforestation within the Amazon.

Desertification detection within Algeria, using Landsat ETM+ satellite data, with a variational autoencoder (VAE) (Verstraete, 1986) is another example of the wide variety of contexts and approaches that have been used with semantic segmentation. Importantly, previous applications of the Attention U-Net have only been within medical contexts, such as brain tumour segmentation (Islam et al., 2021), liver computerised tomography (CT) scan segmentation (Li et al., 2020) and gland segmentation (Zhao et al., 2020). As a result, we believe that this paper represents the first, or one of the first, successful applications of the Attention U-Net to a cover change detection problem.

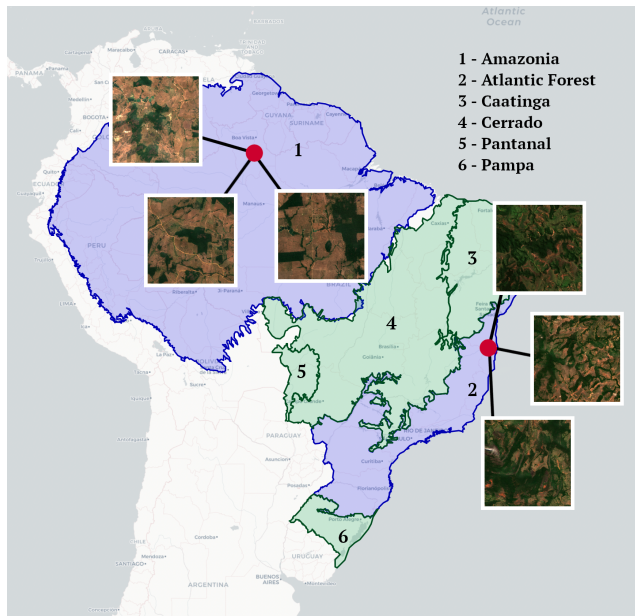


Figure 1: Map of the Amazon Rainforest and Atlantic Forest biomes within South America, including images from the 4-band datasets.

2. Methodology

2.1. Datasets

To evaluate the Attention U-Net for deforestation segmentation, we used three datasets produced from images in the satellite imagery database Sentinel-Hub (Sinergise, 2014). The first dataset is a collection of RGB-converted images and deforestation masks of the Amazon Rainforest ([dataset] Bragagnolo et al., 2019), where 0s and 1s represent deforested and forested areas respectively in the masks. This dataset is composed of 30 training images and 15 validation images. In order to have more unseen data to evaluate our models, we took five images from the training data and added them to the validation dataset.

The other two datasets are both composed of 4-band RGB + near-infrared imagery, one containing images from the Amazon Rainforest and the other from the Atlantic Forest (Mata Atlantica) ([dataset] Bragagnolo et al., 2021). Figure 1 shows the location of these biomes, as well as example images, and clearly

shows that the images are highly concentrated within two geographically distinct regions. These datasets had 499 and 485 training images, respectively, 100 validation images and 20 test images. For training, we randomly selected 250 training images due to memory limitations.

Throughout this paper, we refer to the first dataset as the 'RGB' dataset and the latter two datasets as the 4-band Amazon and 4-band Atlantic Forest datasets. Within each dataset, each image is of shape (512, 512, 3) in the RGB dataset and (512, 512, 4) in the 4-band datasets, and each image has a corresponding (512, 512, 1) deforestation mask. In order to produce the images and masks found within each dataset, the author of the dataset split a large satellite image into sub-images and produced masks using a modified version of the k-means classification algorithm with the GRASS-GIS 7.6.1 software suite (GRASS Development Team, 2020). Images were repeatedly re-classified until the corresponding masks had 'a satisfactory rating'.

2.1.1. Difference Between U-Net and Attention U-Net

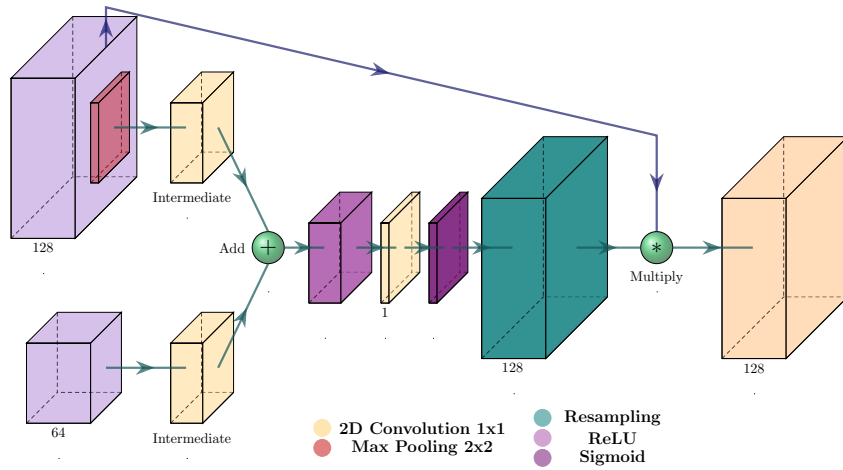


Figure 2: A network architecture diagram for the Attention Mechanism. The AG marker signifies the location of the attention gate, or attention mechanism.

The architecture of U-Net is similar to that of the Attention U-Net, shown in Figure 3, except the number of filters used within each convolutional layer,

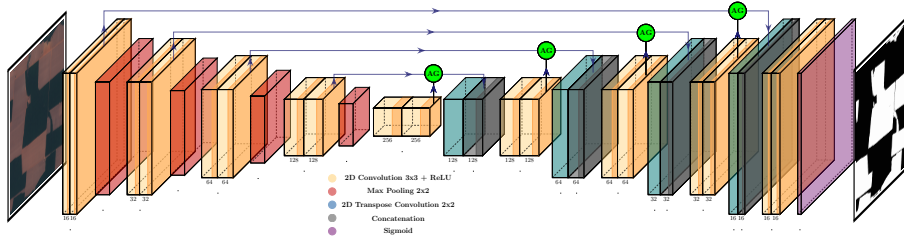


Figure 3: A network architecture diagram for the Attention U-Net.

in each respective block, are 64, 128, 256, and 512, with 1024 filters used in the bottleneck layers, and there are no attention mechanisms. We chose to use 16,32,64, and 128 filters, and 256 filters in the bottleneck layer, as the number of convolutional filters in each block for the Attention U-Net. We found that using the same number of filters as U-Net caused the Attention U-Net to have more parameters and a greater training time, despite being slightly more performant, as the parameters within the attention mechanism as seen in Figure 2 add complexity. In the configuration used within this paper, the Attention U-Net is much more computationally efficient and has greater performance than U-Net.

In order to evaluate the performance of the Attention U-Net, four other models were also tested: U-Net, Residual U-Net, ResNet50 (He et al., 2015) with a SegNet backbone (Badrinarayanan et al., 2016), and FCN32 with a VGG16 backbone (Simonyan and Zisserman, 2015). Each of the models were trained from scratch, including the backbone architecture, in order to provide a fair comparison. Typically, VGG16 and ResNet models are trained using transfer learning, where the model has already been trained on a dataset such as ImageNet (Deng et al., 2009) before being trained on a suitable dataset corresponding to the given task. Transfer learning is particularly useful where limited training data is available, as the features learned from training on a dataset such as ImageNet can provide transferrable knowledge to tasks such as segmentation (Dube et al., 2018).

2.1.2. Training and Testing Procedure

The number of epochs and the learning rate used for each network can be found in Table 1. These values were found to be the values which gave maximal validation accuracy through experimentation. Models were trained using different learning rates and epochs until the highest validation accuracy was obtained. Once the optimal hyperparameters were found, models were not re-trained due to constraints relating to using the Google Colab environment with hardware usage limits. However, our results cover multiple datasets and each show the Attention U-Net to be the stronger performer, therefore reducing the likelihood of the results being down to weight initialisation alone. Also, we used data augmentation on the RGB dataset, including rotation, reflection, zooming and shearing, in order to facilitate the need for a greater amount of training data. With the 4-band datasets, the volume of training data was already sufficiently large. Finally, each image was normalised before training, such that each pixel channel value was in the interval $[0,1]$. For each model, the Adam optimiser (Kingma and Ba, 2017) was used as it provided the greatest peak validation accuracy compared to the stochastic gradient descent (SGD) optimiser. The Binary Crossentropy (BCE) loss function was used as it has been shown to work well within binary semantic segmentation tasks (Jadon, 2020).

Classifier	Learning	Epochs	
	Rate	RGB	4-band
Attention U-Net	0.0005	50	60
U-Net	0.0001	30	20
ResNet50-SegNet	0.0001	40	20
FCN32-VGG16	0.0001	50	50
Residual U-Net	0.0001	40	20

Table 1: Number of Training Epochs and Learning Rate Used for Training Each Classifier on RGB and 4-band Datasets; Values Chosen Through Trial and Improvement.

After training, each of the models were evaluated on the validation data

from the dataset they were trained on, and the models trained on the 4-band datasets were tested on the test dataset also. Due to the existence of 4-band data from two locations, each 4-band model was also evaluated on the data from the opposite location, for instance the models trained on the 4-band Amazon data were also tested on the Atlantic Forest data, and vice-versa. This gives us the ability to test how transferable each model is to imagery from a different location, which could show whether our model could be used successfully for deforestation segmentation within other regions globally. The evaluation process of a model involves generating mask reproductions for each unseen image by passing an image I into it, obtaining a $512 \times 512 \times 1$ output with values being in the range $[0, 1]$. These values are rounded to the nearest integer to create a binary mask. The performance of a model is then evaluated by computing the pixel-wise differences between these masks and the original ground truth masks.

2.1.3. Quantifying Results

To quantify our results, the weighted Precision, Recall, F1-score and Jaccard Index, also known as the Intersection over Union (IoU) score, were used. The IoU score was selected as it describes the similarity of the predicted deforestation polygons with the ground truth, which is a better measure within image segmentation compared to pixel accuracy which only measures the number of accurate pixel predictions. Weighted metrics were used as they account for class imbalance between forest and non-forest pixels (Tague-Sutcliffe, 1992). Precision, Recall and IoU scores were computed for both classes and weighted according to the number of pixels within each class. The positive class is the forest pixels.

Another essential piece of the analysis of a model is determining its computational efficiency; a factor which determines whether it is viable for real-world use. If the training time is too high, it may be more suitable to opt for a slightly less performant model which takes less time to train. Furthermore, models with large parameter spaces are more likely to overfit and have worse generalisability than less complex models (Ying, 2019) despite typically having greater perfor-

mance. Therefore, it is important to evaluate the efficiency of a candidate model to determine whether it should be recommended for wider use so in this paper we compare the number of learnable parameters and the total training time of each model.

To carry out our experimentation, we used the Google Colaboratory Python environment (Google, 2017) as it provides free use of datacenter-grade GPUs. Neural networks within this paper were trained on an NVIDIA Tesla P100 16GB GPU with 12GB of RAM. Each of the models were written with the Keras (Chollet et al., 2015) Application Programming Interface (API) of the TensorFlow machine learning framework (Abadi et al., 2015).

3. Experimental Results

3.1. RGB Dataset

When testing the models on the RGB validation data, the Attention U-Net achieved the highest results overall, as can be seen in Table 4. This is evidenced in Figure 4a where the mask prediction from Attention U-Net is markedly better than those produced by the other classifiers. There is a reduced tendency to incorrectly classify forest as non-forest, false-positives, in contrast to U-Net, which appears to often exaggerate the non-forest polygons. The exception to this is the upper red circle within the upper Attention U-Net reconstructed mask, where the other classifiers fail to identify the extent of the forested polygon being highlighted.

3.2. 4-band Datasets

3.2.1. Amazon Dataset

A similar result is seen with the validation and test metrics for the 4-band Amazon data. Table 2 shows that the Attention U-Net outperforms each of the other three classifiers, with a 0.5% improvement in F1-score over the standard U-Net. We can see this within Figure 4b where the Attention U-Net produces deforestation polygons with greater detail than U-Net and gives less false-positives than ResNet50-SegNet.

Classifier	Dataset							
	Validation				Test			
	IoU	Precision	Recall	F1-score	IoU	Precision	Recall	F1-score
Attention U-Net	0.9581	0.9790	0.9779	0.9785	0.9516	0.9758	0.9748	0.9753
U-Net	0.9530	0.9766	0.9752	0.9759	0.9473	0.9738	0.9724	0.9731
Residual U-Net	0.9432	0.9723	0.9696	0.9720	0.9429	0.9715	0.9703	0.9709
ResNet50-SegNet	0.9467	0.9733	0.9717	0.9725	0.9416	0.9707	0.9694	0.9701
FCN32-VGG16	0.8592	0.9210	0.9214	0.9212	0.8557	0.9212	0.9205	0.9208

Table 2: Quantitative Evaluation of the Performance of Four Classifiers at Classifying Forest and Non-Forest Areas Within the 4-band Amazon Test and Validation Satellite Imagery. Bold Text Identifies the Best Result.

Classifier	Dataset							
	Validation				Test			
	IoU	Precision	Recall	F1-score	IoU	Precision	Recall	F1-score
Attention U-Net	0.9120	0.9563	0.9520	0.9541	0.9199	0.9591	0.9571	0.9581
U-Net	0.8818	0.9387	0.9346	0.9366	0.8883	0.9424	0.9373	0.9399
Residual U-Net	0.9102	0.9544	0.9514	0.9512	0.9073	0.9542	0.9493	0.9517
ResNet50-SegNet	0.9043	0.9514	0.9480	0.9497	0.9026	0.9510	0.9466	0.9488
FCN32-VGG16	0.7182	0.8319	0.8290	0.8304	0.6902	0.8186	0.8077	0.8131

Table 3: Quantitative Evaluation of the Performance of Four Classifiers at Classifying Forest and Non-Forest Areas Within the 4-band Atlantic Forest Test and Validation Satellite Imagery. Bold Text Identifies the Best Result.

3.2.2. Atlantic Forest Dataset

Following on from this, the Attention U-Net once again outperforms other models on the Atlantic Forest data. In particular, Figure 3 shows that the F1-score produced by the Attention U-Net is significantly greater than for the other models. This difference can be witnessed within Figure 4c where the Attention

Classifier	IoU	Precision	Recall	F1-score
Attention U-Net	0.9028	0.9574	0.9526	0.9550
U-Net	0.8888	0.9571	0.9473	0.9522
Residual U-Net	0.9127	0.9539	0.9505	0.9493
ResNet50-SegNet	0.9025	0.9519	0.9470	0.9495
FCN32-VGG16	0.8198	0.8988	0.8978	0.8983

Table 4: Quantitative Evaluation of the Performance of Four Classifiers at Classifying Forested/Deforested Areas Within the RGB Amazon Validation Set Satellite Imagery. Bold Text Identifies the Best Result.

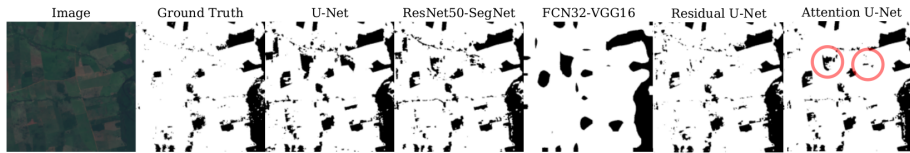
U-Net is again able to identify more complex polygons when compared to U-Net. It can be seen that ResNet50-SegNet can also accurately identify the same polygons in question however it also produces more false-positives, than the Attention U-Net.

Classifier	Training Data Location - Testing Data Location							
	Amazon - Atlantic Forest				Atlantic Forest - Amazon			
	IoU	Precision	Recall	F1-score	IoU	Precision	Recall	F1-score
Attention U-Net	0.8143	0.9222	0.8829	0.9021	0.8722	0.9445	0.9266	0.9355
U-Net	0.8134	0.9169	0.8847	0.9005	0.8254	0.9323	0.8915	0.9115
Residual U-Net	0.7707	0.9164	0.8508	0.8824	0.8709	0.9440	0.9256	0.9347
ResNet50-SegNet	0.7921	0.9156	0.8670	0.8906	0.8453	0.9355	0.9088	0.9220
FCN32-VGG16	0.6797	0.8246	0.7930	0.8085	0.7913	0.8985	0.8750	0.8866

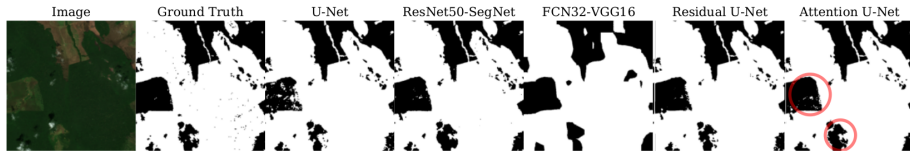
Table 5: Quantitative Evaluation of the Performance of Four Classifiers at Classifying Forest and Non-Forest Areas Within Imagery from the Dataset from the Other Location. Bold Text Identifies the Best Result.

3.3. Testing on non-local imagery

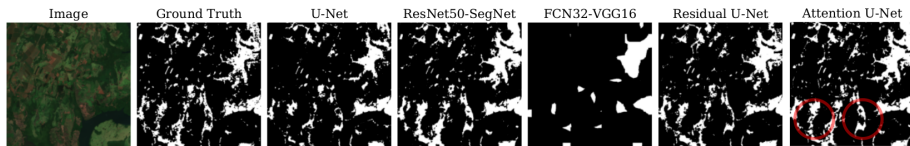
When testing the Amazon-trained models on the Atlantic Forest data, in Table 5, we see that the Attention U-Net is the most performant overall, except



(a) Image from the RGB Amazon dataset.



(b) Image from the 4-band Amazon dataset.



(c) Image from the 4-band Atlantic Forest dataset.

Figure 4: Comparison of ground truth deforestation masks versus classifier-generated mask predictions. Deforested areas appear in black, and forested areas in white. Notable differences are highlighted with red circles.

for the recall score in which it is bested by U-Net, however it more importantly has a higher F1-score and IoU score meaning that the reproduced mask is more spatially similar to the ground truth and has a greater precision and recall overall. When we look at the results from the opposite scenario, the difference in performance between the Attention U-Net and the other models is much greater suggesting that the Attention U-Net provides greater transferability to data from a different location than the other methods.

3.4. Computational efficiency

Figure 6 shows that the Attention U-Net is the most efficient model, containing both the fewest number of parameters, as well as the lowest training time for each of the datasets; it is also vastly more efficient, training between 20% to 56% faster than other models. The training time of both '4-band' models was

identical, due to being trained over the same number of epochs with the same learning rate.

Classifier	Parameters ($\times 10^6$)	Training time (s)	
		RGB	4-band
Attention U-Net	2.01	365	465
U-Net	31.03	366	650
Residual U-Net	31.3	744	975
ResNet50-SegNet	72.27	1092	1475
FCN32-VGG16	134.3	650	2300

Table 6: Comparison of the Computational Efficiency of Four Classifiers, in Terms of the Number of Parameters and Training Time Per Step/Image. Bold Text Identifies the Best Result.

3.5. Attention U-Net versus U-Net

Finally, we compare the ground truth masks to the predictions made by Attention U-Net and U-Net. Figure 5 shows that the Attention U-Net correctly identifies a greater percentage of forest pixels compared to U-Net on both the RGB and Atlantic Forest datasets, by 2.47% and 3.06% respectively and produces 2.47% and 3.16% fewer false-positives, respectively. On the 4-band Amazon dataset, the Attention U-Net produces fewer misclassifications as well as a greater proportion of correct predictions overall compared to U-Net; this is highlighted by the fact that only 2.21% of pixels are mis-classified. When taking into account the correctly identified pixels within each dataset, the Attention U-Net identifies 1.03%, 0.274% and 1.73% more pixels correctly on the respective datasets. When using a model to determine deforested regions in satellite imagery in order to estimate total deforested area, false-positives are more desirable than false-negatives as deforested area being underestimated can potentially caused new deforestation within an area to go undetected. However, in this case, as the Attention U-Net more accurately identifies a greater number of pixels than U-Net the greater number of false-negatives is not an issue.

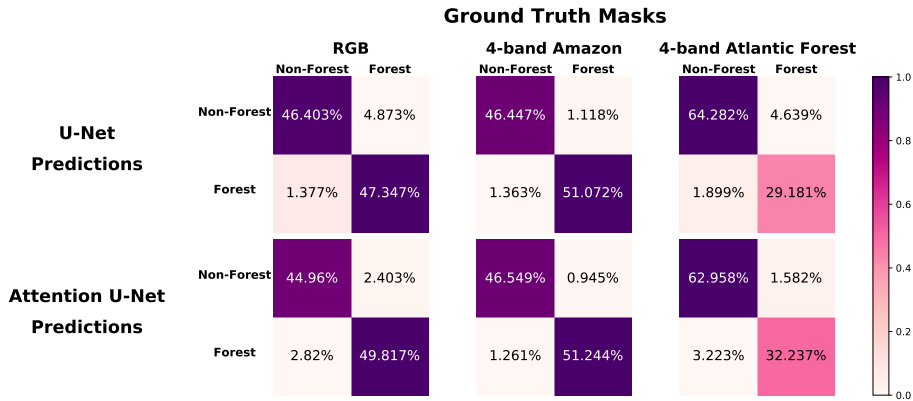


Figure 5: Confusion matrix comparing U-Net and Attention U-Net mask predictions versus ground truth masks within each dataset.

4. Discussion

4.0.1. General comments

Throughout our analysis, the Attention U-Net outperforms the other models. Despite the Residual U-Net providing better results in some cases, the Attention U-Net consistently provides the best, or second-best, results. The improvement of Attention U-Net upon U-Net is likely due to the attention mechanism being able to distinguish high levels of detail in complex polygons, resulting in fewer errors within mask predictions. It was also shown within our experimentation that the 4-band Attention U-Net models are transferable to images from a different region, and this could be further confirmed by testing on a similar dataset from a different forest. We also saw that the Attention U-Net is more efficient than the U-Net, where the training time was up to 30% lower yet had noticeably improved performance. In regards to the datasets themselves, we can see in Figure 5 that the Atlantic Forest dataset has a large class imbalance in favour of non-forest pixels, which accounted for two-thirds of the total number of pixels. This is likely the reason why the Atlantic Forest models performed very well when evaluated on Amazon data.

4.0.2. Limitations

We conjecture that the performance of classifiers is limited by the quality of the ground truth masks as they were produced using an imperfect classification method. It was noted in the dataset author’s paper (Bragagnolo et al., 2021) that some model mask predictions identify deforested polygons which were not picked up within ground truth masks. As a result, it could be useful for future work to update the ground truth masks by adding in the polygons found by our Attention U-Net model.

4.0.3. Future work

To build upon the work from this paper, interested readers could experiment with other loss functions such as Jaccard loss (Bertels et al., 2019), Dice loss (Sudre et al., 2017), or derivatives such as DiceTopK and DiceFocal, as they have been successful with other segmentation tasks (Ma et al., 2021). Also, the addition of regularisation layers such as Dropout and Batch Normalisation could reduce overfitting and validation loss. These were not tested in our experimentation, but have been shown to provide improvements to deep learning models in multiple scenarios.

Since the addition of the attention mechanism allows the Attention U-Net to perform to such a degree despite having very few parameters, we believe that others may have success implementing attention mechanisms into less complex versions of existing deep learning methods to a similar effect. One such possibility is the use of a Residual Attention U-Net which would contain more parameters than the Attention U-Net, and perhaps longer training time, but may improve upon the Residual U-Net.

Finally, we suggest that transfer learning could be used with either of the 4-band Attention U-Net models by training on both 4-band training datasets. This could allow for greater transferability to images from a wider set of locations. It was shown in Section 3.3 that the models trained on a single location were transferable, so it is sensible to suggest that transfer learning would further improve this and allow for successful applicability to forest imagery from

around the world.

5. Conclusion

In this paper, we have carried out a quantitative analysis of the performance of the Attention U-Net at the semantic segmentation of South American tropical rainforest imagery for detecting deforestation. We found that the addition of an attention mechanism to a less complex version of U-Net provides greater performance than the standard U-Net architecture, as well as several other state-of-the-art methods. The attention mechanism enables the network to retain high levels of spatial information despite containing layers of much lower dimensionality than U-Net. Due to the successful application of an attention mechanism to a deep neural network for this task, we can recommend the use of an Attention U-Net for other land cover segmentation tasks in the field.

CRedit authorship contribution statement

David John: Data curation, Formal analysis, Investigation, Methodology, Software, Visualisation, Writing - original draft, Writing - review & editing. **Ce Zhang:** Conceptualisation, Writing - review & editing, Supervision, Project Administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Network architecture diagrams were produced using PlotNeuralNet (Iqbal, Haris, 2018). The Amazon basin shapefile is from Harvard WorldMap (Villegas, J.R) and the South American biome shapefiles are from TerraBrasilis

(Assis et al., 2019). Dataset augmentation code was modified from (Bragagnolu, Lucimara, 2021), and U-Net code was inspired by (Xuhao, Zhi, 2018). ResNet50-SegNet code was used from (Bhatnagar et al., 2020) and code inspiration was used from (Dwivedi, 2019). The code for our FCN32-VGG16 model was modified from (Gupta et al., 2021).

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. URL: <https://www.tensorflow.org/>. software available from tensorflow.org.
- Assis, L.F.F.G., Ferreira, K.R., Vinhas, L., Maurano, L., Almeida, C., Carvalho, A., Rodrigues, J., Maciel, A., Camargo, C., 2019. TerraBrasilis: A Spatial Data Analytics Infrastructure for Large-Scale Thematic Mapping. ISPRS Int. J. Geo-Inf. URL: <http://terrabrasilis.dpi.inpe.br/en/download-2/>, doi:<https://doi.org/10.3390/ijgi8110513>.
- Badrinarayanan, V., Kendall, A., Cipolla, R., 2016. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. [arXiv:1511.00561](https://arxiv.org/abs/1511.00561).
- de Bem, P.P., de Carvalho Junior, O.A., Fontes Guimarães, R., Trancoso Gomes, R.A., 2020. Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks. Remote Sens. 12. URL: <https://www.mdpi.com/2072-4292/12/6/901>, doi:10.3390/rs12060901.

- Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., Blaschko, M.B., 2019. Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, 92–100. URL: http://dx.doi.org/10.1007/978-3-030-32245-8_11, doi:10.1007/978-3-030-32245-8_11.
- Bhatnagar, S., Gill, L., Ghosh, B., 2020. Drone Image Segmentation Using Machine and Deep Learning for Mapping Raised Bog Vegetation Communities. *Remote. Sens.* 12, 2602.
- Bragagnolo, L., da Silva, R., Grzybowski, J., 2021. Amazon forest cover change mapping based on semantic segmentation by U-Nets. *Ecol. Inform.* 62, 101279. URL: <https://www.sciencedirect.com/science/article/pii/S1574954121000704>, doi:<https://doi.org/10.1016/j.ecoinf.2021.101279>.
- [dataset] Bragagnolo, L., da Silva, R.V., Grzybowski, J.M.V., 2019. Amazon Rainforest dataset for semantic segmentation. URL: <https://doi.org/10.5281/zenodo.3233081>, doi:10.5281/zenodo.3233081.
- [dataset] Bragagnolo, L., da Silva, R.V., Grzybowski, J.M.V., 2021. Amazon and Atlantic Forest image datasets for semantic segmentation. URL: <https://doi.org/10.5281/zenodo.4498086>, doi:10.5281/zenodo.4498086.
- Bragagnolu, Lucimara, 2021. Deforestation mapping using UNets. URL: <https://github.com/bragagnololu/UNet-defmapping>. accessed: 2021-07-10.
- Chollet, F., et al., 2015. Keras. URL: <https://keras.io>. accessed: 2021-07-01.
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. doi:10.1109/CVPR.2009.5206848.

- Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* 162, 94–114. URL: <http://dx.doi.org/10.1016/j.isprsjprs.2020.01.013>, doi:10.1016/j.isprsjprs.2020.01.013.
- Dube, P., Bhattacharjee, B., Petit-Bois, E., Hill, M., 2018. Improving Transferability of Deep Neural Networks. [arXiv:1807.11459](https://arxiv.org/abs/1807.11459).
- Dwivedi, P., 2019. Deep Learning. URL: https://github.com/priya-dwivedi/Deep-Learning/tree/master/resnet_keras.
- Galassi, A., Lippi, M., Torrioni, P., 2020. Attention in Natural Language Processing. *IEEE Trans. Neural Netw. Learn. Syst.*, 1–18 URL: <http://dx.doi.org/10.1109/TNNLS.2020.3019893>, doi:10.1109/tnnls.2020.3019893.
- Garcia-Ayllon, S., 2016. Rapid development as a factor of imbalance in urban growth of cities in Latin America: A perspective based on territorial indicators. *Habitat Int.* 58, 127–142. doi:10.1016/j.habitatint.2016.10.005.
- Gong, P., Miller, J., Spanner, M., 1994. Forest canopy closure from classification and spectral unmixing of scene components-multisensor evaluation of an open canopy. *IEEE Trans. Geosci. Remote Sens.* 32, 1067–1080. doi:10.1109/36.312895.
- Google, 2017. Colaboratory. URL: <https://colab.research.google.com>. accessed: 2021-07-01.
- Graf, L., 2020. Semantic Segmentation and Multi-Temporal Analysis of Sentinel-2 Imagery for Mapping Centre Pivot Irrigation Systems. Ph.D. thesis. doi:10.13140/RG.2.2.20441.95849.
- GRASS Development Team, 2020. Geographic Resources Analysis Support System (GRASS GIS) Software. Open Source Geospatial Foundation. USA. URL: <https://grass.osgeo.org>. accessed: 2021-08-01.

- Gupta, D., Jhunjhunu, R., Juston, M., 2021. Image Segmentation Keras : Implementation of Segnet, FCN, UNet, PSPNet and other models in Keras. URL: <https://github.com/divamgupta/image-segmentation-keras>. accessed: 2021-07-06.
- Harris, N., Gibbs, D., Baccini, A., Birdsey, R., de Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M., Herold, M., Houghton, R., Potapov, P., Requena Suarez, D., Román-Cuesta, R.M., Saatchi, S., Slay, C., Turubanova, S., Tyukavina, A., 2021. Global maps of twenty-first century forest carbon fluxes. *Nat. Clim. Change* 11, 1–7. doi:10.1038/s41558-020-00976-6.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. URL: <https://arxiv.org/abs/1512.03385>, arXiv:1512.03385.
- Hubau, W., Lewis, S., Phillips, O., Affum-Baffoe, K., Beeckman, H., Sanchez, A., Daniels, A., Ewango, C., Fauset, S., Mukinzi, J., Sheil, D., Sonké, B., Sunderland, T., Taedoumg, H., Thomas, S., White, L., Abernethy, K., Adu-Bredu, S., 2020. Asynchronous Carbon Sink Saturation in African and Amazonian Tropical Forests. *Nature* 579.
- Hubbell, S.P., He, F., Condit, R., Borda-de Água, L., Kellner, J., ter Steege, H., 2008. How many tree species are there in the Amazon and how many of them will go extinct? *Proc. Natl. Acad. Sci.* 105, 11498–11504. doi:10.1073/pnas.0801915105.
- Iqbal, Haris, 2018. PlotNeuralNet. URL: <https://github.com/HarisIqbal88/PlotNeuralNet>. accessed: 2021-07-20.
- Isaienkov, K., Yushchuk, M., Khramtsov, V., Seliverstov, O., 2021. Deep Learning for Regular Change Detection in Ukrainian Forest Ecosystem With Sentinel-2. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 364–376. doi:10.1109/JSTARS.2020.3034186.

- Islam, M., VS, V., Jose, V.J.M., Wijethilake, N., Utkarsh, U., Ren, H., 2021. Brain Tumor Segmentation and Survival Prediction using 3D Attention UNet. [arXiv:2104.00985](https://arxiv.org/abs/2104.00985).
- Jadon, S., 2020. A survey of loss functions for semantic segmentation. 2020 IEEE Conf. Comput. Intell. Bioinform. Comput. Biol. (CIBCB) URL: <http://dx.doi.org/10.1109/CIBCB48159.2020.9277638>, doi:10.1109/cibcb48159.2020.9277638.
- Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- Li, C., Tan, Y., Chen, W., Luo, X., Gao, Y., Jia, X., Wang, Z., 2020. Attention Unet++: A Nested Attention-Aware U-Net for Liver CT Image Segmentation, in: 2020 IEEE International Conference on Image Processing (ICIP), pp. 345–349. doi:10.1109/ICIP40778.2020.9190761.
- Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L., 2021. Loss Odyssey in Medical Image Segmentation. *Med. Image Anal.* 71, 102035. URL: <https://www.sciencedirect.com/science/article/pii/S1361841521000815>, doi:<https://doi.org/10.1016/j.media.2021.102035>.
- MapBiomas, 2020. Annual Report on Deforestation in Brazil. URL: https://s3.amazonaws.com/alerta.mapbiomas.org/rad2020/RAD2021_-_MapBiomasAlerta_EN.pdf. accessed: 2021-07-08.
- Maretto, R.V., Fonseca, L.M.G., Jacobs, N., Körting, T.S., Bendini, H.N., Parente, L.L., 2021. Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest. *IEEE Geosci. Remote Sens. Lett.* 18, 771–775. doi:10.1109/LGRS.2020.2986407.
- Müller, C., 2020. Brazil and the Amazon rainforest: Deforestation, biodiversity and cooperation with the EU and international forums URL: <https://op.europa.eu/s/pk8M>, doi:10.2861/520925.

- Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D., 2018. Attention u-net: Learning where to look for the pancreas. *arXiv:1804.03999*.
- Pinheiro, P.O., Lin, T.Y., Collobert, R., Dollàr, P., 2016. Learning to Refine Object Segments. *arXiv:1603.08695*.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR abs/1505.04597*. URL: <http://arxiv.org/abs/1505.04597>, *arXiv:1505.04597*.
- Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556*.
- Sinergise, 2014. Sentinel Hub. URL: <https://www.sentinel-hub.com/>. accessed: 2021-07-01.
- Souza, Jr, C.M., Siqueira, J.V., Sales, M.H., Fonseca, A.V., Ribeiro, J.G., Numata, I., Cochrane, M.A., Barber, C.P., Roberts, D.A., Barlow, J., 2013. Ten-Year Landsat Classification of Deforestation and Forest Degradation in the Brazilian Amazon. *Remote Sens.* 5, 5493–5513. URL: <https://www.mdpi.com/2072-4292/5/11/5493>, doi:10.3390/rs5115493.
- Statistics Canada, 2011. Population and Dwelling Count Highlight Tables, 2011 Census. URL: <https://www12.statcan.gc.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/Table-Tableau.cfm?LANG=Eng&T=201&S=3&O=D&RPP=150>. accessed: 2021-07-08.
- Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M., 2017. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. *Lect. Notes Comput. Sci.* , 240–248 URL: http://dx.doi.org/10.1007/978-3-319-67558-9_28, doi:10.1007/978-3-319-67558-9_28.

- Tague-Sutcliffe, J., 1992. The pragmatics of information retrieval experimentation, revisited. *Inf. Process. Manag.* 28, 467–490. URL: <https://www.sciencedirect.com/science/article/pii/030645739290005K>, doi:[https://doi.org/10.1016/0306-4573\(92\)90005-K](https://doi.org/10.1016/0306-4573(92)90005-K). special Issue: Evaluation Issues in Information Retrieval.
- Tollefson, J., 2015. Stopping deforestation: Battle for the Amazon. *Nature* 520, 20–23. URL: <https://doi.org/10.1038/520020a>, doi:10.1038/520020a.
- Verstraete, M.M., 1986. Defining desertification: A review. *Clim. Change* 9, 5–18. URL: <https://doi.org/10.1007/BF00140520>, doi:10.1007/BF00140520.
- Villegas, J.R., . Amazon basin polygon. Harvard University. URL: http://worldmap.harvard.edu/data/geonode:amapoly_ivb. accessed: 2021-07-28.
- WWF, 2020. Amazon. URL: <https://www.worldwildlife.org/places/amazon>. accessed: 2021-07-08.
- Xuhao, Zhi, 2018. Implementation of deep learning framework – Unet, using Keras. URL: <https://github.com/zhixuhao/unet>. accessed: 2021-07-01.
- Ying, X., 2019. An Overview of Overfitting and its Solutions. *J. Phys.: Conf. Ser.* 1168, 022022. URL: <https://doi.org/10.1088/1742-6596/1168/2/022022>, doi:10.1088/1742-6596/1168/2/022022.
- Zhao, P., Zhang, J., Fang, W., Deng, S., 2020. SCAU-Net: Spatial-Channel Attention U-Net for Gland Segmentation. *Front. Bioeng. Biotechnol.* 8, 670. URL: <https://www.frontiersin.org/article/10.3389/fbioe.2020.00670>, doi:10.3389/fbioe.2020.00670.
- Zupanc, A., Peressutti, D., Aleksandrov, M., Milcinski, G., Lubej, M., Burja, A., Batic, M., 2019. Spatio-Temporal Deep Learning: An Application to Land Cover Classification. doi:10.13140/RG.2.2.17961.65128.