Deep Learning model comparison regarding forest canopy height regression

M. Risse[1,+], **B. Carmier**[1,+], **and E. De Labarrière**[1,+]

[1]Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[+]All authors contributed equally to this work

ABSTRACT

This study explores the application of deep learning, specifically a U-net model, to the estimation of canopy height using Sentinel-2 imagery. This report explore the comparison of 4 different customs models, and an existing one. Ending up with a best model performing a mean absolute error (MAE) of 4.27 meters, despite a slight performance gap compared to other studies using both Sentinel-1 and Sentinel-2 data. Key finding reveal that, contrary to expectations, the best model was the one using only 3 spectral bands without data augmentation, while models using 12 bands or data augmentation performed less effectively. Future work should focus on integrating relevant spectral bands, optimising feature selection, and improving model generalisation for broader applicability. The results of this study provide valuable insights into the development of accurate canopy height performing a linear regression task.

The link to the project's GitHub repository is provided below for reference: IPEO-Project Canopy height regression.

1 Introduction

Forests are critical to the global carbon cycle and the stability of the climate (Mitchard, 2018). They act as carbon sinks and influence the atmospheric CO_2 concentration. However, carbon fluxes between the biosphere and the atmosphere are imprecise, with standard deviations as high as 47%; for example, the global terrestrial CO_2 sink has an absorption of 3.2 ± 0.6 per year (Friedlingstein et al., 2019).
Aboveground biomass (AGB) is a key factor in carbon estimates, indeed 50% of AGB is carbon (Rodríguez-Veiga et al., 2017). But global biomass products have significant differences in spatial patterns (Mitchard et al., 2013).
Therefore, AGB shows large inaccuracies in global biomass mapping and is an obstacle to estimating the impact of deforestation and therefore atmospheric CO_2 concentrations (Kunreuther et al., 2014). In addition, AGB is an important indicator for mapping multiple ecosystem services other than climate regulation, such as regulation of water flows, life cycle maintenance, greenhouse gas emissions, carbon storage and sequestration, and oxygen supply (Egoh et al., 2012).

Remote sensing allows mapping of the above-ground portion of total forest biomass over large geographical areas, overcoming problems of under-representation of field data. Canopy reflectance, measured by passive optical sensors and radar backscatter, has been shown to correlate with field estimates of AGB density. Therefore, the maximum height of the canopy can be used to estimate AGB (G. Zhang et al., 2014).
Furthermore, canopy height is a key factor influencing carbon storage, vegetation productivity and biodiversity in forests, as well as an indicator of key processes such as biomass allocation (J. Zhang et al., 2016). Thus, canopy height is a fundamental variable necessary for estimating carbon fluxes, understanding biodiversity, ecosystem services and many other applications, so it is very useful to estimate it.

As a state-of-the-art method in remote sensing, a LIDAR was used to determine the original canopy height: NASA's Global Ecosystem Dynamics Investigation (GEDI). GEDI samples about 4% of the Earth's land surface during its two-year nominal mission, acquiring over 10 billion cloud-free shots with a footprint resolution of 25 m. GEDI lidar observations are used to produce canopy height data sets (Dubayah et al., 2020).

Traditional approaches to LIDAR waveform processing have relied on conventional signal processing, but

GEDI presents a challenge to these methods for several reasons. Firstly, it will produce an unprecedented number of fine spatial resolution observations, leading to structural variability in the waveforms. Second, GEDI will use both high and low power beams, which, combined with the structural diversity, variations in noise levels and potential instrument artefacts, will make calibration difficult.
Therefore a deep learning approach has been used to process the data and find the canopy height rather than using traditional processing. This deep learning model has demonstrated the ability to navigate the complex task of deriving global canopy heights from on-orbit L1B GEDI waveforms with favorable accuracies that may exceed traditional waveform processing (Lang et al., 2022a).

Our goal was to find a way to determine canopy height using Sentinel-2 data instead of GEDI data, conducting an approach based on image processing rather than LIDAR waveform processing. The idea was to use the canopy height data derived from GEDI observations as ground truth to train our model for estimating canopy height from Sentinel-2 data. We choose to develop a deep-learning model based on the U-Net architecure, as presented in (Ronneberger et al., 2015) but adapted to our need.Few studies already show the potential to estimate the canopy height from Sentinel-2 images and by using a U-net model.
As an example, the Centre for Research and Technology Hellas (CERTH) have developed a model based on the U-net architecture that uses Sentinel-2 images for estimating canopy height which is the same work we would like to implement. In the Bohemian forest this model had an MAE of 2.29 meters and an RMSE of 3.15 meters (Alagialoglou et al., 2021).
Another example is a deep learning model that was developped by the Climate and Environmental Sciences Laboratory of the "Université Paris-Saclay". They also used a U-net model, associated with Sentinel-1 and 2 images to predict canopy height in the Landes forest. They managed to find an MAE and RMSE close to the one found by the CERTH, their MAE is 2.02 meters .
With this approach we would like to create a model able to compute canopy height in a cost-effectiveness way and that is still able to have good performance.

2 Data

The data is a set of 9852 images with the corresponding labels (masks), each pair associated with a single id number. The data were already split into train/val/test set with the proportion 60/20/20 according to the CSV file given with the dataset.
Each pair of image/label has a size of 32x32 pixels. The images are Sentinel-2 multi-spectral 12-bands patches, all bands upsampled to a resolution of 10m. The labels are segmentation masks, composed of values from 0 to 255 coding the canopy height in meter. The value 255 represent "no-data" values. Dataset is accessible here: canopy height dataset.

The labels have been generated by processing waveforms captured by NASA's Global Ecosystem Dynamics Investigation (GEDI) mission. GEDI is a space-borne LIDAR system used to measure vertical forest structures and estimate canopy heights. This methodology, described in (Lang et al., 2022b), has achieved robust results for canopy height estimate with a root mean squared error (RMSE) of 2.7 m. Those results makes it a reliable tool for further forest structure studies.

Figure 2 shows the first image of the dataset. Showing the RGB image, while the histogram present the distribution of pixel values across RGB bands. The mask isolates regions of interest and the corresponding histogram summarizes the pixel values of the mask wich represent the height of the canopy. These kind of visualizations ensure data quality and support further analysis.

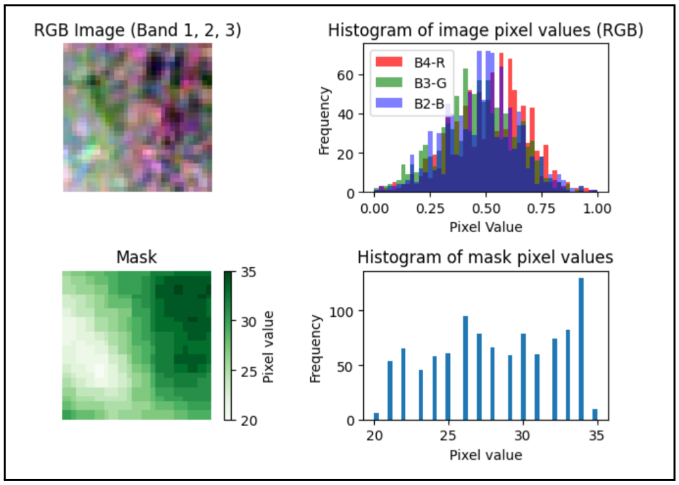

Figure 1. Example of a pair image/mask with histograms

3 Methods

We tried three different models to solve the problem given. The first one is a full convolution network adapted for 12 bands as input and a linear regression as output. The second is the same model but with only the 3 bands B02 (blue), B03 (green) and B04 (red) as input. Those bands are selected from Sentinel-2 assuming that bands haven't been changed when creating the dataset. This has been decided in order to measure how the increased number of bands that can affect the performance of the model. The final model employed a ResNet architecture initialized with pre-trained weights, using the aforementioned three bands (RGB) as input since the pre-training has been made with RGB images and ResNet only accept 3 bands as input.

3.1 Data-loader

In this study, we implemented a custom Dataset class called Sentinel2 using PyTorch's Dataset and DataLoader classes to handle the loading and processing of the Sentinel-2 images and their corresponding canopy height masks. The dataset is already splited following a CSV file, *data_split.csv*, which contains the paths to the images and their labels, along with the dataset split allocation train/validation/test. Based on the specified split, the appropriate images and labels can be loaded for training, validation or testing.

The images are loaded using the Rasterio library (Gillies et al., 2013–), which is particularly suited to reading geospatial raster data such as Sentinel-2 images. Rasterio allows the extraction of 12 bands images. However, a flag in the Sentinel2 class is implemented to trigger only RGB bands when this is more suitable as training on 3 bands models is done. The corresponding canopy height masks are also loaded using Rasterio, the height values is read on the unique band.

After loading and transforming (subsection 3.2) the image and mask, they are converted to PyTorch tensors. This conversion is essential for compatibility with PyTorch models, which expect inputs in the form of tensors. The image tensor retains its shape (C, H, W), where C is the number of channels and H, W the Height and Width of the image (32x32). As the mask is only values, it is reshaped to include a channel dimension leading to a tensor shape (1, H, W).

3.2 Data Augmentation

Data augmentation is a widely used technique in deep learning to improve model generalization by artificially increasing the variability of the data. This is a good method to improve the model's ability to generalize and reducing the risk of overfitting (Shorten & Khoshgoftaar, 2019).

As the images are in 12 bands, customs transforms is implemented to accept as much bands as needed. including 3 and 12 bands images. Thus, Sentinel2 class was created supporting those custom transformation that are applied to both the images and the masks. These transformations are defined through a set of classes that perform three data augmentations: random horizontal flip, random vertical flip and random rotation.

For the two first ones the transformation randomly flips the image and mask along the horizontal or vertical axis, with a probability set to 50% of flipping. The last one, correspond to a random rotation of the image and mask by 90°, 180° or 270° again with a probability of happening of 50%.

Those transformation were implemented using the basics of numpy to allow customization for X bands and so both images and masks are converted to tensor before passing to the model.

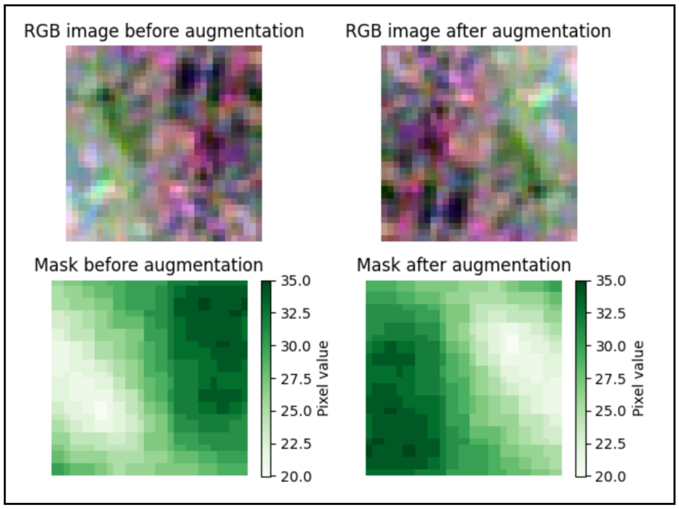

Figure 2. Example of a pair image/mask data augmented

3.3 Convolutional neural network architecture

The challenges for the convolutional neural network architecture were to be able to have 12 or 3 bands images as input and a height regression as output. The implementation must accept the number of bands as parameters to train different models. We decided to adapt an U-Net architecture (described in (Ronneberger et al., 2015)) to our needs.

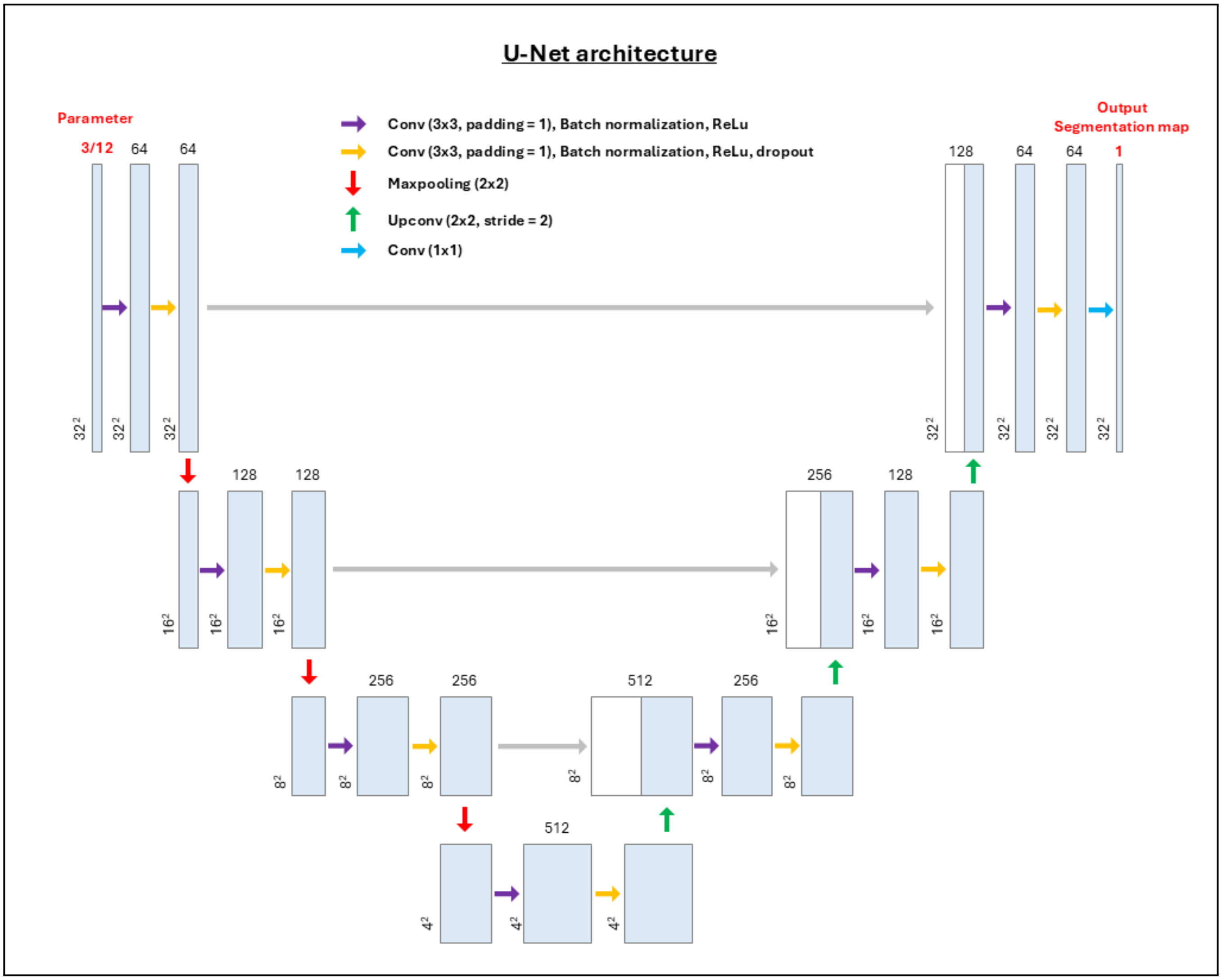

Figure 3. Architecture of the model

Figure 3 describes how we adapted the model to our problem. First, the input of the system can be adapted to be either 3 or 12 spectral bands. For all convolutional layers, we choose padding of 1 to maintain the original image dimensions, as the input size of 32x32 is already small. Additionnaly, it was decided to use a 4 layer architecture. The last layer of a U-Net network is important for overfitting, as it encodes the most abstract and spatially compressed representation of the input image. In this case, a 2x2x1024 image size would have been the last layer in a 5 layers U-Net network, and we tought that this size of activation map was too small and could lead to overfitting due to limited spatial information.

3.4 Loss function

A Smooth L1 has been choosen as a loss for our model. This was decided empirically after different loss testing. The smooth L1 loss behaves like a linear loss if the error is more than β, and like a quadratic loss if the error is smaller than β. In our problem, the canopy height error is mostly over 1, which means that the L1 Loss operates in the linear regime. This reduces the impact of the outliers on the training of the model, while encouraging the refining of the prediction below the threshold of 1. Using $\beta = 1$, the loss is defined as:

$$l_n = \begin{cases} \frac{0.5 \cdot (x_n - y_n)^2}{\beta}, & \text{if } |x_n - y_n| < \beta, \\ |x_n - y_n| - 0.5 \cdot \beta, & \text{otherwise.} \end{cases}$$

All the "no-data" values (255) are ignored in the computation of the loss for each batch. The mean loss is calculated solely based on the valid loss values.

3.5 Metrics

To assess our results and to be able to compute the errors, we decided to use the MAE metrics. It allows us to know the average error for the prediction of the canopy height in meters as done in by the CERTH and the "Université Paris Saclay" alagialoglou2021canopy and schwartz2024high. The MAE is calculated as below:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

3.6 Hyper-parameter Optimization Process

We choose to vary primarily the learning rate of the gradient descent, and the weight decay parameter. To narrow down the best values for those parameters, we implemented a training loop with different learning rate and weight decays, eith each combination being trained on 30 epochs. After multiple training sessions with fewer epochs, we were able to identify the optimal parameters and then train the model for more epochs. We plan to choose the optimal parameters by computing the average loss on the validation set for the last 20 epoch of the training. This method ensures that the model reaches a steady state, reducing the fluctuation of the noise earlier in the process. However, the selected model, which will be saved, will be the one corresponding to the epoch with the best performance on the validation set, based on the chosen hyper-parameters.

3.7 Pre-trained model

Then to assess the usefulness of our model we decided to compare it with a pre-train model, which is a model that have already been trained on a large dataset. The model ResNet-101 was used to compare its performance with our model, it has been trained on the ImageNet dataset which contains millions of images of different categories. It is a convolutionnal neural network (CNN) designed to process RGB images, which are 3 bands images: red, blue and green. However, this model has been trained to extract and classify useful features from those channels such as textures, edges or colors. Using this raw model to determine canopy height linear regression could lead to interesting results. In our case, the model was sligthly modified to give a linear output. By using this model and comparing both MAE, it allows us to know if it is useful to train a model or if a pre-train can have the same effectiveness see documentation.

4 Results

4.1 Training results and best models

During the training of the model, we computed the Smooth L1 Loss over the training and validation dataset. The results for each model is printed as a curve representing the evolution of the training and validation loss across the epochs. For the four models, it has been found that the best learning rate and the best weight decay were each time the same, respectively $l_r = 10^{-3}$ and $w_d = 10^{-5}$. This combination of hyperparameters arise the best mean smooth L1 loss on the validation set, listed as below:

Model Configuration	**Smooth L1 Loss**
12 bands model data augmented $l_r = 10^{-3}$, $w_d = 10^{-5}$	5.30
12 bands model $l_r = 10^{-3}$, $w_d = 10^{-5}$	4.62
RGB bands data augmented $l_r = 10^{-3}$, $w_d = 10^{-5}$	4.87
RGB bands $l_r = 10^{-3}$, $w_d = 10^{-5}$	4.53

Table 1. Performance on validation using different configurations with Smooth L1 Loss.

On each of those hyper-parameters, we choose to save the model from the epoch that performs the best on the validation set. All the training session results have been recorded and are listed in the appendix *results_1001_v3.pdf* available on the git repository of this project.

4.2 Performances on test set

Afterward, we used the best model to estimate the performance on the unseen test data. We used the MAE metric over the whole test set. Table 2 present the associated mean absolute error in meters of the four customs models depending on their configuration and the ResNet-101.

Model Configuration	**MAE [m]**
12 bands model data augmented	5.29
12 bands model	4.84
RGB bands data augmented	4.85
RGB bands	4.27
ResNet101 pretrained model	136.19

Table 2. Performance on test set using different configurations.

The evolution of the training can be observed on the Figure 4. We evaluated a batch of four images, generating predictions from the four models and comparing them with the corresponding ground truth.

Each model is also used to predict a random images of the test set, this is shown in Figure 5 with the corresponding mask (Ground Truth). The MAE associated with each prediction is also printed under each image.

5 Discussion

To begin with, the pre-trained model has a very high error corresponding to more than 4 times the values that is supposed to be predicted. This result is not surprising because this model has not been trained for that application. This highlights the importance and benefit of developing a model designed for our application.

In constrast to the pre-trained model, our custom U-Net showed more encouraging results. During the training, we observed on the Figure 4 that the *training set* and *validation set* errors both start decreasing. We therefore have also a decrease of the bias and variance. This decreasing tendency of the error stops at around the 20^{th} epoch (Figure 4), where the *validation set* loss starts oscillating around the point reached as a minimum. At this point, *training set* error continues to decrease very slowly. We identified this point as being a local minimum reached by our model, because the validation error is not increasing but rather oscillating around it local minimum.

The local minimum may be due to the hyper parameters: learning rate and weight decay, indeed maybe our learning rate is too high, but for lower learning rate the convergence was too slow. So to decrease the MAE we could continue to optimize our learning rates. Another interesting direction for the optimization would be taking a lower learning rate and train the model over many more epochs.

Another way to improve our results would have been coupling our data with relevant features from other satellites. This has been done by the "Université Paris Saclay" which also used the bands from the satellite Sentinel 1 (Schwartz et al., 2024).

With the mean MAE of the last twenty value, we have determined that our best model was RGB (3 bands) without data augmentation. In addition, the worst model is the one with 12 bands with data augmentation. This leads to a counterintuitive conclusion: having only RGB bands and not using data augmentation are two factors that benefit our model because each of those parameter results in a lower MAE.

RGB bands were chosen for the implementation of the three band model because of their simplicity, but other bands could be chosen. Theoretically, the more important bands for canopy height would be IR and NIR bands because they allow good vegetation detection and also discriminate chlorophyll. Perhaps twelve bands as input to the model is too many. There are some irrelevant features that add noise. This noise prevents the model from improving it's accuracy. Too many features can actually be bad for the model as it can lead to overfitting, the model captures noise instead of the underlying patterns, resulting in poor generalisation to new data (Hanka & Harte, 1997). Also, a PCA to determine which bands are the most useful and relevant bands could have been implemented to focus on only those bands and train the model on them.

We also have to explain why MAE is increasing with data augmentation although it is supposed to do the opposite. We suspected two main reasons for this behavior. Firstly, the transformations applied can introduce inconsistencies, making the synthetic data less representative of reality. This can degrade model performance (Lin et al., 2024). Secondly,some data augmentation methods, which are effective for classification tasks, may not be suitable for regression tasks, such as canopy height prediction. Inappropriate use of these techniques can lead to increased error (Hwang & Whang, 2021). This is however surprising, as the augmentation methods we chose were only based on mirroring and right angle rotations. It's possible that by learning with some transformed images, the model lost some locality in favor of robustness on the validation/test set, and this trade-off may not have worked in our favor.

5.1 Prediction comparison

As stated above, our best model during the training and validation is the one using RGB and no data augmentation. However, Figure 5 state the opposite, please keep in mind that the color scale is different for each features. Although the global shape pattern is well detected for the four customs models, the mean absolute error per image

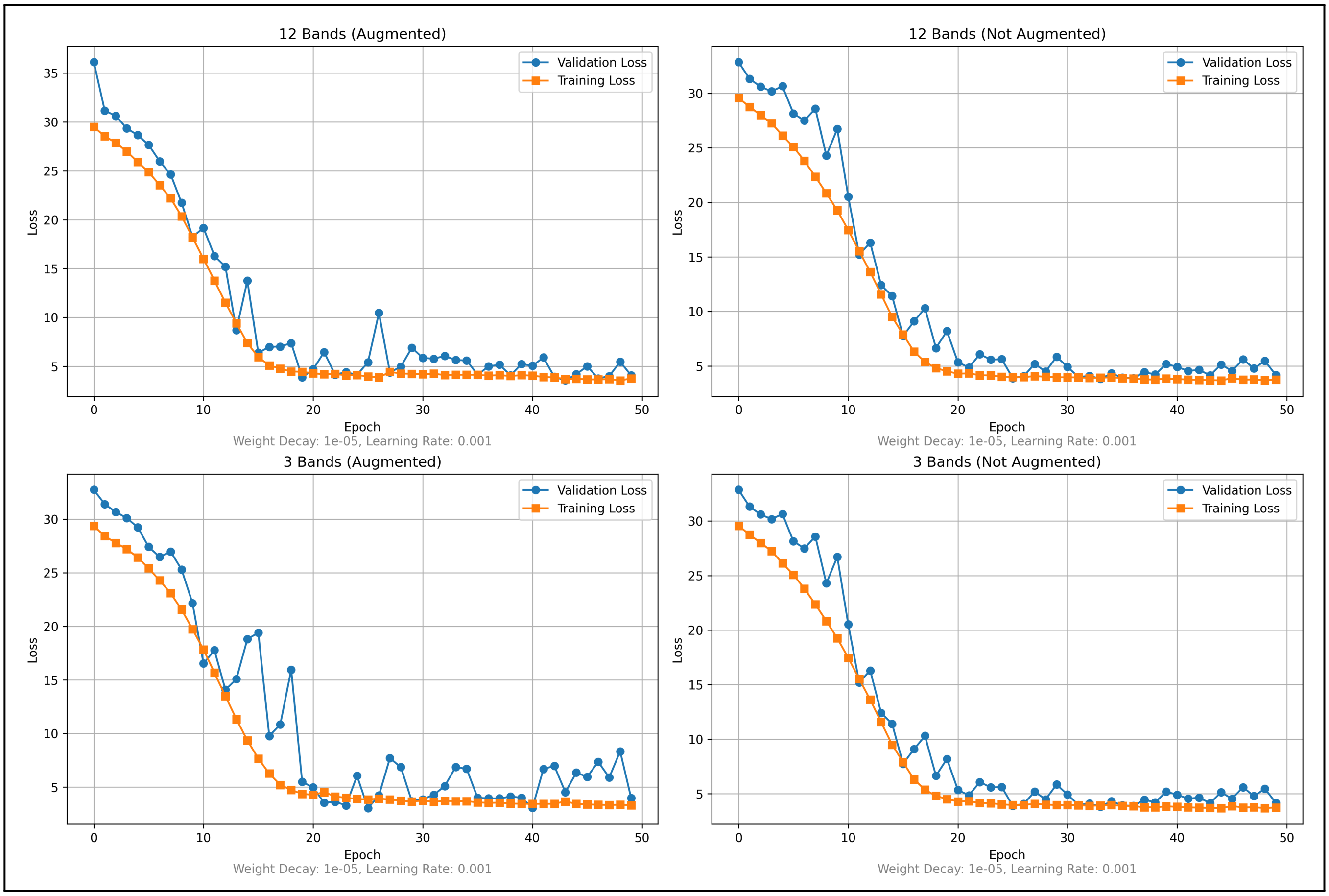

Figure 4. Results of the training

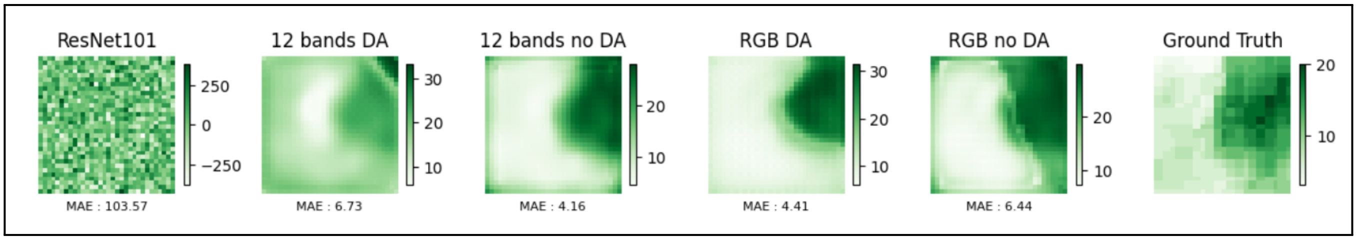

Figure 5. Image predictions

draw a surprising conclusion. Indeed, when looking at the MAE the best model would be the one with 12 bands and without data augmentation, the second one would be the RGB data augmented one, and then come the RGB model without data augmentation. This is not representative of the all dataset, as the lower MAE for each sample, differ from model to model for the four images. This kind of random pattern is probably comes from that some model are more adapted to certain features while other are more adapted to others. In the end, there is always at least one prediction that reflect quite good the ground truth. Thus to reduce the variability in between samples, a combination of the four customs models could be implemented. This super-model should then be able to take the strength of each model individually and avoid their weakness.

6 Conclusions

This work shows the effectiveness of deep learning, here based on U-net model architecture for estimating the canopy height with Sentinel-2 images. By comparing with pre-trained, non specialize models the mean absolute error (MAE) has significantly decreased, which shows the importance of developing our own model. The best MAE was 4.27 m which is more than for others models like the one developed by the "Université Paris Saclay" (Schwartz et al., 2024) which has a MAE of 2.02 m but also uses the data from Sentinel-1. Our model could also have been improved by doing PCA as a pre-processing step, so as to select only relevant features and reduce the noise, indeed there are probably noise comming from irrelevant bands.

The use of data augmentation has not shown expected benefits and has even shown a negative effect, which could be attributed to the possibility of inefficiency of these methods when applied to linear regression.

Further analysis to remove irrelevant features with PCA, and the add of other relevant features for examples some spectral bands from Sentinel-1 could be set up to increase the precision of our model.

In conclusion, this study presents promising prospects for cost-effective and accurate canopy height estimation while highlighting challenges to be addressed for greater generalization and optimization.

References

Alagialoglou, L., Manakos, I., Heurich, M., Červenka, J., & Delopoulos, A. (2021). Canopy height estimation from spaceborne imagery using convolutional encoder-decoder. *MultiMedia Modeling: 27th International Conference, MMM 2021, Prague, Czech Republic, June 22–24, 2021, Proceedings, Part II 27*, 307–317.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., et al. (2020). The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography. *Science of remote sensing*, *1*, 100002.

Egoh, B., Drakou, E. G., Dunbar, M. B., Maes, J., Willemen, L., et al. (2012). *Indicators for mapping ecosystem services: A review*. European Commission, Joint Research Centre (JRC) Ispra, Italy.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., . . . Zaehle, S. (2019). Global carbon budget 2019. *Earth System Science Data*, *11*(4), 1783–1838. https://doi.org/10.5194/essd-11-1783-2019

Gillies, S., et al. (2013–). *Rasterio: Geospatial raster i/o for Python programmers*. https://github.com/rasterio/rasterio

Hanka, R., & Harte, T. P. (1997). Curse of dimensionality: Classifying large multi-dimensional images with neural networks.

Hwang, S.-H., & Whang, S. E. (2021). Regmix: Data mixing augmentation for regression. *arXiv preprint arXiv:2106.03374*.

Kunreuther, H., Gupta, S., Bosetti, V., Cooke, R., Dutt, V., Ha-Duong, M., Held, H., Llanes-Regueiro, J., Patt, A., Shittu, E., et al. (2014). Integrated risk and uncertainty assessment of climate change response policies. In *Climate change 2014: Mitigation of climate change: Working group iii contribution to the fifth assessment report of the intergovernmental panel on climate change* (pp. 151–206). Cambridge University Press.

Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., & Wegner, J. D. (2022a). Global canopy height regression and uncertainty estimation from gedi lidar waveforms with deep ensembles. *Remote sensing of environment*, *268*, 112760.

Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., & Wegner, J. D. (2022b). Global canopy height regression and uncertainty estimation from gedi lidar waveforms with deep ensembles. *Remote Sensing of Environment*, *268*, 112760. https://doi.org/https://doi.org/10.1016/j.rse.2021.112760

Lin, C.-H., Kaushik, C., Dyer, E. L., & Muthukumar, V. (2024). The good, the bad and the ugly sides of data augmentation: An implicit spectral regularization perspective. *Journal of Machine Learning Research*, *25*(91), 1–85.

Mitchard, E. T. (2018). The tropical forest carbon cycle and climate change. *Nature*, *559*(7715), 527–534.

Mitchard, E. T., Saatchi, S. S., Baccini, A., Asner, G. P., Goetz, S. J., Harris, N. L., & Brown, S. (2013). Uncertainty in the spatial distribution of tropical forest biomass: A comparison of pan-tropical maps [Cited by: 191; All Open Access, Gold Open Access, Green Open Access]. *Carbon Bal-*

ance and Management, *8*(1). https://doi.org/10.1186/1750-0680-8-10

Rodríguez-Veiga, P., Wheeler, J., Louis, V., Tansey, K., & Balzter, H. (2017). Quantifying forest biomass carbon stocks from space. *Current Forestry Reports*, *3*, 1–18.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical image computing and computer-assisted intervention – miccai 2015* (pp. 234–241). Springer International Publishing.

Schwartz, M., Ciais, P., Ottlé, C., De Truchis, A., Vega, C., Fayad, I., Brandt, M., Fensholt, R., Baghdadi, N., Morneau, F., et al. (2024). High-resolution canopy height map in the landes forest (france) based on gedi, sentinel-1, and sentinel-2 data with a deep learning approach. *International Journal of Applied Earth Observation and Geoinformation*, *128*, 103711.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, *6*(1), 1–48.

Zhang, G., Ganguly, S., Nemani, R. R., White, M. A., Milesi, C., Hashimoto, H., Wang, W., Saatchi, S., Yu, Y., & Myneni, R. B. (2014). Estimation of forest aboveground biomass in california using canopy height and leaf area index estimated from satellite data. *Remote Sensing of Environment*, *151*, 44–56.

Zhang, J., Nielsen, S. E., Mao, L., Chen, S., & Svenning, J.-C. (2016). Regional and historical factors supplement current climate in shaping global forest canopy height. *Journal of Ecology*, *104*(2), 469–478. https://doi.org/https://doi.org/10.1111/1365-2745.12510