A Research Proposal on Improving Convolutional Neural Network Predictions of Deforestation Using Geographic Information Systems Data

Jord K.J.F. Beek

August 2022

Supervisor: M. Cochez
Second reader: M. Movahed

A thesis completed to obtain the degree of Bachelor of Science in Artificial Intelligence at the Vrije Universiteit Amsterdam

Contents

1 Introduction

1.1 Climate Change and Deforestation

Currently, a significant issue facing mankind is climate change. This exists in various forms such as biodiversity loss, heating temperatures and deforestation. The balance of life on Earth consequentially becomes increasingly damaged until a non-reversible threshold is crossed. These issues are all highly interconnected and while minimizing them all are essential, it is much too grand to address as a whole. Therefore, the scope of this research will focus specifically on deforestation. Coupled with the fact that deforestation rates do not appear to clearly be reducing, it seems that this would be highly appropriate to focus on [1]. In fact, according to the Global Forest Watch, forest loss has remained "stubbornly high" in 2021, losing "11.1 million hectares of tree cover" in tropical areas [2]. The knock on effects that this has for the ecological balance of earth are severe. Commonly referred to as the 'lungs of the Earth', large tropical forests absorb billions of tons of carbon dioxide a year, serving as a significant carbon sink to balance the high volume of carbon sources humans are responsible for [3].

1.2 Machine Learning as a Solutional Aid

While machine learning (ML) may not directly reduces emissions or our carbon footprint like sustainable energy technology does, it can still be used as indirect aid to eventually arrive at the same outcome. ML may also for example serve us as a guide through which oversight can be gained and plans from which appropriate action can be taken. Through this oversight, economic and sociological problems stemming from widespread illegal logging can also be addressed [4]. To achieve this, remote sensory data in the form of satellite imagery over large areas could be used as input on which predictions and classifications can be made. This approach is not novel either. ML models are continually being improved and used in various ways to track and potentially predict deforestation using remote sensing technologies already [5]. Various studies so far have shown ability and promise in classifying deforestation already through newer deep learning methods as well as the established classical ML methods.

One such example can be seen in a study conducted by Mayfield et al. [6] in which freely available georeferenced land use images (images locatable through an internally embedded coordinate system) are evaluated on their quality as neural network inputs. With these land use images, risk maps are created with the goal of predicting the risk of future deforestation in Mexico and Madagascar. This implementation is done through both classical and deep learning machine learning methods such as Bayesian networks, artificial neural networks and Gaussian processes. In practice, the risk maps of all three of these methods scored an area under the curve (AUC) value above 0.8 which - according to Platts et al. [7] - indicates a significant forest/deforestation distinction can be made. Overall, this study found machine learning can serve as a reliable alternative to traditional statistical methods in deforestation risk modelling.

In another study, conducted by de Bem et al. [8], convolutional neural network (CNN) architectures are used along with two older ML methods, namely random forests (RF) and multilayer perceptrons (MLP).

The CNN architectures chosen were Sharpmask, U-Net and ResUnet; three architectures that, at the time (2020), were all considered state-of-the-art. The older ML methods at use would serve as a control group on which the deep learning architecture could be compared against. What was found that while all model approaches were effective (ResUnet being most effective), the older ML models also required additional post-processing noise removal to get the performance up to par with the deep learning architectures.

Table 1: Performance measures of the models used by de Bem et al.

Model	2017–2018						2017–2019					
	F1	Kappa	mIoU	Precision	Recall	Overall Accuracy	F1	Kappa	mIoU	Precision	Recall	Overall Accuracy
RF	0.8014	0.8003	0.8332	**0.9414**	0.6976	0.9979	0.8902	0.8892	0.9000	0.8877	0.8928	0.9979
MLP	0.8926	0.8920	0.9024	0.9282	0.8597	0.9987	0.9101	0.9093	0.9167	0.9314	0.8898	0.9983
Resunet	**0.9432**	**0.9428**	**0.9459**	0.9252	**0.9619**	**0.9993**	**0.9465**	**0.9460**	**0.9487**	**0.9358**	**0.9574**	**0.9990**
Unet	0.9112	0.9106	0.9179	0.9223	0.9003	0.9989	0.9339	0.9332	0.9373	0.9175	0.9508	0.9987
Sharpmask	0.9223	0.9218	0.9274	0.9173	0.9274	0.9990	0.9337	0.9331	0.9372	0.9218	0.9460	0.9987

Table 1 shows the performance of each of the algorithms on two different datasets. What is apparent is that the older machine learning algorithms underperform relative to the deep learning algorithms in almost every performance measure. The only exception being the RF algorithm which obtains the highest precision score but lowest score in other measures. The study further shows that ML is a suitable approach for recognizing deforestation. Furthermore, it additionally demonstrates that the new deep learning ML methods in particular are more optimal over older methods.

In a 2004 study by Mas et al. [9], a multilayer perceptron was used to classify deforestation. This was done using Landsat data ranging from 1974 through to 1991. What makes Mas et al.'s research unique is that they decided to formulate their classification with not only the Landsat data, but also using variable geospatial data found in the images such as roads and rivers. Similarly to the approach by Mayfield et al., risk maps are created where deforestation is identified to have occurred or not. However, the approach used by Mas et al. utilizing spatial data was sub-optimal (by today's standards) as the accuracy of their model came out at a mere 69%.

1.3 Personal approach

The scope of the research conducted in this paper aims to pick up concepts from such studies mentioned prior. A considerable length of time has passed since the study by Mas et al. for instance and in the world of machine learning, progress is continuously made and the standard of what is considered state-of-the-art is rapidly evolving. As seen in other studies such as those performed by de Bem et al. and Mayfield et al., deforestation classification has already come a considerable way. The central question we thereby hope to answer has two parts. Firstly, whether a CNN can be used to obtain optimal results in predicting deforestation based on remote sensed data as input. Considering that de Bem et al. managed to achieve performance measures well over 90% for the CNN's, this is defined as the minimum threshold for an optimal performance. From there, the focus turns to laying out a preliminary framework to creating a multispectral extension to this ResNet model whereby geospatial data is also utilized from geographic information systems (GIS). The question for this framework is whether this could aid the model's performance.

2 Finding a Dataset

2.1 Selection Process

2.1.1 Requirements and Challenges

With remote sensing technologies becoming more advanced and thoroughly developed, the availability of forest satellite imagery is both growing in accessibility and in quality. This can be seen through freely accessible high resolution images from the continuously developing Landsat program initially started in the 1970s [10] along with newer innovative startups like Planet.com since 2010 [11]. While these dataset sources in their raw form are cutting edge in quality, there are other requirements that must also be fulfilled in order for the data to be suitable for the scope of this research.

The first criterion that must be met is image quality. This is viewed in terms of resolution and visibility. The implication this has is that the volume of data obtained in the 1970s and 1980s when imaging programs were still in their infant years have since become obsolete in favor of the more detailed and informative imaging programs introduced later on. Presently, the standard for ideal image resolution is 30x30 meters per pixel [12]. A second criterion is that the data needs to be labeled. That is, for a given area in the satellite image, there needs to be an indication of whether deforestation has occurred or to what extent. This greatly narrows down the number of suitable datasets as satellite images on their own will not suffice. Furthermore, as we hope to extend our CNN to be able to utilize geographic information too, dataset must be georeferenced for an identifiable location.

2.2 Considered Datasets

One dataset that showed to be promising was a 2017 Kaggle competition with Planet satellite imagery [13]. Participants would have to classify images based on whether they fell under categories such as primary forest, roads, cloudy, cultivation and others. This resulted in a vast labeled dataset being accessible that also happened to be well documented thanks to the community taking part in the competition. The problem with this dataset though was that georeferenced satellite images were not accessible, meaning that the image locations could not be retrieved.

Another consideration was to obtain satellite imagery from Planet directly [11]. Through their API, their database of images could be accessed and images of chosen rainforest areas could be requested. The benefits of this approach is that the images would be high resolution, recent, regularly updated and could be located through georeferencing. However, because Planet does not offer labeled information in regards to forest gain or loss, labels would have to be manually created. Being too costly in terms of time and effort, this dataset was subsequently abandoned.

Eventually, a dataset by Hansen et al. from the University of Maryland [14] was found which seemed to fit all the criteria. This dataset uses recent Landsat imagery and overlays it with a deforestation label layer. A visual representation of the deforestation depicted by the dataset can be seen in Figure 1. This source contains deforestation data of all non-ocean areas on earth. It can be accessed through a global grid of mosaic tiles in which each tile is tailored specifically to forest change identification. This can be seen in Figure 2. Each tile is of size 10x10 degrees (latitude and longitude) which corresponds to roughly 1000x1000km at the equator. Data available per tile includes satellite imagery (captured in 2000 as well as another dataset for 2021), tree coverage loss (between 2000 and 2021, which can be seen in the background of figure 2), tree coverage gain (between 2000 and 2012) and overall tree coverage (in 2000). The first datasets (2000 and 2021) are RGB format satellite images while the latter three datasets are boolean maps that indicate forest presence or absence with respect to forest loss, forest growth and forest presence, respectively. The datasets of particular interest in this research are the satellite imagery of 2021 along with the tree coverage loss dataset. In the corresponding paper by Hansen et al. [14], it was found that the Amazon showed most clear deforestation trends. Along with the Amazon having a high level of publicity in the domain of climate change served as motivation behind our tile choice. This tile spans between Brazil, Bolivia and Paraguay, with the coordinate (15S, 55W), at the tile's center. Figure 2 shows this tile outlined in red.

Figure 1: Global deforestation with respect to the year it occurred, as depicted by Hansen et al. 2021 is represented by turquoise, 2020 by red, 2000 by yellow and all years in between 2000 and 2020 in orange.

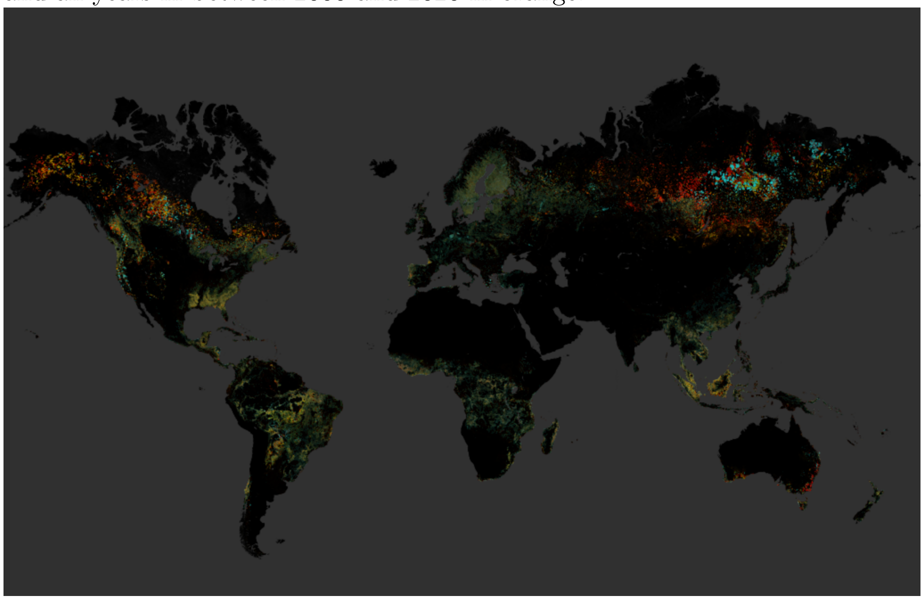
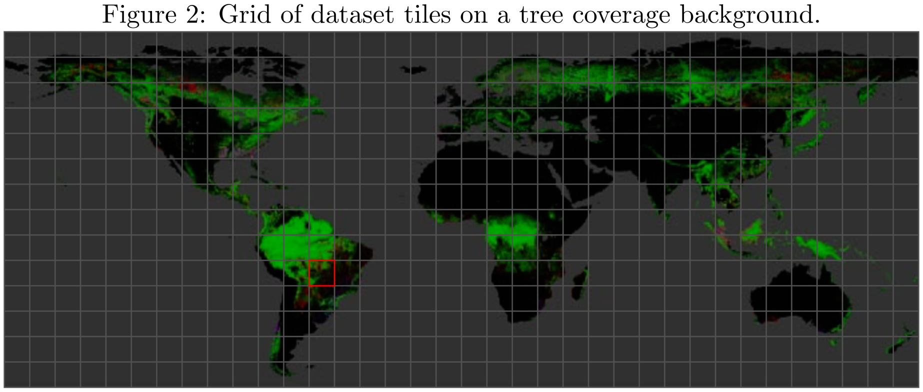


Figure 2: Grid of dataset tiles on a tree coverage background.



Some key strengths of using this data is that cloud interference has already been dealt with (images with cloud interference are replaced with images from one year prior), images are recent (Landsat 8) and resolution is up to standard.

2.2.1 Satellite Image Dataset and Loss Labels

Figure 3: Entire satellite imagery dataset* as provided by Hansen et al. before preprocessing. *Embedding the dataset into this paper has reduced the resolution of the dataset in contrast to the original dataset used in practice.

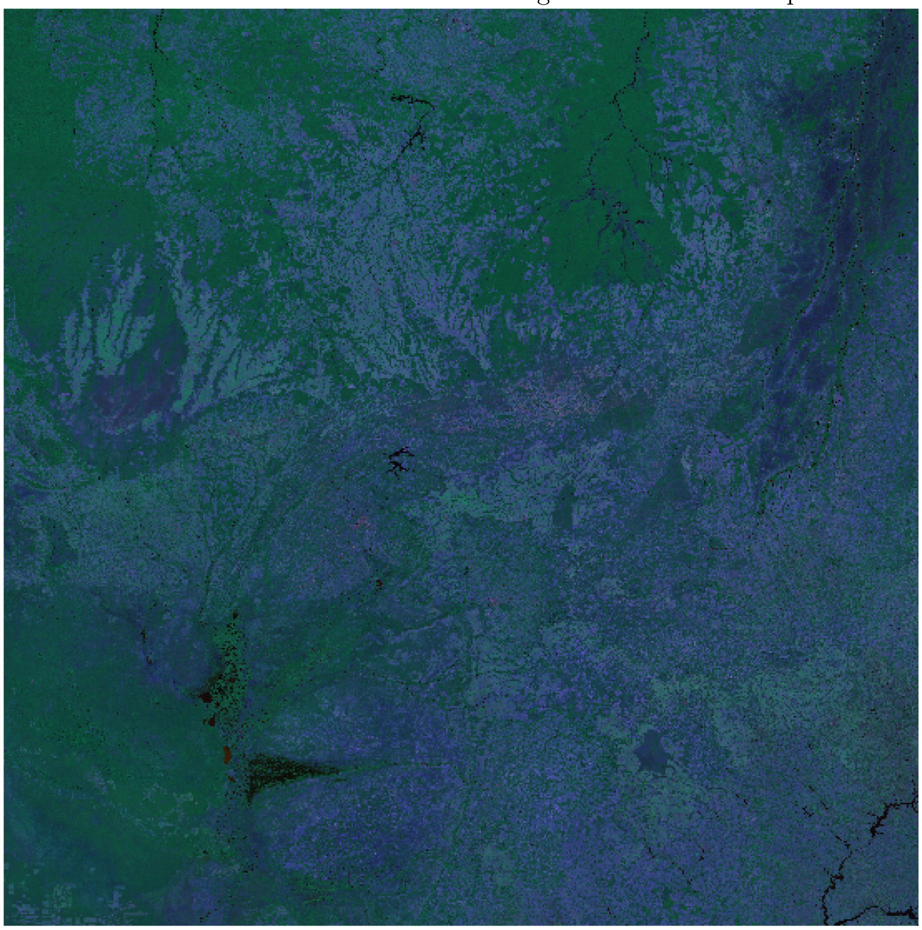
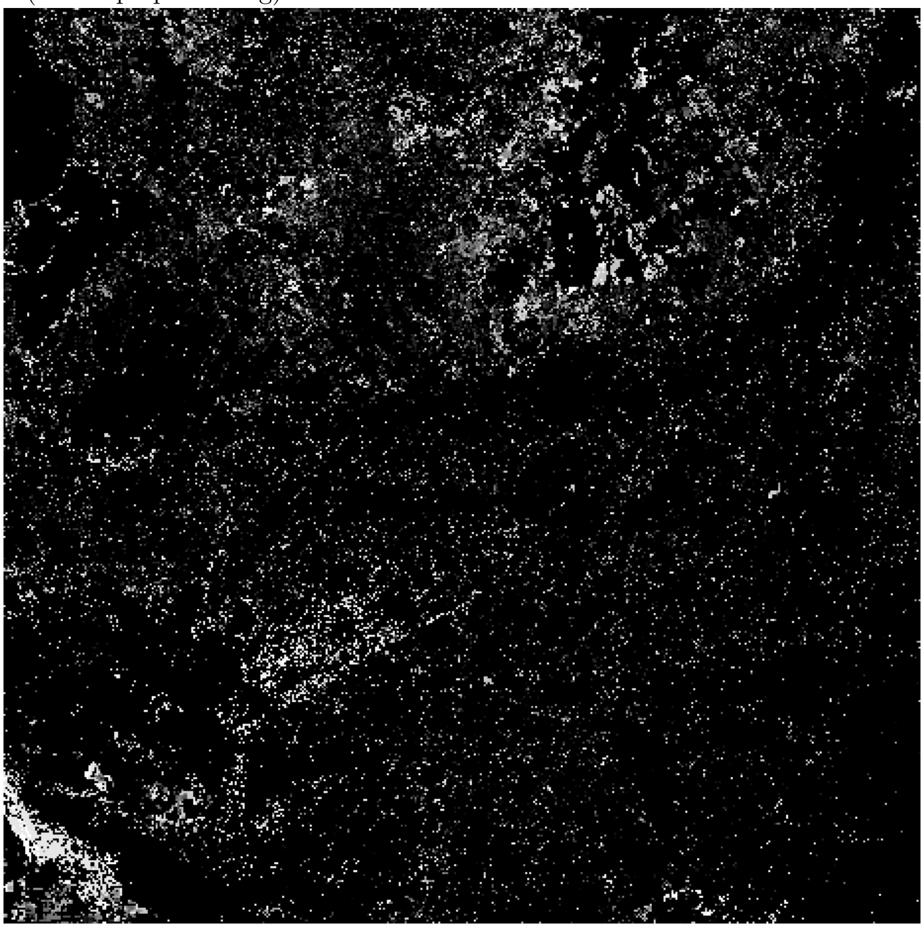


Figure 4: Raw labeled data corresponding to deforestation occurrence in figure 3 (before preprocessing).



Figures 3 and 4 show the raw datasets that are eventually used to train the convolutional neural network before any pre-processing. Figure 3 shows an image of 40,000x40,000 pixels of where forest coverage exists. While appearing low in resolution, the image becomes sharper and more refined once it is pre-processed. Note, this image also appears lower resolution due to file type changes made to include it as a figure. Figure 4 serves as the labels corresponding to the figure 3. It can be seen that pixels are either black, white or some shade of grey. Black indicates that no deforestation having occurred relative to 2000 (whether forestry is present to begin with or not), white indicates forest cover loss having occurred in 2021 and other shades of grey depicting deforestation in the years between 2000 and 2021 (the darker the shade, the closer the year to 2000).

2.3 Preprocessing with QGIS

Using QGIS (a geographic information systems interface that provides image manipulation capabilities), the dataset and dataset label images are preprocessed. Reducing the dataset image to images of 256x256 pixels, a dataset directory can be created with approximately 24,000 images that can be used to train and test our CNN. With these lower pixel images, the finer details can then also be identified, showing geographic features such as rivers, cultivation and mountain ridges. The dataset was then split into a training set (80%) and a test set (20%). This process was carried out with a random number generator.

To convert the dataset label image into a usable format follows the same process with a few extra steps. The image has the same pixel dimensions and is reduced to 256x256 pixel images. Each image in the label directory thereby corresponds to another image in the dataset directory. After this has been completed, the label images are stripped of unnecessary data such as when deforestation has occurred. Each pixel in the image either has an integer value of 0 or greater than 0 for the year of deforestation. Using python, each of the images are iterated through. Per image iteration, integer values are converted to boolean values with respect to deforestation occurring at some point or not. Next, the proportion of an image showing deforestation is calculated. So, all the pixels with a value of 1 in an image are counted. These are then added up and divided by the number of pixels in the image (65,536 pixels) for a value that represents the proportion of deforestation. This allows us to determine that in one image, deforestation occurred in 44% of the total area while in another image that value may be 2%. These values are then written to a csv file that can be fed into the CNN.

2.4 GIS Layer and Feature Querying

After having created and processed the dataset along with the labels corresponding to it, the focus shifts to creating a GIS embedded dataset for the geospatial CNN framework. Embedding geospatial data into the original dataset means creating one or more layers containing geographic feature information (the geospatial multi-layer architecture), and overlaying it on top of the original dataset as a new image. To do this, the original dataset image (before preprocessing) is loaded into QGIS. Using QGIS, OpenStreetMap's Overpass API can be queried for the presence of any selected features in the given area. This query will return any features that are present as a newly created image layer. Depending on what is appropriate for the feature, the layer indicates feature presence as lines, polygons, or pinpoints. Making these layer backgrounds transparent allows the original dataset to be visible as a base layer under the GIS features. This approach with QGIS was advantageous as experience in SQL and image manipulation was limited prior to this research.

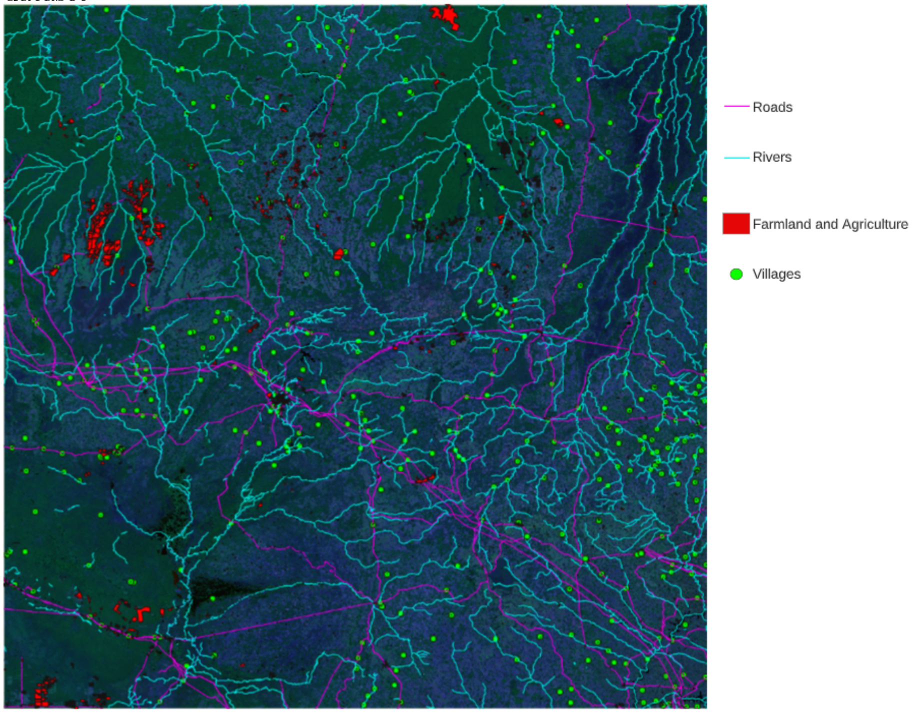
The architecture of the geospatial multi-layer architecture could be created in different ways, each adopting their own strengths and weaknesses.

One approach would be to create one single band boolean layer where each pixel in an image corresponding to the presence of queried geographic features. The advantages of this approach are its simplicity and ability to take in an endless number of features. All features can be compressed into a single layer, producing dataset images of just four bands that still incorporate many features. The downside to this approach is that no representational distinctions can be made between the presence of one feature versus another. This also limits any further research that might be interesting with respect to how influential some features are versus others.

A different approach would be to query geographic features and add each as their own layers. This would allow distinctions to be made between the presence of one feature in comparison to another through RGB colors. However, a tradeoff then exists between the quantity of feature information of an image and how computationally demanding it becomes [15]. So while having differentiable features would be optimal for more informative data, it also implies that it becomes increasingly expensive to train for each additional feature added. For each feature queried (roads, land-use, rivers, etc.), another three bands would be added to the multispectral image. To illustrate further, assuming an image was overlaid with three GIS features, it would quickly turn the three band image into a staggering twelve band image. Therefore, with a limit on the quantity of selectable features, the most influential features must be chosen in regards to how well they correlate with deforestation in order to maximize deforestation predictability and minimize computational cost.

Research on the most significant factors driving deforestation could provide insight as to what GIS features would be the most powerful indicators worth selecting. One study in regards to Myanmarese rainforests [16] found that a highly prominent factor indicating deforestation is the distance of nearby settlements from the location of forest cover loss. With this in mind, it would make sense to query for village locations. Bax et al. found in their research [17] that paved and unpaved roads were the strongest indicators of deforestation. In contrast, Armenteras et al. actually found that in the Colombian Amazon, rivers were particularly correlated with areas of forest loss [18]. Another study focusing on the Brazilian Amazon [19] found that 23% of forest clearances in 2003 were credited to land cultivation. With this information, we add villages (classified under "Settlements" in GIS), roads, rivers and streams (categorized under one umbrella term of "rivers") and farm land uses (under "Farmland and Agriculture") to the list of selected features.

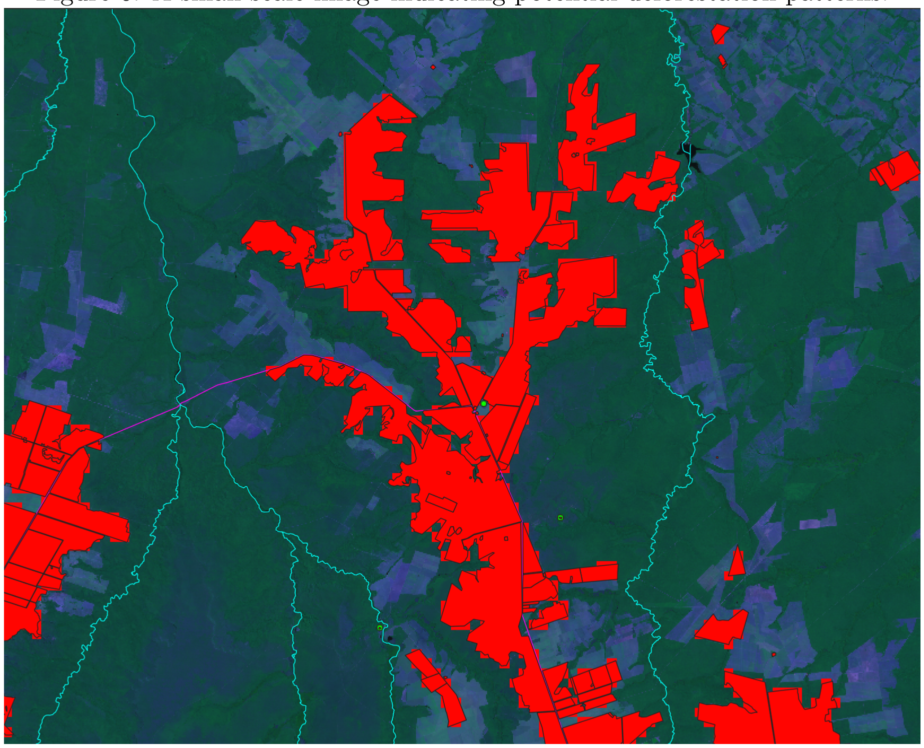
Figure 5: GIS features overlaid as additional image channels over the used dataset



In figure 5, the geospatial dataset can be seen. Note, the largest features are visible when the entire dataset is viewed. Through the same pre-processing procedure as before with the image dataset, smaller scale feature occurrences also become visible. Colors given to feature classes are deliberately made distinct from each other with little potential for color overlap. This was done to ensure different features would not fall under the same classes. This dataset is then also reduced to 256x256 images and matched to the same deforestation labels used prior.

Ideally, the network would then be able to identify patterns of where deforestation occurs in the presence of one or more features. For instance, in figure 6, an example of what a dataset instance looks like can be seen. A river system is visible running throughout the image, surrounded by farmland and singular village occurrences by a road. What can also be seen is that these features occur in the presence of patches of dense forest. The impression this gives is that forest cover loss has taken place to clear space for the village and farmland with the aid of the river and road. In Appendix A, another example can be seen of where different patterns could be found indicating forest loss.

Figure 6: A small scale image indicating potential deforestation patterns.



3 Methods

3.1 Convolutional Neural Networks

Selecting a model, it needs to be suitable for predicting deforestation as well as acting as the base model for the geospatial model. This means that the model needs to be flexible in number of image channels it can take as input, enabling geographic features to be loaded in. The conditions of the experiment consists of a large volume of images which contain pixel values that represent some potential patterns of deforestation. Neural networks have shown to be well suited for this application due to their ability to handle large volumes of data as inputs and mapping it to one flattened tensor output as a classification or regression value. Various approaches could be taken. We deliberately choose a CNN over other options such as multilayer perceptrons and traditional ML methods for various reasons. Firstly, MLP's make use of vector inputs while for CNN's, this input type is in the form of a tensor, allowing it to make spatial inferences which is beneficial for image analysis. While MLP's and CNN's are similar in the sense of being multilayered feedforward networks, each layer in an MLP is fully connected with the prior and following layer. In contrast, CNN's are sparsely connected (locally with nearby neurons in the preceding and following layer) that are only fully connected in the final layers. The resulting difference of this high connectivity in MLP's is that they become more inefficient and perform less well when faced with large numbers of feature inputs, as is the case in this application. CNN's in contrast, are designed specifically for big data applications [20].

This has been proven as CNN's perform well and thrive in large scale images or sensory data contexts [21]. Furthermore, because CNN's make use of convolutional layers, they end up working better with image inputs due to a bigger receptive field of the local pixels in a given area of an image.

Looking at the general architecture of what comprises a CNN, there are various defining aspects involved. Convolutional neural networks are made up of of multiple layer types, namely the convolutional layer, pooling layer, non-linearity layer and finally the fully connected layer. Making use of convolutional layers, the network ensures a drastic increase in efficiency as neurons compute the parameter weights only with the local neurons around it as opposed to all neurons which can be seen happening in a fully connected network. These local neurons then each pass forward calculated parameter weights to the connected neurons in the next layer. Applying this to images, this could be thought of as particular pixel values in a given local area of the image. After these convolutions are performed, a non-linear activation function is used. A common activation function is the Rectified Linear Unit (ReLU) function. In these layers, any negative values are identified and made zero, ensuring that the model does not simply become a linear classifier. This is important to maximizing predictive and classification capabilities in the network as non-linear patterns can then be identified that may be present in the input data. Pooling layers then occur along with these convolutional and activation layers. Pooling can be done in a number of ways. One such method is average pooling. Groups of pixels are merged into one pixel and given the average value of the collective pixel values. Using a filter size of two (2x2 pixels are then looked at per iteration) and a stride of two for example has the effect of reducing an image size by 75% as four pixels are replaced by a singular pixel with a value averaging the prior four. Finally, the network has the fully connected layers. In the fully connected layers, properties of traditional neural networks can be seen to return as all nodes are connected with one another as opposed to locally in previous layers. In the fully connected layers, low-dimensional features extracted in convolutional and pooling layers are fed forward. Being connected to all nodes in the prior layer, the fully connected layer can then extract the non-linear high-level feature combinations that may be present in the flattened data. Multiple other fully connected layers are connected as back-propagation occurs in which weights are reviewed and adapted and fed to these next fully connected layers. Finally, the model creates an output classification or regression value.

3.2 ResNet

The specific convolutional neural network being implemented is a ResNet model architecture. The ResNet architecture works similarly to other CNN's with a few specific details that distinguishes it. These details allow it to address particular issues that deep neural networks are commonly faced with [22]. One such issue is accuracy degradation. It was found that although deeper networks could perform better than its shallower counterparts, the training error (and subsequently testing error) was actually higher. Initially, overfitting was thought to be the cause as this too causes reduced accuracy in response to increased network depth. However, this was dismissed and the accuracy degradation was found to be due to a different reason.

Figure 7: The ResNet architecture

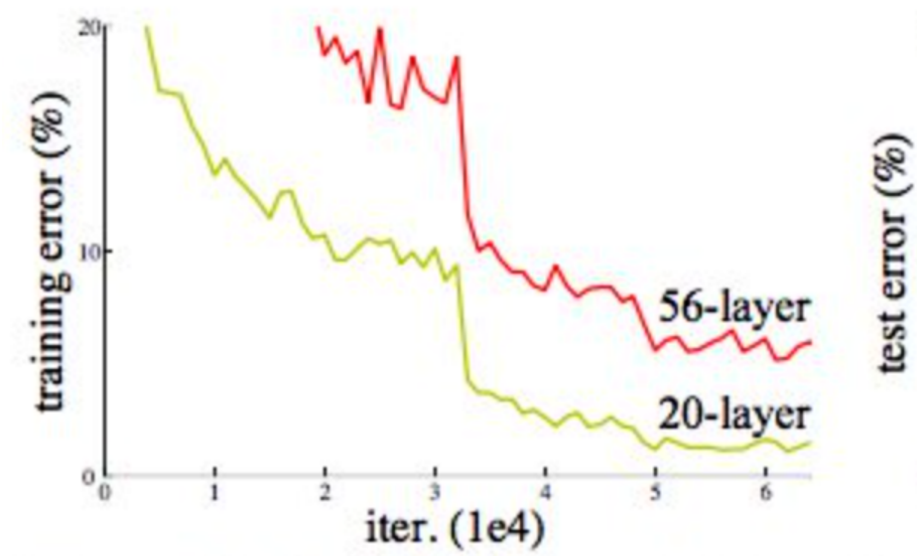
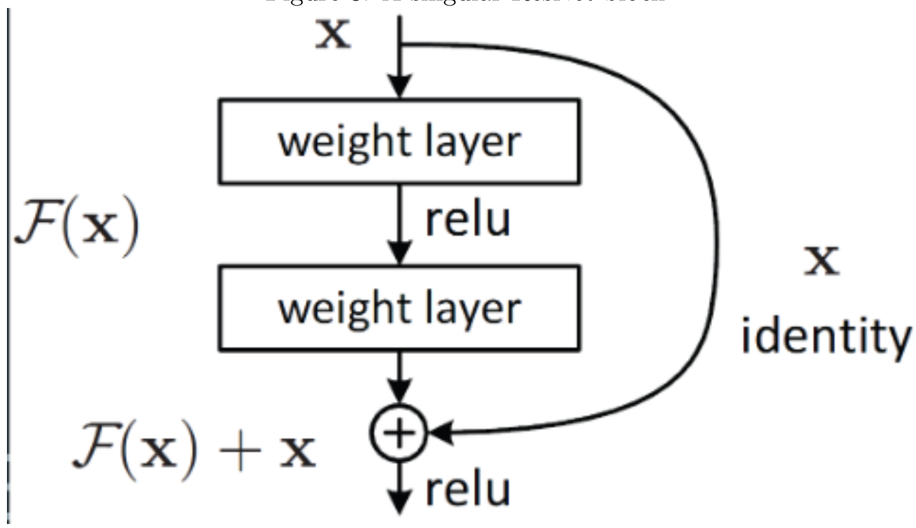


Figure 7 shows an example from the ResNet paper that illustrates this. Although the 56-layer network is deeper, it has a considerably larger validation error in comparison to the 20-layer network.
Another reason the ResNet architecture is chosen is due to the vast amount of documentation available about it online. Having minimal experience implementing convolutional neural networks, it felt more ideal to choose a model that has shown good performance, versatility in applications along with an abundance of research than to aim for what is state-of-the-art or at this current moment.

The solution ResNet poses is through a system of residual "blocks" and skip-steps. Figure 8 shows one such block. Input (X) can be seen coming from the previous block into the next. However, this weight value is then also passed forward several layers, skipping the convolutional layers in between (given by "+

Figure 8: A singular ResNet block



X"). This way, when the two distinct weights are passed through the activation function, the greatest value can be selected between the weight computation that passed through the convolutional layers in between and the weight computation that skipped past these layers. As a result, the training error should at the least, never get worse and prevent the accuracy from saturating. Lastly, ResNet uses a ReLU activation function, which poses further benefits as the vanishing gradient problem can also be avoided.

For the scope of this research, we are interested in creating a prediction (as opposed to classification) for the proportion of deforestation that has occurred in a given area, and so we include an additional final fully connected regression layer. Because we choose regression, this also means that a different error function is used. As opposed to cross entropy loss, mean squared error loss is used.

3.3 Adding GIS Input to ResNet

An advantage of using ResNet is that its parameters can be modified to take in additional image channels. As mentioned prior, adding geospatial data to an image would be through such additional channels. Ordinarily, the model assumes a three channel image input corresponding to the red, green and blue layers that collectively make up a colored image. By adapting the model to take more than three channels, GIS data can be provided to the model through the additional channels. The handling of these additional channels would not happen in a special or different manner to the RGB layers already in place. The only difference that can be seen now is that the deforestation label is associated with a different overall combination of data.

Figure 10 (see Appendix B) shows a diagram can be seen from a study by Fang et al. [23] where a hyperspectral ResNet model is created. The model takes in hyperspectral images as both spatial input (RGB) and spectral input (additional image channels depicting other electromagnetic wavelengths). In practice, this approach is conceptually the same as the approach studied here as the GIS input too, is another form of input that is combined in the fully connected layers alongside the spatial input.

4 Results

The ResNet model was run in 100 epochs and varying batch sizes at a depth of 50 layers. This was done to find what batch size variation would allow the model to perform most optimally. Batch sizes used were 4, 8, 16 and 32. The learning rate of the model was set at $1 * 10^{-3}$.

Figure 9: Performance of the ResNet model with a batch size of 4

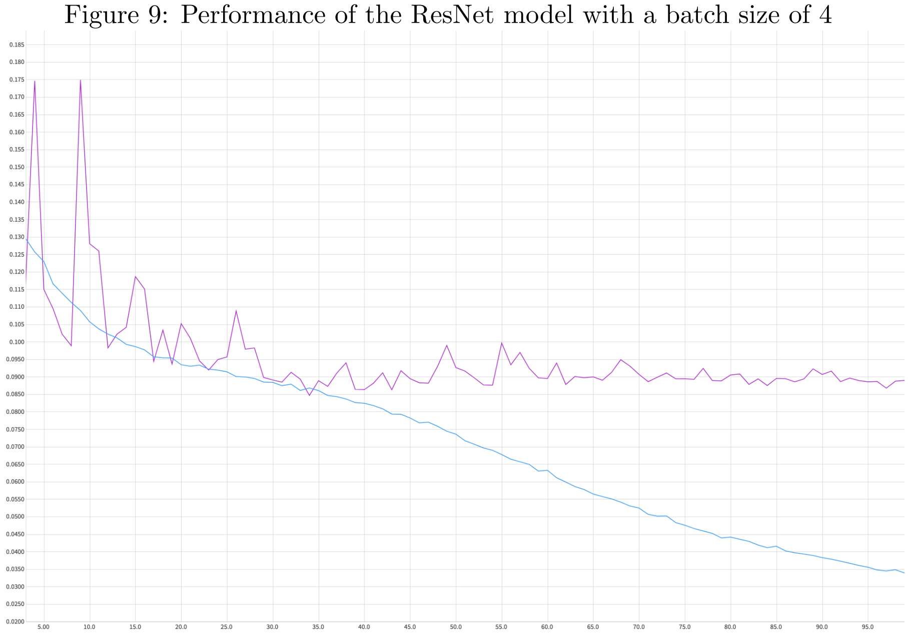


Figure 10: Performance of the ResNet model with a batch size of 8

Figure 11: Performance of the ResNet model with a batch size of 16

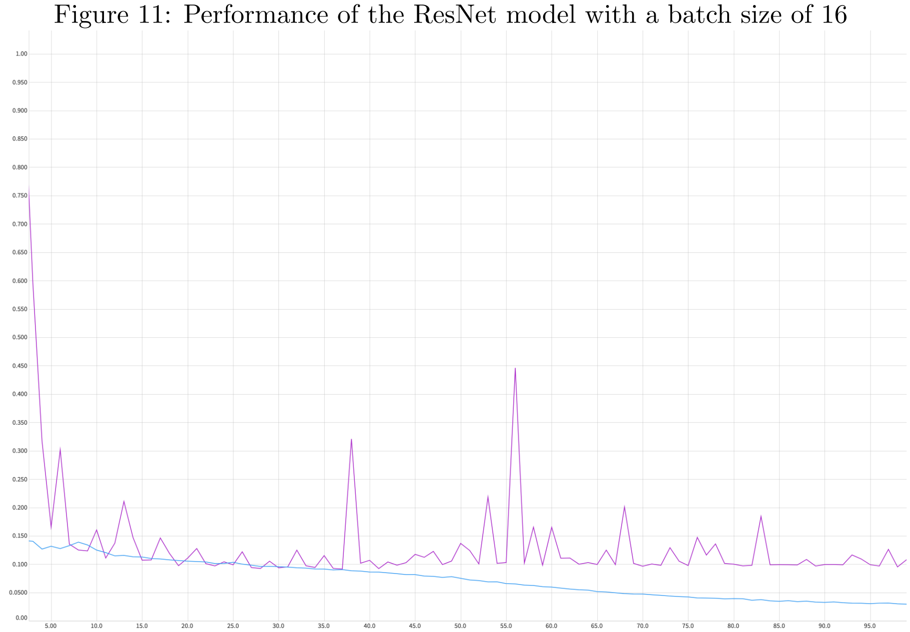
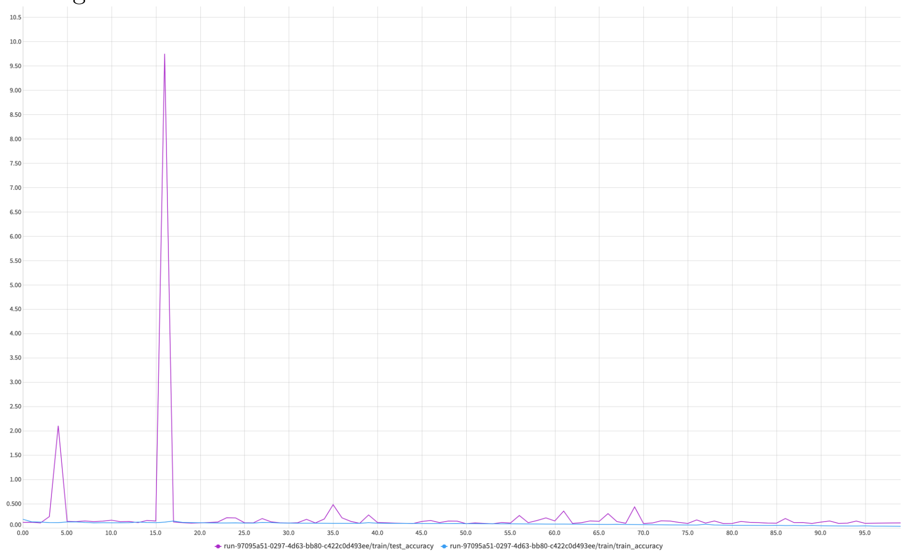


Figure 12: Performance of the ResNet model with a batch size of 32



Figures 9 through 12 show the ResNet model's root mean squared error (RMSE) performance on the train and test datasets (blue for training and purple for testing). Appendix C shows an adjusted version of figure 12 with the performance spike removed for a clearer view. Overall, there does seem to be a general trend that is followed throughout the results where the model is in fact able to make predictions that are significant. The lowest RMSE values the model achieves are 0.085 (batch size 4), 0.089 (batch size 8), 0.092 (batch size 16), 0.095 (batch size 32). Thereby, if the model were to be run again throughout all batch sizes, a safe assumption could be made that the highest performing test RMSE would fall somewhere between 0.08 and 0.1, with lower batch sizes falling closer to 0.08 and higher batch sizes toward 0.1.

Although the model shows strength in being able to exhibit predictive capabilities, it does demonstrate overfitting during training. The RMSE of the model during training becomes steadily lower but during testing on the other hand, the performance stubbornly stays around a particular threshold. What is apparent too is that the point at which overfitting becomes dominant can be identified at roughly the same point throughout all batch sizes. At around 40 epochs, the model starts to overfit on the training set, largely separating the train and test performance of the model. For the test set, the performance also generally worsens slightly after 40 epochs, although being much more subtle.

Furthermore, when the batch size is increased the model generally becomes more inconsistent. At a batch size of 4, the model rapidly increases in performance and fluctuates much less after 20 epochs. With a batch size of 16, a clear difference can be seen as frequent spikes can be seen to occur which exaggerate further with a batch size of 32. Along with the fluctuations, the extreme of the worst performing instance of the model increases with the increasing batch size. A possible explanation could be the model being more sensitive to unusual images in the dataset.

5 Discussion

5.1 Creating a Deforestation Predicting Model

One initial goal of the research in this paper was to explore the possibility of creating a CNN model that could take a satellite image of a forest and be able to accurately deduce whether deforestation has occurred. The results were not perfect as there was overfitting that undermined the model's performance. However, the model still showed a significant capability to predict deforestation on the test set. On the other hand, if the model were not able to make a distinction between deforested and non-deforested areas, the model could be expectd to be accurate only half the time. This can be said as each pixel of a given dataset image has a Boolean classification of depicting deforestation or not. One mentioned benefit of creating such a model that could identify the areas of deforestation was that it could allow for oversight to be gained. For the model to generalize optimally for such oversight, this overfitting would have to be minimized to a larger degree. It was seen that the larger the batch size, the larger the difference between training and testing performance and thus, the worse the model generalizes. It thereby becomes clear that a batch size of no more than 16 would be optimal. If for some reason a higher batch size is required (to speed up training for example), a potential work-around could be to increase the learning rate of the model.

With this issue of overfitting, it seems that adding depth would likely make it worse. In fact, making the model shallower could actually be beneficial. The hope then would be that the model might run more epochs where in performance on the test set can increase without the training set being drastically overfit on.

5.2 Extending the Model to a Geospatial ResNet Model

Another goal of the research posed in this research was to lay out a framework on which a geospatial ResNet model could be implemented. The goal with such a model is that it could improve the performance of a spatial CNN. Carrying this out has shown to be possible through deliberate geographic feature selection for the dataset and adapting the input shape that the ResNet model takes in. With the performance of the ResNet model being positive yet showing overfitting, the stage is set for potential improvement using a geospatial dataset. As mentioned prior, this dataset contains vibrantly outlined geographic features that the model can pick up on. This might subsequently aid the model in separating the signals in the dataset images from the noise.

6 Limitations and Future Work

6.1 Limitations

6.1.1 Dataset

The dataset used by Hansen et al. contains imagery that was obtained using the Landsat program. The imagery used as baselines for forest levels (relative to loss or gain years later) dates back as early as 2000 in the dawn of Landsat 7's launch. Simultaneously though, the Landsat program is also continuously evolving. Currently, forest levels are still being updated annually, now using the Landsat 9 satellites. The effect this has is that because different remote sensing technology is used, not all imagery in the datasets are exactly uniform and minor variations in map tiles occur.

Furthermore, conflicting views exist on the reliability of the Hansen et al. dataset. One study by Bellot et al. [24] claimed that the dataset actually overestimates deforestation in Indonesia, outlining occurrences of "phantom deforestation" where forest cover loss takes place in areas where there was either no loss or no forest to begin with. An accuracy assessment performed by Mitchard et al. [25] found that the dataset performed with high accuracy in Brazilian forests but failed to consistently detect forest change in Ghana.

The exact definition of deforestation or forest loss by Hansen et al. is a complete removal of forest canopy coverage at a pixel scale. This creates a grey area though as the Landsat resolution is 30 meters per pixel and forest cover loss may occur in patches of less than 30 meters. Selective logging for example is a widespread occurrence but occurs at a scale of significantly less than 30 meters. This implies that while macro levels of forest loss can be easily captured, much more precise remote sensing technology would have to be used to truly capture micro levels of forest cover loss. Generalizing the potential of this dataset, it would mean that issues such as illegal selective logging would go undetected.

6.1.2 ResNet Model

The ResNet model used was a model that was created from scratch and therefore not pre-trained. Not using a pre-trained model and creating one from scratch raised questions as to whether the performance of the model could otherwise have been higher. Furthermore, ResNet, being a deep learning architecture, is by nature prone to taking a long time to train. Using a pre-trained ResNet model would have allowed training time to be minimized more. The model would not have had to train from scratch and would already have some kind of foundation by training on a public database (ImageNet for example). Our task would then be to apply it to our own application.

6.1.3 GIS Dataset and Modeling

The main issue in regards to the geospatial dataset is data sparseness. This sparseness arises in the form of incompleteness and non-occurrence. Data from OpenStreetMap tends to be highly documented and complete when it comes to urban and highly populated areas but unfortunately not in rural areas. Appendix D provides a visualization of what this sparseness looks like using location markers. For instance, some rivers are not labeled, causing them not to be outlined at all (non-occurrence) and in other cases, the rivers are merely labeled to a certain point (incompleteness). Farmland is also prone to this problem as land cultivation can clearly be seen taking place but not identified as farmland or agriculture. Non-occurrence of data refers to dataset instances where the image is identical to the original dataset without GIS features as there is no occurrence of these features. In some areas like the deep Amazon that has barely been penetrated, data sparseness is more extreme. In other areas, this is less the case. However, training possibilities consequentially become limited with the reduced exposure to geospatial features.

Furthermore, creating the geospatial dataset involves assumptions to be made which cannot always be quantified exactly. The features that were chosen were selected as they were influential phenomena that are commonly seen to precede or co-occur with deforestation. Knowing that the number of features the network uses cannot be endless due to parameter constraints, the question is raised as to whether a neural network could perform better given a different - more influential - set of features that better fit the area of interest.

6.2 Proposals for Future Work

The next steps ensuing the framework described and presented in this paper would be to make necessary changes and implementing them. Key limitations mentioned prior would be addressed in order to make the model feasible. Ensuring this feasibility entails optimizing the ResNet model for better performance and the dataset quality. A ResNet model could potentially be used that has already been pre-trained. This pre-training could be performed on public ML datasets such as SpaceNet (offered by Amazon Web Services) [26] or the SAT-6 Airborne Dataset [27]. There also exists the Forest Type Mapping Dataset specifically for classifying forestry [28]. However, this dataset contains a mere 326 instances compared to more than 17,500 and 405,000 instances offered in the other mentioned respective datasets.

With the pre-trained model, it would be wise to do preliminary runs on GIS datasets where features are added incrementally. Beginning with two additional features (9 band images), the ease with which this trains can be assessed and additional features can subsequently be added. This would make it possible to find a middleground in the tradeoff between computational cost and how informative a dataset instance can be.

Realistically speaking though, it is still possible that the geospatial model is run on the GIS dataset and still struggles to separate noise from signals (overfitting). In this case, making the model layer depth shallower is a path worth taking. This implies moving from a ResNet50 model to a ResNet34 or ResNet18.

6.2.1 Pre-Training on Different Richer Datasets

One path that could be taken would be to take the principles of this research proposal and apply it to urban areas. For example, monitoring urban land use with a CNN could be optimized by combining it with a geospatial data approach. This would be done as a way of pre-training the geospatial ResNet model that would eventually then be applied back to the Amazon. The benefit of this approach is that it addresses what is arguably the most prominent limitation of using GIS data in rural areas. Urban areas are documented to a greater extent and are more likely to be complete in regards to feature occurrences than rural areas are. Additionally, selected features such as roads will be more comparatively more prevalent in urban areas. While the exact manner in which the model learns is hard to predict (due to the black box problem), this could potentially allow the model to have more opportunities to learn as exposure to feature instances will increase.

6.2.2 Semantic Segmentation Using Dual Datasets

Another potential extension to this research could be to create an annually updating model aided by semantic segmentation. In semantic segmentation, pixels are individually analyzed in an image and assigned a class. As the name suggests, the result of this process is that objects or regions in an image that may show some kind of similarity in contextual meaning are grouped with each other as collections of pixels, also sometimes referred to as superpixels. This approach could be a manner in which geographic features such as rivers, roads and settlements could be identified and differentiated without using actual GIS data. For example, semantic segmentation in regards to a river could be that a river always looks like an elongated shape containing dark pixels, or in regard to farms as monochromatic squares or rectangles. The benefit of this is that it can thereby avert issues relating to overloads of parameters from multispectral images.

To make this model annually updating, a particular benefit of the Hansen et al. dataset comes to light. The dataset is updated each year, providing a new set of input to the model. Semantic pixel segmentation could be performed on both the baseline year and the new year's dataset to subsequently be compared with each other. What may become apparent then is that a superpixel classified as forest in one prior image may be a much smaller superpixel in the next image. This would then influence the model's prediction of how much deforestation has occurred. Other events could also be extracted such as when settlements expand (by segment area) or when forests become less dense (by segment value), further bolstering this prediction. Creating and training such a model might be challenging however. The model would have to have some training in recognizing what superpixels fall under what classes in the first place. A potential solution could be to use the Kaggle competition dataset that was considered for this research, mentioned earlier [13]. This dataset is thoroughly labeled and contains images that all fall under what these superpixels could be classified as. Namely, rivers, logging, primary rainforest, roads and other classes. Having this information would not only allow oversight to be gained over deforestation level differences as time passes, but also to uncovering the potential causes and trends.

6.2.3 Expanding the Focal Area

Lastly, a research extension is proposed that aims to improve the caliber of the model. A possible issue that could occur is that for a given image in the dataset, features present in the direct surroundings outside of the image might influence what can be seen in the image. Because these features occur outside of the visible pixels, they are not detected by the model. For instance, sparsely present forest might be visible in a corner of the image due to a highway that passes by it outside the view of the image. To solve this, a modification to the way in which the CNN parses through the dataset would be recommended. Currently, there is deliberate order in which dataset instances are analyzed. Instead, it might be more beneficial to, per dataset image, create a 9x9 grid of images that are analyzed as a larger image. The focal image would be at the center, with the corresponding eight images surrounding it. A visualization of this can be seen in Appendix E. The focal image is given in green (initially at [2, 2] and blue images surround it. These 9 images are then analyzed.

This would allow the model to analyze what occurs at the edges of the image with complete information beyond the edges. The convolutional stride of the model also therefore does not increase. Thus, referring to Appendix E, the initial focal image [2, 2] then becomes a surrounding image for the next focal image [2, 3] in the next convolution. This is to ensure that this process occurs for each dataset image. Should the stride be increased, then the initial problem remains as there is no overlap between the 9x9 grid of one dataset image and that of the image next to it on the map. Without overlap, the model functions identically to before and information between dataset instances are not shared.

7 Conclusion

The research done in this paper served two goals. To in one part, present a convolutional neural network that could predict deforestation on satellite imagery. A ResNet model of 50 layers was created and trained and appeared to generally be successful, deviating less than 1% from the true deforestation label of dataset instances at its best. The most optimal batch size to run this model was seen with highest performance on lower batch sizes. Another goal of this paper was to present a geospatial convolutional neural network that could improve on the results of the created network. By tweaking the ResNet model's input data shape and the model's in-channels, this has been seen to be possible. Along with careful feature selection, multispectral data could be fed into the model and achieving better performance is a possible outcome. While the research done in this paper was positive overall, there has been areas that were lacking and other paths for the future could be of greater interest moving forward. For instance, overfitting showed to be occurring very clearly in the data to a degree that minor changes to the model may not be enough. Thus, alternative steps moving forward could be to train the model not from scratch but to pre-train on a richer dataset. Or, to use a dual-dataset approach with semantic segmentation. Finally, a third option is to make the model less prone to inaccurate data representations where relationships cannot be inferenced from a given image due to being outside the focal view.

References

[1] C. H. Silva Junior, A. Pessoa, N. S. Carvalho, J. B. Reis, L. O. Anderson, and L. E. Aragão, "The brazilian amazon deforestation rate in 2020 is the greatest of the decade," *Nature Ecology & Evolution*, vol. 5, no. 2, pp. 144–145, 2021.

[2] L. G. Mikaela Weisse, "Forest loss remained stubbornly high in 2021," Available at https://globalforestwatch.org/blog/data-and-research/global-tree-cover-loss-data-2021/ (2022/07/03).

[3] J. G. Canadell and M. R. Raupach, "Managing forests for climate change mitigation," *science*, vol. 320, no. 5882, pp. 1456–1457, 2008.

[4] V. H. Gutierrez-Velez and K. MacDicken, "Quantifying the direct social and governmental costs of illegal logging in the bolivian, brazilian, and peruvian amazon," *Forest Policy and Economics*, vol. 10, no. 4, pp. 248–256, 2008.

[5] M. de Araújo, E. Andrade, and F. Machida, "Performance analysis of machine learning-based systems for detecting deforestation," in *2021 XI Brazilian Symposium on Computing Systems Engineering (SBESC)*. IEEE, 2021, pp. 1–8.

[6] H. Mayfield, C. Smith, M. Gallagher, and M. Hockings, "Use of freely available datasets and machine learning methods in predicting deforestation," *Environmental modelling & software*, vol. 87, pp. 17–28, 2017.

[7] P. J. Platts, C. J. McClean, J. C. Lovett, and R. Marchant, "Predicting tree distributions in an east african biodiversity hotspot: model selection, data bias and envelope uncertainty," *Ecological Modelling*, vol. 218, no. 1-2, pp. 121–134, 2008.

[8] P. P. De Bem, O. A. de Carvalho Junior, R. Fontes Guimarães, and R. A. Trancoso Gomes, "Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks," *Remote Sensing*, vol. 12, no. 6, p. 901, 2020.

[9] J.-F. Mas, H. Puig, J. L. Palacio, and A. Sosa-López, "Modelling deforestation using gis and artificial neural networks," *Environmental Modelling & Software*, vol. 19, no. 5, pp. 461–471, 2004.

[10] "Landsat satellite missions," USGS, NASA, 1999-2022. [Online]. Available: https://www.usgs.gov/landsat-missions/landsat-satellite-missions

[11] P. Team, "Planet application program interface: In space for life on earth," Planet, 2017-2022. [Online]. Available: https://api.planet.com

[12] I. M. McCullough, C. S. Loftin, and S. A. Sader, "Tb207: A manual for remote sensing of maine lake clarity," 2013.

[13] Planet, "Planet: Understanding the amazon from space," https://www.kaggle.com/competitions/planet-understanding-the-amazon-from-space/overview.

[14] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland *et al.*, "High-resolution global maps of 21st-century forest cover change," *science*, vol. 342, no. 6160, pp. 850–853, 2013.

[15] J. J. Senecal, J. W. Sheppard, and J. A. Shaw, "Efficient convolutional neural networks for multi-spectral image classification," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[16] M. S. Mon, N. Mizoue, N. Z. Htun, T. Kajisa, and S. Yoshida, "Factors affecting deforestation and forest degradation in selectively logged production forest: A case study in myanmar," *Forest Ecology and Management*, vol. 267, pp. 190–198, 2012.

[17] V. Bax, W. Francesconi, and M. Quintero, "Spatial modeling of deforestation processes in the central peruvian amazon," *Journal for nature conservation*, vol. 29, pp. 79–88, 2016.

[18] D. Armenteras, G. Rudas, N. Rodriguez, S. Sua, and M. Romero, "Patterns and causes of deforestation in the colombian amazon," *Ecological indicators*, vol. 6, no. 2, pp. 353–368, 2006.

[19] D. C. Morton, R. S. DeFries, Y. E. Shimabukuro, L. O. Anderson, E. Arai, F. del Bon Espirito-Santo, R. Freitas, and J. Morisette, "Cropland expansion changes deforestation dynamics in the southern brazilian amazon," *Proceedings of the National Academy of Sciences*, vol. 103, no. 39, pp. 14 637–14 641, 2006.

[20] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P. M. Atkinson, "A hybrid mlp-cnn classifier for very fine resolution remotely sensed image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 133–144, 2018.

[21] A. Loh and K. Soo, "Amazing amazon: Detecting deforestation in our largest rainforest," *MS CD'17, Standford University*, 2017.

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

[23] B. Fang, Y. Li, H. Zhang, and J. C.-W. Chan, "Semi-supervised deep learning classification for hyperspectral image based on dual-strategy sample selection," *Remote Sensing*, vol. 10, no. 4, p. 574, 2018.

[24] F. Bellot, M. Bertram, P. Navratil, F. Siegert, and H. Dotzauer, "The high-resolution global map of 21st-century forest cover change from the university of maryland ('hansen map') is hugely overestimating deforestation in indonesia," *FORCLIME Forests and Climate Change Programme, Jakarta, Indonesia*, 2014.

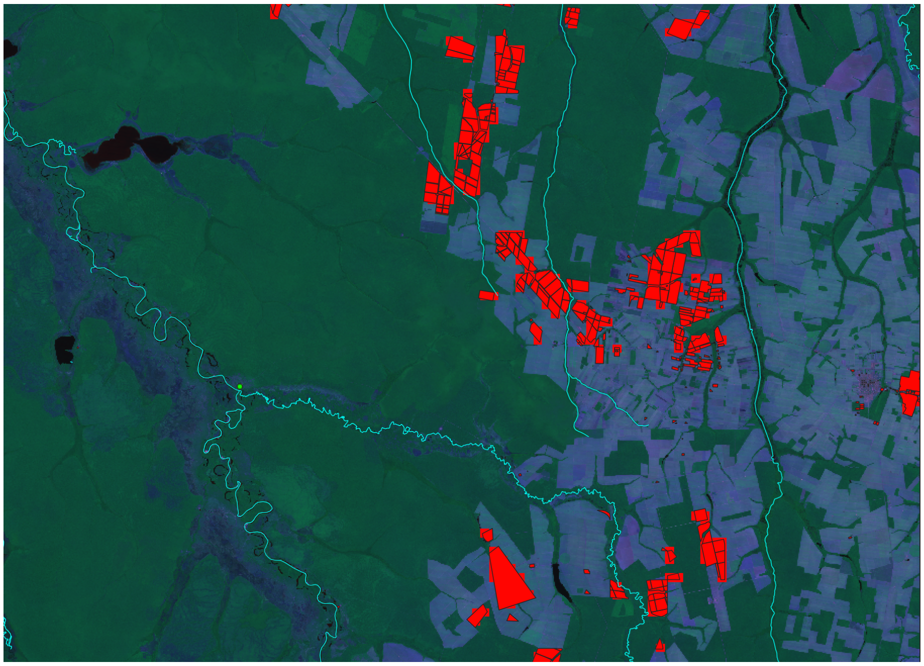
[25] E. Mitchard, K. Viergever, V. Morel, and R. Tipper, "Assessment of the accuracy of university of maryland (hansen et al.) forest loss data in 2 icf project areas–component of a project that tested an icf indicator methodology," *Ecometrica. The University of Edinburgh*, 2015.

[26] A. W. S. (AWS), "Datasets," https://spacenet.ai/datasets/ (2022/08/01.

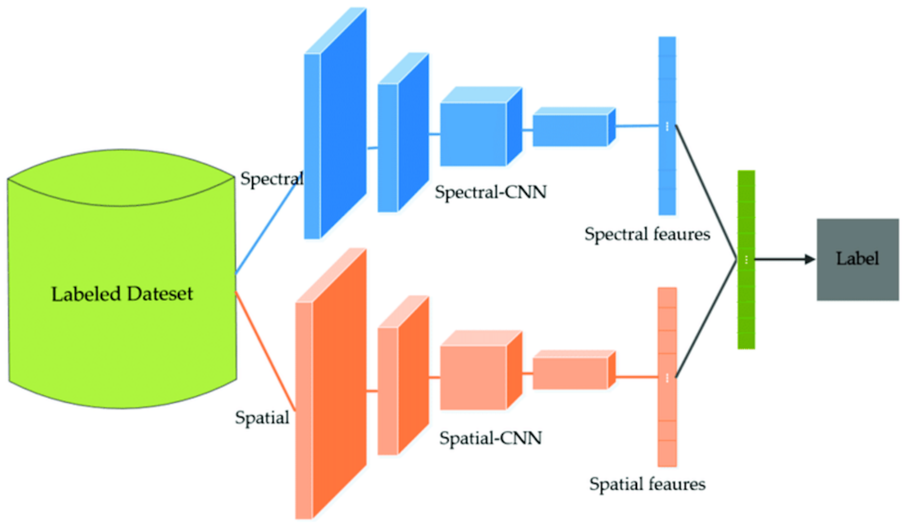
[27] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "Deepsat: a learning framework for satellite imagery," in *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, 2015, pp. 1–10.

[28] S. Chatterjee, S. Ghosh, S. Dawn, S. Hore, and N. Dey, "Forest type classification: A hybrid nn-ga model based approach," in *Information systems design and intelligent applications.* Springer, 2016, pp. 227–236.

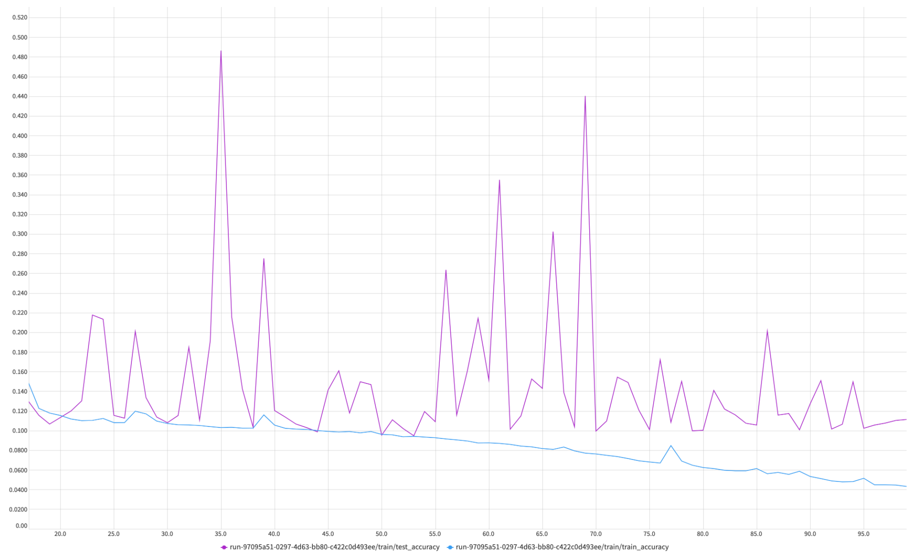
Appendix A: Another small scale image indicating potential deforestation patterns



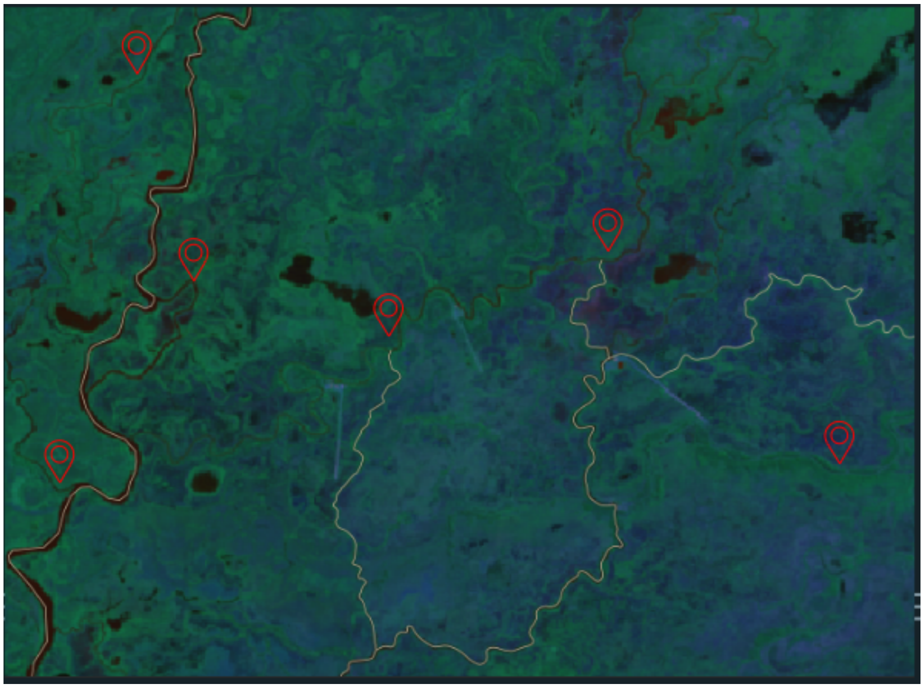
Appendix B: Fang et al.'s Hyperspectral ResNet architecture representation



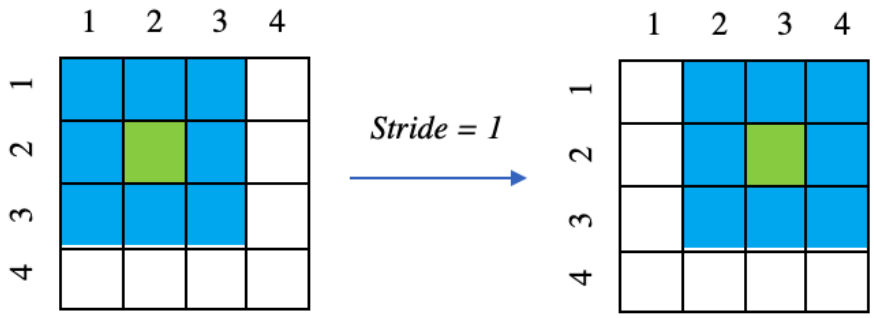
Appendix C: Adjusted view of the ResNet model performance with a batch size of 32



Appendix D: A representation of GIS data sparseness



Appendix E: A visualization of the proposed parsing method to increase the focal area



Appendix F: Access to code used for this research

All code used in this research can be found on the following github profile: https://github.com/jbeek00/Thesis