CS230

Deep Learning and Deforestation

Ian Avery Bick
ianbick@stanford.edu

Dennis Wang
dwang22@stanford.edu

Ben Mullet
bmullet@stanford.edu

Abstract

Accurate classification of satellite imagery is a critical task for understanding the scope and manifestation of deforestation. Commonly used classification methods such as mean-shift, SVM, and random forest classifiers ultimately depend on a high degree of human correction. We further automate the classification process by using a fully convolutional neural network for image segmentation via a U-Net model architecture. Specifically, we seek to ease classification of forest boundaries and forest density to support wildlife habitat and deforestation analyses. Our model is tested on both binary (2-class) and multi-class problems related to drawing forest boundaries and labeling density of forest vegetation. Results indicate the fully convolutional neural network approach can achieve similar or superior accuracy compared to existing techniques, and underscore the importance of data augmentation to successful training on limited training data.

1 Introduction

Accurate monitoring of the world's forests is a task of immense importance. Studies of current rates of deforestation indicate that around two million square kilometers of forest were lost between 2000 and 2012 [6]. Losses include some 7.2% of the total intact forest landscape, and rates of deforestation are increasing [15]. Understanding the forest-civilization boundary is of utmost importance, because most deforestation occurs at the forest boundary and because forest "edge effects" can have damaging implications for conservation of forest species [13]. Indeed, it is hypothesized that HIV originally migrated to humans in a human-species interaction facilitated by the forest boundary [19]. With the recent explosion in abundance of high-quality satellite imagery, accurate monitoring and characterization of forests on a global scale is a more tractable problem than ever. Nevertheless, modern algorithms for forest imagery classification require significant human interaction, limiting the scope and potential value of any research endeavor.

This project was originally inspired by a series of conversations with Laura Bloomfield, a Stanford Ph.D. student in the E-IPER program who researches the intersection of forest boundaries and vector-born diseases. The success of her project hinges on the accurate classification of the boundaries of the Kibale forest in Uganda, where she has been investigating human-forest interaction.

Utilizing RapidEye satellite imagery of Kibale Park which had been overlaid with a binary forest mask a forest classification tool was trained in U-Net using semantic segmentation which reached 87.8% pixel accuracy in the test set after data augmentation. LANDSAT 7 imagery paired with continuous forest density mask was utilized to create a multi-class forest density classifier in U-Net, reaching 68.5% pixel accuracy.

2 Related Work

2.1 Semantic Segmentation

Semantic segmentation is the task of object classification at the pixel by pixel level. For example, given an image, label each pixel as one of several classes. Long et al. first introduced the concept of a fully convolutional neural network to do this task [11]. In a normal neural network, a series of convolutions is used, ending with one or more fully connected layers, which essentially converges all parameters into a single output value. In a fully convolutional neural network, there are no fully connected layers, so the output is multi-valued, and in the case of semantic segmentation the output is the same pixel size as the input image. In the paper, this is achieved by converting the fully connected layers into equivalent convolutional layers. Improvements were made on this architecture in U-Net [17] and SegNet [2], which use a mirrored upsampling or decoder network after the normal downsampling network, which combines upsampled results with previous layers. This architecture better captures the spatial locality of each individual pixel to be recovered in the final output image.

2.2 Remote Sensing Classification

Remote sensing classification techniques are generally divided into supervised and unsupervised methods. Supervised methods demand training information whereas unsupervised methods classify pixels based on clustering techniques like K-Means. Commercially available platforms, such as Google Earth Engine, include options for both supervised and unsupervised classification [5]. [9] used a segment-based classification (SBC) technique with majority-rule to segment multi-class forest boundaries in satellite imagery, acheiving pixel-wise accuracy of 77%. [16] achieved similar results by using a "region-merging" approach, where segments are built up by merging individual pixels into regions in a iterative manner.

3 Dataset and Features

3.1 Imagery

High resolution RapidEye imagery of the Kibale region from Laura was accompanied by binary training labels - created via unsupervised classification and manually corrected utilizing ArcGIS. LANDSAT 7 imagery was obtained from Hansen's Global Forest Change map [6]. The corresponding forest density map from 2000 was presented as continuous (0-100) but was converted into a five-class label. Training and test set data were prepared by splitting and labeling 256x256 pixel clips of the study area via ArcPy.

3.2 Training Sets

For the purposes of training our model, we first split our set of images into three sets: train, dev, and test. For both the RapidEye and LANDSAT datasets, we did an 80%-10%-10% split. Because our datasets are small, this resulted in small training sets (see Table 1), which we address in the next section. We trained our model to minimize loss on the train set and used the dev set to tune hyperparameters and get a sense of the degree of under- versus over-fit. Finally, our overall model performance metrics were derived from the test set.

3.3 Data Augmentation

To address issues with our small datasets and overfitting, we used several data augmentation techniques. We expanded the training set by a factor of 16 by taking each image and flipping it (x2), rotating it in 90 degree increments (x4), and adding noise from a random Gaussian distribution (x2). For the RapidEye dataset, this also allowed us to use a 64%-18%-18% split to increase the number of images in the dev and test sets.

Dataset	Train	Dev	Test
RapidEye	270	34	34
RapidEye (Aug)	3456	122	122
LANDSAT (Aug)	9088	71	71

Table 1: Number of images in each of the dataset splits

4 Methodology

Our model relies on the "U-Net" architecture as proposed by [17], who proposed the methodology to perform semantic segmentation of biomedical images. The network was particularly appealing for this task because of its ability to train on a limited, albeit highly augmented, amount of training data and for its speed. The architecture is fully convolutional, such that there are no fully connected layers. The "U-Net" label is derived from the characteristic sequence of downsampling blocks coupled to an equal number of upsampling blocks, such that an output of the same size as the original input is returned. Each downsampling block includes two 3x3 convolutions ("same" padding) with batch norm and ReLu activation, followed by 2x2 maxpool for downsampling. In this way each downsampling block doubles the number of channels. Upsampling is accomplished by 2x2 transposed convolutions. At each stage of the upsampling, important high-resolution information that is lost in the downsamping process is incorporated by direct concatenation of the similar stage downsampling output. The resulting concatenated information is then fed through two sequences of 3x3 convolutions ("same" padding), with batch norm and ReLu activation. The final class is assigned by a pixel-wise softmax classification layer. Figure 1 visually explains this architecture, as presented by [17].

We employ a cross-entropy loss function,

$$\mathcal{L} = -\sum_{c=1}^{N} y_c \log(\hat{y}) \tag{1}$$

We report accuracy through two metrics, pixel-wise accuracy and mean intersection over union (IoU). IoU is a particularly appealing metric for this project because it penalizes both false positive and false negatives. Because some input images are covered to a significant extent by forest, a high level of pixel-wise accuracy could be achieved by naively outputting the same class label for every pixel. IoU is calculated as

$$\text{IoU} = \frac{F_l \cap F_m}{F_l \cup F_m} \tag{2}$$

where F_l is the set of labeled forest-classified points and F_m is the set of model output forest-classified points.

5 Experimental Results

5.1 Hyperparameter Tuning

We tuned our hyperparameters by training on the train set and optimizing for the dev set metrics. After surveying the search space, we arrived at learning rate 1e-4, minibatch size 32, L2 regularization constant 10.0, starting filter size 32, and 4 layers for the binary classification task. Similarly, for multi-class classification, we arrived at learning rate 1e-3, minibatch size 32, L2 regularization constant 0.1, starting filter size 32, and 5 layers.

5.2 Binary Classification

For the binary classification task, we started with high training metrics: 0.963 pixel accuracy and 0.789 IoU (see Table 2). However, we were having trouble generalizing this error to the dev set, which had 30% and 40% lower pixel accuracy and IoU, respectively. We used early stopping and L2 regularization to address the overfit but with limited success. This led us to expand the size of the training set through data augmentation. As described in Section 3, we expanded the training set by

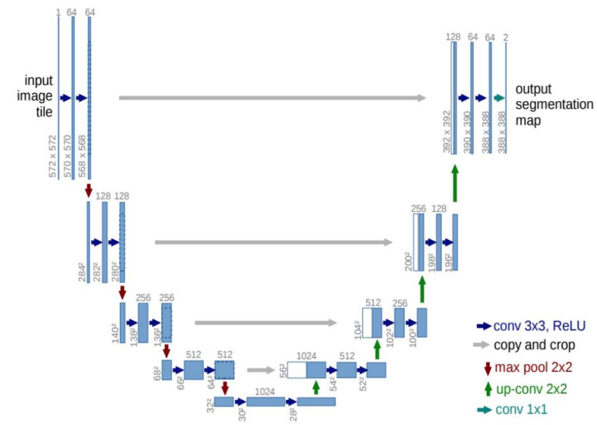
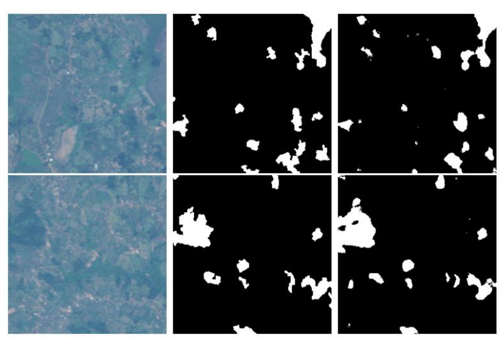


Figure 1: Example U-Net, as appears in [17]. Our model shares essential characteristics with regards to operations (arrows), but dimensions of input/output differ. Furthermore, our model applies "same" padding to convolutions.

Model	Train Acc.	Dev Acc.	Test Acc.	Train IoU	Dev IoU	Test IoU
Binary	0.963	0.651	0.733	0.789	0.326	0.366
Binary Aug.	0.946	0.901	0.878	0.828	0.809	0.768
Multi-Class	0.724	0.692	0.685	0.475	0.462	0.462

Table 2: Pixel accuracy and mean IoU for the binary and multi-class models

16 times. This gave a significant improvement in the dev and test metrics. We ended up with test accuracy 0.878 and test IoU 0.768. In terms of semantic segmentation tasks, the IoU is in the range of recent IoU's achieved in general object classification (0.627) [11] and similar work in satellite image classification (0.829) [20]. As for prior land classification literature, an accuracy of 85% is generally considered sufficient for pixel-by-pixel classification [18] [4].

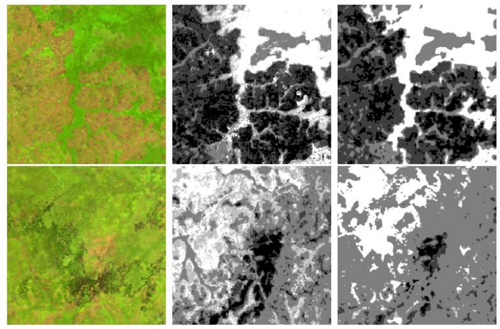


(a) Input Image (b) Label (c) Prediction

Figure 2: Sample images from binary classification (white is forest)

5.3 Multi-Class Classification

For the multi-class classification task, we started directly with data augmentation and the other regularization techniques we used for binary classification. As expected, the multi-class task is much harder than the binary task, and we saw a decrease in both metrics across train, dev, and test sets. To address the underfit, we experimented with increased number of layers and filters in the model. This gave no more than a single digit percent increase in our metrics. Given more time, there are more complex models that we would like to try (detailed in Future Work). Our final metrics were 0.685 accuracy and 0.462 IoU. From the sample images, it is clear that the model is learning the general class structure, but it is not able to do so with high resolution.



(a) Input (b) Label (c) Prediction

Figure 3: Sample images from multi-class classification (lighter is denser forest)

6 Future Work

As mentioned in our results section, we believe there is a lot of room for improvement for multi-class forest density classification. From our reference images, we see that the model is still only learning general characteristics. For this, we would like to use a deeper architecture like the one found in Tiramisu [8]. This architecture borrows the idea of DenseNets [7], networks with shorter and more skip connections that allow for deeper training, and applies them to the U-Net architecture. This, however, would significantly increase the parameter space and therefore require us to collect more data to counteract overfitting. An alternative path of work is to use transfer learning on a architecture like the basic FCN in [11]. This way, we could see how well pretrained weights from object classification work for forest density classification without needing to collect more data.

Incorporation of the model into a Land Use/Land Cover Change (LULCC) model would be particularly exciting, as it would allow for prediction of future changes in forest cover. Achieving this goal would demand a data set with significant temporal variation and additional model complexity, although relatively simple LULCC models have shown significant promise [14].

The disease transmission potential model developed by Laura and her co-authors [3] is a function of percent habitat loss, edge effects, and pathogen characteristics. Our forest classification model could aid in the implementation of this transmission model on geospatial scales.

7 Conclusion

We applied a fully-convolutional network (FCN) to solve the problem of semantic segmentation of forest boundaries in the Kibale National Forest, Uganda. The U-Net model achieved a high degree of accuracy as measured by both test IoU and pixel-wise classification accuracy on the binary (Forest/No-Forest) task, but performed less well on the multi-class problem.

8 Contributions

Dennis Wang was the technical lead on the project, and is responsible for the majority of code development and model testing.

Ian Avery Bick shouldered the responsibilities of data gathering and preprocessing.

Ben Mullet worked with Avery in data preprocessing and led an (ultimately unsuccessful) attempt to deploy the model in Google compute cloud.

All code described for this project can be found at:
https://github.com/denny1038/cs230-semantic-seg-forest/tree/master/tensorflow/vision
Several open source works were referenced[12][1][10].

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.

[3] Christina L Faust, Hamish I McCallum, Laura SP Bloomfield, Nicole L Gottdenker, Thomas R Gillespie, Colin J Torney, Andrew P Dobson, and Raina K Plowright. Pathogen spillover during land conversion. *Ecology letters*, 2018.

[4] Giles M Foody. Status of land cover classification accuracy assessment. *Remote sensing of environment*, 80(1):185–201, 2002.

[5] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27, 2017.

[6] Matthew C Hansen, Peter V Potapov, Rebecca Moore, Matt Hancher, SAa Turubanova, Alexandra Tyukavina, David Thau, SV Stehman, SJ Goetz, TR Loveland, et al. High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853, 2013.

[7] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016.

[8] Simon Jégou, Michal Drozdzal, David Vázquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. *CoRR*, abs/1611.09326, 2016.

[9] So-Ra Kim, Woo-Kyun Lee, Doo-Ahn Kwak, Greg S Biging, Peng Gong, Jun-Hak Lee, and Hyun-Kook Cho. Forest cover classification by optimal segmentation of high resolution satellite imagery. *Sensors*, 11(2):1943–1958, 2011.

[10] Mo Kweon. U-net implementation in tensorflow. `https://github.com/kkweon/UNet-in-Tensorflow`, 2018.

[11] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1411.4038, 2014.

[12] Oliver Mondroit, Guillaume Genthial, and Surag Nair. Cs230 code examples. `https://github.com/cs230-stanford/cs230-code-examples`, 2018.

[13] Carolina Murcia. Edge effects in fragmented forests: implications for conservation. *Trends in ecology & evolution*, 10(2):58–62, 1995.

[14] Bryan C Pijanowski, Daniel G Brown, Bradley A Shellito, and Gaurav A Manik. Using neural networks and gis to forecast land use changes: a land transformation model. *Computers, environment and urban systems*, 26(6):553–575, 2002.

[15] Peter Potapov, Matthew C Hansen, Lars Laestadius, Svetlana Turubanova, Alexey Yaroshenko, Christoph Thies, Wynet Smith, Ilona Zhuravleva, Anna Komarova, Susan Minnemeyer, et al. The last frontiers of wilderness: Tracking loss of intact forest landscapes from 2000 to 2013. *Science advances*, 3(1):e1600821, 2017.

[16] Rastislav Raši, Catherine Bodart, Hans-Jürgen Stibig, Hugh Eva, René Beuchle, Silvia Carboni, Dario Simonetti, and Frédéric Achard. An automated approach for segmenting and classifying a large sample of multi-date landsat imagery for pan-tropical forest monitoring. *Remote Sensing of Environment*, 115(12):3659–3669, 2011.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[18] John R Thomlinson, Paul V Bolstad, and Warren B Cohen. Coordinating methodologies for scaling landcover classifications from site-specific to global: Steps toward validating global map products. *Remote Sensing of Environment*, 70(1):16–28, 1999.

[19] Bruce A Wilcox and B Ellis. Forests and emerging infectious diseases of humans. *UNASYLVA-FAO-*, 57(2):11, 2006.

[20] Wei Yao, Dimitrios Marmanis, and Mihai Datcu. Semantic segmentation using deep neural networks for sar and optical image pairs. 2017.