Improving the mapping outcome of land usage

Prashan Somapala
Stanford center for professional development
Stanford University
prashans@stanford.edu

Abstract

In human history land has played a very crucial role. It is an essential material for humans' survival. Land use is one of the key reasons for the progression and development of mankind. Therefore, it is essential that it is understood, mapped and managed appropriately. This research endeavor aims to carry out classification on images of land, land cover, and develop a methodology to classify land cover examples into several different categories. Analysis of land use has a critical impact on the way we treat our surroundings and environment and could lead to more efficient planning of development projects to build human habitation in future. The data source for this project is DeepGlobe Land Cover Classification Data set (Kaggle) that was published in 2018. This contains several different land use and land cover classifications. I observed that these images were taken using satellite technology and they appears to be a very structured and rich source of information. This research aims to deploy CNN models such as Unet with several different variations to effectively classify these satellite images.Unet with Adam optimizer was able to classify these images with an 0.8939 accuracy.

1 Introduction

The motivation behind this research is to enhance the visibility of the environment surrounds us. These publicly available data sets have given access to people to investigate and find out solutions to their issues or to improve their knowledge of the environment surround them. Satellite data is such a powerful source of information, but it is not well investigated area compared to regular image usage for machine learning purposes. Increased attention and analysis could critically change the way we understand the environment and could lead to significant improvements in our usage of this scares resource. The input to the project is a satellite image and it's mask , which helps to identify its classification of the image in to one of the 7 land use cases by pixel level segmentation. For this semantic segmentation task, a CNN model (modified U-Net architecture) is used with this data to build a learning model that can correctly identify the use case of a new satellite images. The model will out put a mask for the new image that we could compare with the ground truth mask. This accurate categorization will be essential for sustainable development and urban planning or agriculture.

2 Related work

The U-net architecture was introduced in 2015[3] The as a CNN model that is used to do pixel wise segmentation in medical imagery classification . This is designed to be used in conditions where the researcher do not have a huge amount of training data . I have used this model in a scenario where the availability of the training data isn't an issue. One other more recent advancement in this

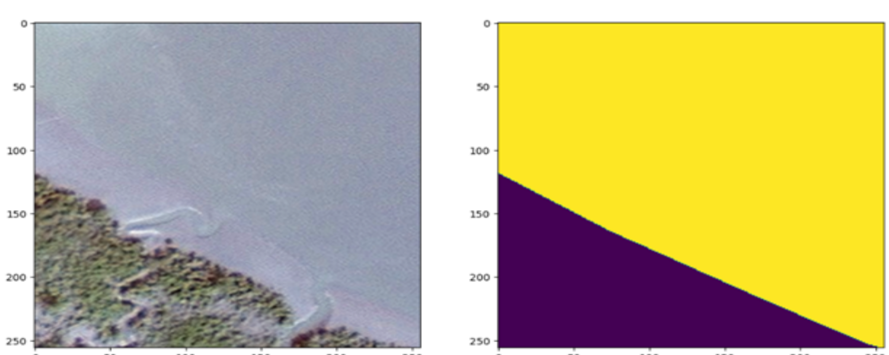
CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

field is to use 3D-CNN for feature extraction (multi scale convolution kernel) [4]. This has yielded slightly better results than U-Net ,SegNet and Conv-LSTM. Another approach was introduced very recently which does not follow the standard paradigm of semantic segmentation. This approach uses Swin Transformer as the backbone to extract the context information and design a novel decoder of densely connected feature aggregation module (DCFAM) to restore the resolution and produce the segmentation map.[5] More recent years researches are deviating from the standard of fully convolutional network (FCN) with an encoder-decoder architecture and gravitating more towards transformer based architectures. Another more recent paper suggested that the semantic segmentation can be achieved by a 2-D semantic transformer model (2DSegFormer) for semantic segmentation on aerial images. [6]("In 2DSegFormer, we design a novel 2-D positional attention to accurately record the 2-D position information required by the transformer. Furthermore, we design the dilated residual connection and use it instead of skip connection in the deep stages to get a larger receptive field.") Above mention transformer models are novel concepts of tackling the semantic segmentation problem and in this project I decided to stick with the standard method as it is more familiar to the subject matter that was taught in the class. SVM, random forest, logistics regression and other similar methods was used in the past to classify land cover data as just another image classification. Development of CNNs and the rise of GPU availability gave chance of CNN to compete for this activity and emerged as a successful classification model for satellite data. According to a recent paper [8] published ("It has been demonstrated that CNN have about 12 % greater predictive ability than SVR with regard to spatial prediction[9]) Therefore I have chose the U-Net related architectures to explore this classification task.

3 Dataset and Features

This dataset is obtained from Land Cover Classification Track in DeepGlobe Challenge in kaggle . The training data for Land Cover Challenge contains 803 satellite imagery in RGB, size 2448x2448.The imagery has 50cm pixel resolution, collected by DigitalGlobe's satellite.Each satellite image is paired with a mask image for land cover annotation. The mask is a RGB image with 7 classes of labels, using color-coding (R, G, B) as follows.

- Urban land: 0,255,255 - Man-made, built up areas with human artifacts (can ignore roads for now which is hard to label)
- Agriculture land: 255,255,0 - Farms, any planned (i.e. regular) plantation, cropland, orchards, vineyards, nurseries, and ornamental horticultural areas; confined feeding operations.
- Rangeland: 255,0,255 - Any non-forest, non-farm, green land, grass
- Forest land: 0,255,0 - Any land with x
- Water: 0,0,255 - Rivers, oceans, lakes, wetland, ponds.
- Barren land: 255,255,255 - Mountain, land, rock, dessert, beach, no vegetation
- Unknown: 0,0,0 - Clouds and others
- File names for satellite images and the corresponding mask image are id _sat.jpg and id _mask.png. id is a randomized integer.
- The values of the mask image may not be pure 0 and 255.'



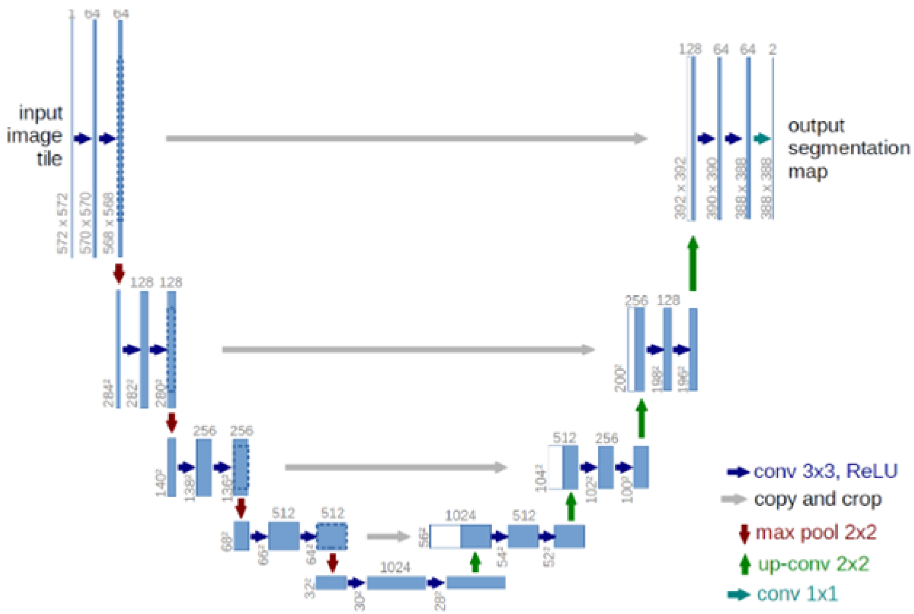
These images were pre-processed by slicing them in to 256x256 patches. corresponding masks are also patched using the same method. Class distributions in the DeepGlobe land cover classification data set.

class	pixel count	proportion
Urban	642.4M	9.35%
Agriculture	3898.0M	56.76%
Rangeland	701.1M	10.21%
Forest	944.4M	13.75%
Water	256.9M	3.74%
Barren	421.8M	6.14%
Unknown	3.0M	0.04%

I chose 10 images that was patched to generate 800 patches for training process. this was split in to 85/15 for training and testing sets.

4 Methods

I followed the Unet model and made changes based on the problem at hand. The image fed to the model is 256X256. Unet is an encoder decoder architecture, this has two key units, contracting first section which down samples the input image to a smaller one by applying 3X3 valid padding convolution layers . it is fed to a RELU activation function before halving of the image by the 2x2 max pooling layer. Then the second section of the architecture is applied to up sample the image . this time by 3x3 valid convolutions with RELU. But then apples a 2x up sampling convolution which uses fractional striding to double the image size. These formed a U shape when stacked together with an added skip connection from each contracting layer to the expanding layer with same input image shape. Apply 1 X 1 convolution at the output layer. The network architecture is as follows.



Segmentation output results are analysed using the following loss functions

Focal loss =

$$FL = -\alpha_c(1-\hat{y}_i)^{\gamma}.log\hat{y}_i$$

Focal Loss presents a better solution to the unbalanced dataset problem. It adds an extra term to reduce the impact of correct predictions and focus on incorrect examples. The gamma is a hyper-parameter that specifies how powerful this reduction will be.it was very clear by the table in previous page that this dataset is unbalanced.

Dice Loss this is also called F1 loss

$$\text{dice\_loss} = 1 - \frac{1}{c}\sum_{i=0}^{c}\frac{\sum_j^N 2y_i^j\hat{y}_i^j + \epsilon}{\sum_j^N y_i^j + \sum_j^N \hat{y}_i^j + \epsilon}$$

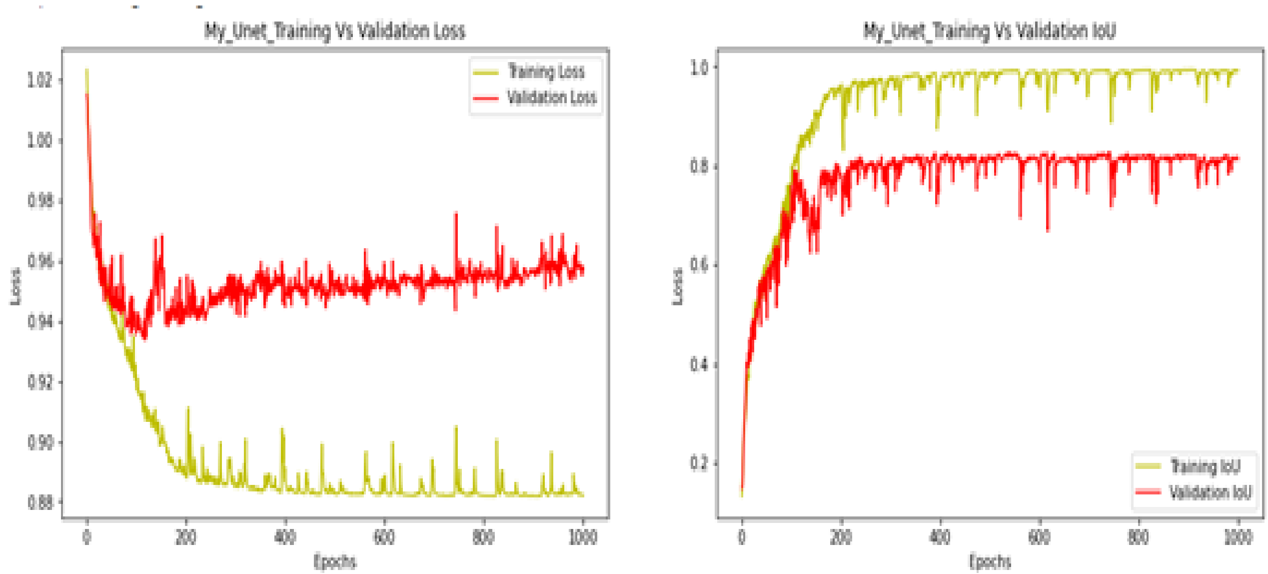
Total loss = Dice Loss + (Focal Loss*1)

5 Experiments/Results/Discussion

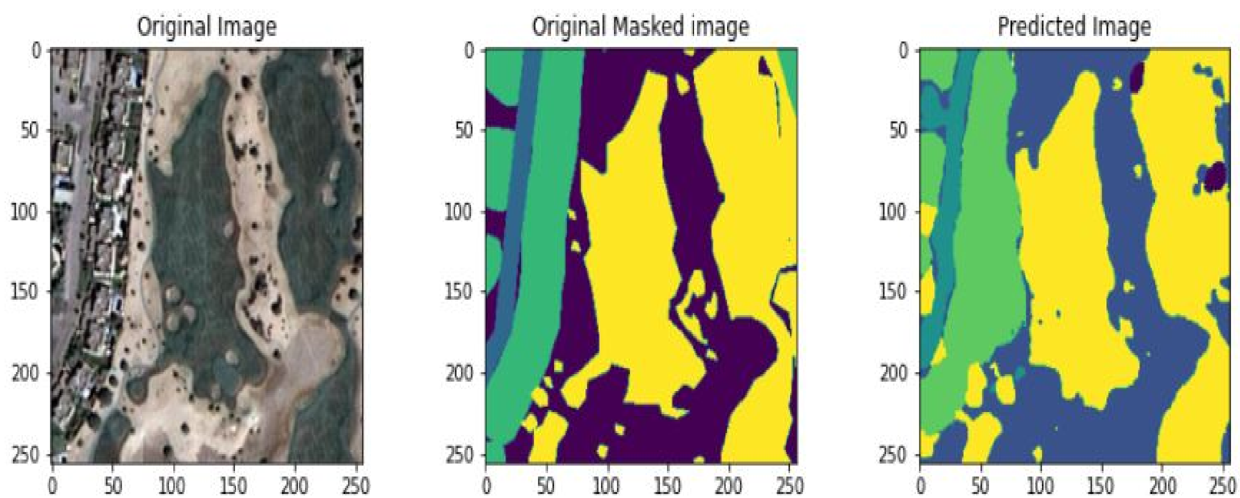
Following table shows the conducted experiments . Based on the results Adam optimizer with 0.01 learning rate produced the most accurate predictions.

Model	Optimizer	Learning rate	Batch size	Val Accuracy
UNet	Adam	0.001	16	0.8939
Unet	Adagrad	0.001	16	0.5051
Unet	Adagrad	0.0001	16	0.5049

following graphs further shows the results of the successful experiment (Unet,Adam,001)



Following figure shows a prediction result with the chosen experiment



6 Conclusion/Future Work

The goal of the project was to experiment with CNN to efficiently and accurately classify satellite images for Land use classification. Unet with Adam optimizer was able to reach 0.8939 accuracy. i would prefer to further improve this model by comparing the results with other Unet models using different backbones (resnet34 ,inceptionv3,vgg19). I already have the code written for the above section but never really get to conduct through experiments. Further I would like to investigate more newer architectures suggest by more recent papers where using transformer based approaches yielding better results.

7 Contributions

Individual project. code can be found at https://github.com/prashan-1212/CS230

References

[1] Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., amp; Raskar, R. (2018, May 17). DeepGlobe 2018: A Challenge to parse the Earth through satellite images. https://arxiv.org/abs/1805.06561//

[2] H I Dewangkoro and A M Arymurthy 2021 IOP Conf. Ser.: Earth Environ. Sci. 704012048//

[3] Olaf Ronneberger, Philipp Fischer, Thomas Brox,Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234–241, 2015,arXiv.org. from https://arxiv.org/abs/1805.06561 //

[4] Rui Li, Shunyi Zheng, Chenxi Duan, Libo Wang Ce Zhang (2022) Land cover classification from remote sensing images based on multi-scale fully convolutional network, Geo-spatial Information Science, 25:2, 278-294, DOI: 10.1080/10095020.2021.2017237

[5] L. Wang, R. Li, C. Duan, C. Zhang, X. Meng and S. Fang, "A Novel Transformer Based Semantic Segmentation Scheme for Fine-Resolution Remote Sensing Images," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6506105, doi: 10.1109/LGRS.2022.3143368.

[6] X. Li, Y. Cheng, Y. Fang, H. Liang and S. Xu, "2DSegFormer: 2-D Transformer Model for Semantic Segmentation on Aerial Images," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-13, 2022, Art no. 4709413, doi: 10.1109/TGRS.2022.3223416.

[7] D. Fitton, E. Laurens, N. Hongkarnjanakul, C. Schwob, L. Mezeix,Land cover classification through Convolutional Neur-al Network model assembly: A case study of a local rural area in Thailand,emote Sensing Applications: Society and Environment,2022,

[8] D. Fitton, E. Laurens, N. Hongkarnjanakul, C. Schwob, L. Mezeix,Land cover classification through Convolutional Neur-al Network model assembly: A case study of a local rural area in Thailand,Remote Sensing Applications: Society and Environment,2022,

[9] ahdi, M., Mótyán, J.A., Szojka, Z.I. et al. Analysis of the efficacy of HIV protease inhibitors against SARS-CoV-2s main protease. Virol J 17, 190 (2020). https://doi.org/10.1186/s12985-020-01457-0