Spatial-Temporal Deep Neural Networks for Land Use Classification of LandSat Images
Honors Prospectus

Cameron Vahid Farzaneh

June 28, 2018

Land cover, the physical material that covers the surface of our planet, is an important form of data that is essential to global environment sustainability. Its use has many applications in the fields of environmental science, land management and geography. Because of the importance of land cover information, there have been many efforts to generate large amounts of land cover datasets. But simply having large amounts of the data is not enough; the data must be processed and observed. This is generally achieved through ground surveys and/or remote sensing. Ground surveys, although very accurate, are limited by logistic constraints making it not a feasible option to cover most of our large planet. Remote sensing, on the other hand, is the favored method because it takes advantage of cameras and sensors that are installed on satellites and observe large areas of land surface of Earth over time. This allows an efficient and affordable method to map land cover patters. Luckily, there is an abundance of publicly available and free of charge remote sensing data at medium-resolution such as Landsat imagery (USGS, 2016a). This data includes aerial photographs, satellite imagery, thermal imagery, hyper-spectral imagery, radar and lidar datasets; which vary in spatial, spectral, radiometric and temporal resolutions. Using the vast amount of data available, both visual interpretation and computer-based digital classification can be used to extract information about land cover. Generally, digital pattern classification is preferred over visual interpretation for mapping land cover in large areas. Now that there is remote sensing data with improved spectral, spatial and temporal resolutions available to the community, the real problem is to effectively and accurately use this data for image classification for land cover use.

There has been many advances with conventional methods such as maximum likelihood classifier (MLC) (Strahler, 1980) and clustering (Huang, 2002); it has now moved towards more advanced techniques such as decision trees (Xu, Watanachaturaporn, Varshney, & Arora, 2005), random forests (RF) (Pal, 2005), neural networks (NN) (Kavzoglu & Mather, 2003; Mas & Flores, 2008), support vector machines (SVM) (Gidudu, Hulley, & Marwala, 2007; Mountrakis, Im, & Ogole, 2011; Pal & Mather, 2005), convolutional neural networks (CNN)

(Castelluccio, Poggi, Sansone, & Verdoliva, 2015; Romero, Gatta, & Camps-Valls, 2016), and recently, patch-based convolutional neural networks (Sharma, 2017) to preform classification on the remote sensing data. Although there has already been research done using CNNs, they have been using smaller datasets with high-resolution images. This is because classification of high-resolution images is similar to object recognition in computer vision. As a result, these CNNs have been shown to work well on high-resolution remote sensing images. While this is a great achievement, the architecture is limited in that it depends on high-resolution images. It is currently difficult to obtain high-resolution images that not only cover all areas of our planet, but also span a very large area. On the other hand, there is plenty of medium-resolution remote sensing data available from the Landsat satellite that is public and free. It also provides the longest continuous observations of Earth's surface from space. The Landsat system provides copious amounts of highly calibrated, multi-spectral data of global coverage. If there exists an accurate form of classifying medium resolution imagery, this would be a large advantage as it combines the power of CNNs along with the abundance of ready-to-use medium-resolution images. There have been previous forms of using medium-resolution images, such as Sharmas patch-based approach, however, the accuracy is only 85.60% after 149,999 iterations (Sharma, 2017). It would be ideal to have a different architecture that can achieve better accuracy. Such a system would be able to provide more reliable and efficient classification of remote sensing data over large areas.

The proposed research method takes as a starting point Sharma's patched-based CNN architecture. This is because the model has been adapted for medium-resolution multidimensional data. His architecture features five convolutional layers for feature extraction, along with a fully connected dense layer for classification. Sharma's proposed architecture uses patched-based samples by acknowledging the spatial relation of a pixel in regard to its neighboring pixels. Using this same approach, patches of size 5x5x8 will initially be extracted from the Landsat dataset as part of our training set. Because this is a supervised learning problem, to train the network, we also need to be provided labels. Luckily, The Florida Cooperative Land Cover Map (CLC), partnered with the Florida Fish and Wildlife Conservation Commission (FWC) and Florida Natural Areas Inventory (FNAI), has provided a reference map for the purpose of land cover analysis. Included in the reference map is a set of all the classes and super-classes. There are a total of ten classes and forty one super-classes. These classes include: Water, Agriculture, Forested Wetland, No Forested Wetland, Barren Land, Mixed Forest, Evergreen Forest, Shrub Scrub, High Intensity Urban, and Low Intensity Urban. We can use the reference map as our labels by aligning the Landsat images with projection coordinates. Once the data is prepared and the reference map has been aligned, the architecture must be created. Once the architecture is implemented, we will use it to gather preliminary results. Afterwards, the architecture must be improved upon in attempt to increase the validation accuracy. There are many different approaches we can take in order to do this. One of these methods is adapting the architecture to support multi-scaled input. This means using different size patches per class.

For some classes, like a body of water, it would be ideal to use smaller patched samples. This is because a body of water, unlike other classes, is easier to classify because it can generally be found in large areas (such as a lake or ocean). As a result, a large spatial relation is not necessary in order to accurately and efficiently classify it. On other classes, like Low intensity Urban, it would be better to use larger patched samples as these areas are harder to classify. They are more compact, and are generally surrounded by other classes as well. Making the architecture dynamic in input size is the challenge and theoretically can improve the validation accuracy. Having one universal patch size for input is a limitation because some classes are harder to classify, and require larger samples, while other classes are easier, and do not need as large of a patch in order to classify.

There are multiple ideas that can be implement to allow variable sized input into the neural network. One method involves using the largest size patch for all classes, for example 11x11x8, and using a non-sequential architecture to reduce the dimensionality if needed. For example, skip connections, or parallel computations, can be implemented allowing different paths in which the input data can propagate through the network. Another method involves having the same class be passed into the neural network with different sized patches in a parallel structure. They would then be combined together through a weighted sum. This layer would contain trainable parameters in which the model can then learn what the most effective patch size is needed in order to classify.

After we have implement the new architecture, we will train the model and test the results. We have listed different methods in which we can experiment to allow variable sized patches to be used as input. Realistically, we would have to experiment with these methods and the outputs we expect may not be the same. Part of the nature of machine learning is to adapt the model based on the output results we obtain. With this being said, the project has been planned, and now it time to begin implementation, obtaining results, and making changes along the way when needed.

References

[1] Atharva Sharma, *A patch-based convolutional neural network for remote sensing image classification.*

[2] Castelluccio, Marco, Poggi, Giovanni, Sansone, Carlo, & Verdoliva, Luisa (2015), *Land use classification in remote sensing images by convolutional neural networks.*

[3] Gidudu, Anthony, Hulley, Greg, & Marwala, Tshilidzi (2007), *Classification of images using support vector machines.*

[4] Huang, Kal-Y. I. (2002), *A synergistic automatic clustering technique (SYNERACT) for multispectral image analysis.*

[5] Kavzoglu, T., & Mather, P. M. (2003), *The use of backpropagating artificial neural networks in land cover classification.*

[6] Pal, M. (2005), *Random forest classifier for remote sensing classification.*

[7] Strahler, Alan H. (1980), *The use of prior probabilities in maximum likelihood classification of remotely sensed data.*

[8] USGS. (2016a), *Landsat Data Access.*

[9] Xu, Min, Watanachaturaporn, Pakorn, Varshney, Pramod K., & Arora, Manoj K. (2005), *Decision tree regression for soft classification of remote sensing data.*