DEEP LEARNING NEURAL NETWORKS FOR LAND USE LAND COVER MAPPING

Christopher D. Storie[1], Christopher J Henry[2]

[1]Department of Geography, The University of Winnipeg, Winnipeg, Manitoba, Canada
[2]Department of Applied Computer Science, The University of Winnipeg, Winnipeg, Manitoba, Canada

ABSTRACT

The importance of accurate and timely information describing the nature and extent of land resources and changes over time is increasing. This research examines the application of deep learning neural networks (DLNN) to the analysis of satellite imagery with specific focus on the production of land use/land cover maps. DLNN have made considerable strides in pattern recognition and machine learning over the last several years. However, their application to remote sensing is less well developed as the technology was originally designed for simple photographs and not satellite imagery. This research presents the results of an experimental study conducted that developed a DLNN to generate land use/land cover maps of the southern agricultural region of Manitoba, Canada. The results of this approach demonstrate a clear advantage in processing time once the DLNN is properly trained when compared to human based semi-automated process.

Index Terms— Neural Networks, Big Data, Machine Learning, Land Cover Mapping

1. INTRODUCTION

The importance of accurate and timely information describing the nature and extent of land resources and changes over time is increasing. Consequently, the accurate production of these land use/land cover maps (LULC) is of critical importance in satisfying this need. The creation of LULC maps, using a supervised or unsupervised image classification approach is one of the most common applications of remote sensing. However, the accuracy and consistency of the maps can vary over time and between different analysts. As such, the automation of LULC map production is of interest to a variety of different stakeholders, as the standardization of production of these products allows for more accurate comparisons of change over time. An emerging solution to this need for automation is the deep learning neural network (DLNN)

LULC maps have many applications, including flood forecasting, urban and rural land-use planning, resource management, and disaster management and planning [3]. GeoManitoba is a Department in the Province of Manitoba (POM), Canada's Sustainable Development Branch, and is mandated to create provincial land-use maps on a regular basis to assist in these activities. Due to budget/labor restraints, limits of current technology, and the ability to obtain relatively cloud free imagery, the process of creating LULC maps from satellite images of the POM can take as long as 2-3 years to produce and can involve as much as 4800 work hours. As a result, the aim of this work was to train a deep neural network to automatically produce LULC maps from Landsat 5/7 data using GeoManitoba's existing set of labelled LULC maps.

This research examines the application of deep learning neural networks (DLNN) to the analysis of satellite imagery. This paper presents a fully convolution deep learning neural network developed to produce 18-category LULC maps for the southern agricultural region of Manitoba.

2. BACKGROUND

In 2012, Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) - by a significant margin - using a deep learning convolutional neural network [4]. DLNNs have made considerable strides in pattern recognition and machine learning over the last several years [1]. A neural network is a mathematical abstraction based (very) loosely on the behavior of neurons in the brain. A neural network is composed of layers of neurons, and the term deep learning characterizes a neural network with many more layers and many more parameters than a traditional neural network. These algorithms "learn" to perform tasks by extracting patterns from large labelled (classified) datasets. Raw data and their related labelled tiles are presented to the neural network for training. Through repeated training on large datasets, the network can extract patterns and perform the correct action for a given input, and increasingly producing error rates much lower than humans [2]. The important observation here is that a human has not explicitly programed the system behavior. Instead, the algorithm "learns" during training by finding the optimal path from raw to labelled data. The outcome is a system, once properly trained, that can classify an image the exact same way each time, thereby removing any human induced bias.

The final solution resulting from this pilot project is based on the work by Jonathan Long, Evan Shelhamer, and Trevor Darrell in using fully convolutional neural networks for performing semantic segmentation of images [5], and subsequent contributions their work inspired [6] [7]. Semantic segmentation is the process of partitioning an image such that each pixel in the image is assigned a unique label corresponding to the perceptual content within the image. The prototype presented in this paper applies this approach to satellite images by modifying the existing network (FCN-8) to accept six channels of data (R, G, B, N-IR, SW-IR, SW-IR). Full a full discussion on the current state of the art of deep learning for remote sensing refer to [8] [9].

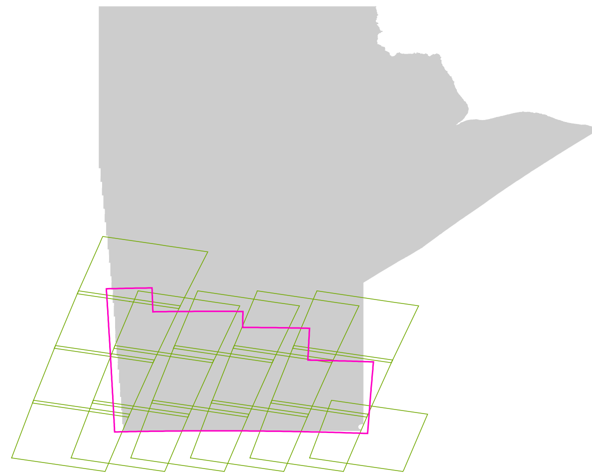
3. DATA AND STUDY AREA

This project was undertaken using LULC data provided by the Province of Manitoba, Canada for 1994, 2000, and 2004. These maps represent the southern agricultural growing region of the province (Figure 1). For each LULC year, the corresponding raw Landsat 5/7 data originally used to the produce the LULC data were acquired. For each year, 18 Landsat scenes are needed to ensure seamless coverage of the study area (Figure 1a and b).

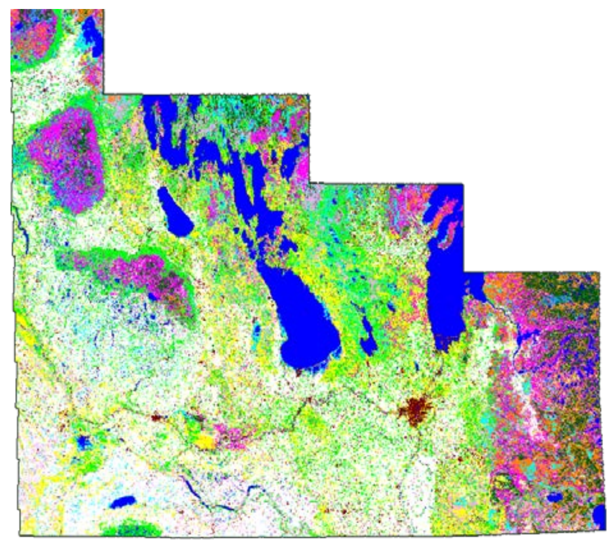
4. METHODS

The first critical step in the development of the neural network was the data pre-processing that needed to occur in preparation for training. The basic steps were as follows:

1. For each LULC year a 6-band mosaic was generated and masked so that the image matched the spatial extent of the corresponding LULC for that year.
2. All input mosaics were named using the following convention – YEAR_U.tif (unlabelled) – while the LULC map for that year was renamed – YEAR_L.tif (labelled). This allows the neural network to know what inputs belong with which classified products.
3. Additionally, a few data adjustments during the pre-processing steps needed to be done, including:
    a. Clouds – The three data years provided did not contain any cloud cover, however, if clouds are present in future years, the neural network needs to have a cloud category to for classification. As such an 18th category was created through the addition of clouds onto the raw imagery. During the mosaicking process, scenes with clouds were placed on top to "keep" the clouds exposed as opposed to being averaged with the underlying scene.
    b. For all mosaics and LULC map products, the same datum and projection were required. Where needed, data was re-projected to WGS 1984, UTM Zone 14.

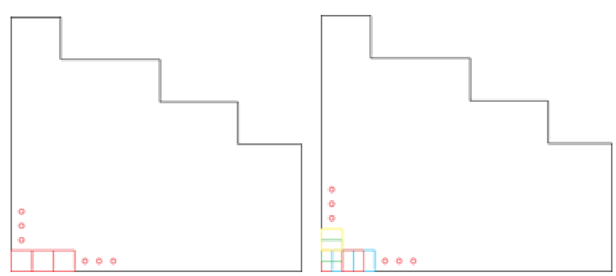


a



b

Figure 1: (a) Province of Manitoba (grey), Southern Agricultural Region (red). Landsat Scene Boundaries (green); (b) 2004 LULC Map producing using a standard supervised classification technique



a b

Figure 2: Tiling process. (a) non-overlapping tiles, and (c) ½ tile overlap (224 / 2 = 112)

4.1 Data Tiling

The deep neural networks considered for this work all were designed around a specific input image resolution (224x224). To produce a viable solution, the original input image resolution for these networks was selected in order to produce the best possible result. In other words, the input resolution of the candidate network was not increased.

This decision was made based on the following observations. First, these networks require a very large number of images to train. Keeping the full resolution of the images mosaics would mean that only 3 training images were available. Furthermore, despite the methods reported in this section, there was simply not sufficient data to train a full network from scratch. As a result, a transfer learning technique needed to be employed [9] (also called fine-tuning), which takes a deep neural network that has been pre-trained on the ILSVRC problem and uses a smaller dataset to train the network for a related, but different problem. Finally, by keeping the same input resolution, the images (both the raw satellite data and associated LULC maps) could be divided into tiles which dramatically increased the number of images (raw and labelled) available for training.

The first set of results were produced by the tiling method shown in Figure 2a, namely the tiles were non-overlapping. To improve these results, the tiling process in Figure 2b was used to produce more tiles. In this case, tiles were overlapped by half the size of the network input resolution, i.e. 224 / 2 = 112. Further, by starting this process in each of the four corners of the full mosaic the total number of tiles generated was increased by more than four times.

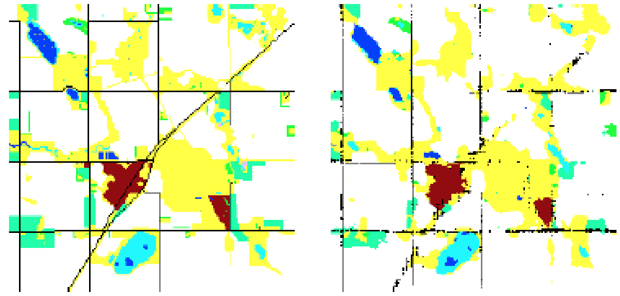
In those situations where a portion of a full tile was created the tiling algorithm "squared" the tile by adding a "no data" value to extra area that the DLNN was designed to disregard. Results using a data augmentation approach where tiles are rotated ninety degree as detailed in [10] were attempted but this technique reduced the training accuracy when compared against the tile shifts for this DLNN.

4.2 Training

The important consideration about training is that labelled examples are used to adjust the network weights in a manner that ultimately finds a global minimum of the error between the output of the neural network and the labels provide by GeoManitoba. It is this reason why an existing large labeled dataset is so important to producing an automated system based on neural networks.

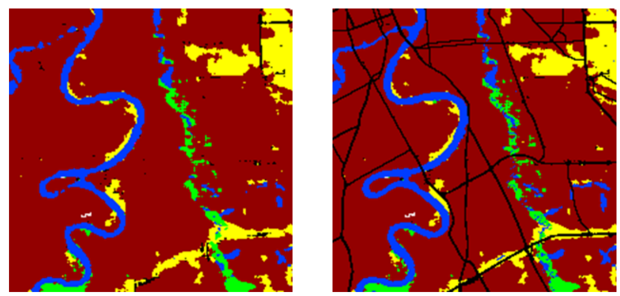
5. RESULTS

The results of this research are promising. In the end, the 2004 data year was used for training as the 1993, and 2000 LULC data did not have a high enough level of accuracy and were confounding the training once the 2004 data was entered When the three years were considered together the accuracy derived was 82.5% (1993), 81.2% (2000) and 79.5% (2004).



a b

Figure 3: Sample results from network trained on 2004 Landsat 5/7 data. (a) GeoManitoba LULC map, and (b) results produced by the FCN-8 network trained using the 2004 Landsat 7 dataset.



a b

Figure 4. Example of improvement due to road & water post-processing. (a) Original output of from FCN-8 VGG-16 network on 2004 Landsat-7 data, and (b) post-processed

To use the 2004 data solely with the tiling method presented in section 4.1, a dataset of 19,039 images were generated, where 18,054 images were used for training and 958 images were used for validation. The best result achieved with the FCN-8 VGG-16 network produced an average accuracy of 88% (Figure 3).

Some interesting observations can be made about these results. First, the initial results took only 2-3 days to train based on non-overlapping tiles. The increased number of images and the push to get the best accuracy possible increased this training time to 7-10 days. However, the training is only required to create the trained weights for the network. Once satisfied with the results, training is no longer required and the only the classification time is considered. The next observation regards the average accuracy. There are two reasons for this loss of accuracy. One is the fact that FCN-8 is not sensitive to fine details that exist at the single pixel level. Due to the spatial resolution of the data being 30m, many of the roads and rivers, features that exist as a single pixel width (or length) disappear as a result of the fine detail loss (Figure 4a). The second reason for the low accuracy is that there may be cases when the FCN-8 output is correct and the original LULC maps is incorrect. Finally, it should be also noted that the fine details associated with roads

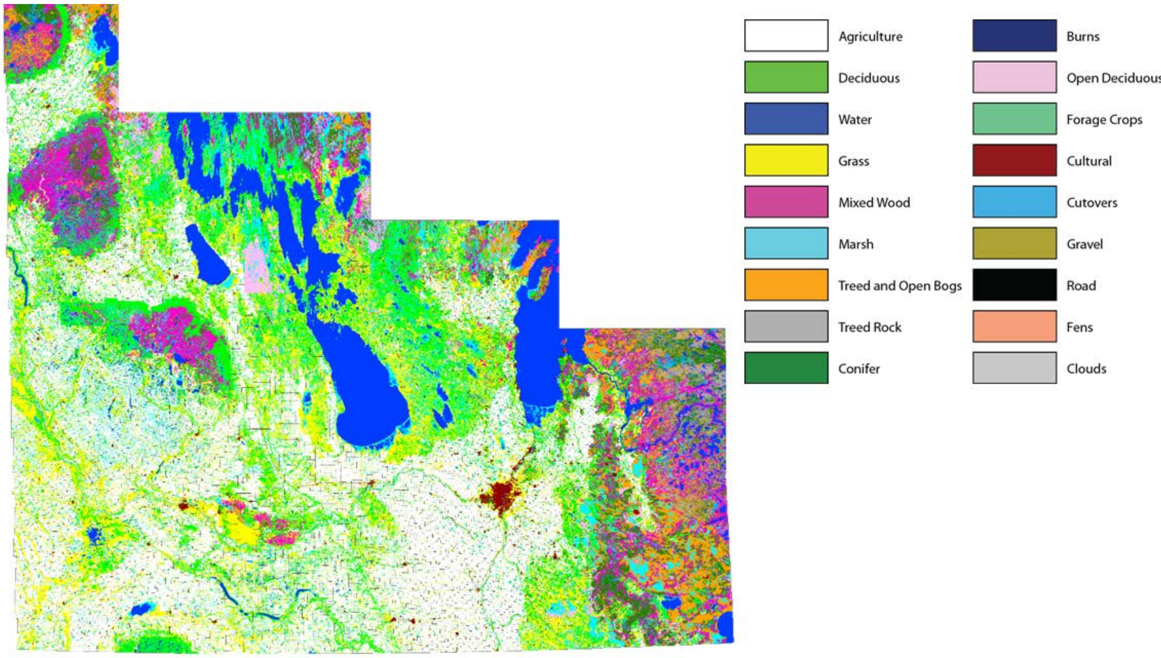


Figure 4: 2010 DLNN Produced LULC Map

and small bodies of water can be easily be added in post-processing through the application of image masks. This process can improve the average accuracy by ~0.25% (Figure 4b).

Based on the successful development of the DLNN a fully neural network based LULC map for 2010, using Landsat 5/7 data was generated (Figure 4). Depending on system configuration and hardware (type of GPU) the average time to produce this LULC map is between 15-25 minutes.

6. CONCLUSIONS

The result of the work presented here is a very powerful technique for quickly producing LULC maps from satellite images. While this particular solution can only be applied to generate a LULC map with eighteen categories as defined by the Province of Manitoba, it does demonstrate that a DLNN can be successfully trained to classify satellite imagery accurate and consistently. Several networks were originally considered for this task with the best results obtained from the FCN-8 deep neural network. While the reported results are impressive, given the lack of sensitivity to single pixel variations in LULC, the loss of these details in the final classified product needs to be overcome through the development of newer DLNN.

7. REFERENCES

[1] K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, “Towards better exploiting convolutional neural networks for remote sensing scene classification,” *Pattern Recognit.*, vol. 61, pp. 539–556, 2017.

[2] K. He, X. Zhang, S. Ren, and J. Sun, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification.”

[3] P. Treitz and J. Rogan, “Remote sensing for mapping and monitoring land-cover and land-use change-an introduction,” *Prog. Plann.*, vol. 61, no. 4, pp. 269–279, 2004.

[4] A. Krizhevsky, I. Sutskever, and H. Geoffrey E., “ImageNet Classification with Deep Convolutional Neural Networks,” *Adv. Neural Inf. Process. Syst. 25*, pp. 1–9, 2012.

[5] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07–12–June, pp. 3431–3440, Oct. 2015.

[6] S. Zheng *et al.*, “Conditional Random Fields as Recurrent Neural Networks.”

[7] F. Yu and V. Koltun, “Multi-Scale Context Aggregation by Dilated Convolutions,” 2015.

[8] X. X. Zhu *et al.*, “Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, 2017.

[9] L. Zhang, L. Zhang, and B. Du, “Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art,” *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, 2016.

[10] X. Yu, X. Wu, C. Luo, and P. Ren, “Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework,” *GIScience Remote Sens.*, vol. 54, no. 5, pp. 741–758, 2017.