

Transfer Learning approach to parking lot classification in aerial imagery

D. Cisek, Y. Lin

Submitted to the 2017 New York Scientific Data Summit (NYSDS) Conference
to be held at New York, NY, USA
August 06 - 09, 2017

Computational Science Initiative
Brookhaven National Laboratory

U.S. Department of Energy

USDOE Office of Science (SC), Advanced Scientific Computing Research (SC-21)

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

A Transfer Learning approach to parking lot classification in aerial imagery

Daniel Cisek
Geospatial Science
Stony Brook University
Stony Brook, USA
dcisek15@gmail.com

M. Mahajan
Technology and Society
Stony Brook University
Stony Brook, USA
manoj.mahajan@stonybrook.edu

Jedidiah Dale
Department of Anthropology
University of Pennsylvania
Philadelphia, PA
jedd@sas.upenn.edu

Susan Pepper
Nonproliferation Department
Brookhaven National Laboratory
Upton, NY
pepper@bnl.gov

Yuewei Lin
Computational Science Initiative
Brookhaven National Laboratory
Upton, NY
ywlin@bnl.gov

Shinjae Yoo
Computational Science Initiative
Brookhaven National Laboratory
Upton, NY
sjyoo@bnl.gov

Abstract—The importance of satellite imagery analysis has increased dramatically over the last several years, keeping pace with the rapid improvements seen in both remote sensing platforms and sensors. As this field expands, so too does the interest in using machine learning methods to automate parts of the imagery analyst’s workflow. In this paper we address one aspect of this challenge: the development of a method for the automatic extraction of parking lots from aerial imagery. To the best of our knowledge, there has been no prior work conducted on the development of an end-to-end pipeline for this particular task. Due to the limited size of our dataset and to accommodate the potentially limited size of future datasets, we propose a deep learning approach using transfer learning. This process hinges upon the use of state of the art Convolutional Neural Networks (CNNs), trained on general image classification datasets. These networks were then fine-tuned on our custom dataset, to establish a comprehensive benchmark for this task. Our method exhibits promising results for automatic parking lot extraction, and is generalizable enough to work with different input types, including high resolution aerial orthoimagery, satellite imagery, full motion video (FMV), and UAV imagery.

Keywords—*Deep Learning; Neural Network; Geospatial; Automation; Satellite Imagery*

I. INTRODUCTION

The use of satellite imagery to aid analysis tasks is a well developed science. In the past, the low resolution and lengthy revisit times of satellites made their usefulness limited. However, over the last several years the capabilities of both sensors and satellites alike have improved to the point where

highly advanced machine learning algorithms can be applied to imagery to identify, delineate, and track objects of interest. Within this field, we specifically look at one problem, the so-called “Parking Lot Problem” [1], a longstanding challenge within the computer vision and deep learning community that focuses on identifying parking lot regions and the vehicles within them. The applications for solutions to this problem are many. Large companies have been known to use reports of how many vehicles are in their parking lots as a major variable in the development of models to predict a stores earnings over a given period. Other fields glean similar insight from aspects of the parking lot problem in the detection of undeclared nuclear facilities [2], the management of parking facilities in large cities [3], and in many other domains.

Transfer learning hinges on the idea that the knowledge gained through the training of a neural network (the source domain) can be harnessed to improve the training process of a network on a new, but similar, task (the target domain). The use of Transfer Learning is currently a popular topic within deep learning community. When done properly, transfer learning can reduce the amount of data you need for training, and often results in a shorter and less computationally intensive training process. Some studies have also reported improvements in generalization performance when initializing with transferred features [4]. However, the use of transfer learning is not a one size fits all approach. Much like the use of deep learning, it is important to recognize that it has its limitations. Determining when it is worth applying and how much apparent benefit it will have is still an open research question, and is one of the primary motivations for this paper.

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Recent papers have found that transfer learning has proven useful in similar visual recognition tasks. By using a 152-layer variant of the popular ResNet neural network trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012, researchers developed a new transfer learning scheme designed to improve the classification accuracy of networks with insufficient data on various deep learning tasks [5]. Despite the shortcomings in dataset size, they recorded results on-par with other state-of-the-art networks in the current research literature. However, the trial and error nature of this approach is made apparent in this paper. They acknowledge that the relationship between source and target domains is critical to the success of transfer learning. But determining what constitutes the best source domain for a given problem is more akin to an art than a science. It remains an active problem within this field, and one that must be solved if transfer learning is to be effectively harnessed for future imagery classification tasks.

Given the uncertain connection between a target domain and the determination of the best source domain, this paper focuses specifically on the evaluation of transfer learning on different pre-trained neural networks for aerial imagery classification. The purpose of this focus is two-fold.

First, we look to provide another approach to solving the “Parking Lot Problem”. We conducted previous research to determine the effectiveness of using a Convolutional Neural Network (CNN) in conjunction with morphological operations for classifying parking lot regions in aerial imagery [6]. In this paper we attempt to build upon the work that was presented in our previous paper, and address shortcomings in our methodology.

Second, testing the effectiveness of transfer learning on aerial imagery classification has the wider goal of providing a benchmark for future imagery classification problems, such as car detection and land cover investigations. As mentioned before, determining the most effective source domain for a given problem is done on a case by case basis. Through our work we hope to shed more light on the effectiveness of several popular pre-trained networks as a source domain, to guide future deep learning approaches to aerial imagery classification.

II. METHODS

Convolutional Neural Networks have seen a burst in popularity over the last few years in classification and pattern recognition problems, following improvements in Graphics Processing Units (GPUs). These increased the amount of computing power that could be harnessed for learning, which allowed for deeper and more complex networks to be trained and evaluated. Today we see CNN’s being used for many different classification tasks, including networks like LeNet for digit classification, and ResNet, GoogLeNet, and AlexNet for object recognition in imagery. The last three of those networks are trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), a dataset composed of over 1 million images spanning 1000 unique classes [7]. Differentiating

between the sometimes slight differences between the different ILSVRC classes requires networks that have very deep and complex layer architectures. Training such networks effectively requires significant computing resources and large amounts of time and data. When dealing with such a large dataset, this isn’t a major problem. However, when looking at a problem set with only sparse amounts of data available, overfitting starts to become a serious issue. In an attempt to reduce the amount of training data needed and jump-start the training process, transfer learning has become a popular means of overcoming some of the shortcomings associated with the traditional training method for neural networks.

Transfer learning applies to the process of carrying over knowledge learned on one task to a related task. In the case of neural networks this process usually takes the form of fine-tuning the weights of each fully-connected layer in the network [8]. In comparison, when a network is trained from scratch the weights are usually randomly initialized, although other weight initialization methods exist.

Transfer learning is based on the idea that each hidden layer of a CNN has distinct feature representation related characteristics [9]. The features of lower layers are responsible for more general features, such as color, texture, and edge detection, and are primarily linked to more local object characteristics [10]. Successively higher levels show increased specialization to the specific learning task the network is trained for. Thus, the goal of our research was to harness the general features learned in the lower layers of several popular CNNs, and use these features to kickstart the learning process of these networks on our custom aerial imagery classification dataset.

To implement transfer learning, we followed the approach followed in [11] and [12]. Their approach relies upon fine-tuning each convolutional layer in the network at a lower learning rate than the network’s default learning rate. Additionally, the final fully-connected layer, responsible for the classifications, was randomized and resized to accommodate the number of classes our problem set deals with. In our experiments and the resulting charts communicated in the following section, the networks trained on this approach are labeled “Network Name: Transfer Learning”.

To demonstrate whether implementing transfer learning had any effects on the classification accuracy of our networks, we established a baseline. This took the form of randomly initializing the same base network that transfer learning was conducted on, so that the changes in accuracy could be compared. Carrying out this approach entailed randomizing the weights in all of the convolutional layers across the network, as well as the final fully-connected layer. This final layer was again resized to fit the number of classes used. Any networks trained in this method will be labeled as “Network Name: Traditional ML”.

To test the effectiveness of transfer learning for aerial imagery classification, we relied upon two different CNNs, AlexNet, and LeNet. Arguably very similar networks in terms of architecture, AlexNet has been previously described as a wider and deeper LeNet, having more weights, and different

filter shapes throughout the network [13]. Preventing overfitting is an ever-present challenge in neural network design. In an effort to prevent overfitting from occurring in these two networks, we more than doubled the amount of data used in training from our last paper, and utilized multiple data augmentation strategies in crafting our dataset. These include random cropping, contrast adjustment, rotation, and mirroring. Incorporating these strategies into the creation of our dataset reduces the chance of overfitting, and can also improve the robustness of our classifier.

LeNet was originally designed for digit classification, and was trained to recognize handwritten digits in 28x28 grayscale images from the MNIST dataset [14]. It has composed of two convolutional layers and two fully-connected layers. LeNet’s architecture is described in more detail in Figure 1. Based upon the task that LeNet was originally trained on, we believed that transfer learning would show little to no improvement in classifying aerial imagery versus being trained from scratch. Although LeNet was revolutionary for its time, its development was before GPU computing truly came into its own, and thus was limited in size and depth. Although well suited for smaller inputs, like digits, its effectiveness at classifying larger images is questionable at best.

The second network that was used for our benchmarking was AlexNet [15]. It is composed of five convolutional layers and three fully-connected layers. A more comprehensive look at its architecture is provided in Figure 1. Given its increased size and complexity, it is more suited to object recognition in imagery. AlexNet is trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset, which is a standard object recognition dataset for computer vision challenges. We expected to see moderate improvement in classification accuracy by applying transfer learning to the AlexNet architecture, given that it was trained for object identification in imagery. Despite the fact that the imagery in the ILSVRC dataset differ substantially from the aerial imagery used in this project, we expected that the lower level features between problems should exhibit enough similarity for us to see improvements.

Our experimental data was obtained through the New York State GIS Clearinghouse [16]. We relied upon high resolution aerial imagery, obtained in 2000 ft. x 3000 ft. tiles, with 1 foot spatial resolution. Each tile was broken down into 60 x 60 pixel patches, each overlapping by 30 pixels. Each of these patches was hand labeled as background (1), parking lot (2), or none (3). The data labeled “none” was obtained from the ILSVRC dataset, to improve the variety of data used. By using the data augmentation strategies described previously, we created a 12,000 image dataset for training and validation.

This process is further explained in Figure 2.



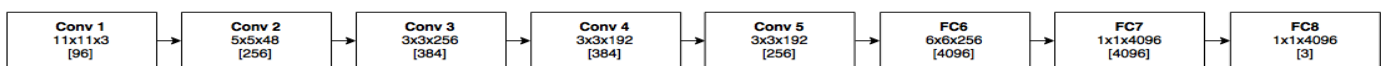
Fig 2. A visual guide to the data augmentation strategies employed for the creation of our custom data set. The Green image represents the original image. Blue represents the Rotated images. Red is the Mirrored image. Black are the images after Contrast Adjustment, and Brown are Randomly Cropped images.

III. RESULTS

For each network, two experiments were carried out. First, training the network from scratch, what we identify as “traditional machine learning”. Second, using transfer learning on the given network, identified as “transfer learning”. Both these experiments are outlined in more detail in the Methods section.

Our networks were implemented in MatConvNet [17], a MATLAB-based deep learning framework, which allowed for easy access to the weights of the convolutional layers. Included in this framework are several graphing tools, which enabled us to display our classification results in a highly accessible manner. We plot the change in two variables over successive epochs. First, we measure Objective. The concept of Objective is tied very closely to that of loss and network optimization, in that the overall goal of training a network is to minimize your given loss function. We can track attempts at minimizing our loss function over time through the Objective. Second, we look at Top-1 Error. Top-1 Error refers to what percentage of the time your network’s highest probability class is different from the target label. By determining our Top-1 Error we also then know the “accuracy” of the network, at least on the validation data set. Finally, we examine the generalization/test accuracy of each network, which refers to the accuracy exhibited on the unseen testing dataset.

AlexNet



LeNet-5

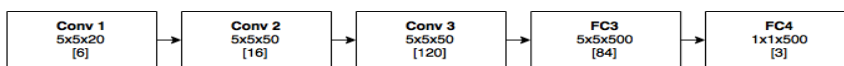


Fig 1. A flow chart representing the weights of both AlexNet and LeNet-5. The number inside the lower brackets indicates the number of filters with the dimensions mentioned above it.

Doing so provides a more accurate look at the network’s accuracy, and also will identify if significant overfitting has occurred. The final accuracy indicates what percentage of the testing dataset the network correctly identifies.

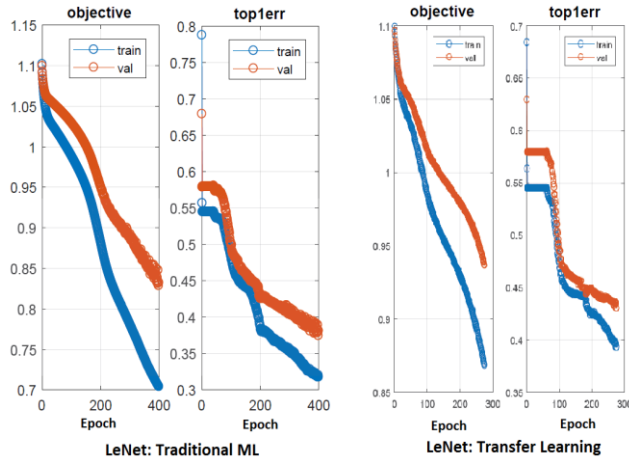


Fig 3. The results of our experiment, performed on LeNet.

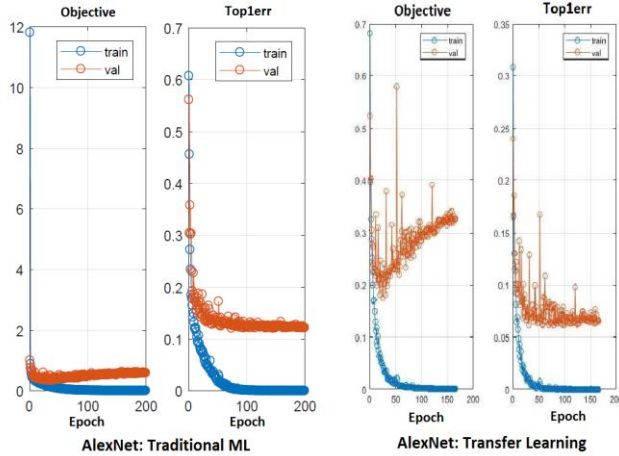


Fig 4. The results of our experiment, performed on AlexNet.

Network	LeNet: Traditional ML	LeNet: Transfer Learning	AlexNet: Traditional ML	AlexNet: Transfer Learning
Epochs	400	300	200	150
Batch Size	100	100	100	100
Learning Rate	0.0003	0.0003	0.0003	0.0003
Train Obj.	0.74	0.87	0.003	0.001
Train Top1Err.	34%	39%	0%	0%
Val. Obj.	0.841	0.936	0.591	0.336
Val. Top1Err.	37.4%	44.1%	12.5%	7.3%
Val. Accuracy	62.6%	55.9%	87.5%	92.7%
Test Accuracy	56.3%	49.7%	85.2%	94.3%

Table 1. Implementation details relevant to both networks, as well as the accuracy exhibited in all experiments.

The primary results of our tests are shown in red in Table 1. As expected, transfer learning produced no improvements in LeNet imagery classification accuracy versus random initialization. In fact, its accuracy actually decreased. However, when looking at AlexNet, we note a 10.7% increase in generalization accuracy by using transfer learning. This result indicates that using ILSVRC as a source domain for imagery classification worked well in this instance.

IV. CONCLUSIONS AND RECOMMENDATIONS

The main focus of this paper was testing the effectiveness of transfer learning for aerial imagery classification. The goal of this research was two-fold. First, to build upon our prior approach to solving the parking lot problem via CNN. Second, to provide a basis for future transfer learning approaches to overhead imagery object identification. By testing the effectiveness of transfer learning on several CNNs we provide groundwork to support the relationship between a target domain and choosing an effective source domain for it. We demonstrated that using transfer learning on AlexNet showed an increase of 10.7% in generalization accuracy over its trained from scratch equivalent.

Future research should implement a similar testing procedure for other popular CNNs, like GoogLeNet and ResNet. Testing the effectiveness of these networks in a similar environment will further establish the connection between source and target domains, and allow for a more conclusive answers to be drawn about using transfer learning for future overhead imagery classification tasks.

ACKNOWLEDGMENT

We wish to thank the Computational Science Initiative, the Nonproliferation and National Security Department, and the Office of Educational Programs at BNL for their guidance and assistance throughout this project.

REFERENCES

- [1] Al-Kharusi, H and Al-Bahadly, I, "Intelligent parking management system based on image processing," World Journal of Engineering and Technology, vol. 02, no. 02, pp. 55–67, 2014.
- [2] Niemeyer, I. and M.J. Canty (2001). "Knowledge-based interpretation of satellite data for safeguards purposes by object-based and multi-scale image analysis". In: Proceedings, 23rd Symposium on Safeguards and Nuclear Material Management, ESARDA 2001, Bruges, Belgium, 8-10 May 2001.
- [3] Valipour, S., Siam, M., Stroulia, E., & Jagersand, M. (2016). "Parking-stall vacancy indicator system, based on deep convolutional neural networks". 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT). doi:10.1109/wf-iot.2016.7845408.
- [4] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. "How transferable are features in deep neural networks?" In Advances in Neural Information Processing Systems 27 (NIPS '14), NIPS Foundation, 2014.
- [5] Ge, W. and Yu, Y. "Borrowing Treasures from the Wealthy: Deep Transfer Learning through Selective Joint Fine-Tuning". In IEEE Conference on Computer Vision and Pattern Recognition 2017 (CVPR 2017).

- [6] Cisek, D., Dale, J., Pepper, S., Mahajan, M., and Yoo, S. (2016). Parking lot delineation and object detection using a localized Convolutional Neural Network. Proceedings of IEEE 2016 New York Scientific Data Summit (NYSDS), New York University, New York, NY. doi:10.1109/nysds.2016.7747821.
- [7] Russakovsky, O., *et al.* "ImageNet Large Scale Visual Recognition Challenge." IJCV, 2015.
- [8] Han, S., Pool, J., Tran, J., Dally, W. "Learning both weights and Connections for Efficient Neural Networks". In Proceedings of The 28th International Conference on Neural Information Processing Systems (NIPS '15), 2015.
- [9] Akcay, S., Kundegorski, M. E., Devereux, M., & Breckon, T. P. (2016). "Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery". 2016 IEEE International Conference on Image Processing (ICIP). doi:10.1109/icip.2016.7532519.
- [10] Ng, J. Y., Yang, F., & Davis, L. S. (2015). "Exploiting local features from deep networks for image retrieval". 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). doi:10.1109/cvprw.2015.7301272
- [11] Shin, H., *et al.* (2016). "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning". IEEE Transactions on Medical Imaging, 35(5), 1285-1298. doi:10.1109/tmi.2016.2528162.
- [12] A. S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition", Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 512-519, 2014.
- [13] Chen, Y., Emer, J.S., Sze, V., & Yang, T. (2017). "Efficient Processing of Deep Neural Networks: A Tutorial and Survey". CoRR, abs/1703.09039.
- [14] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
- [15] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). "ImageNet classification with deep convolutional neural networks". Communications of the ACM, 60(6), 84-90. doi:10.1145/3065386.
- [16] New York State. "NYS interactive mapping gateway,". [Online]. Available: <http://gis.ny.gov/gateway/mg/>. Accessed: August, 2017.
- [17] Vedaldi, A. and Lenc, K. "MatConvNet – Convolutional Neural Networks for MATLAB". In Proceedings of the ACM International Conference on Multimedia, 2015.