

Land-Use Classification for Modeling Urban Population Dynamics

Daway Chou-Ren, 2017

Advisor: Szymon Rusinkiewicz

October 8, 2016

1 Motivation and Goal

A proliferation of geo-referenced images in easy to search online stores such as Flickr, combined with the rise of extremely powerful deep learning algorithms for image classification, means that it is now possible to make new geographic knowledge discoveries. By augmenting powerful pretrained image classifiers, we can classify thousands of classes of images and even understand actions occurring within scenes, such as dancing or eating. By mapping an image's location to its content, we can generate location-wide maps of human activity. This has the potential for a wide variety of applications, such as discovering population flows, geographic and temporal activity centers or even generating urban sentiment maps by extracting emotion from images. The goal of this project will be to explore geographic knowledge for a particular urban area.

2 Related Work

2.1 Geo-referenced Images

Web applications such as Flickr allow users to upload images along with their geo-tags. This has resulted in a massive dataset of geo-referenced images. As of 2014, Flickr had over 6 billion photos, with 50 million more being uploaded each month. This publicly available information has been used for a variety of interesting tasks, such as summarizing and visualizing similar images, creating virtual flythroughs and searchable databases for tourist locations, and even generating 3D models of entire cities. [7] [2] [1]

2.2 Deep Learning Image Classification

In the past decade, deep learning models have proven highly adept at various tasks, such as speech recognition, image object detection, and overall pattern recognition in many fields. Backpropagation algorithms are used to modify weight parameters in many layers of input/output nodes. By tuning activation parameters for nodes in each layer, different representations can be detected. Various layer architectures have been invented to reduce the number of parameters that need to be tuned, to allow for input convolutions, and to allow for input image translations. Deep convolutional nets have proven especially adept at producing layer activations which can correspond to object recognition. [8]

There are an abundance of high-quality pre-trained image classifiers, such as the ResNet and VGG implementations, which have been trained on ImageNet, which achieve high accuracies at labeling everyday objects. [9] [6] [5] Datasets such as the Scene Understanding (SUN) and Google Places datasets have been used to train classifiers to recognize places such as zoos, beaches, and other locations. [10] [11]

One subfield of deep learning image classification is transfer learning, where a deep neural architecture is pretrained on a large dataset, such as SUN, Places or ImageNet. Embedded within the trained architecture will be a final output feature layer, which represents image features that have been identified as being able to distinguish between various labels in the training set. If a network has been trained on Google Places, for example, extracting this layer for any inputted image will return a feature vector representation that the network deems will best differentiate whether the image is of a beach, lake, city, etc. This feature vector can then be run through a different network trained on a different dataset, transferring what the first network learned about the image to the second. In this way, powerful image classifiers can be trained to highly specific image recognition tasks. [3]

3 Approach and Plan

We will

1. Do an extensive literature review of deep learning architectures used for image classification as well as general geographic image research
2. Generate a high-quality training set of geo-referenced images and their labels
3. Use this set to train an image classification system for general image understanding/scene labeling, most likely by using a pre-trained classifier and augmenting it with transfer learning to our application-specific image training set
4. Pull an geo-referenced image set for an urban area, most likely from a service such as Flickr
5. Apply the image classification system to our image set in a selected urban area
6. Explore the discovered spatial-temporal patterns of human activity and/or image sentiment
7. Refine our classifier to extract desired information that appears interesting

There are many immediate applications for a spatial-temporal map of human activity or sentiment. One is to identify where and when activities are occurring, such as looking for where soccer is being played on weekends or looking for romantic date spots on Fridays. But we expect that less straightforward sociology and civil engineering applications will reveal themselves. An obvious example of this might be discovering traffic accident hotspots.

Unlike several land-use classification studies which use labeled datasets such as the popular UC Merced, we will need to create a labeled set that better describes recreational activities. [4] The UC Merced set has labels such as agricultural, airplane, denseresidential, freeway, and storagetanks, whereas we will most likely want activity-based labels like eating or dancing or general descriptors like romantic to describe restaurants.

4 Evaluation Plan

As of now, this project's goal is fairly open-ended and will probably be narrowed to a specific knowledge discovery purpose as a data set is prepared for exploration. Successful subgoals will be easy to measure, however. These will include

1. Preparation of a high-quality labeled training image set with geo-referenced mappings
2. Training of a robust image activity or image sentiment classifier
3. Applying the classifier to images in an urban location

References

- [1] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M Seitz, and Richard Szeliski. Building rome in a day. In *2009 IEEE 12th international conference on computer vision*, pages 72–79. IEEE, 2009.
- [2] Shane Ahern, Mor Naaman, Rahul Nair, and Jeannie Hui-I Yang. World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 1–10. ACM, 2007.
- [3] Yoshua Bengio et al. Deep learning of representations for unsupervised and transfer learning. *ICML Unsupervised and Transfer Learning*, 27:17–36, 2012.
- [4] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092*, 2015.
- [5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.
- [7] Alexandar Jaffe, Mor Naaman, Tamir Tassa, and Marc Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 89–98. ACM, 2006.
- [8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [10] Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, pages 3485–3492. IEEE, 2010.
- [11] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*, pages 487–495, 2014.
- [12] Yi Zhu and Shawn Newsam. Land use classification using convolutional neural networks applied to ground-level images. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 61. ACM, 2015.