Aerial Land-Use Image Classification

Alexander Archer, Linda Li, Brody Vogel, Jinghao Yan

Abstract

This project explores applications of Convolutional Neural Network (CNN) models in the classification of land-use images. Models were trained with and without transfer learning to classify images into one of 21 possible categories. The models made use of various pre-existing architectures and kernels to maximize the accuracy. Overall the methods detailed here achieved relatively strong levels of success – with some models attaining accuracies of over 90 percent.

Introduction

With the world in constant change, having the ability to distinguish between various land-uses could be valuable in many fields. Here, we try to do just that – classify images into one of 21 land-use categories. Our goals are two-fold: (1) To achieve the best accuracy we can, and (2) To understand what the various layers and activations in the models are doing with the data. In this respect, half of the project focuses on the internal structure of the models, and the other half the output and results.

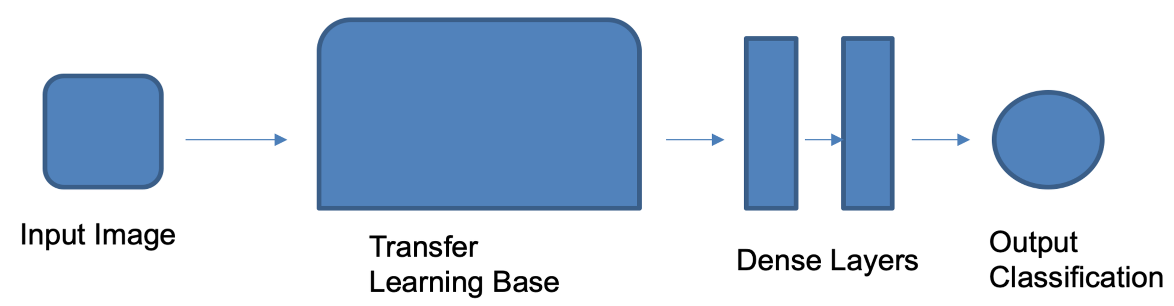
Data

Our dataset consists of 2,100 images belonging to one of 21 classes. The images were manually extracted by the University of California, Merced, from the USGS National Map Urban Area Imagery collection. A sample image from each class is below:

Each RGB image measures 256x256 pixels and the pixel resolution of this imagery corresponds to 1 foot.

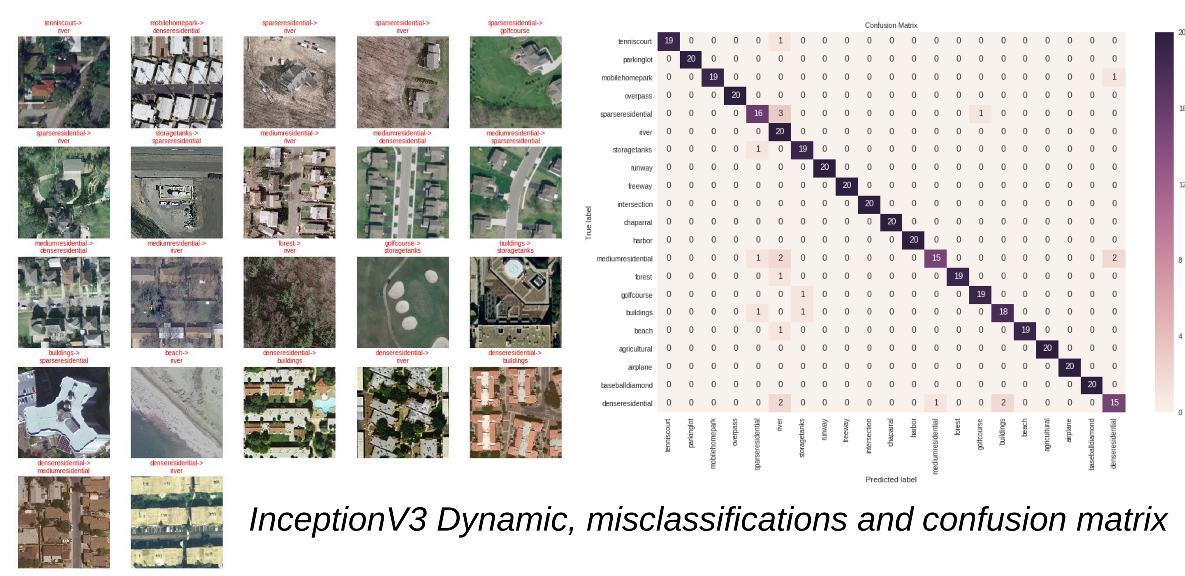
Methodology

In total, 5 models were trained on 80% of the images: a CNN built end-to-end, from scratch; a CNN using the frozen VGG16 weights as the base; a CNN using VGG16 weights that could be fine-tuned as the base; a CNN using the frozen weights of Google's InceptionV3 model as the base; and a CNN using Google's InceptionV3 model weights that could be fine-tuned as the base. A representative model blueprint is shown below:

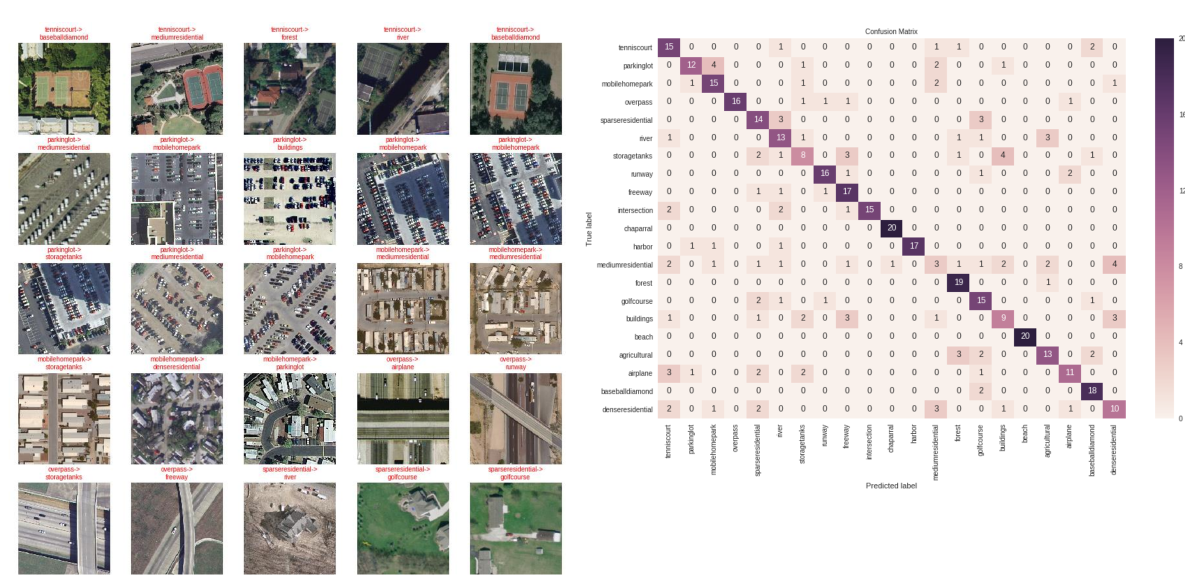


Results

Model	Test Accuracy
VGG16 Static	0.92619
VGG16 Dynamic	0.68095
InceptionV3 Static	0.77857
InceptionV3 Dynamic	0.94761
Full CNN	0.70476



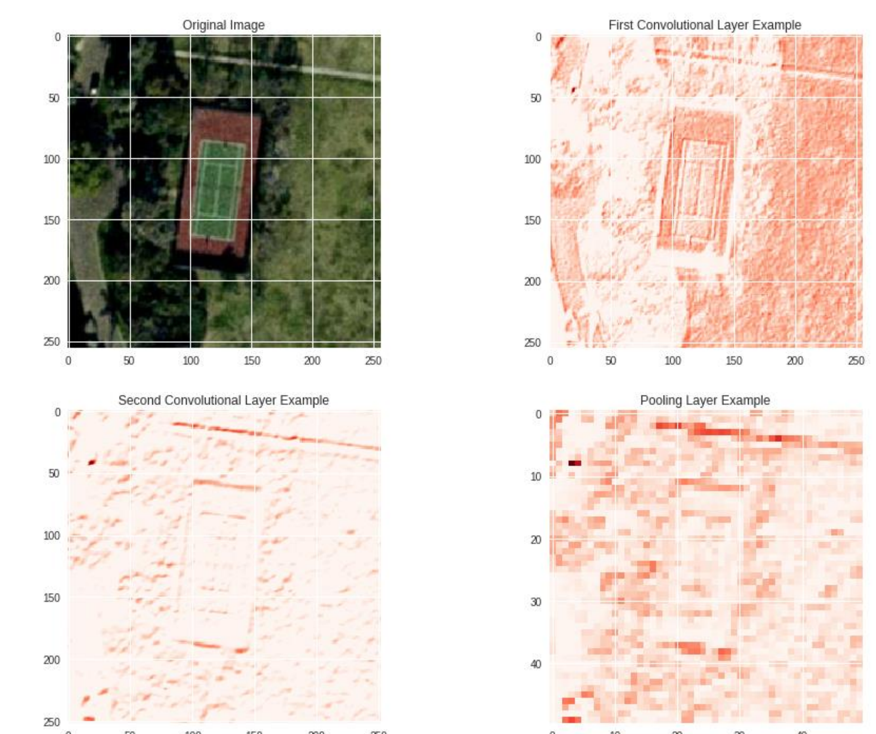
InceptionV3 Dynamic, misclassifications and confusion matrix



Full CNN, misclassifications and confusion matrix

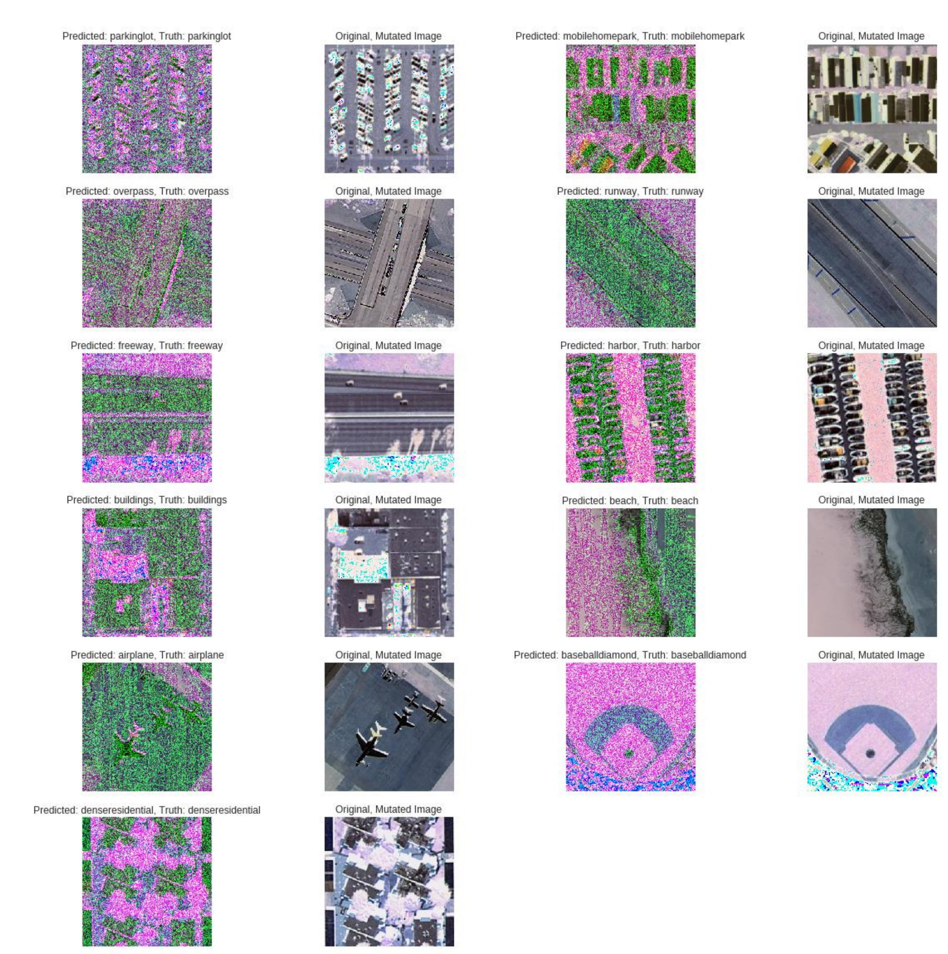
Visualizing Kernel Preferences

Using the CNN built end-to-end, a single image can be passed through the network. Tensors can then be extracted from layers within the network to examine the features being learned at that stage:



The first layer looks for patterns such as lines and densely colored pixels. The latter layer begins to look for finer details in the images, and lastly the final layer picks up details that are unique to the predicted class (e.g. lines on the tennis court).

In a slightly different manner, the *segments* of images that highly influence the predicted class can also be examined:



This suggests the model is capable of recognizing entities like airplanes and baseball diamonds but struggles with patterns and textures within river and agricultural images.

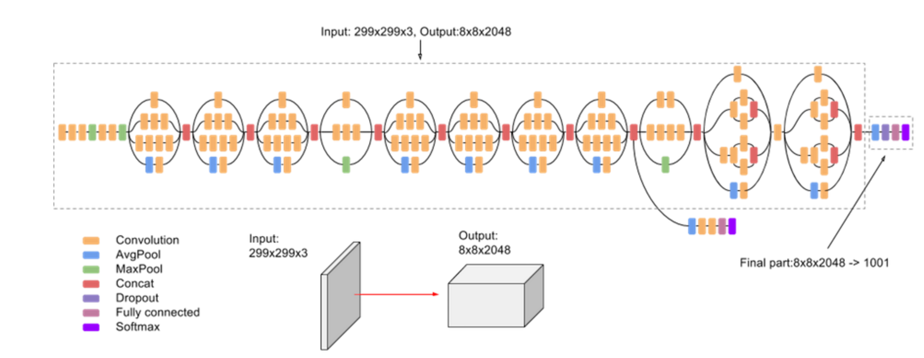
Discussion

Based on the performance results, it seems that the dynamic InceptionV3 model performed best in terms of test accuracy, with the static VGG16 model being a close runner-up.

One possible reason for the InceptionV3 model's success is that it is the most complex model used. The results produced are on par with current research, where InceptionV3 has outperformed many other well-established models in the field.

With the static models, however, the pre-trained VGG16 weights yielded better results than the InceptionV3 weights. The models therefore suggest that fine-tuning does not always improve performance.

Pictured below is the architecture of InceptionV3.



Conclusion and Future Work

This project's best models are strong classifiers of land-use images. If work on this project were to be continued, one may consider modifying some of the image categories. Images of medium and dense residential land-use are similar and can be difficult to distinguish even for a human. Currently, these models can only classify individual images taken from a small range of altitude levels. These models could be made more robust by training on images of a wider range of scales. Another extension of this project could be to develop a detection algorithm that can identify and locate multiple land-use categories within an image. Regardless, this project achieves strong accuracy scores, has important applications, and may be used to propel future modeling.

Acknowledgements

Yi Yang and Shawn Newsam, "Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification," ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010.