

CS461 Fall 2016: Deep Learning For Object Recognition on a Mobile Robot

Miles McCall

October 20, 2017

Abstract

Our senior capstone project is based on a preexisting project being worked on by Dr. Smart and his lab. We will be contributing to their mobile robot platform by modifying and improving the image classification software currently running on the machine. This software incorporates a room navigation system, depth sensors and cameras, and a Convolutional Neural Network (CNN). To improve upon these systems we plan to test an object tagging system to guide the robot's data collection, test the data collection under varying environmental conditions, and analyze the performance of several different neural network models. With all of these efforts combined we hope to create a measurable increase in the accuracy and consistency of the robot's ability to recognize a set of household objects.

1 Definition and Description of Problem

Many current Machine Learning models use batches of input data that are fed through the CNN to train the network. The network is only trained on this particular set of data, however, and new data cannot be added to the set without "forgetting" what the network has "learned" and retraining from scratch.

Part of the way we plan to improve the image classifier is by implementing a Sequential Online Learning neural network model, as opposed to the currently running Batch Learning model. The goal is for the robot to be capable of training itself with an initial dataset, then gathering further data as needed to make the CNN more accurate. By using a more complicated model we should be able to append to the created datasets the robot already recognizes.

Another way we plan on improving the image classifier has to do with a Machine Learning concept known as overfitting. How well trained a CNN is to a specific dataset can be described as how "fit" the model is. If the model isn't trained enough on the dataset, it is under-trained and won't make accurate predictions of the image contents. If a CNN is overfitted to a dataset, however, it has been over-trained to the point that it is too specific, and will recognize objects from the original set extremely well and other data poorly.

We will pursue the idea that this concept of overfitting can be used as a tool instead of a hindrance. The goal is to build classification models that are specific enough to recognize individual objects in its environment. Instead of recognizing that the object on the table is a mug, it will be overfitted to tell my mug apart from Dr. Smart's. Overfitting the CNNs to classify individual objects inherently sacrifices the model's ability to classify other objects, but by creating a hierarchy of object classes such that my mug falls under the category of "cup", we may be able to achieve specific classifications as well as more broad identification.

The last problem the project is attempting to address is controlling for the variance in environmental conditions when collecting data. While humans are able to easily recognize the same object in all sorts of positions, lighting conditions, and from many angles, this task remains difficult for image recognition softwares. Creating a quality dataset to deal with the drawback of analyzing three dimensional objects is not an easy process. Images of an object must be taken from all angles so the initial dataset captures all import features of the object. This would allow the robot to classify the same mug twice even if it was flipped upside down the second try, as it is referencing a bank of images that covers these variations. Obscuring the robot's view of the objects, blurry photos, and dimmer, brighter, or different colored lights, however, are much harder variables for the same CNN to see past.

We plan to address this issue of varying environmental conditions by implementing the Sequential Online Learning model. This will allow us to add onto the original classifier trained on normal conditions datasets gathered in a range of each variable above. By integrating multiple sequential datasets into the same neural network, we plan to improve the robot's ability to recognize the same object on par to a human's ability.

2 Proposed Solution

To address overfitting and the improved accuracy in multiple environmental conditions, we will create multiple datasets using the robot's cameras, depth sensors, and an improved image tagging algorithm. By adding AR code scanning and potentially by guiding lines we plan to improve the robot's ability triangulate where the object it is scanning exists in its environment. The robot can utilize these guides can its depth sensors to take more on-subject pictures with a more accurate 360 view of the object. We will then use this improved technique in an array of lighting conditions to recognize the same object regardless of time of day or whether the object is inside or outside.

Improving the robot's ability to create datasets of objects also allows us to improve the software's ability to then train overfitted CNNs on the specific object. The clearer, more diverse, and better fitted the input images are, the more accurate the CNN will be with its predictions. We can then test our hypothesis that an overfitted network can still be used to identify objects for a more general category.

Lastly, using a Sequential Online Learning model, the robot will be able to build upon these varying datasets and make itself increasingly more accurate. This is still a relatively recent area of interest in Machine Learning research, and poses an interesting question to investigate.

3 Performance Metrics

The final software package should generate multiple datasets through a semi-guided process on several distinct objects. These datasets can be manually checked and compared for quality by a user, and are used for the comparison of multiple neural network models.

There are multiple popular image classification software packages currently available for use that we will likely take advantage of throughout our project. There are also many classes of objects that are typically used by the Machine Learning community when training neural networks, and the algorithms used have published performance metrics available for comparison. We hope to use these preexisting algorithm success percentages as benchmarks to measure our success. If a certain model is known to predict mugs at an eighty percent success rate, we would hope our improved model outperforms the original, and if not investigate why.