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

Michael Rodriguez

October 12, 2017

Abstract

For our capstone project, my team and I will be helping train an autonomous robot to recognize objects using image recognition. We will be using a robotics programming language known as ROS (Robot Operating System) to build upon the already existing robot platform that it currently boasts. The current platform doesn't have near the desired object recognition success rate that Professor Smart finds ideal so we plan on increasing the success rate by implementing Convolutional Neural Networks with online learning. We will also take into account environmental variables such as the time of day or the background that surrounds the object so that our robot can still recognize objects successfully in less than optimal conditions.

1 Definition and Description of Problem

A big issue that we face with the deep machine learning aspect of this project is that it isn't easily managed or compliant when trying to hone previously learned recognition of an object. It mainly has to do with the weights that change when we feed it new data which ultimately leads to a loss of previously held knowledge. This causes what is called the Catastrophic Interference and prevents our robot from building upon it's already known data. This means that whenever new input is received, the long process of training will take the same amount of time (roughly an hour) to learn an object regardless of whether it's already been trained on it or not.

When we met with our client Professor Smart, this was one of the topics that we heavily discussed. He suggested that we approach the problem with a Sequential Online Learning Approach instead of the Batch Learning approach that is normally taken as it's easier to implement. With Batch Learning the robot gets fed a series of training samples that pass through the network at the same time and then get network weights assigned to them so that the neural network can classify the data. Now with the Online Learning Approach, we are able to make the network receive plenty of more data over time while having computational resources not get heavily used. In addition to using this new approach, we will also have to take into consideration environmental factors like trying to recognize a coffee mug with a white background and also with a darker background such as a brown table. Professor Smart advised that we take a "tight crop" meaning that we try to localize a tight region that contains the object in order to ignore irrelevant data that surrounds the object.

Another image recognition problem, or rather question, that we are going to encounter is whether the robot will better recognize objects classifying them exactly as how they exist in 3D space or generally by the shape/color that it has. For example, if the robot classifies a coffee mug, is it going to store it as an indistinguishable general-purpose mug or exactly capture every minute detail of the mug making it unique? Again, this is something Professor Smart talked us through and he suggested that since the robot will likely be placed in a normal room (used his office as an example) where it isn't likely to find more than 2-3 mugs laying out in the open, that we should give each mug a unique classification. However, he did say to experiment a little with a general image container to see how much different of a success rate we could achieve with the robot simply trying to generally classify different coffee mugs.

2 Proposed Solution

In order to best train our robot using a Convolutional Neural Network, we must try to present it with as many examples that pertain to the same overall class. Meaning that we continuously train the robot on the same object but from a variety of different angles, backdrops, and amount of light in the room. This will help us build a better main classifier while also helping build our sub-classifiers concurrently. Our main priority is to get the highest recognition success rate possible while also being able to handle a gargantuan amount of training data to further build its current knowledge.

To successfully build our overarching classifier in addition to the sub clas-

sifiers, it is critical that they're able to distinguish individual mugs from one another. An example of this could be that my two roommates have the same shaped mug with no writing or images on it but are in fact different colors, say one is green and the other is blue. Our robot should recognize that both objects are indeed mugs but be able to classify the blue mug as roommate A's mug and the green mug as roommate B's mug. We can achieve this by training our Convolutional Neural Network by applying different Deep Network architectures that will allow the robot to distinguish general objects that have slightly different features. One of our biggest challenges that we are going to overcome is the Catastrophic Interference that interferes with our robot being capable of continuously learning new data. Some techniques that we plan on using to attempt to solve this problem are Elastic Weight Consolidation, Function Approximation, and Dual Memory Architectures.

3   Performance Matrix

Professor Smart already has some grad students researching Deep Learning with ROS in the OSU Robotics Department. So, we will be collaborating with them to get a feel for what our final project should start looking like by the time we're approaching the finish line. To best evaluate our performance as a team, we need to consider the final training framework that we develop and how our CNN architecture classifies a variety of distinct objects. The training network will consist of leveraging its visual recognition capabilities using a point of reference to best classify the object in its local environment. We can achieve this by either improving the current training network or completely building our own. Either way, the main goal is to improve the final recognition success rate. By the end of our project our classifier should be able to correctly identify 3 different objects with an average success rate of at least 80%. To accomplish this, we will need to do some heavy research in Deep Learning architectures and Convolutional Neural Networks to see which combinations work the best or worst.