

# Deep Learning for Object Recognition on a Mobile Robot

Julian Weisbord

October 18, 2017

## **Abstract**

For our senior design capstone project, we will build an image classifier on top of an autonomous robot. By leveraging ROS (Robot Operating System) and the existing mobile robot platform, we can devote all of our resources to sequentially training a Convolutional Neural Network (CNN) with online learning. In this project, we propose a plan of action and several potential solutions to the issues brought about from sequential learning. Additionally, there are multiple environmental variables that must be addressed during training in order to classify everyday objects in a wide variety of settings. To build a robust classifier, we will pursue three different methods to improve on the current system at the Personal Robotics Lab of OSU. These are: new data capture methods, overfitting to data in different environmental contexts using multiple classifiers, and testing different online learning models to achieve the best classification rate.

# 1 Definition and Description of Problem

One of the biggest issues Machine Learning as a whole faces is that deep neural networks aren't very amenable to online learning techniques. When new data is passed through the network, weights change via backpropagation and previous "knowledge" is lost. This process is known as Catastrophic Interference and without the ability to sequentially receive and classify new training data, intelligent agents will have to be retrained on the entire data set whenever it is presented with a new input sample to predict.

Most artificial neural networks are based on Batch Learning which is the opposite of Sequential Online Learning. In Batch Learning, a group of training samples is passed through the network all at once, the network weights are optimized, and the CNN is ready to classify data. With Online Learning, the network has the potential to receive significantly more data as time goes on and computational resources aren't as taxed all at once.

Catastrophic Forgetting brought out from Sequential Learning is one of many challenges within machine learning. Our group must also understand and be able to implement algorithms that classify images in 3-space. This task is anything but trivial, identifying one object among several can require a lot of human input. To determine how far away and how large an object is, we must establish a coordinate system from a reference object such as an AR tag. This will show how far away the desired object is because we know where it is in relationship to the reference tag.

Another problem is the fact that classifying different objects requires different features to localize and take advantage of. When trying to classify a white floor, depending on the time of day, light may be refracting off of it in a way that makes the floor appear yellow or it could get dirty. It would be a much better choice to determine if a floor is truly a floor based on shape and position in a room. However, when classifying a black mug, it is going to look the same and be invariant to many environmental changes. Therefore, it will be simpler to take a few common household objects that aren't especially difficult to classify and try to make those work with a Sequential Machine Learning architecture. Taking a few different object classes and training the networks to recognize only those with high accuracy is known as overfitting. Overfitting or high variance is the concept that the criteria used to train a statistical model is also the criteria that successful classification is judged on. This will cause the network to successfully classify objects it has been trained on but not new dissimilar objects. The purpose of this experiment is to see if an intelligent agent can overfit its environment over a long period of time, allowing it a specialized knowledge of its surroundings.

Without localizing a tight region containing the object of interest, there could be a lot of extraneous data. For this project, we will try our best to provide the classifier with objects that aren't obscured or obstructed by environmental debris. It is possible that some training data could involve multiple objects within the robots view but the majority of the desired object will be in plain view.

## 2 Proposed Solution

To train the CNN we will select a small dataset and strive to pass in as many different types of objects within the same overarching class, e.g. types of chairs. The more different examples of the same object that are trained, the better the overarching classifier, and sub classifiers. Continually training a large volume of data is important but in order to get high rates of classification accuracy on our training dataset, we will overfit the networks to varying environmental contexts (unusual angles, colors, object orientations, distances).

In addition to classifying a few everyday household objects with high accuracy, this design project will focus on building sub classifiers within an overarching classifier. For example, if one lives with roommates everyone likely has their own mugs, and we can train a CNN to tell individual mugs apart by applying multiple deep networks to classify the mug each honing in on different features. To be able to build classifiers that can continuously be trained on new object data sets, we will also implement different solutions to mitigate Catastrophic Interference. This is still a heavily researched area with no answer that covers all bases, we will experiment with different techniques such as Elastic Weight Consolidation, Function Approximation, and Dual Memory Architectures to achieve best results.

## 3 Performance Metrics

Deep learning with ROS is already being researched in the Personal Robotics Department of OSU, a successful team will collaborate closely with grad students in an effort to build the best classifier. The main performance metrics to consider consist of how our team creates a training framework and how well the neural network architecture classifies several distinct objects. The training dataset should be created with visuals of the object in several different angles/plains of view. The training framework and new data capture methods should supplement the current process done in the Robotics Department of OSU and ideally provide a noticeable increase to the final rate of classification. Specifically, this classifier should correctly classify at least 3 different object classes, with an average success rate of 80% or higher. To guarantee the best results, our team will research how different CNN architectures affect the rate of classification and report the advantages and setbacks of each.