

Group 67 - Technology Review

Deep Learning for Object Recognition on a Mobile Robot

Miles McCall

CS 461 Fall 2017



Abstract

The Technology Review is used to break the project down into its core components with three main points describing each section. For each point, three technology options are compared and contrasted, and the one most likely to be implemented for the project is explained.

CONTENTS

1	Role	2
2	What you are trying to accomplish	2
3	Tech Review	3
3.1	Project Overview	3
3.2	Image Recognition Part 1	3
3.2.1	Overarching Classifiers	3
3.2.2	Sub Classifiers	4
3.2.3	Overfitting in the Algorithms	4
	References	5

1 ROLE

Our group has three members, and we all have equal responsibility across the project. At this point we are all working on every piece of the project to establish a consistent knowledge base. Further along into developing our pipeline we may divide tasks more and gain more expertise in a specific portion of the project, but for now we are all working together equally.

I am particularly interested in the image classification and neural network involved in the project. While the whole project is technically about deep learning and object recognition, the portion that covers the details, implementation, and design philosophy behind these sections is what I would prefer to focus on.

2 WHAT YOU ARE TRYING TO ACCOMPLISH

Our group is working to rewrite and improve upon the current software on the Fetch robot at the Personal Robotics Lab on campus. Through this redesign we aim to recreate the data collection and image recognition systems currently in place.

While the data gathering software currently being used on the robot, or intelligent agent, is likely adequate in its current state, our group would like an improved version to match the new classification design as well. By adding tag sensing to the current object location procedure, we hope to make image localization more accurate. The intelligent agent will be able to scan its environment searching for the tags placed on each test object. With a better measurement of where the object exists in 3D space, the intelligent agent will better center itself around the object and take more on-target images. This will improve the quality of our input and training data sets, and allow the classification pipeline to learn and process on more accurate representations of the objects.

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.

I plan to improve and build upon my own skills and knowledge base throughout the course of the project. While we will all be contributing to every part of the design process, I wish to become especially competent in the implementation

of our full classification pipeline. I'd like to accomplish a full hierarchy of neural networks with sequential learning implemented over the top of the classifiers.

3 TECH REVIEW

In this technology review I will be focusing on the first half of our image classification program, specifically types of image classifiers and sub classifiers, and how overfitting applies to our project.

3.1 Project Overview

- Data Collection
 - ROS for robot actuation
 - Image Tagging - Object Selection with Interface
 - Data Gathering
- Image Recognition Part 1
 - Overarching Classifiers
 - Sub Classifiers
 - Overfitting in the algorithms
- Image Recognition Part 2
 - Backpropagation Methods
 - Types of Online Learning
 - Minimizing Catastrophic Interference

3.2 Image Recognition Part 1

3.2.1 Overarching Classifiers

The most critical program in our pipeline is the classification model, as it performs the most complicated task and actually predicts the object's class. Data prediction is an increasingly large field and collaboration is common practice in the industry to help grow the public knowledge base. Scientific literature on the subject of classification is readily available, and I found an article detailing a broad overview of modern techniques. Image classification is particularly complicated and can be accomplished numerous ways.

In "*A survey of image classification methods and techniques for improving classification performance*", authors D. Lu and Q. Weng "examine current practices, problems, and prospects of image classification". The review notes that choosing an appropriate algorithm and designing a suitable processing procedure is essential to a successful image classifier, making this one of the most critical decisions for the entire pipeline. I am looking for modern approaches to our problem that take full advantage of the processing resources we have available to us, while still having the backing of the community and proper support resources available to developers. The article emphasizes nonparametric classifiers that allow for a dynamic number of model parameters, specifically neural networks, decision tree classifiers, and knowledgebased classification. Based on the analysis in the review, our group has agreed upon a convolutional neural network for image classification. In recent years the algorithm has been widely adopted for the task due to advantages such as "arbitrary decision boundary capability, easy adaptation to different types of data and input structures, ... , and generalization

for use with multiple images". Flexibility and generalization will make training the neural network possible with the number of image classes we want to include in the data set, and may come into play again when we investigate data overfitting. [1]

3.2.2 Sub Classifiers

While my research into literature regarding classifiers compared different methods and algorithms in a concise manner, our project is designed to be more specific than just using pre-existing libraries predict image classes. We are aiming to implement a classification model designed by our group, incorporating two distinct levels of classification to improve prediction accuracy. First, an overarching classifier is designed to be generalized and only recognize different object classes. Processing image data with this layer will result in a prediction as to which class the object belongs, such as stapler or book. This network is trained on all of our desired object classes, but passes its prediction to a more specialized classifier. The second layer is designed to be overfit to the data, or too accustomed to the images it has already seen. We are hoping this allows the classifier to identify individual differences and features in the objects it sees. While there is only one overarching classifier, there will be a specialized second layer network overfitted to each object class to achieve both broad and specific predictions.

As I reviewed more articles on the subject I came across "*Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis*" written by Patrice Y. Simard, Dave Steinkraus, and John C. Platt. The article specifically analyzes convolutional neural networks which are the preferred method to classify images. It presents concise advice on methodologies to adhere to for best results when implementing such a network.

According to their research, the most critical practice is to use the largest training set possible. Increasing the data set improves model accuracy, and our pipeline includes an online learning model to sequentially add new images to the data set over time. The article also notes they expanded their data set with a new form of distorted data which we have incorporated in our design as well. After our initial training set is collected we will generate an auxiliary image set with additional image noise and object variance to diversify the data set.

The second most important practice is to utilize convolutional neural networks instead of fully connected networks for image recognition. Due to the huge amount of data analysis time saved by a CNN not connecting every node in the network, "a simple do-it-yourself implementation of convolution with a flexible architecture is suitable for many visual problems". [2]

3.2.3 Overfitting in the Algorithms

Overfitting has become a recurring problem as the machine learning industry has evolved. If your neural network is over-trained on a data set it will struggle to predict features not present in the values it has already viewed. This is generally a negative result of training, but can be prevented by combining the predictions of multiple models. This isn't a perfect solution, however, as it can be difficult to implement in your algorithm without slowing down as data sets scale.

In "*Dropout: A Simple Way to Prevent Neural Networks from Overfitting*", dropout is the recommended best practice to address overfitting. By randomly dropping units and their connection from the neural network model while training, units are prevented from co-adapting over time. The dropout method "samples from an exponential number of different thinned networks." Then, these subsets' predictions are averaged with one un-thinned using smaller weights. The article presents the method as a straightforward solution to the overfitting problem, and our group will experiment with implementing the technique to differentiate our overarching classifier and specific classifiers. [3]

REFERENCES

- [1] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International Journal of Remote Sensing*, vol. 28, no. 5, pp. 823–870, 2007. [Online]. Available: <https://doi.org/10.1080/01431160600746456>
- [2] P. Y. Simard, D. Steinkraus, J. C. Platt *et al.*, "Best practices for convolutional neural networks applied to visual document analysis." vol. 3, pp. 958–962, 2003.
- [3] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.