Deep Learning for Object Classification in Retail Stores

Idawati Bustan, Mervyn Wee, Timothy Chong {idawati, wyrm, timchong}@stanford.edu

Stanford University

Introduction

Being able to recognize products from images would be useful in automating inventory management for the retail space. We took pictures of products on shelves in stores and investigated the use of a Convolved Neural Network (CNN) in accurately recognizing these products. By examining the way in which we constructed the training data, the CNN obtained good accuracy not only in recognizing products that it had seen before, but also in products that it had never seen before.

Model

The retail stores we took images from had a limited variety of products, leading to a small training set. We used transfer learning to overcome this problem.

Data

Retail stores have thousands of products. To start, we targeted classifying eight classes of products.

We constructed 3 CNN models that each received a different training set. The rationale behind this was to see how a mix of in-store and outside data could help in generalizing the model (for future products) and yet make it specific to classify current products.

All 3 models were tested on the same test set in order to compare which model performed better.

Discussion

We had expected M3 to perform the best, followed by M1 and then M2. We were worried that the lack of variety of products in retail stores meant that M1 would have low accuracy on products it has not seen before. M3 proved that adding images of products from the Internet to the training set would increase the accuracy of such predictions.

We also expected that M2, would give poor predictions on images because images store the Internet were significantly different from images from the store. M3 was a balanced model which did well on seen and unseen test images.

For future consideration, M2 performed best on products that had a specific, shape that was low in variation.

Future

Despite M3 performing the best, we found that many predictions had a probability of 20% to 40%, even if these predictions were correct. The first step we would take would be to increase the confidence in these predictions so that the model would be more well trained. This could be done by training it on more data or increasing the epochs when training the CNN.

Reference
