CS5242 Project Proposal

Group 3: Akankshita, Spatika, Ganeshkumar, Trinh

March 2019

1 Definition

1.1 Objective

Utilise deep learning and neural networks to conduct multi-class classification on images of trash.

1.2 Problem Definition and Literature Survey

Classification of trash items using deep learning and neural networks is an efficient way to improve the recycling process. An experiment by Thung and Yang [4] suggested that a simple model such as support vector machines (SVM) with scale-invariant feature transform (SIFT) features outperforms a convolutional neural network (CNN) in this case (63% vs 22% of test accuracy).

However, we believe that with the use of existing deep convolutional neural network architectures and other performance optimizations, the result of CNN classification can be improved. Work by Rad et al. in [1] on the classification of urban litter using OverFeat-GoogLeNet achieved recalls of 61.02% and 60% for the cigarette butts and leaves classes respectively, which comprised most of their dataset. Sakr et al. [3] reported a test accuracy of 83% with AlexNet architecture of CNN. They created a balanced dataset of paper, plastic and metal images, and used label-preserving transformations to augment their dataset of 2000 images to 6000. They implemented their CNN using Caffe framework and trained it with NVIDIA DIGITS on NVIDIA 740 M GPU.

The objective of our project is to utilise deep learning and neural networks to conduct multi-class classification on images of trash on the dataset created by [4].

1.3 Model

The original 72x72 images from [4] are resized and the model is trained on 150x150 pixel RGB images labelled either as trash or recyclable (includes cardboard, glass, metal, paper, plastic). Output for unseen images in the testing phase is the label - trash or the category of recyclable.

1.4 Methodology

Using the "one-vs-all" [2] approach to multi-class classification, we have N different output nodes (which are analogous to N binary classifiers), where each node outputs the result for a class. The idea is represented in Figure 1, where the classes are [Apples, Bear, Candy, Dog, Egg] and the last layer is a sigmoid function.

Given n classes, the last layer outputs the correct classification for a given image x:

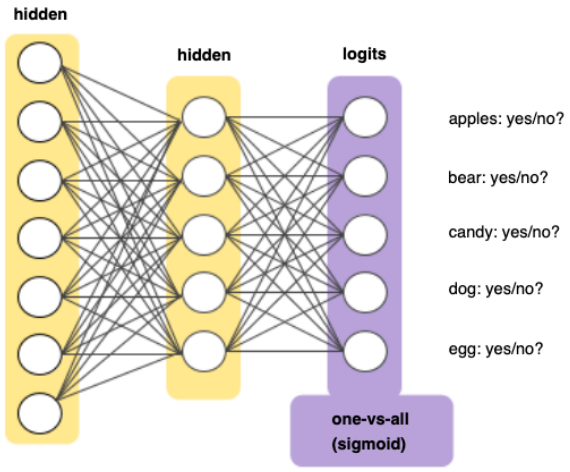

Figure 1: Sample CNN for mutliclass classification

$$f(x) = \arg\max_i f_i(x)$$

where f_i is the i^{th} classifier (in the last layer of the network), where $1 \leq i \leq N$.

2 Motivation

Most existing recycling efforts for trash require manual labour - sorting through piles of garbage to decide what article can be reused or recycled. This work is often tedious and hazardous. By taking images of the trash instead, we attempt to develop a model that can automatically detect the object in question. The end goal is to integrate this model into larger systems (automated recycling schemes by government) and make identifying different types of trash useful. This should help in cost-saving by lowering the amount of manual labour required in recycling efforts.

We make use of deep learning, specifically convolutional neural networks (CNNs). CNNs explicitly assume that the input data are images. Thus, the model can make use of different types of layers (from existing architectures or customised ones) to extract useful information from images.

3 Idea

The dataset [4] consists of labelled images, with about 400 images per category of recyclable material and about 100 images of trash. We plan to augment the original dataset with various image transformations (flipping/cropping/rotation etc.) to produce a more balanced dataset.

We also plan to try multiple approaches to improve the classification results. The following are a few directions that we're considering:

- Hyperparameter tuning through Bayesian Optimization and Grid Search
- Optimizers for parameter updating during training (e.g. SGD, Adam, RMSprop, Adagrad)
- Existing Deep Convolutional Neural Network Architectures and Transfer Learning (ResNet, VGGNet, InceptionNet)

- Use of ensemble models.

Our experiment environment is:

- Keras with TensorFlow backend
- GPU on GCP/NSCC

3.0.1 Initial CNN architecture

We start with a simple CNN architecture to see some initial results that we can further improve upon.

1. Input Layer [*height* x *width* x *depth*] = $[150, 150, 3]$
2. Convolutional layer (filter of 6X6, stride of 1, N filters, ReLu activation)
3. Max Pooling (kernel 2X2, stride 2)
4. K Dense Layers with 256 output nodes, ReLu activation
5. Dropout Layer with rate 0.4
6. Fully Connected Output layer - This layer takes the input volume and outputs an N dimensional vector where N is the number of output classes. This will have softmax activation.

4 Schedule

- Week 8 - Proposal submission, Decide candidate architectures + training system and hardware
- Week 9 - Data Augmentation, Train-Test-Validation Split
- Week 10 - Training, Compare candidate architectures
- Week 11 - Hyperparameter optimization
- Week 12 - Further performance tuning
- Week 13 - Presentation

References

[1] Mohammad Saeed Rad, Andreas von Kaenel, Andre Droux, Francois Tieche, Nabil Ouerhani, Hazım Kemal Ekenel, and Jean-Philippe Thiran. A computer vision system to localize and classify wastes on the streets. In *International Conference on Computer Vision Systems*, pages 195–204. Springer, 2017.

[2] Ryan Rifkin. Multiclass classification. *Lecture Notes, Spring08. MIT, USA*, page 59, 2008.

[3] George E Sakr, Maria Mokbel, Ahmad Darwich, Mia Nasr Khneisser, and Ali Hadi. Comparing deep learning and support vector machines for autonomous waste sorting. In *2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*, pages 207–212. IEEE, 2016.

[4] G Thung. Trashnet. *GitHub repository*, 2016.