**Design Proposal for the Garbage Classification System**

1. **Team Members**

Guo, Yuhua

Jiang, Tianhan

Laditan, Oluwapelumi David

Lawal, Tobi

Zhao, Peiyun

1. **System Specifications**
   1. Approaches

Garbage recognition is a supervised learning task. We will use deep learning libraries to perform a multiclass classification job on an image dataset. We plan to label images to build a single-labelled training. Although the accuracy may improve if we train our model on a multi-labelled training set, we may face much more uncertainties and difficulties if we do so.

* 1. Models to be used

We plan to train and fine-tune multiple convolutional neural network (CNN)-based networks, including pre-trained on the ImageNet database and trained from scratch. Some articles[1] suggested that a pre-trained CNN model may perform better, but for our case, it may not apply well because of the difference in labelling rules.

Since the size of the dataset we will use is relatively small, we plan to use data augmentation[2] and explicit regularization, including dropout and weight decay, to avoid overfitting.

* 1. Metrics to be computed

Since this is a classification job, the most critical metric to tune hyperparameters and to represent training/validation score is accuracy. We will still record other classification metrics such as accuracy, precision, recall, F1-score, ROC, and AUC.

We will use Categorical Cross Entropy (CCE) as a metric in the loss function.

* 1. Problems expected to face

**About image**

* Cell-phone photos have different quality depending on the device used to take the picture.
* Images with different resolutions.
* Model generalization to photos taken using different cell phones will be hard.

**About subjectivity in labelling dataset**

* What degree of contamination is acceptable for recyclable and for composting.

**About limitations of the training set**

* Misclassification of hazardous waste into black bin, due to lack of hazardous waste class.
* Too much variability within classes.
* Misclassifications caused by differences in packaging, shape, or graphic design.

1. **Dataset Creation**
   1. What kind of data will you use to solve the problem?

A dataset including three classes: recyclables (blue bin), composting (green bin), landfill garbage (black bin). We will have as many as needed labels for subclasses in each class, such as glass bottle, plastic bottle, tin container, Tetra Pak container, plastic jug subclasses in the recyclable class.

* 1. How the data should be collected
* The object should be centralized in the photo
* Homogeneous background: we will use a white background
* One object per photo

**Reference**

[1] How to build an image classifier for waste sorting. <https://towardsdatascience.com/how-to-build-an-image-classifier-for-waste-sorting-6d11d3c9c478>

[2] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0