Data Selection Proposal

Project Deliverable 1

1. Dataset

Garbage classification dataset (https://www.kaggle.com/asdasdasasdas/garbage-classification) will be used from Kaggle for this project. The dataset has 6 classifications (cardboard, glass, metal, paper, plastic, and trash) and each category has good amount of datasets. If needed, more data could easily be added, and more categories could be added as well.

2. Methodology

With the dataset of 6 classifications, the model needs to be trained to recognize them using labeled example photos.

i. Data Processing

From the images in each category, pixel data, color histogram, textures, shapes are important data. Preprocessing will be done with CNN (Convolutional Neural Network), where it will take image's raw pixel data as input and learns how to extract the important features/data and infer the classification from it. Also, preprocessing of images can be done so that images have uniform aspect ratio, scale, and etc.

ii. Machine learning model

The estimate from this dataset will be able to predict what kind of garbage/recycling it is. As mentioned above, CNN will be used, which takes 3 operations, convolution, ReLu(Rectified Linear Unit), Pooling. CNN could work on many datasets and categories, more efficient on memory and complexity with weight sharing, and are good feature extractor. However, overfitting could be a problem. SVM(Support Vector Machine) could be an alternative, but time complexity may be a problem.

iii. Final conceptualization

To showcase the project, it will be deployed on a simple web application, where user can upload or show image in real time using webcam to classify the object into one of the recycling categories.