models and ThanosNet which utilizes meta-labels to make classification decisions.

## II. Related Literature and Motivation

Yang and Thung [7] curated the Trashnet dataset in 2016. This dataset contains approximately 2500 images of trash across six classes (cardboard, glass, metal, paper, plastic, and trash). Each class contained approximately 400-500 images that were taken against a monochromatic background. To introduce variance in the dataset, the lighting and pose between images were modified. Data augmentation techniques including random translations, rescaling, shearing, and rotation were applied to further increase the variance of the dataset. The researchers proposed two novel methods for classifying trash: support vector machines and convolutional neural networks. These methods achieved a test accuracy of 63% using a 70/30 random training/testing split.

The small size of Trashnet motivated Knowles et al. [10] to utilize transfer learning techniques with deep CNN models pre-trained on the ImageNet dataset [8]. Transfer learning for image classification uses a pre-trained model as a feature extractor to extract lower-level features such as edges and lines. Trainable fully-connected layers are then added to classify these features. This enables researchers to train large CNN models with millions of parameters using a small dataset like Trashnet. Knowles et al. utilized the pre-trained weights of the VGG-19 [11] network. In addition to the images in Trashnet, Knowles et al. created a non-waste object class by taking images from the Flowers dataset by the Visual Geometry Group and PASCAL VOC 2012 [12].

Aral et al. [13] further experimented with the efficacy of various transfer learning architectures with established CNN-based models such as DenseNet [14], Inception-ResNet-V2 [15], and Xception [16]. Based on their experimental results, DenseNet121 and DenseNet169 performed the best, while Inception-V4 was a close second.

Vo et al. [17] continued the trend of transfer learning-based architectures with their DNN-TC model. DNN-TC utilizes ResNext-101 [18] as a feature extractor with the addition of two fully-connected layers following the global average pooling layer. The team also produced their own VN-trash dataset, which consists of images found online and taken in the surrounding environment. It covers the classes of medical waste, organic, and inorganic wastes.

A publication most similar to this one is White et al. [19] WasteNet uses DenseNet architecture with fully-connected layers added on top. A hybrid tuning method was used by first pre-training the classifier layers. Once the performance of these layers began to converge, the remaining layers were unfrozen and a smaller learning rate was applied to calibrate these lower-level feature extractors. The team chose a 50:25:25 split of training, validation, and testing, respectively, using images from the Trashnet dataset. They used a combination of random translation, zooming, shearing, and rotation to augment the images. After training over more than 1000 epochs, WasteNet achieved state-of-the-art results on TrashNet.

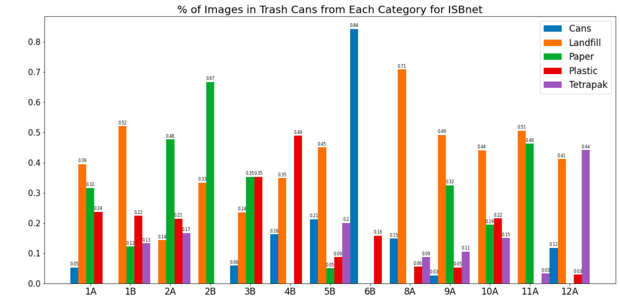


Fig. 1. Distribution of images belonging to each trash class within the individual trashcans. Values are denoted in percentages calculated as the percentage of total photos in the trash can represented by the class.

Previous waste classification systems that incorporate CNNs [13], [7], [17], [19], [20], [10] rely on a purely image-based approach. However, in a real environment, classifiers that sort waste into high-level categories, such as plastic, cans, paper, and landfill, are subject to many diverse features from the variety of objects present. As a result, pure image-based classifiers are vulnerable to low generalizability and feature confusion, explained in a later section. Moreover, purely image-based approaches assume that distributions between classes and objects of that class are uniform across all locations and time. However, intuitively, these factors have a large impact on said distributions. For example, during meal times, we would expect an influx of trash belonging to the landfill class (food scraps, wrappers, plastic utensils). Alternatively, for a trash can next to a printer, we would expect a large amount of recyclable paper being disposed of.

In a response to these limitations, our model utilizes metadata that is associated with the physical trash can including, but not limited to, location, time of day and weight of the trash. Metadata provides context for an image and information that reflects the likely distributions of trash within a trash can at given points in time. As illustrated in Figure 1, there is a distinct difference in trash distribution between the various trash can locations. Therefore, a model that synergizes metadata and image-based features may exploit the information present in inter-trash can variance to enhance classification capabilities.

We envision the approach described in this study to be deployed in a modified consumer trash can. In turn, the smart trash can could effectively sort trash into 5 high-level categories: cans, landfill, paper, plastic, and tetra pak. By using metadata, our model is not as dependent on image features as current state-of-the-art waste classification models, improving precision for object-dense classes and overall accuracy post-implementation.

## III. Exploratory Data Analysis

ISBNet is hand collected by our group at the International School of Beijing. The trash in these images was gathered from trash cans around the school. ISBNet totals 889 images distributed across 5 classes: cans, landfill, paper, plastic, and tetra pak. The distribution of these classes is shown in