ThanosNet: Attention-Based Trash Classification Using Meta-labels*

Alan Sun
University of Maryland, College Park
asun17904@gmail.com

Harry Xiao
Columbia University
hx2310@columbia.edu

***Abstract*— In modern society, waste recycling and classification has become a necessity for reducing resource consumption and economic loss. For an effective waste classification system to be feasible, such as in an intelligent sorting trash can, a robust model must be available. In this study, we propose a novel attention-based trash classification model utilizing deep neural networks and meta-labels. We utilized time of day, location of the trash can, and distance to landmarks as the meta-labels. We collected the ISBNet dataset which contains 889 images and their associated meta-labels, distributed over 5 classes (paper, plastic, cans, tetra pak, and landfill). Afterwards, we developed two different architectures for the attention-based trash classification, BilinearThanosNet and AdditiveThanosNet, both of which used ResNet50 as the feature extractor. By comparing ThanosNet with state-of-the-art image classification models (VGG16, ResNet50, DenseNet169) on ISBNet, we found that ThanosNet displayed the best performance, with an F_1 score of 0.952. This shows the effectiveness of using meta-labels and our ThanosNet model architecture.**

***Keywords*—Trash classification, computer vision, deep learning, waste, recycling**

I. INTRODUCTION

Society today has an ever-increasing awareness towards the importance of classifying trash properly for the purpose of recycling. Waste that is not properly sorted could pose danger to soil, air, and water sources, while effective waste management reduces the pressure on landfills and can create beneficial economic and financial effects. Reducing total waste is crucial in conserving resources and becoming more sustainable, something that is enabled through more extensive recycling. However, a key issue that stands in the way of widespread recycling lies in the accurate classification of recyclable and non-recyclable trash, particularly for consumers. At a consumer level, non-recyclable papers and plastics are often mistakenly placed into the recycling bins, contaminating and thus disqualifying the entire batch from being recycled. Utilizing machine learning to improve classification seemed promising and could vastly improve these inaccuracies present in consumer classification. The development of an automated trash can that could assist in the sorting of trash is a potentially impactful use case.

Utilizing deep learning to classify trash has been proposed numerous times. Salimi et al. created a trash-bin robot that is capable of detecting trash and classifying it. Similarly, Auto-Trash, a trash can that can automatically sort waste into compost and other made its debut at the 2016 TechCrunch Disrupt Hackathon in New York [1].

*This research is affiliated with the International School of Beijing (ISB)

In the same year, Yang and Thung [1] released Trashnet, a dataset that is now used as a benchmark for measuring waste classification performance. Currently, most of the state-of-the-art models on Trashnet use transfer learning to finetune well-known CNN-based models developed for the ImageNet challenge, which includes over 14 million images belonging to 20,000 categories. However, these models, which rely solely on image features to classify waste, are not effective for discriminating between objects with similar features but belonging to different classes.

The rising pressure for effective waste classification has already been seen on the local level. In 2019, the Beijing Municipal government implemented mandatory waste management regulations: households and institutions must sort their waste into recyclables, food waste, other, and hazardous material. "Individuals who fail to follow the regulations repeatedly will be fined a maximum of [~30 dollars]." In accordance with these new regulations, the International School of Beijing (ISB) installed a new trash sorting system that replaced vague landfill and recycle class trash cans with multiple bins (landfill, paper, plastic, cans, tetra pak). Almost a year later, waste audits reveal (Table I illustrates the most recent one of these waste audits) that students are not responding to numerous education initiatives employed by the school's environmental organization. This again stresses the need for a more reliable method of classifying waste at the consumer level.

Class	Percentage Correct
Plastics	28.10
Cans	89.13
Paper	46.67
Other	33.09

TABLE I
JANUARY 2020 WASTE AUDIT RESULTS

In light of these issues and observation, this study develops a deep convolutional network model, ThanosNet, for trash classification which incorporates metadata to improve existing trash sorting systems.

The contributions of this study are as follows.

1) Curated ISBNet dataset that includes 889 images belonging to 5 classes. Each picture contains metadata identifying the location of the trashcan, activity of the trash can with respect to the time of day, and the time which the object was thrown into the trash can.
2) The development of our network ThanosNet.
3) A proof-of-concept that incorporating metadata into the trash classification model can improve precision.

Experiments were conducted to demonstrate the performance differences between current state-of-the-art models and ThanosNet which utilizes meta-labels to make classification decisions.

II. RELATED LITERATURE

Yang and Thung [1] 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. These images 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. Researchers proposed two novel methods for classifying trash: support vector machines and convolutional networks. These methods achieved a test accuracy of 63% using a 70/30 random training/testing split. A more in-depth analysis of the Trashnet dataset will be conducted in later sections of the paper.

The small size of TrashNet motivated Knowles et al. [2] to utilize transfer learning techniques with deep CNN models pre-trained on the ImageNet dataset, which contains over 1000 classes with 14 million images. Transfer learning for image classification uses a pre-trained model as a feature extractor by extracting lower level features such as edges and lines, then adding trainable fully-connected layers 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 network. In addition to the images in TrashNet, Knowles et al. expanded the dataset by creating a non-waste object class from PASCAL VOC 2012 and the Flowers dataset by the Visual Geometry Group at University of Oxford.

Aral et al. [2] further experimented with the efficacy of various transfer learning architectures with established CNN-based models such as DenseNet, Inception-ResNet-V2, and Xception. Training, testing, and validation were done in a ratio of 70:17:13, respectively, using Trashnet. Based on their experimental results, DenseNet121 and DenseNet169 performed the best, achieving 95% test accuracy, while Inception-V4 was a close second with 94%. The team also experimented with Adam and Adadelta optimizers, and found that the Adam optimizer resulted in higher test accuracies. Therefore we reimplement and perform this model with the above settings as a baseline model to demonstrate the effectiveness of our model, ThanosNet, against current state-of-the-art models.

Vo et al. [12] continued the trend of transfer learning-based architectures with their DNN-TC model. DNN-TC utilizes ResNext-101 as a base model 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. This architecture achieved an accuracy of 94% when using a 60:20:20 split for training, validation, and testing, respectively, on Trashnet and VN-trash datasets.

A publication most similar to this one is White et al. They utilized transfer learning to construct WasteNet. 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 begin to converge, the remaining layers are unfrozen and a smaller learning rate is applied to calibrate these lower-lower feature extractors. The team chose a 50:25:25 split of training, validation, and testing, respectively, images from the Trashnet dataset. They used a combination of random translation, zooming, shearing, and rotation to augment their images. After training over more than 1000 epochs, WasteNet achieved a F_1 score of 0.970. These experiments showed again the potential that DenseNet has as a feature extractor for transfer learning, and further encouraged us to evaluate our model against DenseNet.

Previous waste classification systems that incorporate CNNs rely on a purely image-based approach. The approach described in this study is designed to be deployed in a modified consumer trash. This allows our model to use metadata that is associated with the physical trash can including, but not limited to, location, time of day and weight of trash. This first pass sorts trash into 5 high level categories: plastic, paper, landfill, tetra pak, and cans which then allows for further, more specific, differentiation at large industrial plants.

III. EXPLORATORY DATA ANALYSIS

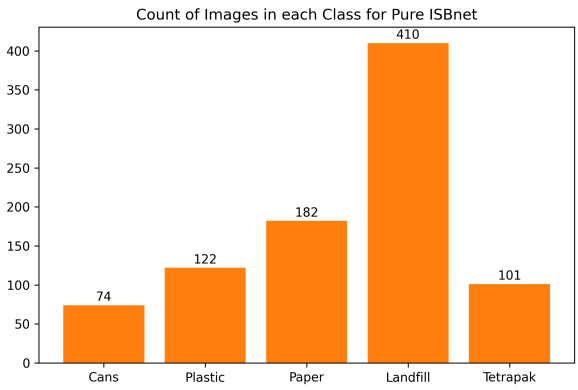

Fig. 1. Class Distribution of ISBNet

ISBNet is hand collected by our research group in the International School of Beijing. These images were gathered from trash cans around the school. ISBNet totals 889 images distributed across 5 classes: paper, plastic, cans, tetra pak, and landfill. The distribution of these classes is shown in Figure 1. The data acquisition process involved using a piece of black poster paper as a background; this would create enough contrast for trash belonging to the paper category. These pictures were taken with an iPhone 8 and an iPhone XS. We recorded the trash can in which the piece of trash

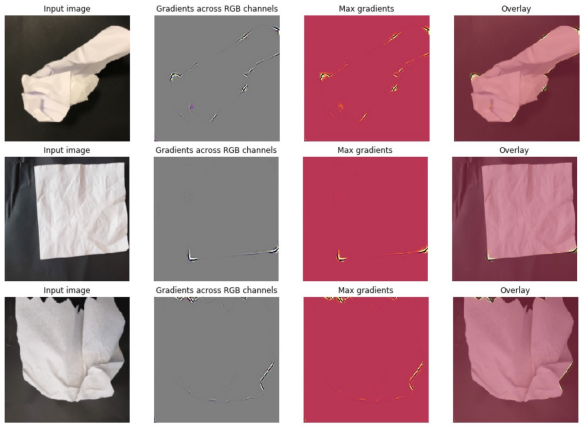

Fig. 2. Saliency maps from a baseline ResNet50 classifier. The images from top to bottom: crumpled piece of paper (paper), a napkin (landfill), cumpled piece of tissue (landfill).

originated from, and any trash-generating landmarks near this trash can. The sample time – the time which the trash was thrown into the trash can – was also recorded. Section IV details the encoding and format of these meta labels. Data augmentation techniques were performed on the images due to the limited size of each class. This included grey-scaling, random rotation, re-scaling, and shearing. Mean subtraction and normalization was also performed on the dataset.

IV. METADATA

Metadata of all kinds can be collected through sensors in a smart trash can where our model would reside, such as location of the trash can and its distance to landmarks, or time of day. The geographical location of a trash can allows us to identify its proximity to certain landmarks. We define a landmark, in the context of trash classification, as an identifiable area that may skew the distribution of either the type of trash or amount of trash found in a trash can in proximity of this landmark. Landmarks may affect trash generation either through the inherent nature of these landmarks, or through the increased foot traffic experienced by these areas. Examples of landmarks that we identified are cafeterias (eating may produce more contaminated and food-related trash), printers (recyclable paper would be more common next to a printer), or entrances/exits (the large flow of people means more trash is likely to be deposited in the nearby bins). Time of day refers to the time when a piece of trash was thrown into the bin. The time of day may also skew the probability that particular categories of trash are thrown away, such as lunch time, which would expect more landfill trash.

Incorporating metadata as extra inputs to an image-based neural network decreases the likelihood for *feature confusion*. Items of trash belonging to different categories may have similar features. A network that solely depends on image features is often not able to differentiate between these objects. This is exemplified in Figure 2.

Saliency maps of an image-based trash classifier were generated for images of paper and tissues/napkins, one of the classes with lowest precision for image-based classifiers. This trained classifier shown in Figure 2 incorrectly falsely predicted both landfill pictures, the napkin and tissue, as belonging to the paper class. The saliency maps illustrate that the image-based model associates rigid edges and crumples to the paper class. However, the napkin and the crumpled tissue contain similar features. As a result, the network falsely associated those with the paper class. Exposing the network to additional time and location information will increase its ability to discern images of similar features, decreasing the likelihood for *feature confusion*.

A. Location and Distance

Thanks to the ISB administration, we acquired detailed, scaled blueprints, with trash can locations marked, of the two floors where ISBNet was collected. These maps are shown in *Figure 3*. Two trash cans that we collected trash from were not marked on the map. Thus, we marked their locations by hand, namely, 11A and 12A which were located near the Sodexo offices and the library, respectively.

Since the scale of the map was not available, the 25 meter pool located on the first floor was used as a reference point of absolute distance to deduce the scale in pixels/meter.

Using these blueprints, we identified 11 types of trash-generating landmarks around the school. These landmarks are listed below.

- Entrance/Exit
- Bathroom
- Lounge
- Stairwell
- Cafeteria
- Theater
- Gym
- Pool
- Printer
- Couch Area
- Library

A trash can's *proximal landmarks* are defined as the landmarks in the set of all landmarks listed above that are within a 70 meters radius of the trash can. Each trash can was annotated with its respective proximal landmarks. The landmarks were also expressed as a one-hot encoding vector for each trash can.

A second vector describes the distances between between the bin and each proximal landmark. The distance was defined as the diagonal distance from the center of the bin to the center of the landmark. A measurement is shown in *Figure 4*.

We converted these two encoded-vectors into inputs for the network by performing the following process:

Let $\overrightarrow{v_{\mathrm{d}}} \in \mathbb{R}^{11}$ and $\overrightarrow{v_{\mathrm{oh}}} \in \{0,1\}^{11}$ be the distance between the landmark and the trash can and the one-hot encoded vector of proximal landmarks, respectively.

$$\overrightarrow{v_{\mathrm{c}}} = \overrightarrow{v_{\mathrm{d}}} \odot \overrightarrow{v_{\mathrm{oh}}}$$

The zero entries of the unit vector of $\overrightarrow{v_c}$, $\hat{\boldsymbol{v}}_\mathbf{c}$ are replaced with a constant β. A component-wise negative logarithmic transformation was applied to the unit vector $\overrightarrow{v_c}$. This transformation is reflective of the behavior of consumers: consumers are most likely to throw away trash in the trash can closest to the trash-generating landmark. The convergence of this logarithmic transformation is shown below.

$$\lim_{\hat{\boldsymbol{v}}_{\mathbf{c}i} \to 0} -\log \hat{\boldsymbol{v}}_\mathbf{c} = \infty$$
$$\lim_{\hat{\boldsymbol{v}}_{\mathbf{c}i} \to \beta} -\log \hat{\boldsymbol{v}}_\mathbf{c} = -\infty$$

The relative influence of a landmark decays exponentially as the distance between the landmark and the trash can increases linearly. Non-proximal landmarks that have an influence of β approximate $-\infty$. These negative values help the network to further discriminate between confusing features.

B. Time

The 8am - 6pm school day was segmented into one hour segments. The time when the trash was thrown into the trash can is represented as a unit vector $\hat{\boldsymbol{v}}_\mathrm{t} \in \mathbb{R}^{11}$, with each column denoting the likelihood that that specific piece of trash was discarded during that one-hour interval.

We were unable to capture the exact time of day when each individual piece of trash was disposed of. Thus, for each trashcan that we collected from, we identified multiple time periods throughout the day where these trash cans would be most active and exposed to the highest level of foot traffic. Each landmark is categorized into one of three categories: multi-modal, normal, and uniform. Landmarks characterized as multi-modal demonstrate increased foot traffic during regular, scheduled times of day. For example, the cafeteria is only accessed during lunch periods and after school, the lounge is accessed by teachers primarily during lunch, and entrances/exits are accessed by all personnel at the start and end of the school day, 8am and 3pm respectively. Landmarks characterized as normal experience activity clustered around a central mean, adhering to a normal distribution. Landmarks that follow such a distribution include the theater which, on regular days, only experiences traffic after lunch time for assemblies. Landmarks that are described by uniform distributions are accessed throughout the day with similar levels of activity, such as bathrooms and printers.

In order to create the probability distributions for each landmark, the time intervals between 8am and 6pm were converted into values between 0 and 1. Namely, 8am was considered 0, 1pm (13:00) was considered 0.5, and 6pm (18:00) was considered 1.0. Multi-modal probability distributions were constructed through the addition of normal distributions. Normal distributions were created with means centered on the high activity times. Standard deviations were determined as the approximate time before the mean that foot traffic begins a significant uptick, or the time after the mean required for foot traffic to significantly decrease. An example is the cafeteria multi-modal distribution. We estimated that the times of highest activity were 12pm (around lunch time, represented as 0.4) and 4pm (after school, represented as 0.8), with a standard deviation of 1 hour (0.1) and 30 minutes (0.05), respectively. This was used to create two normal distributions, which were combined through addition and reduced accordingly such that the bi-modal distribution had a total area of 1. The probability distributions for landmarks that have normally-distributed foot traffic were created using the same procedure outlined above. Uniform probability distributions have a value of 1 through all time intervals.

Each bin was assigned a convolutional probability distribution, which is the average of all waste-generating probability distributions from the bin's proximal landmarks. Consider the example below for the bin b_m which has n proximal landmarks. Its corresponding probability density function (PDF) is defined as $f_\mathrm{m}(x)$. b_m's probability density function with respect to time can be described using the following:

$$f_\mathrm{m}(x) = \sum_i^n \frac{1}{n} f_\mathrm{i}(x)$$

where f_i is the PDF for landmark i. The time vector was determined by evaluating the area under the PDF for each time interval is calculated and recorded into the corresponding column e.g. area between 0 and 0.1 of the PDF would be the first column. Thus, the final time vector is of length 10.

V. EXPERIMENTS

This section is organized into three subsections. The first section, *Weighted Loss Function*, the loss function we employed is explained and justified. In the second section *Baseline Experiments*, transfer learning methods are used with pre-trained ImageNet models to establish a baseline score for comparison against ThanosNet. Later, *Metadata Experiments* goes into detail regarding the architecture and performance of our ThanosNet model.

For all our experiments, we utilized stratified cross-validation training with a fold count of 5, each fold being trained over 50 epochs. The performance of a model was evaluated through average validation macro F_1, and the average validation loss and average validation accuracy in the same epoch as the macro F_1. As the dataset was highly imbalanced, we used macro F_1 as the validation metric to gauge the precision and recall of our model, which is defined by the following,

$$F_1 = 2 \cdot \frac{pr \cdot re}{pr + re}$$

where pr and re represent precision and recall, respectively.

VI. WEIGHTED LOSS FUNCTION

For this multi-class classification model, let X denote the input image and metadata. The scalar $y_\text{label} \in \{0, 1, 2, 3, 4\}$ denote the label representing the class that input X belongs to: cans, landfill, paper, plastic, and tetra pak respectively. $\overrightarrow{y_\text{predict}} \in \mathbb{R}^5$ represent the models predicted class of the input X. Component i of the prediction vector $\overrightarrow{y_\text{predict}}$ is indexed by $\overrightarrow{y_\text{predict}}[i]$. The standard cross-entropy loss function is defined as by the following.

$$L(\overrightarrow{y_{\text{predict}}}, y_{\text{label}}) = -\overrightarrow{y_{\text{predict}}}[y_{\text{label}}] + log\left(\sum_j e^{\overrightarrow{y_{\text{predict}}}[j]}\right)$$

The trained model defined by the loss function above shows poor performance as it is demonstrates a poor recall rate. The reason stems from the fact that ISBNet is imbalanced. Therefore, a weighted loss function was used in place of the standard cross-entropy loss, where the weights of a class are inversely proportional its class size. The weight of class i is defined by

$$\omega_i = \max \frac{\sum_j \|j\|}{\|i\|}$$

where $\|j\|$ is the size of class j and $\|i\|$ is the size of class i.

Thus, the weighted cross-entropy loss function is defined by the following. In our experiments, we discovered that using such a weighted loss function resulted in better performance.

$$L(\overrightarrow{y_{\text{predict}}}, y_{\text{label}}) = \omega_{y_{\text{label}}}\left(log\left(\sum_j e^{\overrightarrow{y_{\text{predict}}}[j]}\right) - \overrightarrow{y_{\text{predict}}}[y_{\text{label}}]\right)$$

A. Baseline Experiments

We experimented with VGG16, ResNet50, and DenseNet169 as feature extractors for our image-based, baseline models. These are networks that have performed well in prior literature and thus are a valuable benchmark for comparison. The pre-trained ImageNet model had its respective classification layers removed and replaced with three fully connected layers with ReLU activation functions in between. During training, regularization techniques including batch normalization, dropout, and l_2 normalization were applied. We trained with a batch size of 32. The Adam optimizer was used with a learning rate of 10^{-5} and weight decay of 10^{-10}.

All image inputs were first resized to 256 $\times$ 256, then center-cropped to 224 $\times$ 224 to match the ImageNet feature extractor input parameters. These images were then normalized based on the mean and standard deviation in the ImageNet training set: $\mu_r = 0.485, \mu_g = 0.456, \mu_b = 0.406$ and $\sigma_r^2 = 0.229, \sigma_g^2 = 0.224, \sigma_b^2 = 0.225$. A random-resized crop with the same image sizes was experimented with during training instead of the center crop. This resulted in an extremely volatile training F_1 score, preventing the model from converging. This suggests that object localization plays an important factor in classification as there was a large variance in trash size. The random resized crop could be leaving out important features.

B. Metadata Experiments

To incorporate metadata into the network, two attachments to the existing baseline networks were proposed: a bilinear attachment, and an additive attachment. The architecture of these two variants of ThanosNet, BilinearThanosNet and AdditiveThanosNet respectively, are shown below in Figure 3 and Figure 4.

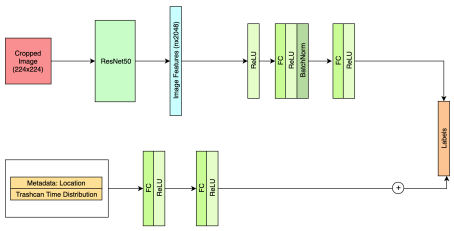

Fig. 3. ThanosNet with additive metadata attachment

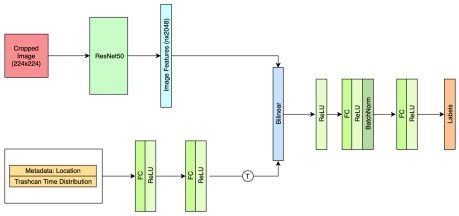

Fig. 4. Thanosnet with bilinear metadata attachement

Each of these two variants seek to replicate an intuition behind small-scale trash classification. The first, ThanosNet with an additive metadata attachment, is defined by the observation that for a given trash can, there exists a bias between the distributions of its categories, the location of the trash can, and the time from which the distribution was sampled from. Thus, an additive attachment uses the metadata to determine the magnitude of the bias for each category, during the final stage of the prediction. The second, ThanosNet with a bilinear attachment, captures the intuition that for a given image, the weight of its features, extracted from ImageNet models, is correlated with the meta labels collected from the trashcan including location and time. Therefore, in this bilinear model the transpose of the features extracted from the metadata were multiplied with the image features. Principal component analysis was then performed on this bilinear matrix. Moreover, this bilinear model uses the meta-labels to create an attention mask, forgetting relatively unimportant, ambiguous features and retaining features that contain more distinguished information. This bilinear layer is defined by the following

$$y = x_1^T A x_2 + b$$

where x_1 are features extracted from the metadata, x_2 are image-based features. A is a parameter matrix, while b is the bias of the layer.

For both networks, the metadata input was created by concatenating the location meta-label and time meta-label.

C. Results

As seen in Table II, ResNet50 achieved the best results out of the three baseline ImageNet models with a 0.9240 validation macro F_1 score, while DenseNet169 and VGG16 achieved 0.8750 and 0.9198, respectively. Therefore, we used

Model	Loss	Macro F_1	Accuracy
VGG16	0.406	0.875	0.875
ResNet50	0.273	**0.924**	0.920
DenseNet169	0.287	0.920	0.918
AdditiveThanosNet	0.290	0.907	0.906
BilinearThanosNet	0.161	**0.952**	0.947

TABLE II

BASELINE MODELS AND THANOSNET MODELS RESULTS

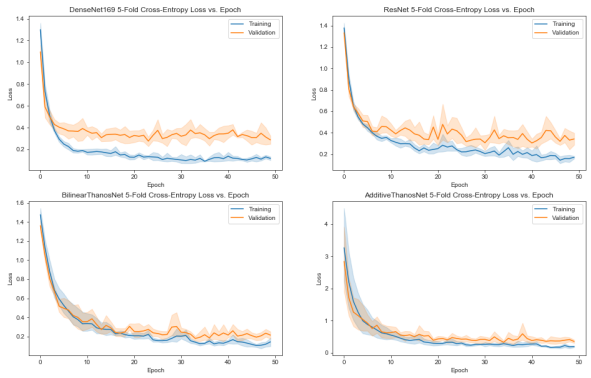

Fig. 5. The loss in the training and validation process for DenseNet169, ResNet50, BilinearThanosNet, and AdditiveThanosNet respectively.

ResNet50 as the feature extractor for ThanosNet. However, it should be noted that the VGG16 and DenseNet169 networks were able to converge at a faster rate, as shown in 5.

Of the two ThanosNet variations, the bilinear variant performed the best, achieving an F_1 of 0.952 compared to the 0.922 from the additive model. We observe that overall, BilinearThanosNet had the best performance in terms of loss, macro F_1 score, and accuracy, while AdditiveThanosNet was comparable with ResNet50 and DenseNet169. We believe that the performance difference between these two variants of ThanosNet is attributed to the relatively subtle variance between the contents of trash cans. There is not enough skew within the distributions of each trash can for the additive bias to significantly impact the predictions of the network positively. Moreover, we believe that the improvements of BilinearThanosNet come from the fact that creating an attention-like mask at the level of the feature extractor gives the network more opportunity to correct the possible lack of discrepancy between estimated time distributions and real time distributions using the classifying fully-connected layers.

As shown in Figures 5 and 6, BilinearThanosNet shows good stability, unlike ResNet50, the second highest-performing model, which struggled with volatility. We believe that this is induced by the noise of the dataset as well as the image augmentations that were applied to the training set. For categories including landfill and paper, where there is an extremely diverse set of objects and features present, the image-based networks struggle to generalize these features sufficiently, especially with the limiting size of the dHowever, since BilinearThanosNet utilizes metadata it is not nearly as dependent on these relatively noisier image

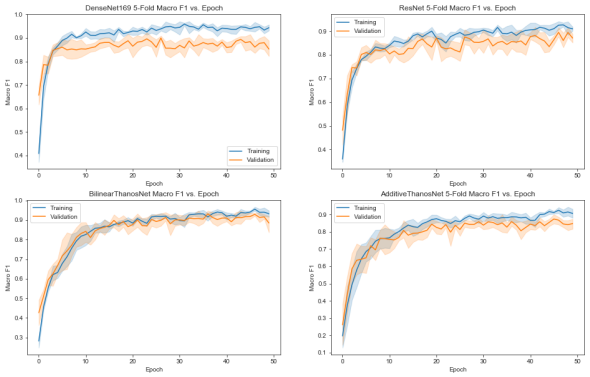

Fig. 6. The macro F_1 in the training and validation process for DenseNet169, ResNet50, BilinearThanosNet, and AdditiveThanosNet respectively.

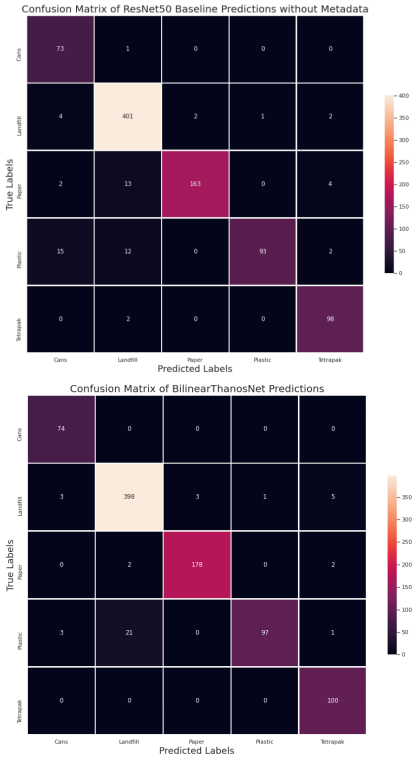

Fig. 7. Confusion matrix of ResNet50 model without incorporating metadata and BilinearThanosNet respectively, generated by combining validation predictions from each model in each fold.

features. Therefore, the usage of metadata not only improves performance, but also stabilizes training by mitigating object-dense classes. The confusion matrix of the experimental models are displayed in Figure 7. From this, we can see that BilinearThanosNet outperforms the best image-based model, ResNet50, in almost all trash categories. In particular, BilinearThanosNet demonstrates a significant improvement in paper classification over ResNet50 with 0.983 precision and 0.978 recall rate. Since the paper class contains a diverse set of features which varies between what is printed on the paper and possesses similar features with other classes,

metadata played an important role in providing additional contextual information that clarifies whether the paper is recyclable or not. Again, this reduces BilinearThanosNet's dependence on potentially noisy image features within the paper category and improves classification performance.

VII. CONCLUSION

In this study, we contributed a novel method of improving trash classification through our ThanosNet network, and compiled the 889 images and associated metadata that forms the ISBNet dataset. We demonstrated the effectiveness of ThanosNet and the associated concepts through comparisons against image-based, state-of-the-art models in TrashNet. The results show ThanosNet outperforming the other models with an F_1 score of 0.952.

ThanosNet has important implications and applications in the field of trash classification. In particular, the creation of a smart trash can for consumer use is promising, as it may be a solution to the difficulties consumers encounter when trying to classify recyclable trash. Successful implementation and deployment of such technology could result in visible improvements in recycling and trash management within various communities. Moreover, the idea of an attention-based classification system can be utilized by trash classification applications at the macroscopic level. Recycling plants could geo-tag incoming trash with landmarks including but not limited to residential communities, industrial parks, shopping centers, and hospitals. These meta labels would further help improve proposed image-based classification methods.

For future experimentation and research, we will continue to expand ISBNet to include more images and relevant fields of metadata. A larger, more comprehensive ISBNet dataset that is balanced will likely improve the performance of models trained and tested on it. Such a dataset would also be beneficial for creating accurate product-time models as it would more closely model the real-world, trash-generating, probability distribution. Furthermore, we believe that for time metadata, landmarks that possess multi-modal distributions cannot be effectively represented with time intervals that are one hour long. Therefore, collecting the "real" time of when the garbage was disposed, possibly as a scalar, would expel the need for estimating complex probability distribution.

ACKNOWLEDGMENT

We thank George Lin Wu, Minkyu Colin Jung, and Andy Kim for labelling and photographing all of the images that constructed ISBNet. We thank Hannah Graham, Hyoree Kim, and Alex Zheng for bringing this issue to light and inspiring the development of this project. This work was supported by the International School of Beijing's branch of the Net Impact organization.

REFERENCES

[1] M. Yang and G. Thung, "Classification of Trash for Recyclability Status," pp. 1–6, 2016. DOI: 10.1145/2971648.2971731. [Online]. Available: http://cs229.stanford.edu/proj2016/report/ThungYang-ClassificationOfTrashForRecyclabilityStatus-report.pdf.

[2] J. Knowles, S. Kennedy, and T. Kennedy, *OscarNet-Using Transfer Learning to Classify Disposable Waste.pdf*.