Computer Vision Based Waste Segregation

Mustafa İzzet Muştu
Department of Computer Engineering
İstanbul Technical University
İstanbul, Turkey
mustu18@itu.edu.tr

***Abstract*—Solid waste generation is growing to be a serious issue that requires immediate attention. Since different types of trash require different methods of disposal, a reliable and accurate categorization process is an essential stage in garbage disposal. Due to the varied architecture networks used, the current deep learning-based waste classification models are difficult to generate accurate results and still require improvement. Their performance on various datasets differs, and there aren't many particular large-scale training datasets that consist of enough labels available. Thus, in this project, we use and image of a single item of waste from 2 different datasets to categorize it into 2 main categories and 8 sub categories. We merge 2 different datasets, apply augmentation and create a model for this task using a SOTA model based on convolutional neural network (CNN). After training, model can predict the main classes with an accuracy of %94,67 and predict the 2 subclasses with an accuracy of %88,65 and %89,98.**

***Index Terms*—computer vision, waste, classification**

I. Introduction

A subfield of artificial intelligence (AI) and computer science called machine learning focuses on using data and algorithms to simulate how people learn, progressively increasing the accuracy of the system and demand for it is huge nowadays. There are several reasons and several outcomes of this demand. Starting with the reasons, smart devices and computers are everywhere around human being. Almost every interaction between a person and a computer leaves a mark, which means data, hence machine learning algorithms try to utilize it for different purposes such as prediction of an event. More usage of computers means more data and more data means better prediction or better problem solving. To better explain this, we need to understand the basic logic behind machine learning algorithms. Machine learning algorithms consist of 2 main steps after creation of a model. First it is trained with a given data for a problem, which means it learns, then we feed it with different data related with the same problem and we expect sensible output. Increasing the data helps the model the generalize and learn the problem better in training. That's why more data is better. If we consider the outcomes, it simply makes human life easier since machine learning help us to solve complex problems. A subset of machine learning called deep learning tries to learn problems by imitating the neural system of the humans with deep neural networks. These neural networks takes an input, dynamic or static, generates an output. For example, we may expect a neural network to generate the digit if we feed the network with the image of the same digit.

Computer vision which is also a subfield of computer science focuses on understanding the given image. Thanks to deep neural networks, computer vision has evolved and solving a lot of different type of vision based problems is now easier than before because of both neural networks and transer learning. To briefly explain transfer learning, some create a deep learning model, trains the model on millions of images for weeks and gets great results in terms of classification, segmentation, detection etc. But prodigiousness here is not getting great results on a certain dataset, it is that just finetuning the model on different dataset you can adopt the model to your different problem. Finetuning is the process in which you freeze most of the weights of the network, replace the last layer according to your problem and retrain network again with the data you have. This training takes less time because the layers closer to the input in the network have already learned the representations of low-level features in the image such as edges.

According to Global Waste Index 2022 [1], Turkey is chosen as the least environmentally friendly waste management country. Mostly, people throw every type of waste into the same garbage can, while some of the waste is biodegradable and others not, because of the lack of proper waste grouping. This situation gives rise to increase in landfills, consumption of toxic waste by animals and pollution. In this work, a computer vision based classifier by using SOTA ResNet-50 [2] as feature extractor model created and finetuned. It can segregate the waste from camera and may help governments like Turkey to improve their waste management systems.

II. Literature Review

There have been a lot of study in image-based waste classification systems in recent years. Some create new architectures for the same dataset and some just finetune the SOTA models.

In [3], they compared 12 different SOTA CNN based model and compared Support Vector Machine (SVM), sigmoid and softmax classifiers for each one. They used TrashNet dataset [4] and got best result from VGG19 model with SoftMax classifier has an accuracy of around 88%.

In [5], for processing the ImageNet database and achieving high classification accuracy, they suggested a new combination classification model based on three pretrained CNN models. The transfer learning model based on each pretrained model is built as a candidate classifier in their suggested model, and the best output of three candidate classifiers is chosen as

the classification outcome. For six categories, their suggested model achieves 96.5% percent classification accuracy.

In another study on waste classification [6], they offered a system using the ResNet-50 model as the extractor and SVM as the classifier. They obtained an accuracy rate of 87% using the TrashNet [4] dataset.

In [7], they also compared different SOTA CNN architectures on TrashNet. They also compared Adam and Adadelta optimizer in their study to see which works better and they got best result by using DenseNet121 [8] with an accuracy rate of 95%.

Similar study with [7] is [9]. They used TrashNet dataset with DensNet121. However, they utilized genetic algorithm (GA) to optimize the fully connected layer of the DenseNet121. This method improved the accuracy upto %99.6.

The literature depicts that the majority of the proposed solutions rely on only 1 dataset (mostly from TrashNet which is a small scale dataset consists of 2527 images with 6 categories) and try to get highest result by using the same dataset. However, in this study, 2 different datasets with different labels are merged to create diversity and see the impact of the combination on the results.

III. Methodology

A. Dataset

To function at its best, a deep learning-based image classification model needs to be trained on a lot of relevant datasets. There are numerous public image datasets available, including MS-COCO with object segmentation and ImageNet7 from the ILSVRC competition. However, these databases are general-purpose datasets which contain large number of classes that is not relevant to waste. To accomplish the objective stated in this paper, different datasets are required.

2 different datasets [10], [11] which are shared publicly on Kaggle are used in this work. First one consists of about 256000 RGB images with resolution of 200x200 and with biodegradable and nonbiodegradable labels. Luckily, this dataset is evenly distributed. Second one consists of about 15400 RGB images with a total of 8 labels (4 for biodegradable and 4 for nonbiodegradable). Unfortunately, most of the images belong to 1 label. Labels of the second dataset are food waste, leaf waste, paper waste and wood waste for the biodegradable class and electronic waste, metal cans, plastic bags and plastic bottles for the nonbiodegradable class. Sample images from both dataset is shown in Figure 1 and distribution of the data is shown in Figure 2.

At first, all of the images from both datasets are rescaled to 3x224x224 to be able to give them to the model. Then, since second dataset is not evenly distributed, augmentation is applied to images in the classes in which number of images are less. As an augmentation policy, suggested methods for the ImageNet dataset is used [12]. Distribution of the images after augmentation can be seen in Figure 3.

Fig. 1. Sample images from both dataset, first row is drawn from [10] and others are drawn from [11].

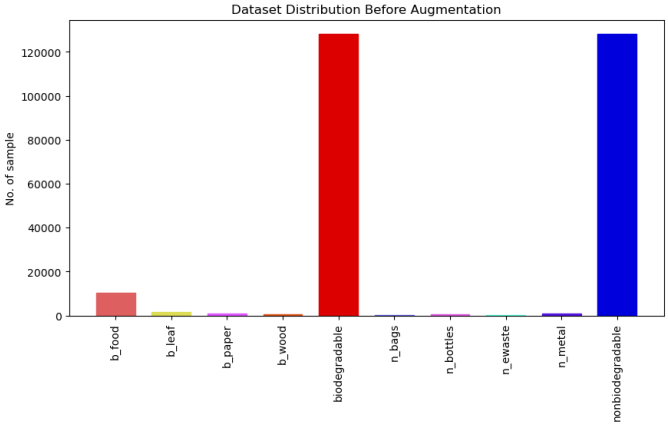

Fig. 2. Dataset distribution before augmentation.

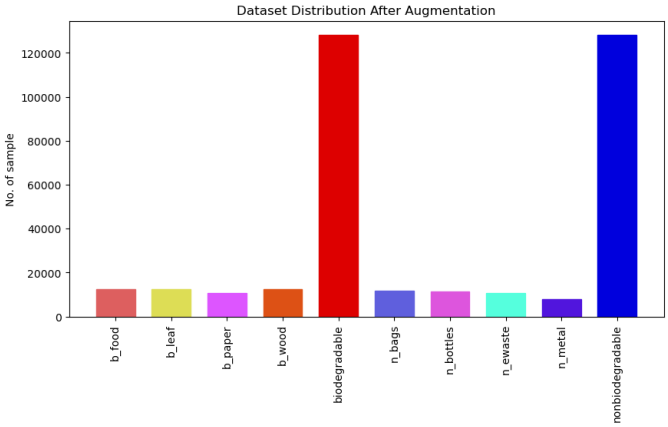

Fig. 3. Dataset distribution after augmentation.

B. Architecture

Since there are 2 different dataset and 2 different classifications, a multi-headed structure with 3 heads are used. For the first head, it classifies if the given image biodegradable or not. Second and third heads tries to predict the subclasses of the main classes. For example, head 2 predicts if the given image is food waste, leaf waste, paper waste or wood waste. Overall structure is shown in Figure reffig:OverallStructure.

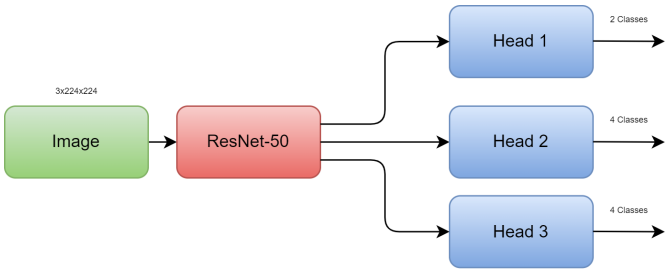

Fig. 4. Overall structure.

After data preprocessing, all of the data are sampled together to finetune ResNet-50. As labels, if the sample drawn from the data is from a subclass, one hot encoding are used for the head 1 and relevant subclass head, others labels are 0. However, if the subclass of the sample is not known, all of the outputs of the one head which represents the subclasses of the main are set to 1.

As feature extractor, ResNet-50 is chosen because it has 22.85 top-1 error rate on the ImageNet validation set and has a lower inference time on GPU with respect to the models that have higher accuracies. Comparisons of models can be seen in Figure 5.

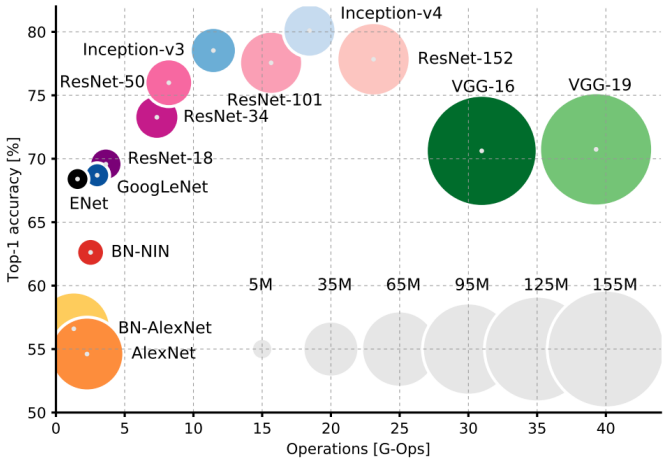

Fig. 5. Top-1 accuracies and operations of different SOTA models [13].

Different combinations of ResNet models are shown in Figure 6. In this work, 50-layer model is used and layers after the "conv4_x" layer are finetuned. Also, the fully connected last layer is replaced with 3 different head. Output of the first head is put into softmax and output of the other heads are put into sigmoid.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Fig. 6. Different combinations of ResNet [2].

C. Evaluation

Model predicts 1 main class and 1 subclass after an image is fed to the network. Evaluation of head 1 can be simply calculated because it is a binary classification. However evaluation of other heads a bit tricky. While target values for head 2 and head 3 can be $[1, 1, 1, 1]$, predicted labels are only in the one-hot encoded format. So, evaluation metrics for head 2 and 3 are calculated only by using drawn sample from subclasses. In addition, for the accuracy, if head 1 predicted the class incorrectly and head 2 predicted correctly, it is assumed that head 2 prediction is also incorrect.

IV. Experiments

A. Training

During training, batch size is set to 128. 80 percent of the training data is chosen as training and the rest is used for validation. Every image normalized using mean and standard deviation of the ImageNet dataset before fed into the model. Adam optimizer is chosen as optimizer with a learning rate of 0.00001, beta values of 0.9 and 0.999. Since we have large dataset, number of epoch is set to 10. Learning rate is multiplied by 0.95 after every epoch. As loss function, cross entropy loss is chosen for head 1, binary cross entropy is chosen for head 2 and head 3. The reason for the binary cross entropy is that when the drawn sample is from the first main class, target values of head 2 outputs are set to 1 and target values of head 3 is set to 0. So, the model act as a multilabel classifier. Same approach also valid for the target values of head 3 under the reverse condition. However, as mentioned in the Evaluation section, model predicts 1 main and 1 sublcass for an image. Training loss for this configuration is shown in Figure 7.

By observing the loss, we can say that 10 number of epoch is more than enough. Training and validation accuracies are calculated for all of the heads differently during training. Accuracy plots are shown in Figure 9 and Figure 10.

B. Testing

Testing accuracies are ended up being lower than training/validation accuracies for all of the heads. Outputs are shown in Figure 11. Especially for the head 2 and head 3, we can see that accuracies can go down to 0.6. This might be caused from that testing and training images are not fairly

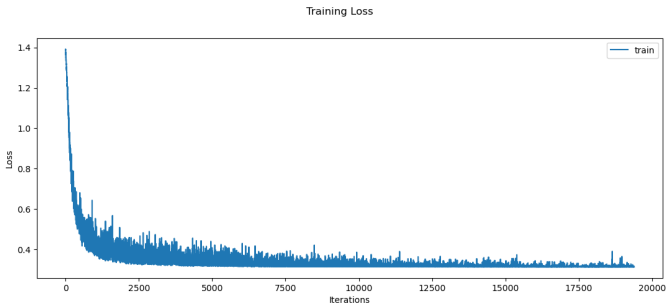

Fig. 7. Training loss plot.

Classification Report

	Precision	Recall	F1
biodegradable	95.96	93.66	94.80
nonbiodegradable	93.35	95.76	94.54
food	79.58	77.42	78.49
leaf	94.46	99.59	96.96
paper	81.57	88.54	84.91
wood	95.40	85.63	90.25
ewaste	82.60	98.89	90.02
metal	84.98	95.94	90.13
bags	95.78	92.08	93.89
bottles	97.67	76.81	86.00

Fig. 8. Classification report.

distributed in terms of representation of a class. Confusion matrices are shown in Figure 12. From these matrices, we can see that head 1 is classified well, head 2 and head 3 are biased to certain subclasses. For example, for the head 2, model tends to predict the given biodegradable image is more likely to be a leaf. Accuracy scores which are calculated from these matrices are 94.67, 88.65 and 89.98 for the head 1, 2 and 3 respectively. Precision, recall and F1 scores are also calculated from the confusion matrices and can be seen for all classes in Figure 8.

V. Conclusion

To conclude all of the work, ResNet-50 works well for this type of problem. Finetuning process is very helpful and make the training time last shorter. The best part of the work is that the model can actually be used in real time in waste management system of a country if a mini computer with camera is placed into garbage cans. Thus, recyclable items can be separated with a low error rate.

VI. Future Work

It can be seen from the Figure 2, merged dataset is not distributed well. Even after the augmentation, there is a large gap between subclasses and main classes. Suggested way to solve this problem is creating a dataset from scratch by using Google Search Engine. This method increases the dataset size and reduces to problem complexity to a single classifier output. Also, we can choose classes and distribute the images in each class as we want. Thus, we can use 1 classifier and there will be no need to 3 classifier.

References

[1] Person, "Global waste index 2022: Ranking the biggest waste polluters worldwide," Mar 2022. [Online]. Available: https://waste-management-world.com/research/global-waste-index-2022-ranking-the-biggest-waste-polluters-worldwide/

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[3] N. Ramsurrun, G. Suddul, S. Armoogum, and R. Foogooa, "Recyclable waste classification using computer vision and deep learning," 08 2021.

[4] Garythung, "Garythung/trashnet: Dataset of images of trash; torch-based cnn for garbage image classification." [Online]. Available: https://github.com/garythung/trashnet

[5] G.-L. Huang, J. He, Z. Xu, and G. Huang, "A combination model based on transfer learning for waste classification," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 19, p. e5751, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5751

[6] O. Adedeji and Z. Wang, "Intelligent waste classification system using deep learning convolutional neural network," *Procedia Manufacturing*, vol. 35, pp. 607–612, 2019, the 2nd International Conference on Sustainable Materials Processing and Manufacturing, SMPM 2019, 8-10 March 2019, Sun City, South Africa. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2351978919307231

[7] R. A. Aral, R. Keskin, M. Kaya, and M. Hacıömeroğlu, "Classification of trashnet dataset based on deep learning models," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 2058–2062.

[8] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016. [Online]. Available: https://arxiv.org/abs/1608.06993

[9] W.-L. Mao, W.-C. Chen, C.-T. Wang, and Y.-H. Lin, "Recycling waste classification using optimized convolutional neural network," *Resources, Conservation and Recycling*, vol. 164, p. 105132, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921344920304493

[10] R. Zamzamy, "Non and biodegradable material dataset," Jun 2021. [Online]. Available: https://www.kaggle.com/datasets/rayhanzamzamy/non-and-biodegradable-waste-dataset

[11] AashiDutt, "Waste segregation image dataset," Sep 2022. [Online]. Available: https://www.kaggle.com/datasets/aashidutt3/waste-segregation-image-dataset

[12] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation policies from data," 2018. [Online]. Available: https://arxiv.org/abs/1805.09501

[13] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," 2016. [Online]. Available: https://arxiv.org/abs/1605.07678

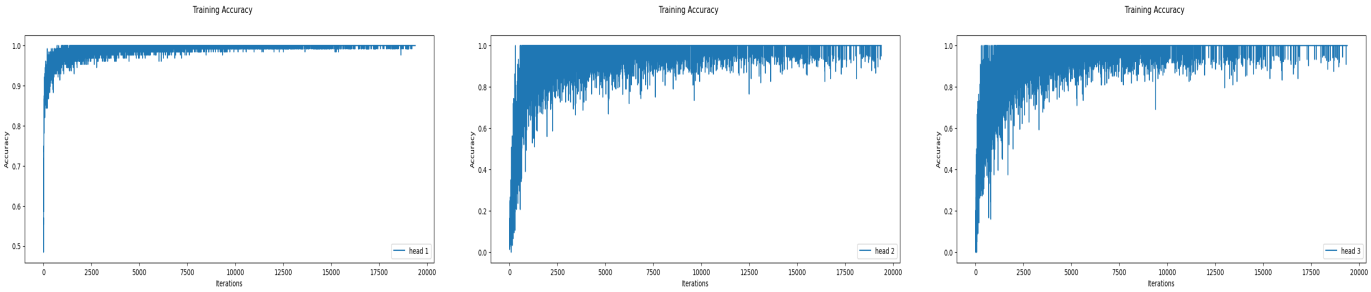

Fig. 9. Training accuracy plots for all heads.

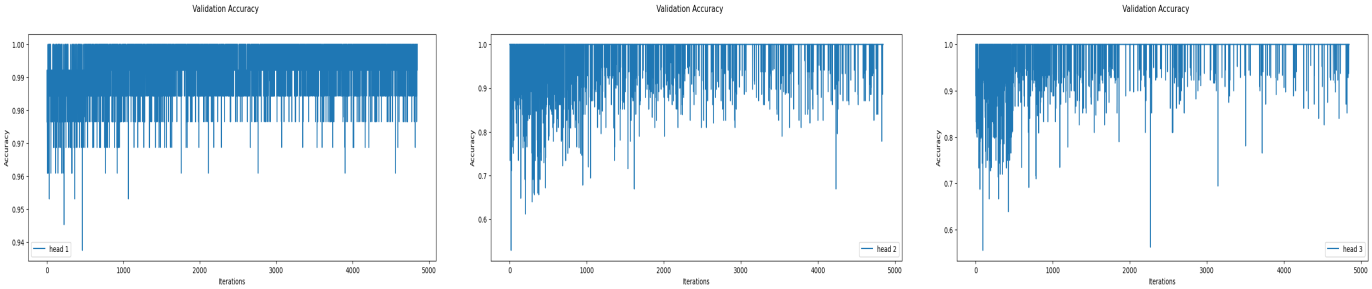

Fig. 10. Validation accuracy plots for all heads.

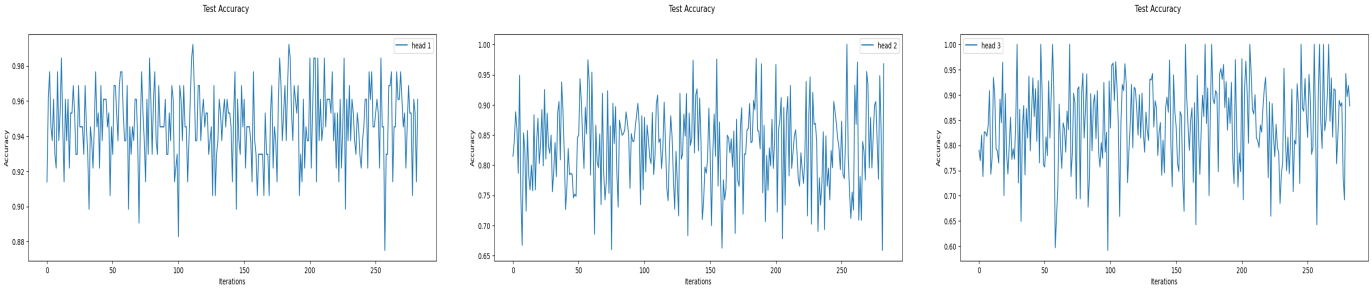

Fig. 11. Test accuracy plots for all heads.

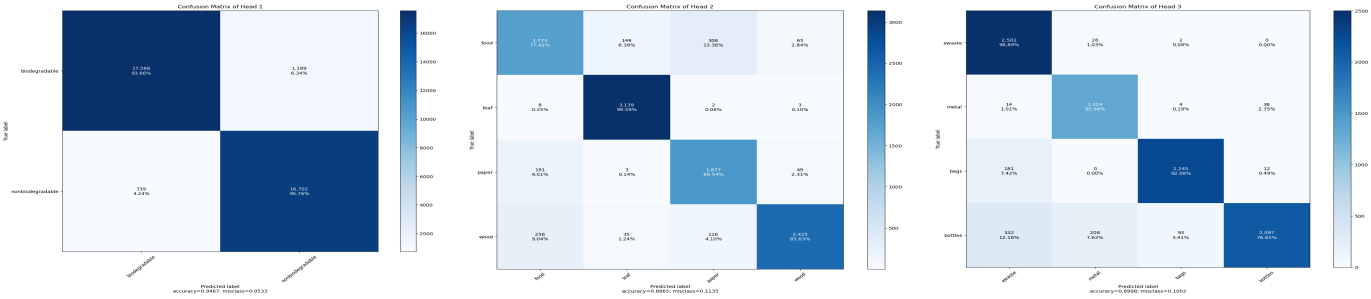

Fig. 12. Confusion matrices for all heads.