RecycleNet: Intelligent Waste Sorting Using Deep Neural Networks

Conference Paper · July 2018

DOI: 10.1109/INISTA.2018.8466276

Cenk Bircanoğlu
Bahçeşehir University

Meltem Atay
Middle East Technical University

Fuat Beser
Yildiz Technical University

Merve Ayyuce Kizrak
Bahçeşehir University

RecycleNet: Intelligent Waste Sorting Using Deep Neural Networks

Cenk Bircanoğlu
Computer Engineering
Bahçeşehir University
Deep Learning Turkey
Istanbul, Turkey
cenk.bircanoglu@gmail.com

Meltem Atay
Neuroscience and Neurotechnology
Middle East Technical University
Deep Learning Turkey
Ankara, Turkey
meltem.atay@metu.edu.tr

Fuat Beşer
Electronics and Communication Engineering
Yıldız Technical University
Deep Learning Turkey
Istanbul, Turkey
fuatbeser@gmail.com

Özgün Genç
Electrical and Computer Engineering
Koç University
Deep Learning Turkey
Istanbul, Turkey
ozgungenc@gmail.com

Merve Ayyüce Kızrak
Electronics and Communication Engineering
Yıldız Technical University
Deep Learning Turkey
Istanbul, Turkey
ayyucekizrak@gmail.com

***Abstract*—Waste management and recycling is the fundamental part of a sustainable economy. For more efficient and safe recycling, it is necessary to use intelligent systems instead of employing humans as workers in the dump-yards. This is one of the early works demonstrating the efficiency of latest intelligent approaches. In order to provide the most efficient approach, we experimented on well-known deep convolutional neural network architectures. For training without any pre-trained weights, Inception-Resnet, Inception-v4 outperformed all others with 90% test accuracy. For transfer learning and fine-tuning of weight parameters using ImageNet, DenseNet121 gave the best result with 95% test accuracy. One disadvantage of these networks, however, is that they are slightly slower in prediction time. To enhance the prediction performance of the models we altered the connection patterns of the skip connections inside dense blocks. Our model RecycleNet is carefully optimized deep convolutional neural network architecture for classification of selected recyclable object classes. This novel model reduced the number of parameters in a 121 layered network from 7 million to about 3 million.**

I. INTRODUCTION

In nature, everything is degraded and reused, this means there is no waste. Plants and animals die and rot into the ground, providing nutrients for successive plants and animals. For ages, both the reality of the waste and terms defining the waste have been evolving together. In many languages, there are several types of word structures defining waste, having common semantic relations [1]. The word *vastum* is the original Latin word has the meaning of to make room. *Vastum* was borrowed from Latin to early French language and Gaelic languages, eventually adapted into the English language in progress of time [1]. Throughout the human history, high population areas developed their approaches to dispose of the human-generated waste. Additionally, human-generated waste management was often confused with well-being and hygiene issues [1]. During the period of Middle Ages, as a consequence of increasing population and developing manufacturing techniques, solid wastes increased enormously in the streets [2]. To prevent the spreading of the diseases, authorities decided to move garbage piles from the streets that were facilitating the growth of the rat populations [2], [3]. This practice could be considered as the first garbage collection attempt [3].

During the 18th century, as a consequence of industrialization waste management became an inevitable part of the foundation of the modern societies. First forms of recycling practice were emerged to salvage reusable materials, especially scrap metals, paper and wood [4], [5].

Modern industrial recycling came into place during the 19th century, especially throughout World Wars I and II. The main reason behind recycling was the shortage of materials and economic crisis [6].

Through the end of the 20th century, especially 1970s, were the years of increasing environmental awareness [7]. First Earth day proposed in 1970 to emphasize the importance of protecting the natural environment. As a result of increased awareness, recycling became an important part of the socio-economical culture all around the world [7], [8].

Recycling has utmost importance from economical and environmental aspects. For more sustainable future, it is vital to maintaining an effective recycling system. Despite the seriousness of this field, currently recycling process is hugely depending on human abilities. As the technology advances products to recycle are also changing and overexposure to waste materials and by-products could be carcinogenic [9].

In order to automatize recycling process, along with advanced composting and incinerators, it is vital to propose

978-1-5386-5150-6/18/$31.00 ©2018 IEEE

intelligent systems to detect the waste components correctly.

The main purpose of this work is to demonstrate an efficient intelligent system to classify selected classes of common waste materials.

II. Related Works

One of the early works to demonstrate intelligent waste classification is based on a classical pattern recognition methodology; Bayesian framework [10]. Despite the well documented mathematical background, this methodology is not capable of perfect automation because it requires hand extracted features. Apart from waste classification, proof of different non-linearities into convolutional neural network architectures made impressing contribution to object recognition literature [11]. This benchmark work demonstrated the efficiency of non-linearities. An advanced computer vision approach to detect recyclability is main task of this work and dataset to experiment on novel approaches is obtained from a student work from Stanford University [12]. In addition to this early work, there is another study reporting usage of one of the latest approaches fast region proposed convolutional neural networks [13]. Recyclable material classification is a challenging problem, it requires meticulous techniques to obtain a reliable dataset. Apart from dataset collection, it is necessary to provide a generalizable approach for industrial applications. Our research results are promising insights on solving this challenge. In this study, we aimed for more efficient application of well-documented convolutional neural network architectures. We employed and compared several algorithms; deep residual convolutional neural network, MobileNet, DenseNet and one of the most efficient mixtures of inception modules with residual networks. In addition to well-known convolutional neural network architectures, we propose a novel architecture specific to recycling material dataset: RecycleNet. Details of RecycleNet is givenin the Models and Methods section.

III. Dataset

Focus of this study is to classify recyclable materials, hence waste management and recycling systems are among the essential parts of a sustainable economy. Using intelligent systems instead of humans as workers in the dump is a vital requirement for economical and safe recycling.

For this reason, our purpose is recognizing some of the most common recyclable materials such as glass, paper, cardboard, plastic, metal and trash. TrashNet dataset incorporates six classes of recycled objects [14]. The images of dataset have white background and pictures of trash and recycling around Stanford. Selected different pose and lighting for each photo, includes several variations in the dataset. Each image resized down to 512 x 384 pixels and the size of the original dataset is almost 3.5GB [12], [14]. Table I below shows number of images for the six classes and Figure 1 shows sample images from the dataset.

IV. Models and Methods

In this research, we experimented on several different deep convolutional neural network architectures as well as different optimization methodologies.

IV-A. Selection of Deep Convolutional Neural Network Architectures for Classification Task

Deep Residual Networks: Deep residual networks or ResNet is 2015 ILSRVC (ImageNet Large Scale Visual Recognition Competition) winner on ImageNet dataset, and ResNet50 is one of the most commonly used architectures of residual networks. Deep residual networks solved the inaccuracy problems and increasing training and test error as the layers stacked for deeper networks, in a scientific manner. Residual blocks allowed the formation of extremely deep convolutional neural networks. Apart from simply stacking convolutional layers, ResNet proposes an elegant formation of residual blocks with three convolutional layers accompanying with batch normalization and rectified linear unit (ReLU) activation function. Recurrently usage of these residual blocks constructs ResNET as deep as possible with fewer hyper-parameters owing to this deep structure and batch normalization between residual blocks, ResNET is able to achieve better feature extraction in comparison to previous versions [15]. In this research, we used ResNet50 variation for our classification task. ResNet50 input size is 224x224.

MobileNet: MobileNets an efficient model proposed by Google's research team for efficient usage of mobile devices. MobileNets perform depth-wise separable convolution after full convolution operation which provides to reach higher accuracy with a small number of hyper-parameters. In addition to depth-wise convolution, MobileNets are thinner with lesser parameters because they use reduced representations of the input and they are depending on model shrinkage parameter to keep the model from producing additional hyper-parameters. Due to smaller nature of MobileNets, they are able to train faster with lower resource requirements which is one of the most useful properties of them for versatility [16]. MobileNets input size is 224x224.

Inception-ResNet, Inception-v4 (InceptionResNetV2): Inception-v4 is the advanced version of the Inception-v3 model and both are hybrid of inception modules and residual connections. Residual connections make deeper and wider inception networks more efficient with lesser hyper-parameters. Earlier inception module implementation required more training resources and time, with residual connection improvement, requirements reduced and model turned out to be more efficient to train. Different from previous Inception-ResNet works, in inception-v4 batch normalization takes place on the top of traditional convolutional layers not before residual blocks, due to this property inception block size is increased. Apart from inception-v3 for inception-v4 contains scaling of residuals before layer accumulation, this scaling results with more stable training and higher accuracy and eventually allows to build an increased model size [17]. Inception-v4 input size is 299x299.

(a) Paper (b) Paper (c) Glass (d) Glass (e) Plastic (f) Plastic

(g) Metal (h) Metal (i) Cardboard (j) Cardboard (k) Trash (l) Trash

Figure 1. Sample Images of Dataset [12]

Table I
DATASET INFORMATION [12]

Recyclable material types	Number of each material	Training (70%)	Validation (13%)	Test (17%)
Paper	594	403	83	108
Glass	501	354	65	82
Plastic	482	347	61	74
Metal	410	286	56	68
Cardboard	403	287	46	70
Trash	137	91	17	29
TOTAL	**2527**	**1768**	**328**	**431**

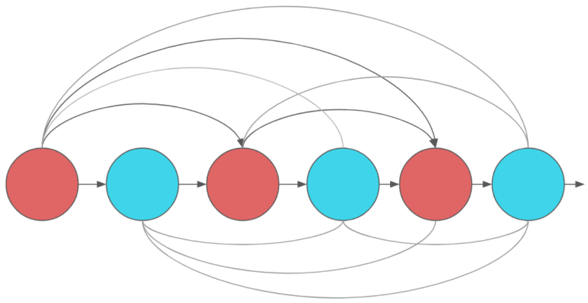

Figure 2. Skip Connections of DenseNet121 [18]

Densely Connected Convolutional Networks: Densely connected Convolutional Networks or shortly DenseNets are one of the most efficient deep convolutional neural network structures, because they contain shorter connections between layers close to the input and output. DenseNets are superior to previous research with their abilities to reduce parameters, strengthening the feature propagation and solving vanishing gradient problem as the network grows . Main different layer architecture of DenseNet is dense connectivity layer and its feature map forms the concatenation of multiple inputs into a single tensor as shown in Figure 2. Comparing with ResNet variants, DenseNet works more efficiently with fewer parameters. DenseNets naturally integrate identity mapping, deep supervision, and depth diversification. It is interesting to note that DenseNet also claimed to work better without data augmentation because of large margin properties [18]. Due to limited data number, our results seem to be confirming this claim. DenseNet input size is 224x224.

Xception: Extreme inception or in other words Xception model uses depth-wise convolution model instead of inception modules. It can be described as a stack of linear depth-wise separable convolution layers with residual connections. Comparing to inception-v3 model Xception gives better results with faster training [19]. Xception input size is 229x229.

RecycleNet

For this classification, task we want classification models that can achieve high accuracy when trained on relatively fewer number of classes (6 classes in our case), good at discriminating at low level features such as textures and materials, and most importantly suitable for real-time usage.

Analyzing the performances of various models we have tested, we observed that we are able to achieve the ideal classification performances with DenseNet family of networks when they were pre-trained on ImageNet dataset. One disadvantage of these networks, however, that they are slightly slower in prediction time. To enhance the prediction performance of the models we altered the connection patterns of the skip connections inside dense blocks. One particular modification that seems to work is to halve the number of skip connections. More precisely, Equation (2) of [18] becomes,

$$x_l = H_l([x_{l-1}, x_{l-m-1}, x_{l-2m-1}, x_{l-3m-1}, ..., x_0]), \quad (1)$$

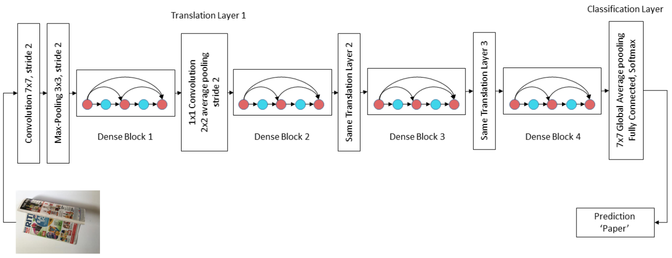

Figure 3. Purposed Model: RecycleNet

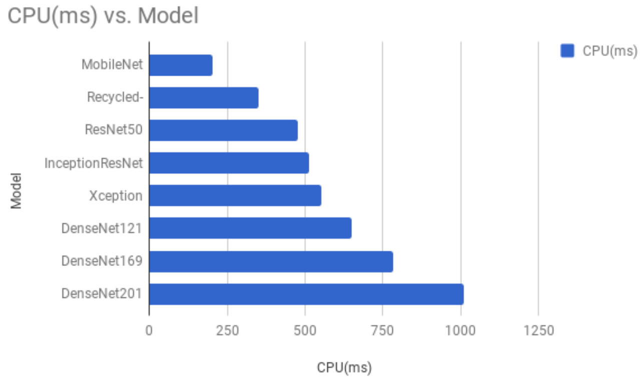

Figure 4. Performance of CPU(ms) vs Model

where x_l is the output of l^{th} layer, H_l is the composite function of operations, m is the hyper-parameter controlling the density of the skip connections ($m = 2$ in the case of halving the skip connections). This change reduced the number of parameters in a 121 layered network from 7 million to about 3 million. This scheme also gives as the ability to control the density of different dense blocks separately, by changing m. Figure 3 show the end-to-end model.

Different Optimization Approaches

Therefore our dataset is not uniform, and validation set is to small it could be expected for proposed algorithms to stick to one of the local minimas, in order to deliberately search for the global minimum and to explore the most optimized approaches and to demonstrate the outcomes of the different learning rates we experimented on two different optimizers.

Adam: Adam is an extension algorithm to stochastic gradient descent that has recently gained a wide range of use for deep learning applications in computer vision. The name Adam is derived from adaptive moment estimation. The parameters come with estimates of first and second moments of gradients can be computed individual adaptive learning rates. The method of optimization is invariant to diagonal rescaling of gradients, is computationally powerful, is straightforward to implement, has insufficient memory necessity, and is appropriate for problems that are wide in terms of data and/or parameters. Empirical results demonstrate that Adam works fine in the proceeding and compares approvingly to other stochastic optimization methods. [20]

Adadelta: Adadelta is a per-dimensional learning rate method using first order data and has minimize computational continuing cost of operation beyond vanilla stochastic gradient descent. Adadelta method to overcome the sensitivity to the hyper-parameter selection as well as to avoid the continual decay of the learning rates [21].

V. Experiments and Results

Main experiments of this research were performed using Keras library with TensorFlow back end (version 2.1.4) on Google Colaboratory platform [22]–[24]. Most of the local experiments were performed using GTX1080 Ti. During the training of 200 epochs of Inception-ResNet, Inception-v4 training each epoch took about 700 ms. and whole training of lasted for lasted for one and half hour. Longest training session took 3 hours with Inception-ResNet, Inception-v4 module for 300 epochs. For the main experiments, we performed training models from scratch, on training data, while training we also performed validation and using these weights we performed test experiments. Figure 4 shows performance of CPU vs model. In order to take advantage of model capacity, we also performed fine-tuning experiments on the selection of the best-performed models. Learning rate of Adam and Adadelta optimizers kept as default in Keras model definitions, 0.001 and 1 respectively without weight decay. For the transfer learning experiments, we performed fine-tuning of the weights of the pre-trained model on ImageNet dataset, after a small step of pre-training of ImageNet trained weights containing model, fine-tuning applied with stochastic gradient descent with 0.9 Nesterov momentum and learning rate was set to 0,0001. To improve generalization capability of neural network, we also tried simple data augmentation methodologies such as horizontal and vertical flip and 15 degrees random rotation. For all the training experiments, batch size was selected as 32.

Prediction time experiments are conducted on a dual-core Intel(R) Xeon(R) CPU @ 2.20GHz for the CPU measurements and on a GTX 980 GPU for the GPU measurements.

As a performance metric, we measured training, validation and test accuracies and multi-class cross-entropy loss, as well as real-time prediction performances on typical CPU and GPU systems. Figure 5 shows comparison of GPU, CPU and model performance.

VI. Conclusions

Classification of real-world examples for waste classification is not an easy computer vision task. Apart from environmental factors such as lighting, waste could be described as a shape-shifting material. For instance, one can compress a plastic bottle, or tear a paperboard. These materials do not lose their material properties but they lose their key properties to be identified as an intact object. Besides, virtually any object can be an input to a waste sorting system, but the available training samples are limited.

Table II
RESULTS OF TRAINING FROM SCRATCH

Model	Test Accuracy	Optimizer	Data Augmentation	Epochs	CPU (ms)	GPU (ms)
ResNet50	75%	Adam	-	100	478	15.4
MobileNet	76%	Adam	-	500	202	8.3
InceptionResNetV2	80%	Adam	-	100	513	39.9
InceptionResNetV2	87%	Adadelta	Vertical and horizontal flip, 15 degree rotation	100	”	”
InceptionResNetV2	90%	Adadelta	Vertical and horizontal flip, 15 degree rotation	200	”	”
InceptionResNetV2	88%	Adadelta	Vertical and horizontal flip, 15 degree rotation	300	”	”
DenseNet121	83%	Adam	-	100	649	22.2
DenseNet121	85%	Adam	-	200	”	”
DenseNet121	75%	Adam	Vertical and horizontal flip, 15 degree rotation	100	”	”
DenseNet121	84%	Adadelta	-	100	”	”
DenseNet121	76%	Adadelta	Vertical and horizontal flip, 15 degree rotation	100	”	”
DenseNet169	82%	Adadelta	Vertical and horizontal flip, 15 degree rotation	100	783	31.7
DenseNet201	85%	Adam	-	200	1010	39.9
DenseNet201	80%	Adam	- Vertical and horizontal flip, 15 degree rotation	200	”	”
Xception	85%	Adam	- Vertical and horizontal flip, 15 degree rotation	100	553	17.1

Table III
RESULTS OF FINE-TUNING EXPERIMENTS

Model	Test Accuracy	Optimizer	Data Augmentation	Epochs
DenseNet121	95%	Adam for initialization, SGD	Vertical and horizontal flip, 15 degree rotation	10 +200
InceptionResNetV2	87%	Adam for initialization, SGD	Vertical and horizontal flip, 15 degree rotation	10 +200

Table IV
RESULTS OF RECYCLENET MODEL

Model	Test Accuracy	Optimizer	Data Augmentation	Epochs	CPU (ms)	GPU (ms)
RecycleNet	81%	Adam	Vertical and horizontal flip, 15 degree rotation	200	352	15.9

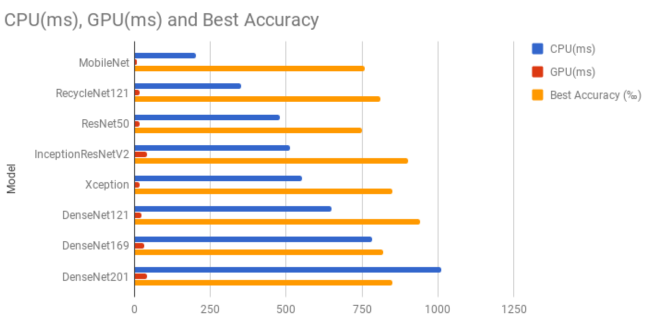

Figure 5. Comparison of GPU(ms), CPU(ms) and Model Performance

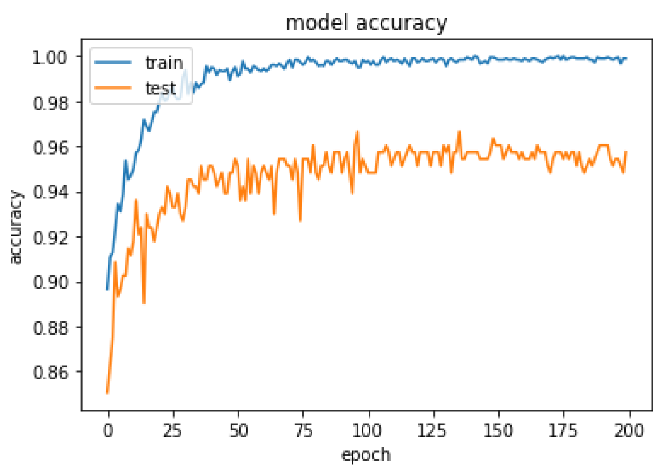

Figure 6. Accuracy for Best Result

This requires the system to generalize extremely well when trained with a relatively small training set. We believe that when the state-of-the-art convolutional neural networks are meticulously trained, they would be capable of producing industrial-grade results to solve these types of problems. This was the main motivation behind RecycleNet research. The results of this research prove this claim. When trained from scratch with the dataset of just 1768 images, all CNN based models we tried were able to score more than 75% in test accuracy. Some models like InceptionResNetV2 scored as high as 90%. With fine-tuning a 121 layered DenseNet model (pre-trained on ImageNet), we achieved 95% of test accuracy. We conclude that learning of waste sorting is feasible with modern deep learning based approaches.

Apart from different model architectures, we also compared the effects of different optimization methodologies (namely Adam and Adadelta) and data augmentation strategies. In some

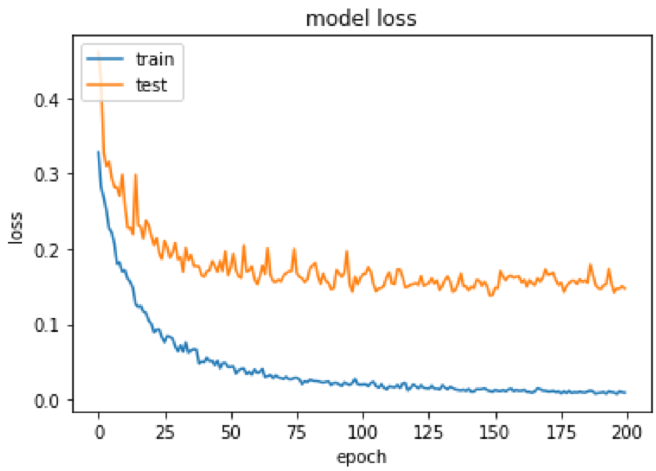

Figure 7. Loss for Best Result

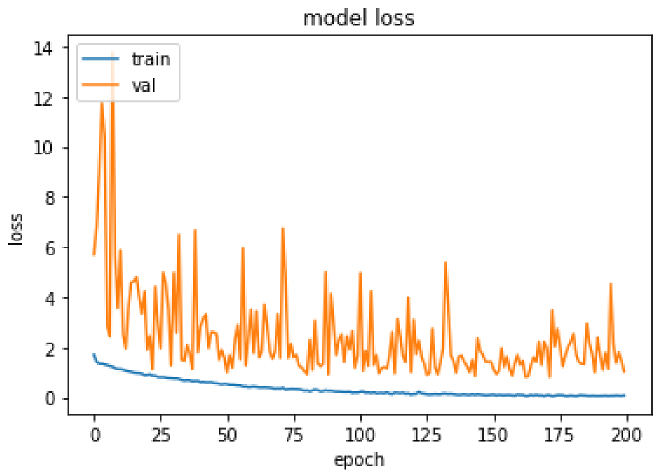

Figure 8. RecycleNet Train and Validation Loss

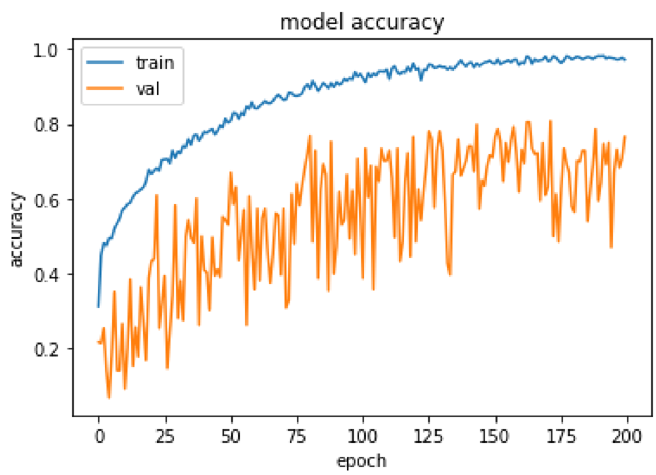

Figure 9. RecycleNet Train and Validation Accuracy

cases due to the complexity of the model data augmentation resulted with over-fitting, but mostly data augmentation found to be very useful. Especially for the dataset that we used where the class representations are relatively low. Table II and Table III show the results of training from scratch and fine-tuning experiments. Figure 6 and Figure 7 show accuracy and loss curves for the best results.

There are two novel approaches in this research. Usage of transfer learning on DenseNet architecture rather than training from scratch resulted in highest possible accuracy on trashnet dataset. On the other hand, main novelty of this research is the unique architecture of proposed model, namely RecycleNet. With this unique architecture we obtained 81% test accuracy for this limited dataset of recyclable materials. RecycleNet model configuration shown in Figure 3. Figure 8 and Figure 9 show that comparing accuracy graphs of best results RecycleNet is relatively consistent, considering the fact that our best performing model is using transfer learning methodology. Proposed model result as shown in Table IV.

One particular performance metric for the waste sorting problem is the real-time implementation potential. A typical deployment target of our models is either a smart bin application that automatically detects and segregates recyclable materials or an industrial-grade sorting system that can be installed over a conveyor-belt in a Material Recovery Facility. For the former scenario low-power implementation should be feasible and for the latter, the system should work at very high speeds. To address this requirement, We measured the prediction times of all our models. We see that our accuracy-wise top performing architecture was relatively slower in prediction time. We modified the network architecture, to gain 46% prediction-time performance improvement on CPU. We refer to this new model as RecycleNet throughout this paper. This new model is both faster and more flexible without much penalty in accuracy, hence more suitable for our specific task of material recognition and for real-time deployment.

To conclude, RecycleNet is a promising example of usage of artificial intelligence or more specifically deep learning on behalf of ecological awareness.

VII. FUTURE WORK

As a future work, we plan to refine our proposed novel architecture named RecycleNet for deformations and natural adversarial examples, using latest advances on convolutional neural network research. We will experiment with more sophisticated connection patterns and different optimization algorithms. Pre-training the new network on a very large scale general image dataset such as ImageNet was not performed in this work, but as our result with the best accuracy suggests, this step tends to increase the performance of the waste classification.

ACKNOWLEDGEMENT

This work is part of a collaborative research project by Deep Learning Turkiye, a broad non-profit organization dedicated to deep learning research in Turkey.

REFERENCES

[1] S. Barles, "History of waste management and the social and cultural representations of waste," in *The Basic Environmental History*. Springer, 2014, pp. 199–226.

[2] P. T. Williams, *Waste treatment and disposal*. John Wiley & Sons, 2005.

[3] S. A. Alchon, *A pest in the land: new world epidemics in a global perspective*. UNM Press, 2003.

[4] J. A. Palmer, "Environmental thinking in the early years: Understanding and misunderstanding of concepts related to waste management," *Environmental Education Research*, vol. 1, no. 1, pp. 35–45, 1995.

[5] D. S. Amick, "Reflection on the origins of recycling: a paleolithic perspective," *Lithic Technology*, vol. 39, no. 1, pp. 64–69, 2014.

[6] T. H. Witkowski, "World war ii poster campaigns–preaching frugality to american consumers," *Journal of Advertising*, vol. 32, no. 1, pp. 69–82, 2003.

[7] L. Davidson Cummings, "Voluntary strategies in the environmental movement: recycling as cooptation," *Journal of Voluntary Action Research*, vol. 6, no. 3-4, pp. 153–160, 1977.

[8] C. H. Lipsett, *100 Years of Recycling History: From Yankee Tincart Peddlers to Wall Street Scrap Giants*. Atlas Publishing Company, 1974.

[9] C. R. Álvarez-Chávez, S. Edwards, R. Moure-Eraso, and K. Geiser, "Sustainability of bio-based plastics: general comparative analysis and recommendations for improvement," *Journal of Cleaner Production*, vol. 23, no. 1, pp. 47–56, 2012.

[10] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring features in a bayesian framework for material recognition," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 239–246.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[12] M. Yang and G. Thung, "Classification of trash for recyclability status," *arXiv preprint*, 2016.

[13] O. Awe, R. Mengistu, and V. Sreedhar, "Smart trash net: Waste localization and classification," *arXiv preprint*, 2017.

[14] G. Thung, "Trashnet," *GitHub repository*, 2016.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[16] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[17] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning." in *AAAI*, vol. 4, 2017, p. 12.

[18] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. 1, no. 2, 2017, p. 3.

[19] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *arXiv preprint*, 2016.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[21] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.

[22] A. Google. (2017) Google colaboratory. [Online]. Available: https://colab.research.google.com/

[23] F. Chollet *et al.*, "Keras," 2015.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.