CS3244 Group 49 - Domestic Trash Classification

Yang Zi Yun (A0184682U), Ye Tong(A0183760B), Yu Xiaoxue (A0187744N), Liu Han (A0194490X), Wang Hanbo (A0204762J), Zhang Yuhao (A0194564U)

Department of Computer Science, National University of Singapore, Singapore

CS3244 Group 49

{e0313671, e0310555, e0323328, e0376920, e0424685, e0376994}@u.nus.edu

Abstract

Waste management is a crucial part of our daily lives and one of the biggest growing challenges faced by Singapore (Lai 2020). Waste classification based on its impact on the environment and human health is important for effective waste management. Although Singapore has not implemented any compulsory domestic waste classification, we foresee a need for Singapore to impose a stricter waste classification policy in the future. In this study, an automatic waste classification application is proposed to help with domestic waste classification. Different image classification machine learning models were tested and ResNet 50 was chosen based on the result after considering accuracy, model size and inference time. Code implementation can be found on https://github.com/ziyun99/CS3244.

1. Background information

1.1 Domestic Waste Classification Around the World

Over the past few years, we have witnessed a trend in which a greater number of countries are implementing compulsory domestic waste classification. For example, a new municipal solid waste (MSW) classification strategy was launched by the Chinese government in 2017 and Shanghai was selected to be the first pilot city (Zhou et. al. 2019). The Volume Based Waste Fee (VBWF) System has been implemented in South Korea and it requires residents to sort out their waste according to waste sorting rules. Fines of up to $1000 are imposed for violating garbage disposal rules. The system has managed to increase the recycling wastes by a significant amount, which proves the efficiency of the system (Henam 2019).

1.2 Waste Management in Singapore

Currently, the waste management in Singapore is handled by the National Environment Agency (NEA). There are 3 main categories of waste in Singapore: packaging waste, food waste and electronic waste. Under the category of packaging waste, single-use packaging containers such as those containing single-serving yoghurts, orange juice and soft-drink are one of the main culprits of packaging waste. However, these containers cannot be recycled due to contamination by food residue (Lai 2020). NEA research has shown that Singapore's recycling rate is generally maintained at a low level and has been decreasing in recent years. Lack of efficiency in the waste recycling process caused by improper waste classification leads to low recycling rates and hence results in unsustainable development.

Countries like Japan and South Korea have successfully increased their recycling rate over the years by implementing domestic waste classification, which alludes to the success of domestic waste classification. As Singapore is trying to achieve self-reliance, we foresee a need to impose a stricter waste classification policy in the future to help increase the recycling rate in Singapore.

According to our survey conducted on 80 Singaporeans aged between 20 to 50 years old, 75% of people feel that there is a need for Singapore to implement domestic waste classification measures as they help to promote environmental sustainability. 76.25% of people are unable to correctly classify certain domestic waste and have an accuracy of below 50% in the waste classification test. Therefore, an application that helps with domestic waste classification is necessary to prepare us for the future.

1.3 Domestic Waste Classification Application

Our group proposed a computer-vision based domestic waste classification application. The target group is domestic households and the aim is to help them classify domestic waste correctly. The input of the application will be an image of a waste and the output will be the category that it belongs to. The waste will be classified into 4 main categories: hazardous waste, residual waste, recyclable waste and compost waste.

Machine learning provides a smart and efficient way to analyse big datasets (Priyadharshini 2021). A good machine learning model will be able to extract meaningful insights by learning the dataset iteratively and handle complex and data-rich problems in a fast and accurate way. When an unknown image is input into the model, the model can classify the image to its corresponding category according to the features extracted and output the result with high accuracy.

The application makes use of machine learning models to implement image classification, which is one of the applications of computer vision where the computer is able to identify and process objects in images in the same way as the human vision system (Dexlab 2020). Image classification is to predict the category of only one object in an image (Brownlee 2021). Another consideration is to implement object detection, which is to detect instances of objects of a certain category within an image. However, the application is not required to identify and locate multiple objects in an image and hence, we choose to implement image classification in the application. Furthermore, there are very few pre-labeled datasets for object detection of waste and manual labelling of data can be extremely tedious considering that we need a large dataset. Therefore, we choose to go with image classification.

We considered three different image classification machine learning models: Vision transformer, Residual Network and EfficientNet. After sufficient experiments, Residual Network is chosen to be the most suitable machine learning model in terms of accuracy, inference time and model size.

2. Related Work

In the study "Optimization of CNN-based Garbage Classification Model", a new classification algorithm, DSCR-Net, which borrows from Inception-V4 and ResNet is used. The dataset contains 4 different classes: recycle waste, organic waste, harmful waste and residual waste. The algorithm was optimized and tested on the dataset with a 94.38% accuracy rate. The study shows that convolutional neural network models such as ResNet can be particularly useful in waste classification (Song et. al. 2020).

In another study "Classification of Trash for Recyclability Status" (Thung et. al. 2016), their research collected a dataset TrashNet which is one of the most complete trash dataset available online. Their research is able to achieve 75% test accuracy using CNN models, which is a great start in the area of trash classification. We benefited from the dataset collected to train our own model.

3. State-of-the-Art Models

3.1 Residual Network (ResNet)

ResNet is a specific type of neural network introduced in the paper "Deep Residual Learning for Image Recognition" (He et. al. 2016). In the traditional approach of solving complex problems, additional layers have to be stacked into the deep neural network in order to improve accuracy and performance. The intuition is that adding more layers will allow these additional layers to progressively learn more complex features. (Mujtaba 2020) However, addition of more layers leads to the problem of vanishing gradients. The gradients of the loss function approaches zero with increasing number of layers, making the network hard to train (Wang 2019). ResNet implements a deep residual learning framework, which uses identity shortcut connections that skips one or more layers to reduce the vanishing gradient problem (He et. al. 2016). ResNet-50 is able to achieve 83.2% top-1 accuracy on ImageNet.

3.2 EfficientNet

EfficientNet is a form of convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth, width and resolution using compound coefficient. Conventional practice for model scaling is to arbitrarily increase the CNN depth or width or use larger input image resolution for training and evaluation. Although these methods do increase the accuracy, they may require tedious manual tuning, and still often yield suboptimal performance. According to the paper "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", EfficientNet is able to achieve state-of-the-art 84.3% top-1 accuracy on ImageNet, while being much smaller and faster on inference than the best existing ConvNet in 2019 (Tan et. al. 2019).

3.3 Vision Transformer (ViT)

The fundamental flaws of CNN is that CNN fails to encode relative spatial information. It is able to detect certain features based on the image but it does not consider their position with respect to each other (Bhuiya 2020). For instance, CNN will classify an image as a human face even if the mouth, nose and eyes are out of place. Transformer was initially designed for natural language processing tasks. It uses the concept of self-attention, where it will look at each word of the sentence and compare its position in the sentence with respect to the position of all the works present in the same sentence, including itself (Vaswani et. al. 2017). ViT makes use of the same concept of self-attention where it divides the whole image into small patches and feeds into the transformer along with their positions in the image. It will go through alternating layers of multi-headed self-attention, multi layer perceptron and layer normalizations (Agarwal 2020). According to the statistics on Image Classification on ImageNet, ViT-H/14 is able to achieve 88.55% top-1 accuracy on ImageNet.

ViT makes use of transformers which were initially designed for natural language processing tasks. However, in contrast to words, individual pixels do not convey any meaning by themselves. The solution to this is relatively interesting, which is to divide the whole image into small patches or words, then apply the operation onto each of these groups of pixels. It is also interesting to see how the algorithm designed for another learning field is able to be applied to image classification.

4. Experiments

4.1 Dataset

We utilized several online datasets: which are the TrashNet (Thung et. al. 2016) and CompostNet (Frost et. al. 2019). TrashNet consists of 2527 images out of 6 classes: recyclable classes (glass, paper, cardboard, plastic, metal) and residual waste. CompostNet was built on TrashNet and added a compost class. Our final dataset is further built upon CompostNet and added 2 classes which are fabric and hazardous class. We collected the images from online search based on available items found in various Singapore stores, the hazardous class consists of items such as paint bucket, nail polisher remover, battery and insect repellent spray bottle. Hence, we classify all possible domestic waste into **9 classes** as shown in Table 1.

Table 1. Number of Images in each Waste Classes

Waste Classes	**Number of Images**
Recyclable	Glass(501), Paper(594), Cardboard(403), Plastic(482), Metal(410), Fabric (39)
Hazardous	206
Compost	177
Residual	184

4.2 Train-test Split

The train-test split is performed using random selection and stratified selection to prevent bias in data and ensures equal distribution of classes in each dataset. Train-test split is necessary because a model might overfit in the train data and performs badly on unseen data in the real world. We split our dataset into **80%** train data, **15%** validation set and **5%** test set.

4.3 Data Augmentation

Image augmentation is performed on training dataset to increase the diversity and size of train dataset. This helps to improve the performance and ability of the model to generalize and reduce overfitting. New images are created on existing images by performing transformations such as random resized cropping, color jitter, random rotation and flipping. The images are resized to the dimension (224, 224) and normalized according to the settings in ImageNet.

4.4 Model Training

We choose to do **transfer learning** based on image classification models that have weights pretrained on ImageNet. These models have learnt general visual features such as edges, shapes and patterns, hence are able to transfer the knowledge to a new dataset much more efficiently than training the models from scratch.

We first load the **pretrained models** from https://github.com/rwightman/pytorch-image-models and remove the last classification layers at the top because **Imagenet** models are default to output 1000 classes. We then add a fully connected layer with ReLU activation, followed by a dropout layer with probability 0.4, and another fully connected layer with 9 output units and softmax function corresponding to our 9 waste classes.

We experimented both feature-extraction and fine-tuning method on ResNet50 and found out that feature-extraction method has limited performance which has a highest accuracy of 89%, while fine-tuning allows the model to reach 94% accuracy. This is because the backbone of the model is freezed during feature extraction and hence are not able to generalize the features learnt from ImageNet to our dataset which are considerably different from ImageNet. Therefore, fine-tuning which retrain all weights of the models is more suitable for our use case.

We performed a comparative analysis to select a suitable **fine-tuning** model for our application. We compared the performance ResNet, EfficientNet and Vision Transformer (ViT) under the same training parameters, with batch size of 64, SGD (Stochastic Gradient Descent) as optimizer, cross entropy loss as loss function, learning rate of 0.001 and weight decay of 0.9. To prevent overfitting, **early stopping** is implemented to terminate the training when validation loss stops to decrease for 3 epochs although training loss might still be decreasing. All experiments above are carried out on four GeForce RTX 2080 GPUs and using Pytorch as the main framework.

For example, Figure 1 shows the loss (left) and accuracy (right) versus number of epochs in the training process of ResNet50. Early stopping happens at the 18th epoch where the validation loss no longer decreases.

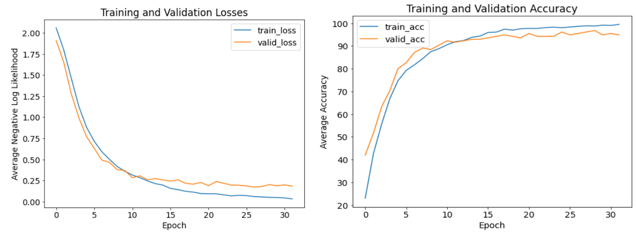


Figure 1. Loss (left) and accuracy (right) versus number of epochs in the training process of ResNet50.

4.5 Model Evaluation

The trained models are tested on unseen validation and test dataset, and assessed based on evaluation metrics such as accuracy and confusion matrix as shown in Figure 2. We also recorded other results including training time per epoch, inference time and model size, which are listed in Table 2 in details.

Table 2. Experiment results on ResNet50, ResNet101, EfficientNetB0, EfficientNetB1 and ViT.

Model	Train accuracy (%)	Validation accuracy (%)	Test accuracy (%)	Training time per epoch (secs)	Inference time (ms)	Model size (total parameters)
ResNet 50	98.62	95.48	95.31	6.79	13.46	24,034,889
ResNet 1001	98.7	**96.13**	96.87	10.3	25.35	43,027,017
EfficientNetB0	97.7	93.55	92.18	**5.65**	18.8	**4,337,797**
EfficientNetB1	98.41	92.26	92.18	7.2	26.31	6,843,433
ViT	98.87	94.84	95.31	18.5	**11.63**	7,218,569

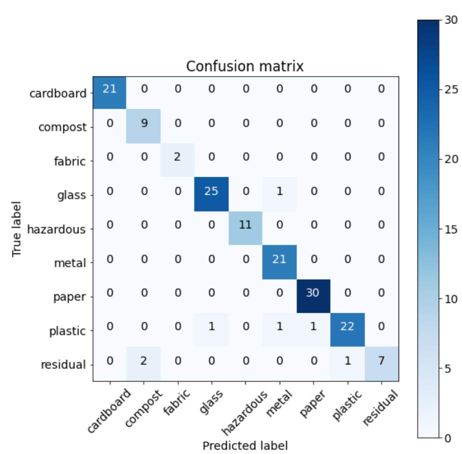


Figure 2. Confusion Matrix of ResNet50 on test dataset

4.6 Model Selection

We consider the 3 factors in selecting our final model, namely validation accuracy, inference time and model size.

4.6.1 Accuracy

Among all the fine-tuning models we tried, **ResNet101** achieves the highest validation accuracy of 96.13%, followed by **ResNet50** (95.48%), and then **ViT** (95.31%).

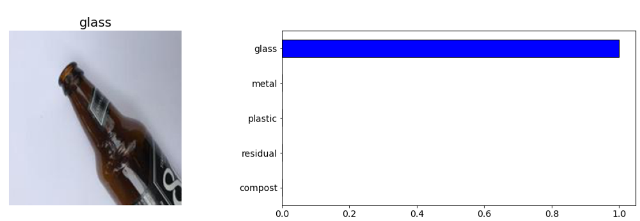
4.6.2 Inference time

Due to the high variance in the runtime on a low latency system, we run inference on 300 examples and take the average time. **ViT** has the shortest inference time of 11.63ms, followed by **ResNet50** (13.46ms) and **EfficientNetB0** (18.8ms).

4.6.3 Model size

Since our application is targeted to domestic users and would be mostly applied on mobile platforms, hence model size is crucial to fit the model into limited size of hardware devices. According to the number of total parameters, **EfficientNetB0** has the smallest size of 4,337,797, followed by **EfficientNetB1** (6,843,433) and **ResNet50** (24,034,889).

After taking into consideration all factors above, we decided to choose **ResNet50** as our final model. This is due to it's 2nd high accuracy (95.48%) and 2nd fast inference time (13.46ms) and 3rd smaller model size (24,034,889 parameters). Although ResNet101 has the highest accuracy (96.13%), ResNet50 is much smaller in size and has smaller hypothesis space, hence it is chosen according to Occam's razor theory. Inference time of ResNet50 (13.46ms) is considerably fast compared to the fastest one ViT (11.63ms). For model size, we can perform quantization to further reduce its size.



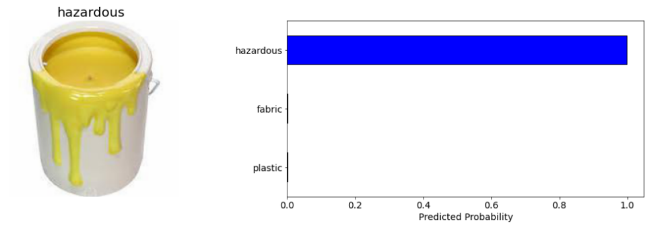


Figure 3. Examples of classification of ResNet50.

5. Ethical Considerations

When photos are uploaded into our application, privacy may be violated due to the access of personal information. However, this could be resolved by asking for user consent and checking if the user agrees to allow the application to use the photos for future model training. It is also important to ensure the security of the image uploaded for classification so that personal information will not be leaked.

6. Conclusion & Future Work

This paper describes the exploration of three different image classification machine learning models: Residual Network, Vision Transformer and EfficientNet. After a series of experiments, we found out that the model with the best tradeoff in terms of model size, accuracy and inference time is ResNet50.

One of the improvements to the application is to collect feedback including misclassified images from users for error checking and further improvement. Our model can also be enhanced if users agree to upload photos for future model training.

Besides that, we could maintain a knowledge graph of waste classes to prepare our application to be more robust to handle more complex combinations of waste classes. We wish to include more related information related to trash processing to educate the public on various implications of proper waste management.

7. Roles of Team Members

Yang Zi Yun: Code implementation and experiments on various image classification models & final result analysis.

Ye Tong: Data collection and research related work on section 2 and machine learning models in Section 3.

Yu Xiaoxue: Generated idea for domestic waste classification. Data collection and research on Section 1.

Wang Hanbo: Data collection and writing on the first part.

Liu Han: Data collection and research on state-of-the-art models.

Zhang Yuhao: Data Collection and research on experiments in Section 4.

8. Learning Points & Manpower Allocation

We divided the work into several parts with respect to our abilities and strengths. Teammates with good research skill focused on research and report writing while teammates with good coding skill focused on implementing the model searched. Data collection was done by every member to increase efficiency. The main learning point for all of us is the vast availability of different state-of-the-art machine learning models. Despite each model having their own working principle, many models are built upon a similar base model. We greatly appreciate this project opportunity to build our first ML application on trash classification that could be impactful to society.

References

Lai, L. (2020, August). Waste Management In Singapore (Learn What Happens To Your Trash). ISwitch. https://iswitch.com.sg/waste-management-in-singapore/

Zhou, M. H., Shen, S. L., Xu, Y. S., & Zhou, A. N. (2019). New policy and implementation of municipal solid waste classification in Shanghai, China. International journal of environmental research and public health, 16(17), 3099.

Henam, S., & Sambyal, S. S. (2019, December). Ten zero-waste cities: How Seoul came to be among the best in recycling. DownToEarth. https://www.downtoearth.org.in/news/waste/ten-zero-waste-cities-how-seoul-came-to-be-among-the-best-in-recycling-68585

Priyadharshini. (2021, March). What is Machine Learning and How Does It Work? Simplilearn. https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-machine-learning

Dexlab. (2020, February). Computer Vision and Image Classification -A study. DexLab Analytics. https://www.dexlabanalytics.com/blog/computer-vision-and-image-classification-a-study

Brownlee, J. (2021, January). A Gentle Introduction to Object Recognition With Deep Learning. Machine Learning Mastery. https://machinelearningmastery.com/object-recognition-with-deep-learning/

Song, F., Zhang, Y., & Zhang, J. (2020, October). Optimization of CNN-based Garbage Classification Model. In Proceedings of the 4th International Conference on Computer Science and Application Engineering (pp. 1-5).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Mujtaba, H. (2020, September). Introduction to Resnet or Residual Network. GreatLearning. https://www.mygreatlearning.com/blog/resnet/ Wang, C. (2019, January). The Vanishing Gradient Problem. Towards Data Science. https://towardsdatascience.com/the-vanishing-gradient-problem-69bf08b15484

Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning (pp. 6105-6114). PMLR. Bhuiya, S. (2020, June). Disadvantages of CNN models. OpenGenus IQ: Learn Computer Science. https://iq.opengenus.org/disadvantages-of-cnn/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.

Agarwal, S. (2020, November). Is this the end for Convolutional Neural Networks? Towards Data Science. https://towardsdatascience.com/is-this-the-end-for-convolutional-neural-networks-6f944dccc2e9

G. Thung, M. Yang. Classification of trash for recyclability status. CS229 Project Report 2016, 2016.

S. Frost, B. Tor, R. Agrawal and A. G. Forbes, "CompostNet: An Image Classifier for Meal Waste," 2019 IEEE Global Humanitarian Technology Conference (GHTC), Seattle, WA, USA, 2019, pp. 1-4, doi: 10.1109/GHTC46095.2019.9033130.