Comparison of Different Neural Network Architectures for Waste Management

Abstract—**Recycling our waste efficiently can reduce our impact on the environment, for example reducing non-biodegradable waste in the oceans. Correctly identifying types of waste helps to make the recycling process efficient and a neural network can be utilised during this process. This report investigates and compares different neural network architectures for identifying types of waste. It will mainly focus on accuracy and runtime. It finds that the pre-trained 'RegNet' network is most accurate but at the cost of having the longest runtime, whereas 'EfficientNet' has a lower accuracy but runs much more quickly.**

I. Introduction

Efficient waste management and recycling is essential for environmental sustainability and resource conservation. In recent years huge efforts have been made to increase our use of recyclable materials and to stop relying on unsustainable one-use materials. In the UK between 2016 and 2018, generation of landfill decreased by 2.8% whilst recyclable waste increased by 4.3% [7]. Despite the positive affects of these efforts on the environment, the improper sorting of waste is becoming a concerning issue; in the United States it is estimated that nearly 25% of all waste in recycling receptacles are non-recyclable [2] [3]. Consequently the separation process takes more work, recycling plants frequently slow down to deal with malfunctioning machinery and recycling companies make less profit; some even have to shut down. The end result is less recycling actually occurs [4]. The increase in contamination of recycling can be attributed to changes in product use and public perception of recycling and landfill, for example: promotion of recycling being a means to save the environment may unintentionally cause people to choose to dispose waste in recycling since it has better connotations. Additionally, other measures created to encourage recycling unknowingly backfire by increasing contamination, such as increasing the size of recycling bins and decreasing the size of landfill bins. The solution to recycling contamination is efficient and accurate sorting of garbage.

Current methods of sorting waste include: manual, mechanical and optical sorting. Manual sorting is very accurate but labour intensive and expensive. Mechanical sorting uses the weight, density and size of waste to sort it but struggles to sort waste with more complex compositions based on physical properties alone. Optical sorting classifies waste by colour and size but lacks versatility when it comes to waste it hasn't seen before. AI computer vision is a solution that is inexpensive and versatile compared to current methods.

This report seeks to explore using Convolutional Neural Networks (CNNs) to accurately classify garbage and prevent contamination and improperly sorted waste. We will evaluate how several CNN models perform and optimise them to be as accurate and efficient as possible. The objective is to find or create a CNN that can identify trash with as few mistakes as possible, to avoid contamination of recycling, and with a small inference time, to allow recycling plants to sort the large volumes of waste produced everyday.

II. Literature Review

Numerous studies have explored the application of computer vision and CNNs for garbage image classification and created effective and robust models, but many of these studies mention challenges mainly relating to the dataset and the number of classes that are used. ResNet is found to be an accurate and successful network at classifying images of garbage but its application to the real world, where waste can include many more categories than in most garbage datasets, may be limited [9]. The report uses ResNet18 and achieves an accuracy of 95.87% on a dataset with 6 classes. From testing multiple ResNet models they find extracting complex features results in a lower accuracy and slower inference time. This suggests there is a balance between complexity and practical constraints, such as efficiency. Another study produces a successful network with high recall and precision, however due to its high complexity it uses a combination of small datasets and few classes [5]. The identification time is optimised to be 9.8s, but this could be lowered using different hardware and a consistent dataset. Considering how many tonnes of waste a recycling plant might have to sort in a day, 9.8s to classify an image could be too long. More advanced attempts manage to create a highly accurate CNN with a low inference time despite using a relatively large number of classes, however using this many classes creates challenges getting sufficient data [8]. This report's CNN reaches 96.96% accuracy, which is incredibly high considering it uses a dataset of 43 classes. The challenges of the paper were addressing the limitations of traditional deep learning techniques in handling noisy data and lack of data, and the need for incremental learning to accommodate new categories of garbage. These challenges necessitated the development of novel methods such as weakly-supervised transfer learning and feature mixup modules. From this it can be inferred that high accuracy models are still possible with many classes, but fewer classes with many samples are far easier for deep learning techniques, especially when some obscure waste categories have insufficient data.

III. Method

A. Dataset

Given the existing literature related to our project, the dataset and number of classes used will have a large impact on the accuracy and complexity of the final model. We chose to use a dataset consisting of 6 classes consisting of approximately 400 images for each class. The classes consist of: cardboard, glass, metal, paper, plastic and trash. This was to keep the complexity and inference time low whilst allowing the network to distinguish between the important types of recycling and landfill. If a recycling plant could separate the garbage they receive into these 6 classes it would sufficiently decontaminate recycling.

B. K-means and PCA

Our preliminary exploration of the dataset used Principal Component Analysis, (PCA) and K-means clustering to use as a baseline for more complex deep learning techniques. PCA reduces the dimensionality of the samples in the dataset into two orthogonal axes that maximises the variance of the samples which allows us to visualise the structure and relationships of the data. K-means clustering is an unsupervised learning algorithm that splits the data into a specified number of clusters, identifying similar samples and grouping them together. By clustering the data based on PCA-reduced features, we can establish a benchmark for subsequent classification tasks using CNNs. This allows us to compare the performance of more complex models (such as CNNs) against a simpler clustering-based approach.

(a) Cardboard (b) Glass (c) Trash (d) Plastic

Fig. 1: Visualisation of 4 images from different classes in dataset.

C. CNNs

Multiple CNN's are compared to explore which are effective at and well adapted to garbage classification. ResNet, short for Residual Network, is a CNN well adapted for computer vision tasks. When neural networks are trained deeper they can suffer from degradation when gradients become very small. ResNet uses skip connections to bypass network layers which helps avoid issues with vanishing gradients, which allows deeper architecture and easier training. Pretrained ResNet CNN's are widely available from PyTorch. Pretrained models are often more accurate than networks trained from scratch on a dataset because they have already learnt to distinguish between classes of images. For example, the models used in this report have been trained on ImageNet, a huge dataset composed of more 1 million images and 1000 classes [6]. The parameters and weights they learn help them capture patterns like edges and textures which are important in all image classification tasks. Two ResNet models are used, ResNet50 and ResNet101, to explore the effect of additional layers/complexity.

EfficientNet is a CNN with exceptional performance and efficiency in computer vision tasks. It introduces an approach called compound scaling, which optimally balances network depth, width, and resolution to achieve superior results while conserving computational resources. By scaling network parameters, EfficientNet achieves high accuracy with reduced computational complexity. Its versatility and adaptability make it an ideal choice for a wide range of applications, offering high accuracy and efficiency in resource-constrained environments. This could be more appropriate for the applications of garbage classification where recycling plants could be constrained by the inference time of a model.

RegNet, short for Regularised Network, is another CNN designed to address challenges in computer vision tasks. It incorporates skip connections like ResNet to combat vanishing gradients but also introduces regularisation techniques for enhanced generalisation and prevention of overfitting. RegNet's scaling capability strikes a balance between model complexity and computational resources, making it an ideal choice for various applications, including garbage classification in resource-constrained environments.

Simple CNN is a CNN made from scratch without pretrained weightings using the fundamentals of what a CNN should have: transformation of dataset into suitable form, image normalisation, a basic CNN class and training function. Simple CNN will act as a baseline for the performance of all other models.

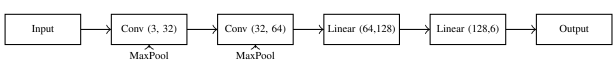


Fig. 2: Simple CNN 6 layers architecture.

Having a simplified version of a CNN offers advantages such as reduced complexity and improved interpretability. These benefits make Simple CNN architecture an appealing choice where computational resources and model interpretability are constraints.

We hypothesise that ResNet50 will be a "middle of the road" option, balanced between inference time and accuracy, whereas EfficientNet and RegNet are more optimised for speed and accuracy respectively due to their difference in complexity. We can also predict that ResNet101 will be more accurate than ResNet50 due to additional layers and complexity but slower for the same reason. Simple CNN is expected to perform worst because of its lack of pretrained weightings and basic architecture.

The loss function used for all CNN's is the cross entropy loss, which is the most common for multi class classification tasks, and can be calculated using

$$H(q,p) = -\sum_{c} q_c \log p_c,$$

where q_c is the true probability of class c in the data and p_c is the probability of class c predicted by the model.

Adam is the most widely used optimiser for loss functions in neural networks due to its adaptive learning rates and incorporation of momentum, preventing slow convergence and oscillation. The selection of a learning rate of $5e^{-5}$ aims to find a middle ground, ensuring the network converges efficiently while avoiding excessive oscillations during training.

D. *Evaluation*

To quantify the performance of each CNN we need to be able to evaluate them, several metrics can be used to measure the effectiveness of NN's in multi class classification. F_1 scores are the harmonic mean of precision and recall and can be calculated for each class in the dataset. Precision is how many images identified in a class are correctly identified, whereas recall is a measure of how many images of that class weren't identified. F_1 scores are calculated using:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2TP}{2TP + FN + FP},$$

where TP is true positives, FN is false negatives and FP is false positives. F_1 scores are chosen over other F scores, for example $F_{0.5}$ and F_2, because F_1 is a measure that weighs false positives and false negatives equally. In the application of garbage decontamination, a false negative is just as harmful as a false positive. The F_1 score ranges from 0 to 1, where 1 indicates no images from that class were misidentified and no images were not identified.

Additionally we can extract confidence values to analyse the models performance. After processing an image the CNN makes a set of logits for the 6 classes, which it uses to choose which class the image belongs to. These are applied to a sigmoid function, fitting them in [0,1], this value can be interpreted as the probability the image belongs to a class. The confidence values indicate the certainty level of the model's predictions. While a network may generate accurate predictions, overlooking a class with a high confidence value or exhibiting generally low confidence values might suggest the network is unreliable.

Mean Average Precision (mAP) another metric for evaluating the performance of multi-class classification models. It provides an assessment of a model's ability to accurately identify classes across various confidence thresholds. To calculate mAP, precision-recall curves are constructed by varying the confidence threshold for classification. The average precision (AP) for each class is then computed by calculating the area under the precision-recall curve. Finally, mAP is obtained by averaging the AP values across all classes. This metric is particularly valuable where class imbalances exist, the dataset we use has an imbalance in samples; most classes have 400 or more images but the trash class has less than 200. mAP score is calculated using:

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i,$$

where N is the number of classes and AP_i is the average precision.

Inference time is the time taken in milli-seconds for a CNN to process and classify an image. The average inference time measurements are taken during training of the model so the actual inference time in an application of the CNN will be shorter than measured, due to backpropagation during training, but still similar.

IV. Experimentation and Analysis

A. K-means and PCA

Figure 3 displays the result of K-means clustering on a PCA of the dataset. K-means has identified the 6 clusters and we can see the clusters aren't very separated from eachother. This suggests samples from different classes are relatively similar. The cluster in the centre, glass, borders all 5 other classes. This can be interpreted as glass samples having similar features to other classes and we would expect it to be the hardest to distinguish in our deep learning techniques. We might also expect later models to struggle identifying the trash class as its data points in Figure 3 have a high variability.

B. CNNs

The hyperparameters, optimiser and hardware are consistent between all NN's to ensure fair comparison in performance.

To analyse the performance of CNN's with individual classes we can look at the average F_1 scores across all CNN's instead of across all classes.

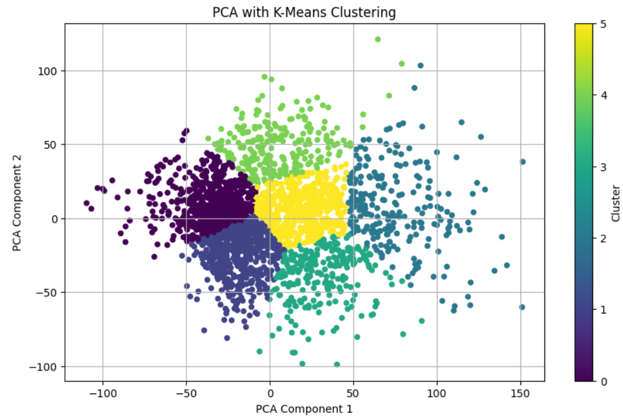


Fig. 3: PCA and K-means clustering of dataset where 0 represents paper, 1 is metal, 2 is cardboard, 3 is trash, 4 is plastic and 5 is glass.

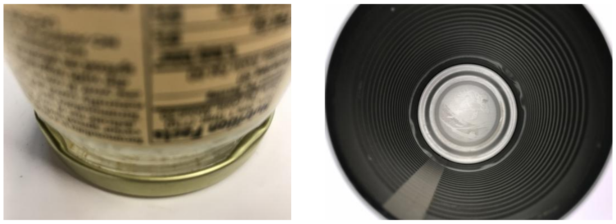
	ResNet 50	ResNet 101	RegNet	Efficient Net	Simple CNN
Avg Acc	0.9487	0.9140	0.9431	0.8707	0.6214
Max Acc	0.9670	0.9497	0.9757	0.9653	0.6335
Avg F_1	0.9253	0.9203	0.9310	0.9501	0.5963
mAP score	0.9758	0.9740	0.9781	0.9862	0.6571
Inf. time (ms)	13	20	71.4	6	9
Avg Confidence	0.7119	0.7423	0.7529	0.6702	0.4503

TABLE I: Results from network trained from scratch and pretrained ResNet50, ResNet101, RegNet and EfficientNet. Avg Acc and Max Acc are the average and maximum validation accuracy over 20 epochs, Avg F_1 is the average F_1 score over all 6 classes, Avg Confidence is the average confidence over all 6 classes and Inf. time is the average inference time in milliseconds.

	Average F_1 score
Trash	0.9729
Paper	0.9688
Plastic	0.9643
Metal	0.9394
Cardboard	0.9394
Glass	0.9230

TABLE II: Average F_1 scores for each class across all pretrained CNN's.

Despite glass being the class with the second most samples in the dataset, glass consistently has a low F_1 score compared to other classes. This was an expected result from the PCA and K-means clustering of the data. Figure 4 visualises examples of ResNet50 mistakes relating to the glass class.



(a) Glass jar incorrectly identified as paper

(b) Tin can incorrectly identified as glass

Fig. 4: False negative and false positive for glass class, the class with the consistently lowest F_1 score.

	Average F_1 score
Trash	0.7152
Paper	0.4607
Plastic	0.5116
Metal	0.6798
Cardboard	0.6010
Glass	0.5000

TABLE III: Average F_1 scores for each class from Simple CNN

On the other hand, paper has the lowest F_1 score among the classes for Simple CNN with it being the class with the most samples. The mention of Glass being expected to have low F_1 scores based on PCA and K-means clustering suggests prior analysis into the characteristics of this class, which might not apply to the paper class. This tells us that the model to the architecture of the Simple CNN model not be well-suited to capture the nuanced features necessary for accurately classifying instances of the paper class, leading to lower performance compared to other classes.

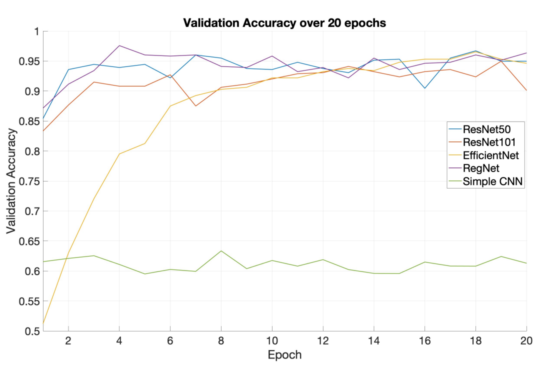


Fig. 5: Validation accuracy of ResNet50, ResNet101, RegNet, EfficientNet and Simple CNN over 20 epochs.

As expected, Simple CNN is the lowest performing of the CNNs at an average accuracy of 62%. It also doesn't seem to improve over the 20 epochs, which can be explained by the simplicity of the CNN's architecture. The CNN only has 6 layers which doesn't allow it to capture intricate details in the images. All the pretrained CNNs apart from EfficientNet reach over 90% average accuracy after just a few epochs. EfficientNet also gets to a high accuracy but after more epochs than other CNNs. EfficientNet is designed to achieve better parameter efficiency by scaling the depth, width, and resolution of the network in a balanced manner. While this improves performance on resource-constrained devices, it could result in taking more epochs to achieve high accuracy. Predictably EfficientNet is the fastest pretrained CNN, however we expected this efficiency to come at a cost to accuracy, which was proved wrong after more epochs. We also hypothesised that ResNet101 would outperform ResNet50, yet it seems ResNet50 sufficiently captures the data's features since ResNet101 actually performs worse. Finally, we accurately predicted RegNet would be one of the best performing CNN's, but also one of the slowest.

V. Discussion

A. *Consistency Across Models:*

The hyperparameters, optimizer, and hardware were kept consistent across all neural network architectures to enable a fair comparison of performance. This approach allowed us to attribute differences in performance to the architectures themselves rather than external variables.

B. *Performance on Individual Classes:*

While overall performance was strong, analysis of the average F_1 scores revealed disparities across different classes. Table II highlights that glass, despite being the second most common class in the dataset, consistently had lower F_1 scores. This suggests a challenge in the model's ability to generalize the features of glass compared to other materials.

C. *Error Analysis:*

Specific instances of misclassification, as seen in Figure 4, provided practical examples of the challenges faced by the ResNet50 model. A glass jar was incorrectly identified as paper, while a tin can was wrongly classified as glass, indicating possible confusion between reflective surfaces and transparency in the model's learning.

D. Validation Accuracy

Figure 5 illustrates the validation accuracy of each network over 20 epochs. It is evident that the EfficientNet and the neural network trained from scratch had a lower initial performance compared to the pre-trained models, however EfficientNet quickly improved. The graph shows that pre-trained models benefit from transfer learning, achieving higher accuracy more quickly than a model trained from scratch.

E. Implications for Waste Management Systems

There is currently a significant proportion of waste is directed landfills with approximately 24% in the UK. Instances of improper disposal, such as placing recyclable material in the wrong bins or co-mingling waste contribute to increased costs and time associated with sorting and separating the trash as well as unintentionally releasing toxins into the surrounding area and affecting natural biodiversity so having an automated waste management system will mitigate these issues.[1].

The implications of these findings are significant for automated waste management systems. High classification accuracy is critical for effective sorting and recycling. Therefore, enhancing the accuracy of classifying challenging categories like glass should be a priority.

F. Limitations

One limitation of this study is the focus on validation accuracy and F_1 score. Other metrics such as precision, recall, and real-time processing speed could be considered in future work. Additionally, the dataset's balance and the number of samples could affect the model's ability to learn certain classes.

VI. Conclusion

This report has compared the effectiveness of different neural network architectures. It is important to consider both accuracy and inference time when choosing an architecture. If the accuracy is too low, some waste will be sorted incorrectly and if the inference time is too long, the network may not be able to sort the waste fast enough.

This report concludes that 'EfficientNet' is best suited for sorting waste because it has a very low inference time and it reaches a high level of accuracy after training. This means that this network will be able to sort pieces of waste the quickest and will do so to a high level of accuracy.

To integrate this model into a real-world waste sorting system, another network would be necessary. This additional network would identify individual pieces of waste from a photo of a collection of waste using a bounding box and feed each item separately into the network we have designed.

Future research could explore the incorporation of additional data augmentation techniques, the use of ensemble methods, or the development of more sophisticated models to improve classification, particularly for problematic categories like glass.

References

[1] *Ai and Waste Recognition - Why It Works So Well*. https://recycleye.com/ai-and-waste-recognition-why-it-works-so-well/. [Accessed 19-03-2024].

[2] Brent Bell. *The Battle Against Recycling Contamination is Everyone's Battle Regional Recycling Waste Reduction District — regionalrecycling.org*. https://regionalrecycling.org/the-battle-against-recycling-contamination-is-everyones-battle/. [Accessed 03-03-2024]. 2018.

[3] Aaron Davis. *American recycling is stalling, and the big blue bin is one reason why — washingtonpost.com*. https://www.washingtonpost.com/local/dc-politics/american-recycling-is-stalling-and-the-big-blue-bin-is-one-reason-why/2015/06/20/914735e4-1610-11e5-9ddc-e3353542100c_story.html. [Accessed 03-03-2024]. 2015.

[4] Claire Groden. *The American recycling business is a mess: Can Big Waste fix it?* https://fortune.com/2015/09/03/waste-management-recycling-business/. [Accessed 03-03-2024]. 2015.

[5] Yuefei Li and Wei Liu. "Deep learning-based garbage image recognition algorithm". In: *Applied Nanoscience* 13.2 (Sept. 2021), pp. 1415–1424. ISSN: 2190-5517. DOI: 10.1007/s13204-021-02068-z.

[6] *Models and pre-trained weights 2014; Torchvision 0.17 documentation — pytorch.org*. https://pytorch.org/vision/stable/models.html. [Accessed 19-03-2024].

[7] Food UK Department for Environment and Rural Affairs. *UK statistics on waste — gov.uk*. https://www.gov.uk/government/statistics/uk-waste-data/uk-statistics-on-waste-recovery-rate-from-non-hazardous-construction-and-demolition-cd-waste. [Accessed 01-03-2024].

[8] Jianfei Yang et al. "GarbageNet: A Unified Learning Framework for Robust Garbage Classification". In: *IEEE Transactions on Artificial Intelligence* 2.4 (Aug. 2021), pp. 372–380. ISSN: 2691-4581. DOI: 10.1109/tai.2021.3081055.

[9] Qiang Zhang et al. "Recyclable waste image recognition based on deep learning". In: *Resources, Conservation and Recycling* 171 (Aug. 2021), p. 105636. ISSN: 0921-3449. DOI: 10.1016/j.resconrec.2021.105636.