Optimizing recycling using automated image-based classification:
A machine learning approach for improved waste management

Lili Freia Annemarie Böhmer

Universidade Católica Portuguesa

Dissertation written under the supervision of professor Pedro Afonso Fernandes and submitted in partial fulfilment of requirements for the MSc in Business Analytics, at the Universidade Católica Portuguesa.

May 30, 2024

Title: Optimizing recycling using automated image-based classification: A machine learning approach for improved waste management

Author: Lili Freia Annemarie Böhmer

Abstract

This thesis covers image-based waste classification using machine learning and discusses its impact on sustainable waste management. To identify the optimal model, the prediction performance of a DenseNet, a state-of-the-art Convolutional Neural Network, and DAtNet model are examined and compared to each other. The DAtNet integrates attention layers on the DenseNet architecture, inspired by the transformer model, known for its success in large language models. Moreover, the impact of transfer learning and augmentation on the test accuracy is analyzed. The performance of these models is evaluated across multiple datasets to examine their generalization capabilities. The findings indicate that while the DAtNet surpasses the DenseNet model in accuracy with large datasets, it faces difficulties with smaller datasets and requires significantly more time to preprocess the images. In contrast, the DenseNet consistently performs well and processes images more efficiently. Therefore, a DenseNet model is recommended for waste management facilities due to its reliability and lower computational demands. However, the further investigation and improvement of attention layers is encouraged. Additionally, the development of more practical, representative datasets is essential for the effective implementation of machine learning models in real world waste management. The deployment of this work could support the achievement of Sustainable Development Goals and the realization of zero-waste cities.

Keywords: Intelligent Waste Classification; Convolutional Neural Networks; Waste Management; Machine Learning.

Título: Otimização da reciclagem através da classificação automatizada baseada em imagens: Uma abordagem de aprendizagem automática para uma melhor gestão dos resíduos

Autor: Lili Freia Annemarie Böhmer

Resumo

Esta tese incide sobre a classificação de imagens de resíduos sólidos urbanos recorrendo a métodos de aprendizagem automática e sobre a discussão do respetivo contributo para a gestão sustentável de resíduos. Para o efeito, é comparada a performance de uma rede neural convolucional de última geração DenseNet com a de um modelo DAtNet. Este último integra camadas de atenção na arquitetura DenseNet, sendo inspirado pela tecnologia "transformer" que é reconhecida pelo seu elevado desempenho em modelos de linguagem de grande dimensão. Em particular, é analisado o impacto da transferência e da ampliação de imagens na precisão das predições. O desempenho desses modelos foi avaliado recorrendo a várias bases de dados de modo a comprovar a respetiva capacidade de generalização. Os resultados obtidos indicam que, embora o DAtNet supere o modelo DenseNet em precisão com grandes conjuntos de dados, enfrenta sérias dificuldades com bases de menor dimensão e requer mais tempo para pré-processar as imagens. Em contraste, o DenseNet tem consistentemente um bom desempenho e processa as imagens com mais eficiência, sendo recomendável para ecocentros devido ao pequeno custo computacional. Tal não impede futuras investigações, nomeadamente, em termos de melhoria das camadas de atenção dos modelos. Paralelamente, disponibilizar bases de dados fiáveis e representativas é essencial para a implementação de modelos de aprendizagem automática em casos reais de resíduos sólidos urbanos. Em última instância, o presente trabalho contribui potencialmente para a concretização de Objetivos de Desenvolvimento Sustentável e para o compromisso de cidades com zero desperdício.

Palavras chave: classificação inteligente de resíduos sólidos urbanos; redes neurais convolucionais; gestão de resíduos; aprendizagem automática

Acknowledgements

Special acknowledgements go to my family and friends for always supporting me along the way, not just during the journey of this thesis, but throughout my entire path. Your constant encouragement has been my source of strength and motivation, enabling me to reach this milestone.

Further, I want to thank my supervisor, Professor Pedro Afonso, who played a crucial role throughout the creation of this thesis by providing guidance and insightful advice.

My appreciation also goes to all the contributors and supporters of open-source platforms. Sharing your knowledge and resources has been instrumental in my academic journey.

Finally, the highest respect goes to the important contributions of individuals committed to environmental protection, particularly in the field of waste management. You make this world a better one.

Contents

List of Figures

List of Tables

Acronyms

BiT	Big Transfer
CNN	Convolutional Neural Network
DAtNet	Dense-Attention Network
DPN	Dual Path Network
FN	False Negative
FP	False Positive
MSDA	Mixed Sample Data Augmentation
MHS	Multilayer Hybrid Method
MLH-CNN	Multilayer Hybrid CNN
NLP	Natural Language Processing
PBA	Population Based Augmentation
SDG	Sustainable Development Goal
SNN	Spiking Neural Networks
SVM	Support Vector Machines
TN	True Negative
TP	True Positive
ViT	Vision Transformer

1 Introduction

The overpopulation, the steady rising consumption and the unawareness of resource recovery in today's urban cities are using up the global finite natural resources. One consequence is the high waste generation, which produces nearly 5% of global greenhouse gas emissions. These could be reduced by 25% through basic improvements in waste collection, waste reduction, reuse of products, recycling and organics waste management (Kaza et al., 2018, Chapter 6). Due to a lack of waste management in especially low- and middle-income countries, solid waste is polluting the oceans, air and soil as well as negatively affecting the associated economy. Additionally, policies and technology in particular are only occasionally available for developing countries, as they depend on the income and the geographical attributes of a city (Kaza et al., 2018, Chapter 4-6).

As recycling is one of the key issues in sustainable waste management, a circular economy model, which aims to recover waste for reuse and recycling, is essential for minimizing resource consumption (Zaman and Lehmann, 2011; Mehedi et al., 2023). However, with special regard to developing countries, the tools, methods and strategies designed for waste recycling management should be affordable, manageable and effective in the case of economy and technology. Further, for a circular economy to work, the end consumer is expected to contribute to repairing, returning, and recycling material products (do Carmo Stangherlin et al., 2023; Chioatto and Sospiro, 2023).

Through policies made by the government this behavior can be encouraged and hence allow the protection of 17% to 24% of raw materials (Zhang et al., 2021b; Chioatto et al., 2023). In addition, Ferronato and Torretta (2019) state that especially in developing countries, common projects in a global manner should be introduced to improve the overall technology and economic investments. Such technology and innovation are required for the implementation of a circular economy with specific regard to the adoption of intelligent waste management systems (Mehedi et al., 2023; Schmidt and Laner, 2023; Akbarpour et al., 2021; Ajwang et al., 2021). Thereupon, machine learning techniques can be used to transform cities into zero waste cities, as they enable automated sorting techniques for a better precision and quality (Chen, 2022; Zaman and Lehmann, 2011). In zero waste cities 100% of waste is recycled and all resources from waste materials are recovered.

This thesis is built on the benchmark of zero waste cities and the personal wish and commitment for a more sustainable future. Based on the above mentioned difficulties and the lack of clarity for individuals on how to properly recycle their waste, the primary objective of this thesis is the discovery of a suited machine learning model for automated waste classification and recycling based on image data. Several models are getting reviewed in order to identify two models for further analysis and comparison. One of the selected models will be a traditional Convolutional Neural Network (CNN), given their widespread usage in image classification tasks and consistently outperforming accuracy throughout the years. Nevertheless, previous research has introduced extensions of CNNs and additional models, such as the Vision Transformer (ViT), which are at least as accurate. As such, the second model under consideration will be one of these newer architectures. Thus, the following research questions are supposed to be answered during this thesis:

How do traditional CNN models and emerging machine learning models differ in their performance in terms of reliability, accuracy and computational efficiency?
Moreover, what implications do the findings have for future image-based waste recycling?

It is assumed that *one machine learning model will outperform the other in image-based waste recycling.* Three datasets and extensions of them are utilized in this thesis, to ensure a broad range of data for improved generalization. The datasets are called TrashNet, which has six classes, Seven Plastics, with seven classes and Household Garbage, with 12 classes. The thesis is limited to solid waste, which excludes heavy industrial, clinical, agricultural, radioactive and mining waste. Additionally, to guarantee the inclusion of most recent technology and discussions, only literature published after the year 2010 is considered. Although there are no restrictions on location, the further application and integration of the findings may vary depending on it, influenced by the technology-level and financial possibilities. In general, variety, quality and time play a crucial role in image-based waste recycling and are going to be considered along the thesis.

The gained insights and findings present valuable opportunities for both governments and non-profit organizations[1] to enhance waste management facilities and initiatives. By leveraging these insights, the establishment of zero waste cities is promoted and developing countries can significantly reduce waste and diseases. Apart from the public sector, a smart home bin can be designed in order to support individuals in accurately recycling their waste within their household. Further, through educational applications, the awareness for sustainable waste management can be enhanced.

The remaining thesis is structured as follows. Section 2 covers the used literature and its findings, followed by the materials and methodology of the thesis in section 3. The code's overall structure and the architectures of the models are explained in section 4. Afterwards, the findings and comparison of the chosen models' key measures are outlined in section 5. Section 6 further discusses and questions the work of this thesis with the help of previous research. The conclusion with recommendations for further steps can be found in section 7.

[1]Such non-profit organizations and associations are CleanHub, International Solid Waste Association (iSWA), Sungai Watch and the Ocean Cleanup.

2 Literature Review and Related Work

In this section, multiple machine learning models and strategies are presented, building the fundamental of selecting the two models for the subsequent analysis. First, the traditional CNNs are getting analysed and discussed in section 2.1, followed by training strategies, extensions of CNNs and additional models in section 2.2. Subsequently, related research on image-based waste recycling is summarized in section 2.3.

2.1 Traditional Convolutional Neural Networks

When it comes to image classification and recognition tasks, CNNs are the most used and reliable, because of their high accuracy and ongoing improvements through breakthroughs in deep learning and machine learning technology. The LeNet-5 model was introduced in 1998 and represents the start of CNN's popularity, but was surpassed by other models the years after (Lecun et al., 1998; Chen et al., 2021). However, the AlexNet model by Krizhevsky et al. (2012), which has a similar architecture as the LeNet-5, won the ImageNet competition in 2012 and represents the benchmark of CNN models nowadays.

Based on the AlexNet, Simonyan and Zisserman (2015) introduce their architecture VGGNet using deep learning. Deep learning is done by continuously increasing the depth of an image by using convolutional filters and layers. Accordingly, Simonyan and Zisserman (2015) create a model that generalises well on other datasets and further prove that the depth is beneficial for improving the classification accuracy. Having a generalized model that can be trained on different kinds of datasets is crucial for a model's performance, as it can reveal limits and potentials. This applies particularly to CNN models, as they tend to vary in their performance across datasets (Krishna et al., 2018).

Built upon the VGGNet, He et al. (2016) implement the ResNet and thereby improve the performance of deep learning through a deep residual learning framework that solves degradation problems. The deeper the network, the more layers information about the input or gradient have to pass, causing difficulties in optimizing a deep network (Huang et al., 2017). This in turn leads to an increase of the train and test error and is called degradation problem (Chen et al., 2021). Consequently, the training error resulted by a deeper model should not exceed the one produced by a shallower counterpart (He et al., 2016).

While the ResNet's architecture uses shortcuts from one layer to another for the information flow, Huang et al. (2017) introduce an architecture, where all layers are getting connected, called DenseNet. The dense connectivity patterns have a regularizing effect, which reduces overfitting on smaller training datasets. Further, a DenseNet is simple to train and requires only one third of the parameters of a ResNet (Huang et al., 2017). To combine the advantages of ResNet and DenseNet, Chen et al. (2017) introduce the Dual

Path Network (DPN) using an architecture that merges effective feature re-usage from the ResNet and re-exploitation from the DenseNet. The DPN achieves lower computational costs and memory consumption, a faster training speed, smaller model size and higher accuracy than usual CNN models. Due to these benefits, DPN is optimization friendly and highly valuable for research and practical tasks.

The above mentioned CNN models represent only a few possibilities. Chen et al. (2021) provide a clear and transparent overview of the evolution of CNNs up to 2021. Nevertheless, they highlight that extensions of CNNs should not be underestimated in their performance, as they gain more popularity and deliver notable potential.

2.2 Training Strategies, CNN Extensions and Other Models

As an architecture built on the ResNet, the Big Transfer (BiT) architecture uses transfer learning by training a network on a large supervised source dataset and fine-tuning its weights on upcoming tasks (Kolesnikov et al., 2019). Thus, task-specific data is replaced by a pre-training phase, providing the BiT with a general recipe to achieve excellent performance in many tasks. Further, BiT prevents per-task requirements and thus high costs for the training of networks to new tasks, as deep learning usually requires a large amount of task-specific data to obtain good performance (Kolesnikov et al., 2019).

Besides the downside of training, CNNs have a high probability of overfitting due to limited data. Through augmentation, rotating, flipping and injecting noises to images, the training dataset can be increased resulting in an improvement in the generalization of the model (Krizhevsky et al., 2012). The primary difficulty with data augmentation is that domain knowledge is needed to guarantee that the newly created data corresponds to legitimate transformations, that is, transformations that would naturally occur in that domain (DeVries and Taylor, 2017). Further, training a CNN model with a learned data augmentation policy automates the data enhancement and can significantly improve accuracy, model robustness and performance (Chen et al., 2021). Ho et al. (2019) introduce the Population Based Augmentation (PBA), which generates nonlinear augmentation policy schedules rather than fixed augmentation policies, allowing it to learn state-of-the-art augmentation policy schedules fast and effectively.

Another strategy to improve a CNN's performance is the Mixed Sample Data Augmentation (MSDA), which randomly mixes two training samples and their labels according to a certain ratio. Thus, MSDA can not only reduce the incorrect classification of some difficult samples but also improve the robustness of the model and make it more stable during training (Liang et al., 2023).

Hazan et al. (2018) find that trained deep neural networks can be converted to Spiking Neural Networkss (SNNs) and maintain strong image recognition performance, highlighting the competitive potential of SNNs with traditional deep learning methods. Mozafari et al. (2018) train an SNN model by using reinforcement learning, in which the network receives a signal of reward if its decision is correct and punishment if incorrect, therefore motivating it to unlearn previous knowledge if needed. Although reinforcement learning improved the overall performance of the model, Mozafari et al. (2018) find that it is not as competitive as the state-of-the-art deep learning techniques. However, according to Alrebdi et al. (2022), reinforcement learning should be used in practical automation tasks since it improves classification accuracy when integrated in traditional deep learning techniques.

Moreover, as the transformer model was a huge breakthrough in the field of Natural Language Processing (NLP) (Vaswani et al., 2017), it is tried to be implemented in the image recognition and classification discipline as well, especially in combination with CNNs. Finding an effective way to apply transformer models into computer vision and to complement their advantages with advantages of CNNs is still challenging and has become one of the current focus areas (Dai et al., 2021b; Chen et al., 2021). One of these challenges is the large database plane transformer models need to be trained and validated on (Dosovitskiy et al., 2020), while they still fall behind compared to traditional CNNs (Dai et al., 2021b). Such hybrid model of transformer and CNN is the CoAtNet by Dai et al. (2021b), where convolution layers and attention layers get vertically stacked. The result is an effective improvement in generalization, capacity and efficiency.

A more practical attempt of a hybrid model is the so called TransMed, by Dai et al. (2021a), which combines the advantages of CNN and transformer to efficiently extract low-level features of images and establish long-range dependencies between the different categories. This is done by first processing the images as sequences and sending them to the CNN model, afterwards using transformer models to learn the relationship between the sequences and to make predictions (Dai et al., 2021a).

The attempt of applying a pure transformer on image classification is successfully done by Dosovitskiy et al. (2020), who introduce the Vision Transformer (ViT). It achieves comparable results to state-of-the-art CNNs, indicating that transformer models potentially have higher capacity at scale than CNNs (Dosovitskiy et al., 2020). Nevertheless, having a limited training dataset size, ViT still largely lags behind state-of-the-art CNNs.

2.3 Related Research and Work

Yang and Thung (2016) were the first ones to use image classification on waste recycling and compare the results of Support Vector Machines (SVM) with CNN approaches. Their self-made dataset, called TrashNet[2], consists of six categories, paper, glass, plastic, metal, cardboard and trash, forming the foundation for subsequent research. Surprisingly, the SVM model outperforms the CNN, possibly due to its simpler algorithm compared to the

[2]This dataset is utilized in this thesis, among others.

CNN. To enhance the CNN's performance, Yang and Thung (2016) augment the dataset, highlighting the necessity of continuously expanding the data source for a more accurate system. A large and comprehensive dataset is highly important for waste classification, since any object can fall into one of the recycling categories listed, but not every object may be represented in the dataset (Adedeji and Wang, 2019).

In regard to CNNs and small datasets, Shi et al. (2021) argue that traditional CNNs, especially those with complex structures or large networks, are unsuitable for waste image classification tasks involving small datasets. However, with the support of augmentation, Nnamoko et al. (2022) use a deep CNN model and achieve promising results in a binary classification task of recycling, differentiating between organic and recyclable goods. They propose two models that differ in the dataset's image resolution size and reveal notable differences in training time and performance between the large model and small model. While the small model exhibits faster training times and higher accuracy, suggesting suitability for applications on low-cost devices with limited memory, the large model demonstrates potential for generalization despite its memory-intensive nature. Additionally, Zhang et al. (2021a) state that CNNs are feasible to apply on waste classification tasks, as they achieve high accuracy with a DenseNet model. Yet, they extend their DenseNet model with transfer learning as they use a pre-trained model on a large image dataset and fine-tune the model's parameters using a waste image dataset. Nevertheless, their experimentation highlights the effectiveness of the DenseNet image classification model based on transfer learning for waste classification tasks.

Another approach for image-based waste recycling is introduced by Adedeji and Wang (2019), by combining SVM with a pre-trained ResNet model. Further, Ahmad et al. (2020) find that leveraging multiple deep models and fusion techniques are significant for a more reliable waste classification system. They investigate the optimization of waste classification performance through the strategic combination of deep learning models using various fusion techniques. By integrating both early and late fusion methods, a double fusion approach, they demonstrate enhanced waste classification capabilities and at the same time outperform individual fusion methods. Ahmad et al. (2020) utilize genetic algorithm as one of their fusion methods, which is also used by Mao et al. (2021) to optimize the neuron number and the dropout rate of fully-connected-layers of the created DenseNet model. The genetic algorithm combined with data augmentation, horizontal flipping, vertical flipping, and random rotation, clearly improve the accuracy (Mao et al., 2021).

Further, Chu et al. (2018) introduce a Multilayer Hybrid Method (MHS) designed for automatic waste classification. The MHS contains three interconnected subsystems, including an image processing system, a numerical sensor system and a multilayer perceptrons system. By joining the information of image and sensor data, the MHS outperforms traditional CNN models, demonstrating enhanced efficiency and effectiveness in waste classification tasks. Moreover, Shi et al. (2021) propose a combination of MHS and CNN, introducing a Multilayer Hybrid CNN (MLH-CNN) for waste image classification. The MLH-CNN model, characterized by its simplified structure similar to VGGNet but with fewer parameters, surpassing existing methods with a superior classification performance on the TrashNet dataset (Yang and Thung, 2016).

Huang et al. (2021) make use of a deep neural network architecture only based on self-attention mechanisms, a ViT for automatic classification of waste on the TrashNet dataset. By leveraging self-attention processes to dynamically allocate weights, the ViT model overcomes CNN limits in global information processing and achieves an outstanding accuracy rate.

In a more practical manner of waste classification, the quality of images used to train a machine learning model during research might differ greatly to the one in practice, as a high resolution camera and the needed technology are very costly and most facilities are financially limited. Thus, Liu et al. (2024) train a model based on ResNet on degraded images in order to have a strong model that predicts well on bad quality, improving the accuracy for practical use. Moreover, Gondal et al. (2021) translate their work into practice, using a multilayer CNN for real-time waste classification and achieve an accuracy of 99%. To conclude, an comprehensive assessment of various available methods for waste detection and classification using machine learning and deep learning is provided by Abdu and Noor (2022).

3 Materials and Methodology

This section outlines the materials and methodology, covering the summary of the experimental dataset and the strategy used to answer the research questions of this thesis. In order to ensure an adequate comparison of the two chosen models across different areas of complexity, various datasets are considered in this thesis. A summary of the datasets can be found in section 3.1. Further, it was decided to analyse the effect of transfer learning via using BiT and augmentation on the two classification models. As for the classification models, a DenseNet model gets compared to a hybrid approach extending the DenseNet architecture by an attention layer inspired by the transformer model. The reason behind this decision and the strategy for executing the thesis to further address the research questions are outlined in section 3.2. The preparation and prediction of the images and the further analysis was carried out in Python.

3.1 Datasets

Table 1: Datasets Considered in this Thesis and their Characteristics

Dataset	**Number Classes**	**Number Images**	**Classes**
TrashNet[3]	6	2,527	Cardboard, paper, metal, plastic, glass, trash
Household Garbage[4]	12	15,150	Battery, biological, cardboard, clothes, paper, metal, plastic, shoes, trash, white-glass, green-glass, brown-glass
Seven Plastics[5]	8	685	PET, PE-HD, PVC, PE-LD, PP, PS, other resins, no plastic
Household Garbage Distinct	6	6,238	Cardboard, paper, metal, plastic, glass (white-glass, green-glass, brown-glass), trash
Glass	3	2,011	White-glass, green-glass, brown-glass

Having several datasets makes it possible to discover differences in accuracy across them depending on the number of images and classes of the dataset. A summary of the

[3]Source: https://github.com/garythung/trashnet
[4]Source: https://www.kaggle.com/datasets/mostafaabla/garbage-classification
[5]Source: https://www.kaggle.com/datasets/piaoya/plastic-recycling-codes/code

datasets, their sources and categories considered in this thesis are provided in Table 1. In contrast to the large dataset size of the Household Garbage dataset, Seven Plastics is a comparably small dataset with a notably high amount of classes for that size. TrashNet represents a middle ground between these two in terms of dataset size.

The Household Garbage Distinct dataset is a modulation of the Household Garbage dataset, where the categories not listed in TrashNet are excluded and the three glass classes are considered together as one class. The Glass dataset is another modulation of the Household Garbage dataset, only containing the three glass classes. A combination of the datasets is not possible, as the characteristics of images vary a lot and some of the datasets partially contain the same images. All datasets consist of images containing mainly one object each, taken on a white background using various exposure and lighting settings.

3.2 The Models and their Comparison

As already mentioned before, a DenseNet, as a traditional CNN model, and an extension of it utilising an attention layer, a transformer approach, are chosen for the further analysis. The choice of a DenseNet model was motivated by their regularizing effect, which helps to reduce overfitting (Huang et al., 2017). As some of the chosen datasets are quite small, this feature is of good use. For the second model, the DenseNet model is getting extended by an attention layer, further simply referred to as the Dense-Attention Network (DAtNet) model. Since the attention mechanism of the transformer models get an increased popularity in the image classification tasks and even outperform traditional CNN models in some examples (Huang et al., 2021), the DAtNet model represents a worth model to analyse the difference in performance of traditional CNN and new approaching classification models.

Further, the impact of including augmentation and BiT on the accuracy are getting analysed. Augmentation supports the performance, especially when it comes to smaller datasets, while the use of BiT is a cost-effective approach and often improves the performance of a prediction model. Thus, these two extensions represent potential benefits for a waste classification model.

Having the DenseNet and DAtNet model defined, the models get trained and validated on the images of the different datasets. Afterwards, the model predicts the class of each image in the according test set, leading to an overall accuracy result of the model. During the training and validation phase, the model processes the provided dataset several times and forwards information after each run to minimize the possible loss of information during training. While the number of run-throughs is defined by the number of epochs, the passing on of information is specified by the optimizer in a model based on the loss-function. This feedback of performance represents a criteria of success for classification tasks (Chollet, 2021).

In this thesis the key measures to compare the models' performances are the accuracy of the test set, information loss, number of epochs, the recall per class and run-time of the code. The run-time of the code includes the time to preprocess the images through the model itself and how long it takes to train and validate the model. These criteria are particularly significant for waste classification tasks, where precision and efficiency are essential attributes.

The accuracy is defined as the ratio of the number of correctly classified images to the total number of images, as shown in Formula 1. Moreover, the recall, also known as the sensitivity of a class, enables the analysis of the model's prediction performance in each class, revealing insights into potential areas for improvement. It represents the proportion of images that are correctly predicted to the total number of images in a specific class, illustrated in Formula 2. Here, True Positive (TP) and True Negative (TN) reflect the correctly predicted images, while False Positive (FP) and False Negative (FN) represent the incorrectly predicted images.

$$Accuracy = \frac{TP + TF}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The ability to achieve high accuracy and recall values while maintaining a reasonable run-time of the code is important for cost-effective waste classification in especially waste management facilities.

Additionally, the impact of including augmentation and BiT in the model architecture on the key performance measures is examined. By analysing the results of the key performance measures, it becomes possible to detect which model architecture is the most suited for waste classification across varying levels of complexity.

The overall findings get discussed within real world applications and compared to existing studies and their findings. Finally, this research will lead to recommendations for further steps and the implementation of this work in different application areas. The aim is to bridge the gap between research outcomes and practical utilization, enabling the translation of findings into solutions for real-world challenges.

4 Experimental Setup

In this section, the structure of this work's Python code is getting discussed more into detail. The overall layout of the code including the BiT and augmentation is described in section 4.1. The architecture of the DenseNet model is shown in section 4.2 and the later on applied attention layer, resulting in the DAtNet model, is presented in section 4.3.

4.1 Overall Structure

The overall structure of the classification procedure is visualized in Figure 1, where the optional steps are shown as ellipses, and described in the following.

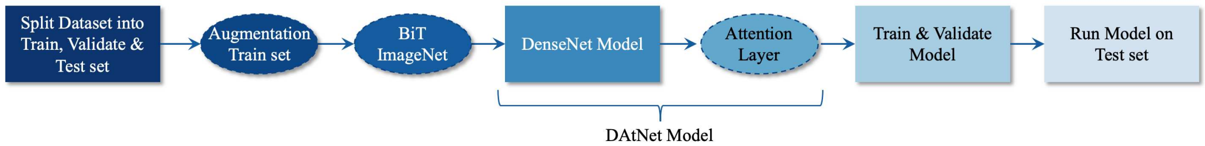

Figure 1: Code Structure Overview

First, the images from the chosen dataset are divided into a train, validation and test set. The distribution of the images into the train, validation and test set depends on the dataset and classification model used. For example, as for a dataset with less images, only a small part of it is used as validation and test set. Accordingly, a higher percentage of images for the test and validation set is chosen in case of a larger dataset. The code for loading and splitting the images in each dataset into the according files is shown in Appendix A.

After dividing the dataset, the images of the train set may get augmented in order to enlarge the dataset, by editing it in terms of zooming in, rotating and changing the lightning on a random basis. The effect of augmentation on an image is illustrated in Figure 2.

Before defining the architecture of the DenseNet model, BiT is utilized. Hence, the model gets pre-trained on a much larger and supervised dataset called ImageNet, which is a large-scale visualization dataset developed to advance the image recognition technology. It includes over 14 million images, covering more than 20,000 classes. Further, if applied, the attention layer allows the DenseNet model to focus on the most important areas of the image, forming the DAtNet model.

Afterwards, the outputted weights of the pre-trained classification model are fine-tuned on the preprocessed train set and validated on the validation set, provided for the classification task. In the end, the classification model is getting evaluated on the test set.

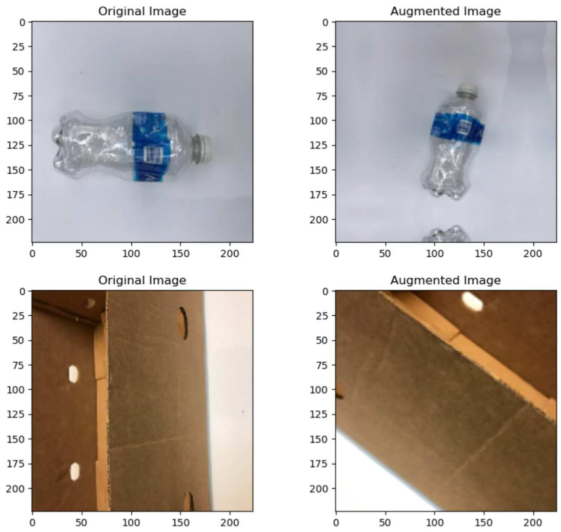

Figure 2: Augmentation Example

4.2 The DenseNet Architecture

The architecture of the DenseNet model is illustrated in Figure 3. The image is loaded into the model with the size 224 x 224 x 3 pixels, where 3 represents the depth, also called the feature maps dimension. First, a convolution block (C) is added, reshaping the image to a 56 x 56 x 64 cube. A dense block (D) then changes the depth to 256, by accessing to its former collective knowledge through their feature maps and adding new information to this collective knowledge, see Formula 3.

$$k_{new} = k_0 + k * l \tag{3}$$

More precisely, k (here $k = 32$) represents the number of feature maps, which include new information and get multiplied by the number of dense layers l (here $l = 6$) within that dense block, increasing the depth of the cube from originally 64 (k_0) to 256 (k_{new}).

After the dense block, a transition block (T) is added, which hovers a 1 x 1 x 128 convolution filter over the cube, leading it to reshape to 56 x 56 x 128. This is followed by a 2 x 2 pooling, dividing the volume of the cube in half. This procedure is repeated until the fourth dense block, leaving a final shape of 7 x 7 x 1024.

Afterwards, global average pooling calculates the average for each feature map, down sampling it to its average value, leading to the shape 1 x 1 x 1024. A dense layer, also called fully connected layer, prepares the data for the final classification and leaves only as many feature maps as there are classes.

$$softmax(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \tag{4}$$

Finally, softmax as an activation function is used, shown in Formula 4, which converts the output feature map values of each class z_i into probabilities. This is done by dividing

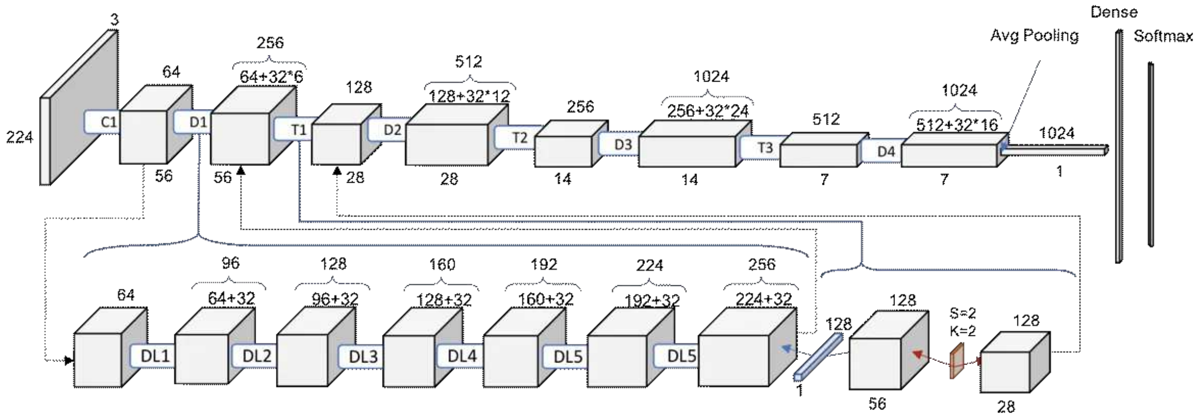

Figure 3: The DenseNet Architecture[6]

the exponent of each feature map value by the sum of the exponents of all feature map values. The final predicted class is the one with the highest probability.

Regarding the forward of information between the epochs, it was decided to use the RMSprop as an optimizer, as it often decreases the time of model training and utilizes the computational resources more effective than other optimizers. Since the waste classification task is a many-class problem, the loss-function used is a categorical crossentropy CE, illustrated in Formula 5. This function quantifies the difference between the model's predicted probability distribution p and the actual label distribution y over the classes. By minimizing this difference, the model learns to generate predictions that are closely aligned with the true labels.

$$CE = -\sum_{j=1}^{N} y_j * log\,(p_j) \tag{5}$$

The code for the creation of the DenseNet model, including the augmentation and BiT definition is shown in Appendix B.

4.3 The Attention Layer

The integration of the attention layer aims to improve the ability to focus on relevant feature maps and regions of each feature map within the input data of the classification model, to enhance the overall classification accuracy. Before the average pooling in the DenseNet architecture, the attention layer gets added on each feature map of the generated cube and extracts the most important feature maps from the input data. Thus, the input data to the attention mechanism is a 7 x 7 x 1024 cube and the output of the DAtNet model maintains the same size accordingly, allowing a smooth integration of it into the existing DenseNet architecture. In basic terms, the attention layer enables the model to

[6]Source: Understanding and Visualizing DenseNets

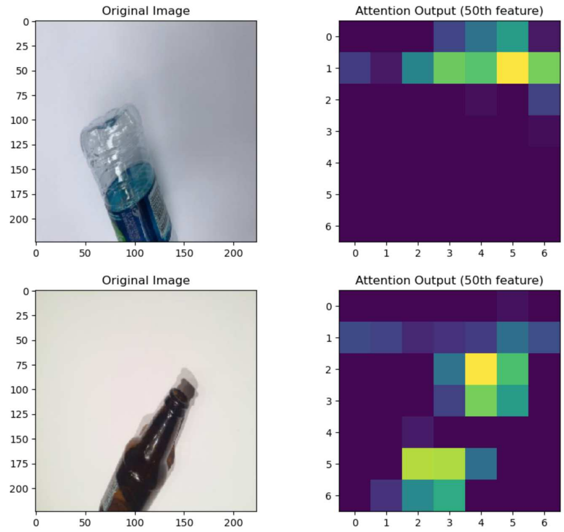

Figure 4: The Attention Layer on the 50th Feature

extract meaningful information from the input data, by identifying dependencies between various feature maps. To illustrate the functionality of the DAtNet model, Figure 4 shows an applied attention layer on the 50th feature map of two chosen images, highlighting the most important regions of that feature map. The code for the attention layer is shown in Appendix C.

It is important to note that since the attention layer is applied after the image was passed through the DenseNet architecture, the image did not retain its original shape, represented by the different scale range in Figure 4. Consequently, it is not unusual for the 50th feature to highlight regions that do not specifically correspond to the areas where the original image contains an object.

5 Findings

In this section, the main findings are summarised. The interpretation of the key measure results are listed in section 5.1, focusing on accuracy, loss, epochs and time. Afterwards, the most typical misclassifications, based on the recall values, and other difficulties are presented in section 5.2.

5.1 Key Measures and Model Architecture

As the DenseNet model without using BiT, already performs much worse than with the BiT, it was decided to always include it in order to achieve the best results. More into detail, without the BiT, the DenseNet model achieves an accuracy of 52.2% in the TrashNet dataset, around 40% worse than with it.

The final accuracy results are listed in Table 2, with the results for the models that include augmentation shown in parentheses. Regarding the prediction of the Seven Plastics dataset, the uploaded images have a size of 400 x 400 x 3 instead of 224 x 224 x 3 to increase the precision in the DenseNet model, since the dataset size is comparably low. However, as for time and computational resource limits, it was not possible to use the size 400 x 400 x 3 for the DAtNet model.

Table 2: Test Accuracy, Loss and Epochs per Dataset and Model

	Accuracy [%]		**Loss**		**Epochs**	
Dataset	DenseNet	DAtNet	DenseNet	DAtNet	DenseNet	DAtNet
TrashNet	87.0 (88.1)	90.3 (91.1)	0.450 (0.419)	0.398 (0.357)	45	25
Household Garbage	95.6 (95.2)	96.1 (96.8)	0.168 (0.177)	0.102 (0.095)	20	20
Seven Plastics	49.0 (52.9)	37.3 (37.3)	1.648 (1.486)	1.845 (1.865)	50	30
Household Garbage Distinct	92.6 (92.0)	93.5 (92.0)	0.188 (0.184)	0.204 (0.194)	20	15
Glass	98.7 (98.2)	98.2 (97.8)	0.091 (0.104)	0.102 (0.112)	10	10

For the TrashNet dataset it was discovered that the augmentation has a great impact on the test accuracy of the DenseNet model, increasing it by 1.1 percentage points and further helping the DenseNet model to prevent overfitting. It had the same effect on the DAtNet model, where it enhances the accuracy of the test set by 0.8 percentage points. As for the Household Garbage dataset, the DAtNet model performs better including the augmentation in both accuracy and loss of information. However, it has no improving impact on the DenseNet model. As for the Seven Plastics dataset, the augmentation leads the DenseNet model to increase its accuracy by around 4 percentage points, but has no effect on the DAtNet model. In case of the Household Garbage distinct dataset, the augmentation leads to a decrease in accuracy in even both, the DenseNet and the DAtNet model. Nevertheless, it prevents overfitting, which is shown in Figure 5 for the DenseNet model. The Figure illustrates the accuracy during the training and validation phases over the epochs, differentiating between the inclusion and exclusion of augmentation. Without augmenting the images, the train accuracy *Train Acc* exceeds the validation accuracy *Val Acc* after approximately 11 epochs, indicating overfitting. Including the augmentation in the model, the train accuracy stays bellow the validation accuracy, preventing overfitting. However, for the Glass dataset, the accuracy decreases by around 0.5 percentage points and the loss of information increases for both classification models when including augmentation. To conclude, the augmentation leads to an increase of accuracy in smaller datasets with especially a higher amount of classes and has the opposite impact on the accuracy for larger datasets, even leading to worse results.

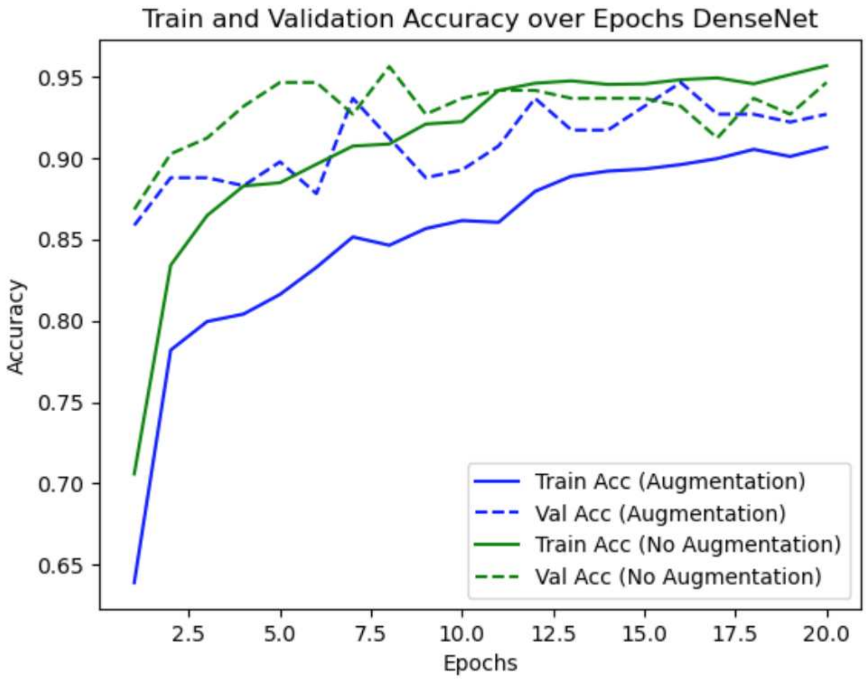

Figure 5: Train and Validation Accuracy of the DenseNet Model on the Household Garbage Distinct Dataset

As for the loss of information during the test phase, the loss is lower the bigger the dataset size for the DAtNet model. Further, including augmentation in the DAtNet model leads to a lower loss in nearly every dataset, except in the smaller datasets, Seven Plastics and Glass dataset. Thus, augmentation has no significant impact on the loss of the DAtNet model when it comes to smaller dataset sizes. The DenseNet model achieves a lower loss of information in larger dataset sizes and datasets with less classes. Including augmentation helps the model to decrease the loss in every dataset, except the Household Garbage and Glass dataset. Thus, the loss of a DenseNet model is lower when including augmentation, besides the utilization of larger datasets, leading to an increase of loss.

Comparing the best accuracy results of the two models in the test set, it can be observed that the DAtNet model achieves a higher accuracy in the TrashNet dataset, with less epochs to run than the DenseNet model. Additionally, the DenseNet model performs nearly as good as the DAtNet model, but has a higher loss of information in the Household Garbage dataset. Nevertheless, having a smaller dataset size in the Seven Plastics dataset, the DAtNet model achieves a noticeable lower accuracy than the DenseNet model and a higher loss of information. Additionally, the DenseNet model achieves a slightly higher accuracy in the Glass dataset, with around the same amount of loss of information as the DAtNet model. In general it can be said that the DAtNet model performs better in larger datasets, but has huge difficulties with smaller datasets, where it performs much worse than the DenseNet model. Further, the DenseNet model demonstrates comparable performance to the DAtNet model, if the dataset contains approximately 650 images per class.

Moreover, it was discovered that although the DAtNet model needs less epochs to achieve the same results, it needs much more time to process the images through the attention layer. The exact values can be observed in Table 3, where s indicates seconds and ms denotes milliseconds. For the DenseNet model, an outlier in preprocessing time of the Seven Plastics dataset can be observed. The image size in this dataset, which is almost double the size of the standard image size used in this thesis, leads to a significant increase in preprocessing time. Therefore, the observed outlier in preprocessing time of the DenseNet model is expected under these circumstances. However, the appearance of the outlier in the train and validation time of the DAtNet model in the Household Garbage distinct dataset is unclear. Possibly, the device running the code was simultaneously used for other applications as well, though this theory is unverified.

Additionally, it can be observed that the DAtNet model on average needs approximately 100 milliseconds longer for preprocessing each image compared to the DenseNet model, despite the inclusion of the outlier in the Seven Plastics dataset. However, compared to the DenseNet model, the train and validation time of the DAtNet model appears noticeably faster in the TrashNet and Household Garbage datasets, while showing no specific pattern in the distribution of the values. In contrast, the time per epoch of the DenseNet model seems to be dependent on the dataset size, requiring less time per epoch the smaller the size of the dataset. Although the DenseNet outpaces the DAtNet model in three out of five cases, on average, the DAtNet appears to be twice as fast.

Table 3: Preprocessing, Train and Validation Time per Dataset and Model

	Preprocessing Time [ms/Image]		**Train and Validation Time [s/epoch]**	
Dataset	DenseNet	DAtNet	DenseNet	DAtNet
TrashNet	80.37	207.67	6.53	1.80
Household Garbage	83.47	210.50	30.07	1.64
Seven Plastics	241.16	177.77	0.24	1.59
Household Garbage Distinct	89.39	237.24	13.60	15.50
Glass	75.98	226.31	3.25	3.36
Average	114.07	211.90	10.74	4.78

5.2 Typical Misclassifications and Difficulties

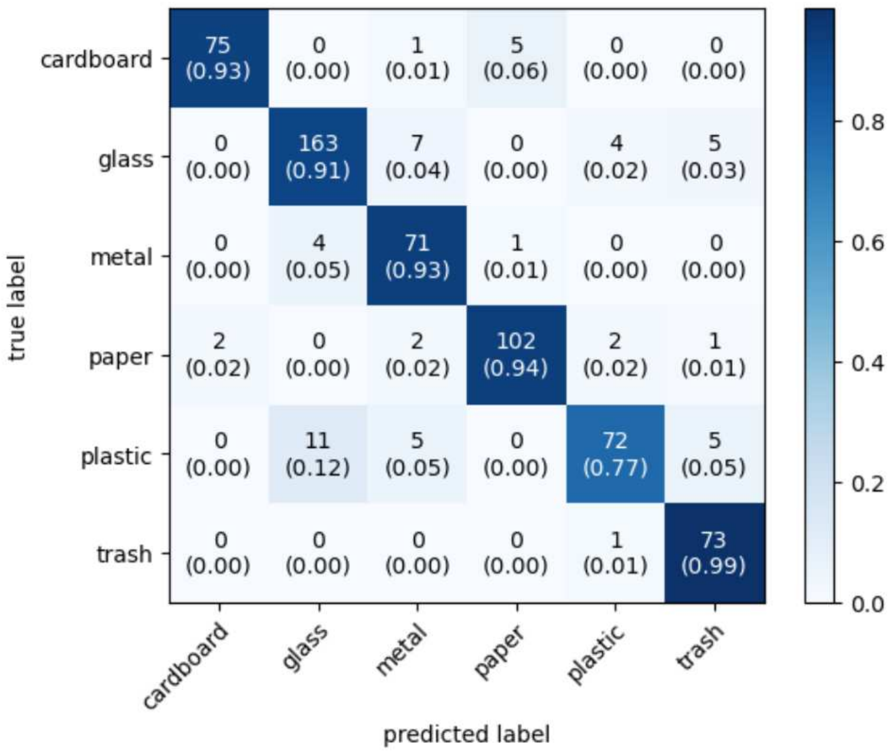

Figure 6: Confusion Matrix of the DenseNet Model on the Household Garbage Distinct Dataset

Depending on the dataset, it can be observed that the models have difficulties in predicting specific classes. This observation underscores the importance of analysing and gaining a deeper understanding of the models' performance across different classes to identify possible pitfalls and to avoid them in future work. For example, in Figure 6, the results of the DenseNet model on the Household Garbage Distinct dataset are visualized, showing the number of images in each category and their percentage share in parenthesis. In this matrix, the predicted labels (horizontal) are compared to the true labels (vertical), while the blue legend represents the level of percentage share of the predicted class. Hence, the darker the color, the higher the percentage share. The images that are correctly classified are represented by the diagonal values, having the recall values as the percentage share, whereas the values outside the diagonal represent misclassified images.

While the DenseNet model achieves a high recall value across most of the classes in the Household Garbage Distinct dataset, it notably struggles in accurately predicting the plastic class. Specifically, the recall only equals to 77% in this class, where plastic is misclassified as glass in 12% of the cases. Comparably, the DAtNet model predicts 11% of plastic images as glass. A reasonable explanation is the large size of the glass folder, which contains more images than the other classes, since it is a combination of the three classes green-glass, brown-glass and white-glass. This specific outlier in both models highlights a specific area where the models' performance falls short, emphasizing the need for further investigation and the potential for improvement of the model and dataset. Figure 7 illustrates some examples of wrong predicted images by the DenseNet and DAtNet model, along with their corresponding true and predicted class.

Figure 7: Misclassified Images

The opposite effect was observed in the Household Garbage dataset, where the DenseNet model predicts 6% of the white-glass images as plastic. Similarly, the DAtNet model predicts white-glass as plastic in 9% of the cases. However, as for the huge size of the Household Garbage dataset, no significant outliers were discovered, the same applies for both models in the Glass dataset.

As for the TrashNet dataset, the DenseNet model classifies trash as paper in 33% of the cases. However, it is worth mentioning that the number of trash images is comparably smaller than in other classes, possibly leading to this tendency. Further, the DenseNet model encounters challenges in separating metal and glass, resulting in a misclassification in around 10% of the cases. Besides that, the DenseNet model experiences similar misclassification tendencies as experienced in the Household Garbage Distinct dataset, predicting 16% of the plastic images as glass. The TrashNet dataset also reveals difficulties of the DAtNet model, which predicts 9% of the cardboard images as paper, along with 7% of glass images misclassified as plastic.

It is worth mentioning that the inclusion of transparent goods in both the glass and plastic classes and the similarities between cardboard and paper validate the models' misclassifications. However, they discover potential improvements of the models in these areas.

In the context of the Seven Plastics dataset, the DAtNet model has the issue of predicting every image either as PET or PP. This is not a surprise, considering the larger number of images in these two classes within the overall dataset and consequently in the training, testing, and validation sets. Surprisingly, the DenseNet model follows the same pattern. However, the DenseNet model was able to recognise every image that was no plastic, which in the end, resulting in the significant difference in the final test accuracy. The confusion matrix of the DenseNet model on the Seven Plastics dataset is shown in Figure 8.

The comparison of Figure 6 with Figure 8, reveals the noticeable difference in predictive

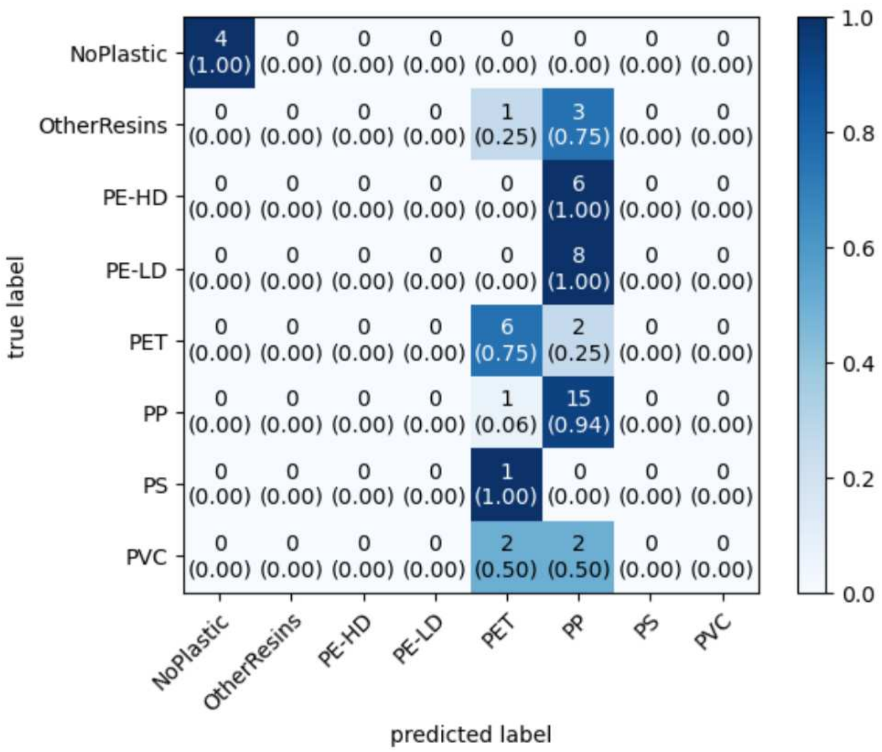

Figure 8: Confusion Matrix of the DenseNet Model on the Seven Plastics Dataset

performance. Figure 6 shows an almost consistent high level of the recall values, evidenced by the presence of the crossed line across the classes. In contrast, Figure 8 lacks this characteristic, indicating less reliable predictive capabilities.

Besides the typical misclassifications of the models, it was observed that the integration of the attention layer into the classification model introduces significant image preprocessing challenges in especially larger datasets. In that regard, it requires considerable time and computational resources, leading to time constraints and storage limitations and therefore errors, hindering the seamless execution of the classification tasks. These challenges underscore the importance of considering computational efficiency and resource utilization when implementing complex architectures such as the DAtNet model in practical applications.

6 Discussion

In this section, the findings of this thesis are discussed more into detail and compared with similar works and their findings. This way, possible extensions and improvements of this work can be identified to avoid typical pitfalls, which are mentioned in 5.2, and to further improve the performance. While transfer learning and augmentation are discussed in 6.1, the models' performance is compared to findings of other studies in 6.2. The discussion of the datasets utilized in this thesis takes place in section 6.3.

6.1 Transfer Learning and Augmentation

Similar as in this thesis, Huang et al. (2021) claim that transfer learning is very effective and even essential in the construction of a waste classification model. Furthermore, Zhang et al. (2021a) use transfer learning combined with a DenseNet model to predict waste. While their accuracy increases by 73 % by using transfer learning, in this thesis it was only around 40%. They achieve an overall test accuracy of 82.8% on a large dataset, having different kinds of image backgrounds, making it more complicated than the ones used in this thesis. Their dataset is not available online and is closer to, but does not represent, the scene of waste classification in real world.

In regards of augmentation, in this thesis it was found to be beneficial for especially smaller datasets and to prevent overfitting, which aligns with the findings of Krizhevsky et al. (2012). The concerns about augmentation addressed by DeVries and Taylor (2017), regarding the accurate construction of augmented data to meet naturally occuring transitions in that specific class, is ensured in this thesis. This can be observed in Figure 2, where the augmented images retained the characteristics of the original class, while enhancing accuracy.

Furthermore, the misclassification problem is something that can be addressed with augmentation in future work. More precisely, with the MSDA method introduced by Liang et al. (2023), features from two different classes are combined into one new generated image, based on a certain mixing method. Subsequently, the class label of the generated image is a combination of the two original class labels. Liang et al. (2023) are able to achieve outstanding results with this strategy, helping the model to prevent overfitting and to better generalize on unseen data.

6.2 Prediction Performance

As for the prediction performance of the models in this thesis, the DenseNet model surpasses the accuracy achieved by the authors of the TrashNet dataset (Yang and Thung, 2016). However, as they do not include transfer learning in their SVM model, they achieve around 10 percentage points higher than the DenseNet model in this thesis without transfer learning. Thus, it is worth exploring whether a SVM model would outperform the DenseNet with transfer learning. Further, similarly to the findings in this thesis, they encounter difficulties in predicting the trash class.

Concerning traditional classification models, Huang et al. (2021) state that these models often achieve low accuracy values, limiting their real-world application. Therefore, investigating more into transformer, especially attention layers, in future work might enhance the application of machine learning models in practice.

In this thesis, the DAtNet model outperforms the DenseNet model in larger datasets, but struggles with smaller datasets, a difficulty already observed by Dosovitskiy et al. (2020). Similarly, Dai et al. (2021a) include attention layers in a DenseNet architecture for medical image datasets, achieving a higher accuracy with them, while encountering challenges with smaller datasets. Additionally, despite its faster training time, the DAtNet needs significantly more preprocessing time and storage resources, resulting in challenges in larger datasets. Equally to these findings, the ViT model of Huang et al. (2021) needs on average 100 milliseconds longer than traditional CNN models to preprocess an image, emphasizing the need to further optimize the transformer methods.

Further, Huang et al. (2021) find that their ViT, which focuses on attention layers, increases the final accuracy, even outperforming the introduced DAtNet model by around 5 percentage points. However, despite their code availability, it was not possible to replicate their results.

6.3 The Perfect Dataset

All the literature utilized in this thesis include the discussion of the lack in dataset quality for waste classification. The importance of large and comprehensive datasets in waste management cannot be overstated, as a dataset that reflects the actual footage in practice is essential for accurate waste classification. Hence, there is a critical need for dataset improvement, through investigating into the sources and quality of waste images (Adedeji and Wang, 2019; Yang and Thung, 2016; Zhang et al., 2021a).

As for the datasets used in this thesis, they mostly had a single white background and only included one object per image. This does not apply to practical utilization, as waste management facilities are mostly out-of-date, with few reliable records and scarce sensory data, particularly in developing countries (Abdu and Noor, 2022). Thus, considering the financial limitations of this sector, which affects the technology procurement, future work should encounter the inclusion of image quality degradation (Liu et al., 2024).

Further, the insufficient data availability leads to small sample sizes and uneven distributions among classes of waste. This and the simple design of the images lead to a low generalization ability of the classification models (Abdu and Noor, 2022; Zhang et al., 2021a). The importance of a balanced distribution of images across classes was observed in this work as well, demonstrated by the issue in the Seven Plastics dataset, where at most only three out of eight classes are getting predicted by both classification models.

Besides the degradation and distribution of the images, it is necessary to consider object detection and segmentation methods for real world application. These methods are further beneficial for the use of video-based detection systems and for the adaption to the specific requirements of waste management facilities.

Another step towards reforming waste classification is the inclusion of material detection. As the recall values of the classes reveal that misclassifications often appear due to the model's inability to distinguish waste materials effectively, integrating advanced technologies such as laser-based material detection could lead to an increase in classification performance. Such approach is done by Chu et al. (2018), who join the information of image and sensor data, introducing the MHS.

To conclude, the available datasets for waste classification lack quality and usage in real world applications. Thus, the addressed challenges regarding the uneven distribution of images among classes, the single backgrounds and the financial impact on the degradation require comprehensive solutions. Developing datasets based on these criteria and the integration of advanced technologies could further enhance the accurate prediction of waste. Such developments are video classification, object detection, object segmentation and laser-based material detection.

7 Conclusion and Further Steps

In this thesis, the performance and implications of two different machine learning models, specifically a DenseNet, as a traditional CNN model, and an extension of it by an attention layer, called DAtNet model, were analyzed in the context of image-based waste recycling. The aim was to answer the research questions concerning the current differences in performance of the two image classification models and their impact on the future of waste management. The findings and recommendations outlined in this section could lead the way for more efficient and environmentally friendly waste management practices.

It was observed that using transfer learning to pre-train the model leads to a high increase in accuracy, thus being essential for a classification model. In most cases, including augmentation reduces overfitting, however, it does not always positively impact the test accuracy. Especially in larger datasets, augmentation leads to a decrease in accuracy. Regarding the two models that were compared, it was observed that despite the enhanced performance of the DAtNet model in larger datasets, the DenseNet model significantly outperforms the DAtNet model when dealing with smaller datasets, indicating that the DAtNet model lacks generalization capability. Additionally, the DenseNet model achieves similar results with a dataset that includes around 650 images per class. Even more important is the measure of time, as the computational costs depend on it. The findings showed that the DAtNet is on average double as fast in the train and validation phase, but at the same time needs twice the time to preprocess an image than the DenseNet model. As each image has to be preprocessed by the model, this phase is even more important, letting the DAtNet model loose reliability for waste classification.

To answer the research questions more precisely, although extending a traditional CNN model with an attention layer improves the test accuracy in some cases, it does not surpass the CNN model's classification reliability. Therefore, especially for time and complexity reasons, it is recommended to prioritize the utilization of DenseNet models for image-based waste recycling in waste management facilities. Nevertheless, the inclusion of transformer, especially attention layers, in image classification hold potential for future work. Thus, overcoming their common challenges discussed in this thesis could lead to an enhancement in classification performance.

Moreover, while it was crucial to include several datasets in this thesis to evaluate the models' generalization capabilities, the quality of available datasets for waste classification represents an overall challenge and requires significant improvements, as discussed in section 6.3. The upcoming datasets should be more practical, featuring realistic backgrounds and multiple objects per image. To achieve this, a partnership with waste management facilities could be established to receive more practice-oriented datasets.

In general, the implementation of classification models in waste management facilities holds significant potential, especially in developing countries. By introducing machine learning to automate the waste sorting process, human contact with waste and potentially harmful materials can be reduced. Further, the adoption of this technology could enhance operational efficiency and enhance the transparency of the recycling status in the specific city.

Additionally, in regions where waste recycling practices are not very common, educational initiatives could play a crucial role in promoting sustainable waste management. One such initiative could be the development of an interactive application that uses the classification model to identify different types of waste. This application could serve as an engaging educational tool, even gamified for younger audiences, to increase the awareness and understanding of waste separation and recycling. An example of how this could work is illustrated in Figure 9, where a website was created that allows users to upload an image and receive a prediction of the waste type. This approach might not only educate but also encourage citizens to actively participate in waste management, leading the way for a more sustainable future.

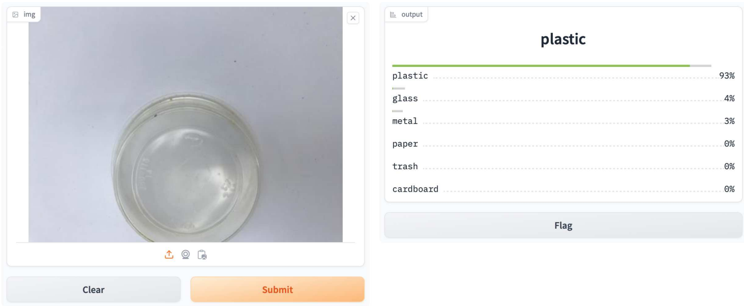

Figure 9: Prediction Example of a Piece of Plastic[7]

Another step towards a more innovative and transparent waste management system could be done by the implementation of smart home bins, which sort the type of waste automatically. These intelligent bins offer a direct approach to waste management at the source, potentially revolutionizing how communities engage with recycling. For example, a monthly report detailing the amount of recycled waste and its environmental impact of a household could foster awareness and encourage citizen participation in a zero waste city. Building upon previous initiatives such as Donovan (2016), which focuses on binary classification, the here proposed system would be more complex, potentially involving five or six classes, offering citizens a deeper understanding of their waste habits. However, it is worth considering whether a smart home bin truly requires six classes.

Moreover, the implementation of such bins presents an opportunity to enhance transparency within waste management facilities, as the data collected from these bins can enable them to optimize their operations. This can be done by minimizing unnecessary

[7]Please note that this is a temporary website created for this thesis' purpose only, therefore no source can be named.

drives and maximizing efficiency in waste collection routes, through the estimation of the volume of waste produced. Thus, the integration of smart home bins holds great promise for predicting and managing waste production in a more sustainable and efficient manner. In this context, extending the DenseNet model with an attention layer to increase the accuracy might be a better option, as the budget for training and implementation may vary depending on the product's premium level.

The further steps discussed in this thesis, based on the automation of waste recycling using machine learning models, significantly contribute to several Sustainable Development Goals (SDGs)[8], especially to the SDG 12. This goal aims to ensure sustainable consumption and production patterns, which are critical for reducing the environmental footprint and promoting resource efficiency. Through the implementation of machine learning models, the amount of waste sent to landfills can be reduced and the recovery of recyclable materials can be increased. Notably, target 12.5 gets supported, which aims to reduce waste generation through prevention, reduction, recycling, and reuse. Additionally, by improving education and public awareness of sustainable waste management, a more sustainable consumer behavior is encouraged. Thus, aligning with target 12.8 and supporting the SDG 13, which try to ensure that all humans have the relevant information and awareness for sustainable development.

To conclude, this thesis could foster the improvement of waste management, especially in developing countries, by making use of its results in order to increase awareness and education for waste sorting purposes. Furthermore, the development of a smart home bin, as an extension of this thesis, could be a significant step towards achieving zero waste cities. Hence, brightening the prospects of future generations and underscoring the importance of sustainability in our everyday lives. Thus, this thesis could support the way towards a more sustainable future.

[8]See: https://sdgs.un.org/goals

A Loading and Splitting the Images

In order to load and split the images into the according test, validation and train set, a file called *prepare_dataset* was created. Thus, it was possible to execute the code for each dataset, without the need to make an extensive amount of manual adjustments. The code included in the file is shown bellow.

```python
def create_excel(base_dir, dictionary={'cardboard':0, 'glass':1, 'metal':2, 'paper':3, 'plastic':4, 'trash':5}):
    # Initialize an empty list to store image and folder information
    data = []

    # Iterate through each folder in the dataset
    for folder_name in os.listdir(base_dir):
        folder_path = os.path.join(base_dir, folder_name)
        if os.path.isdir(folder_path):
            # List all image files in the current folder
            image_files = [f for f in os.listdir(folder_path) if f.lower().endswith(('.jpg', '.png', '.jpeg'))]
            # Append image names and folder names to the list
            for image_name in image_files:
                data.append({'image': image_name, 'category': folder_name})

    # Create a DataFrame from the list
    labels = pd.DataFrame(data)

    # Save the DataFrame to a CSV file
    labels.to_csv('labels.csv', index=False)
    print("CSV created")

    labels.category = labels.category.map(dictionary)
    return labels

def subdirectories(base_dir = "./waste_split", categories = ['cardboard', 'glass', 'metal', 'paper', 'plastic', 'trash']):

    train_dir = os.path.join(base_dir, 'Train')
    test_dir = os.path.join(base_dir, 'Test')
    validate_dir = os.path.join(base_dir, 'Val')
    train_test_val = [train_dir, test_dir, validate_dir]

    for directory in [base_dir, train_dir, test_dir, validate_dir]:
        if not os.path.exists(directory):
            os.mkdir(directory)
            print(directory, "created")

```

```python
    directories = []
    subdirectories = {}

    # Loop through each subdirectory (train, test, validate)
    for subdirectory in [train_dir, test_dir, validate_dir]:
        for category in categories:
            category_path = os.path.join(subdirectory, category)
            if not os.path.exists(category_path):
                os.mkdir(category_path)
                directories.append(category_path)
                print(f"{category_path} created")

    # Initialize an empty dictionary to store all directories for each category
    subdirectories = {}

    # Iterate through each category
    for category in categories:
        # Filter directories containing the current category
        category_dirs = [d for d in directories if category in d]
        subdirectories[category] = category_dirs
    return [train_test_val, subdirectories]

# get every image to the right directory (train, test, validate)
def copy_images(original_dataset_dir, subdirectories, category_codes, labels, list_images):
    # original_dataset_dir: where the data comes from. i.e. "dataset"
    # subdirectories: created in subdirectories()
    # category_codes: the names according to the numbers in the list (should be a dictionary)
    # labels: the table created in create_excel()
    # list_images: three tables which contain images from each train, test and validation

    # Loop through each subdirectory (train, test, validate)
    for ix, image in enumerate(list_images):
        category_numimages = {}
        for i in category_codes.keys():
            category_numimages[i]=0

        for file in image:
            category_code = labels.loc[labels['image'] == file]["category"].values[0]
            category_name = next(key for key, value in category_codes.items() if value == category_code)

            category_numimages[category_name]+=1
```

```

                src = os.path.join(original_dataset_dir, category_name,
                    file) # source
                dst = os.path.join(subdirectories[category_name][ix], file)
                    # destintaion
                shutil.copyfile(src, dst)

        # print out the number of files for each category
        for category in category_numimages:
            print(category_numimages[category], f"{category} copied
                to:", subdirectories[category][ix])
    return 'FINISHED'
```

An example of how the code of the prepare_dataset file can be imported and used is outlined in the following.

```
import prepare_dataset as pre_dat

dictionary = {'cardboard':0, 'glass':1, 'metal':2, 'paper':3,
    'plastic':4, 'trash':5}
dataset_dir = "dataset/"
labels = pre_dat.create_excel(dataset_dir,dictionary)
categories = dictionary.keys()
directories = pre_dat.subdirectories("test_split_1", categories)
train_dir = directories[0][0]
test_dir = directories[0][1]
validate_dir = directories[0][2]
subdirectories = directories[1]
original_dataset_dir = "./dataset"
images = os.listdir(original_dataset_dir)
images.remove(".DS_Store")

# taking 90% as train data from labels list
x_train, x_test1, y_train, y_test1 = train_test_split(labels["image"],
    labels["category"] ,test_size=0.10, random_state=42)

# from 10% data, splitting 70% as test data and 30% as validation data
x_test, x_val, y_test, y_val = train_test_split(x_test1, y_test1
    ,test_size=0.3, random_state=42)

# creating a dataframe containing image names and their correct values
train_df=pd.DataFrame({'image':x_train.values,
    'category':y_train.values})
val_df=pd.DataFrame({'image':x_val.values, 'category':y_val.values})
test_df=pd.DataFrame({'image':x_test.values, 'category':y_test.values})

pre_dat.copy_images(original_dataset_dir, subdirectories, dictionary,
    labels, [x_train, x_test, x_val])
```

B Augmentation, BiT and DenseNet Code

```
# Including transfer learning: using a pretrained DenseNet model
conv_base = DenseNet121(weights = "imagenet", include_top=False,
    input_shape=(224, 224, 3))

def get_features_and_labels(dataset):
    all_features = []
    all_labels = []
    for images, labels in dataset:
        preprocessed_images =
            keras.applications.densenet.preprocess_input(images)
        features = conv_base.predict(preprocessed_images)
        all_features.append(features)
        all_labels.append(labels)
    return np.concatenate(all_features), np.concatenate(all_labels)

train_features, train_labels = get_features_and_labels(train_dataset)
val_features, val_labels = get_features_and_labels(validation_dataset)
test_features, test_labels = get_features_and_labels(test_dataset)

from tensorflow.keras.utils import to_categorical

train_labels = to_categorical(train_labels, num_classes=6)
val_labels = to_categorical(val_labels, num_classes=6)
test_labels = to_categorical(test_labels, num_classes=6)

# Code for the augmentation:
data_augmentation = keras.Sequential([
        layers.RandomFlip("horizontal_and_vertical"),
        layers.RandomRotation(0.80),
        layers.RandomZoom(0.4),
        layers.RandomContrast(factor=0.5),])

inputs = keras.Input(shape=(7, 7, 1024))
x = data_augmentation(inputs) # Here, augmentation is included
x = layers.Flatten()(x)
x = layers.Dense(256, activation='relu')(x)
x = layers.Dropout(0.5)(x)

outputs = layers.Dense(6, activation="softmax")(x)
model = keras.Model(inputs, outputs)

model.compile(loss=losses.CategoricalCrossentropy(),
              optimizer='Adam',
              metrics=['acc'])
```

```

callbacks = [
        keras.callbacks.ModelCheckpoint(
        filepath="DenseNet121_TrashNet_model_a.keras",
        save_best_only=True,
        monitor="val_loss")]

history = model.fit(train_features, train_labels,
                    epochs=45,
                    batch_size=64,
                    validation_data=(val_features, val_labels),
                   callbacks=callbacks)

# Evaluating the model on the test set
test_model = keras.models.load_model("DenseNet121_TrashNet_model.keras")
test_loss, test_acc = test_model.evaluate(test_features, test_labels)
print(f"Test accuracy: {test_acc:.3f}")
```

C Attention Layer and DAtNet Code

Defining the layer:

```
class ChannelGate(layers.Layer):
    def __init__(self, gate_channels, reduction_ratio=7, **kwargs):
        super(ChannelGate, self).__init__(**kwargs)
        self.gate_channels = gate_channels
        self.reduction_ratio = reduction_ratio
        self.mlp = keras.Sequential([
            layers.Dense(gate_channels // reduction_ratio,
                activation='relu'),
            layers.Dense(gate_channels)
        ])

    def call(self, inputs):
        avg_pool = tf.reduce_mean(inputs, axis=[1, 2])
        channel_att_raw = self.mlp(avg_pool)
        channel_att = tf.nn.sigmoid(channel_att_raw)
        channel_att = tf.reshape(channel_att, [-1, 1, 1,
            self.gate_channels])
        return inputs * channel_att

class SpatialGate(layers.Layer):
    def __init__(self, gate_channels, **kwargs):
        super(SpatialGate, self).__init__(**kwargs)
        self.spatial = keras.Sequential([
            layers.Conv2D(1, kernel_size=17, padding='same'),
            layers.Activation('sigmoid')
        ])

    def call(self, inputs):
        scale = self.spatial(inputs)
        return inputs * scale

class AttentionModule(layers.Layer):
    def __init__(self, gate_channels, reduction_ratio=20, **kwargs):
        super(AttentionModule, self).__init__(**kwargs)
        self.ChannelGate = ChannelGate(gate_channels, reduction_ratio)
        self.SpatialGate = SpatialGate(gate_channels)

    def call(self, inputs):
        x_out = self.ChannelGate(inputs) + self.SpatialGate(inputs)
        return x_out
```

How the code was implemented:

```
# this definition is "nearly" the same as for the DenseNet, BUT:
def get_features_and_labels(dataset, conv_base, attention_module):
```

```python
    all_features = []
    all_labels = []
    for images, labels in dataset:
        preprocessed_images = keras.applications.densenet.preprocess_input(images)
        base_features = conv_base(preprocessed_images)
        # Here, the attention layer gets added
        features = attention_module(base_features)
        all_features.append(features)
        all_labels.append(labels)
    return np.concatenate(all_features), np.concatenate(all_labels)

# Initialize conv_base and attention_module
attention_module = AttentionModule(gate_channels=1024)  # Assuming the output of conv_base has 1024 channels

# Get features and labels
train_features, train_labels = get_features_and_labels(train_dataset, conv_base, attention_module)
val_features, val_labels = get_features_and_labels(validation_dataset, conv_base, attention_module)
test_features, test_labels = get_features_and_labels(test_dataset, conv_base, attention_module)
```

References

H. Abdu and M. H. M. Noor. A survey on waste detection and classification using deep learning. *IEEE Access*, 10, 2022. ISSN 21693536. doi: 10.1109/ACCESS.2022.3226682.

O. Adedeji and Z. Wang. Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manufacturing*, 35:607–612, 2019. ISSN 23519789. doi: 10.1016/j.promfg.2019.05.086. URL `https://linkinghub.elsevier.com/retrieve/pii/S2351978919307231`.

K. Ahmad, K. Khan, and A. Al-Fuqaha. Intelligent fusion of deep features for improved waste classification. *IEEE Access*, 8, 2020. ISSN 21693536. doi: 10.1109/ACCESS.2020.2995681.

S. O. Ajwang, J. O. Abila, and I. Tejero-Dakay. Adoption conceptual model for intelligent waste management in smart cities: Theoretical review. *International Journal of Computing Sciences Research*, 5:426–440, 1 2021. ISSN 25460552. doi: 10.25147/ijcsr.2017.001.1.50.

N. Akbarpour, A. Salehi-Amiri, M. Hajiaghaei-Keshteli, and D. Oliva. An innovative waste management system in a smart city under stochastic optimization using vehicle routing problem. *Soft Computing*, 25:6707–6727, 4 2021. ISSN 1432-7643. doi: 10.1007/s00500-021-05669-6. URL `http://link.springer.com/10.1007/s00500-021-05669-6`.

N. Alrebdi, S. Alrumiah, A. Almansour, and M. Rassam. Reinforcement learning in image classification: A review. pages 79–86. IEEE, 1 2022. ISBN 978-1-6654-3605-2. doi: 10.1109/ICCIT52419.2022.9711620.

L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao. Review of image classification algorithms based on convolutional neural networks, 2021. ISSN 20724292.

X. Chen. Machine learning approach for a circular economy with waste recycling in smart cities, 2022. ISSN 23524847.

Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng. Dual path networks. 7 2017. URL `http://arxiv.org/abs/1707.01629`.

E. Chioatto and P. Sospiro. Transition from waste management to circular economy: the european union roadmap. *Environment, Development and Sustainability*, 25:249–276, 1 2023. ISSN 1387-585X. doi: 10.1007/s10668-021-02050-3. URL `https://link.springer.com/10.1007/s10668-021-02050-3`.

E. Chioatto, M. A. Khan, and P. Sospiro. Sustainable solid waste management in the european union: Four countries regional analysis. *Sustainable Chemistry and Pharmacy*, 33, 2023. ISSN 23525541. doi: 10.1016/j.scp.2023.101037.

F. Chollet. *Deep Learning with Python, Second Edition*. Manning Publications Co., 2 edition, 2021. ISBN 9781617296864.

Y. Chu, C. Huang, X. Xie, B. Tan, S. Kamal, and X. Xiong. Multilayer hybrid deep-learning method for waste classification and recycling. *Computational Intelligence and Neuroscience*, 2018:1–9, 11 2018. ISSN 1687-5265. doi: 10.1155/2018/5060857.

Y. Dai, Y. Gao, and F. Liu. Transmed: Transformers advance multi-modal medical image classification. *Diagnostics*, 11, 2021a. ISSN 20754418. doi: 10.3390/diagnostics11081384.

Z. Dai, H. Liu, Q. V. Le, and M. Tan. Coatnet: Marrying convolution and attention for all data sizes. 6 2021b. URL http://arxiv.org/abs/2106.04803.

T. DeVries and G. W. Taylor. Dataset augmentation in feature space. 2 2017. URL http://arxiv.org/abs/1702.05538.

I. do Carmo Stangherlin, J. Thøgersen, and M. D. de Barcellos. Behavioral spillover in the circular economy: The importance of consumer goals. *Journal of Environmental Psychology*, 91, 2023. ISSN 15229610. doi: 10.1016/j.jenvp.2023.102123.

J. Donovan. Auto-trash sorts garbage automatically at the techcrunch disrupt hackathon., 9 2016. URL https://techcrunch.com/2016/09/13/auto-trash-sorts-garbage-automatically-at-the-techcrunch-disrupt-hackathon/?guccounter=1.

A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. 10 2020. URL http://arxiv.org/abs/2010.11929.

N. Ferronato and V. Torretta. Waste mismanagement in developing countries: A review of global issues. *International Journal of Environmental Research and Public Health*, 16:1060, 3 2019. ISSN 1660-4601. doi: 10.3390/ijerph16061060.

A. U. Gondal, M. I. Sadiq, T. Ali, M. Irfan, A. Shaf, M. Aamir, M. Shoaib, A. Glowacz, R. Tadeusiewicz, and E. Kantoch. Real time multipurpose smart waste classification model for efficient recycling in smart cities using multilayer convolutional neural network and perceptron. *Sensors*, 21, 2021. ISSN 14248220. doi: 10.3390/s21144916.

H. Hazan, D. J. Saunders, H. Khan, D. Patel, D. T. Sanghavi, H. T. Siegelmann, and R. Kozma. Bindsnet: A machine learning-oriented spiking neural networks library in python. *Frontiers in Neuroinformatics*, 12, 12 2018. ISSN 1662-5196. doi: 10.3389/fninf.2018.00089.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. volume 2016-December, pages 770–778. IEEE, 6 2016. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90. URL `http://ieeexplore.ieee.org/document/7780459/`.

D. Ho, E. Liang, I. Stoica, P. Abbeel, and X. Chen. Population based augmentation: Efficient learning of augmentation policy schedules. volume 2019-June, 2019.

G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger. Densely connected convolutional networks. volume 2017-January, 2017. doi: 10.1109/CVPR.2017.243.

K. Huang, H. Lei, Z. Jiao, and Z. Zhong. Recycling waste classification using vision transformer on portable device. *Sustainability (Switzerland)*, 13, 2021. ISSN 20711050. doi: 10.3390/su132111572.

S. Kaza, L. C. Yao, P. Bhada-Tata, and F. V. Woerden. *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Washington, DC: World Bank, 9 2018. ISBN 978-1-4648-1329-0. doi: 10.1596/978-1-4648-1329-0. URL `https://openknowledge.worldbank.org/handle/10986/30317`.

A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and N. Houlsby. Big transfer (bit): General visual representation learning. 12 2019. URL `http://arxiv.org/abs/1912.11370`.

M. M. Krishna, M. Neelima, M. Harshali, and M. V. G. Rao. Image classification using deep learning. *International Journal of Engineering and Technology(UAE)*, 7, 2018. ISSN 2227524X. doi: 10.55041/ijsrem27484.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. volume 2, 2012.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86:2278–2324, 1998. ISSN 00189219. doi: 10.1109/5.726791. URL `http://ieeexplore.ieee.org/document/726791/`.

W. Liang, Y. Liang, and J. Jia. Miamix: Enhancing image classification through a multi-stage augmented mixed sample data augmentation method. *Processes*, 11:3284, 11 2023. ISSN 2227-9717. doi: 10.3390/pr11123284.

H. Liu, W. Wang, H. Liu, S. Yi, Y. Yu, and X. Yao. A degradation type adaptive and deep cnn-based image classification model for degraded images. *Computer Modeling in Engineering Sciences*, 138, 2024. ISSN 1526-1506. doi: 10.32604/cmes.2023.029084.

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*, 164, 2021. ISSN 18790658. doi: 10.1016/j.resconrec.2020.105132.

M. H. K. Mehedi, I. Arafin, M. Hasan, F. Rahman, R. Tasin, and A. A. Rasel. A transfer learning approach for efficient classification of waste materials. pages 0636–0640. IEEE, 3 2023. ISBN 979-8-3503-3286-5. doi: 10.1109/CCWC57344.2023.10099127. URL https://ieeexplore.ieee.org/document/10099127/.

M. Mozafari, S. R. Kheradpisheh, T. Masquelier, A. Nowzari-Dalini, and M. Ganjtabesh. First-spike-based visual categorization using reward-modulated stdp. *IEEE Transactions on Neural Networks and Learning Systems*, 29:6178–6190, 12 2018. ISSN 2162-237X. doi: 10.1109/TNNLS.2018.2826721. URL https://ieeexplore.ieee.org/document/8356226/.

N. Nnamoko, J. Barrowclough, and J. Procter. Solid waste image classification using deep convolutional neural network. *Infrastructures*, 7, 2022. ISSN 24123811. doi: 10.3390/infrastructures7040047.

S. Schmidt and D. Laner. Environmental waste utilization score to monitor the performance of waste management systems: A novel indicator applied to case studies in germany. *Resources, Conservation Recycling Advances*, 18:200160, 10 2023. ISSN 26673789. doi: 10.1016/j.rcradv.2023.200160. URL https://linkinghub.elsevier.com/retrieve/pii/S2667378923000329.

C. Shi, C. Tan, T. Wang, and L. Wang. A waste classification method based on a multilayer hybrid convolution neural network. *Applied Sciences (Switzerland)*, 11, 2021. ISSN 20763417. doi: 10.3390/app11188572.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. 2015.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. 6 2017. URL http://arxiv.org/abs/1706.03762.

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

A. U. Zaman and S. Lehmann. Urban growth and waste management optimization towards 'zero waste city'. *City, Culture and Society*, 2:177–187, 12 2011. ISSN 18779166. doi: 10.1016/j.ccs.2011.11.007.

Q. Zhang, Q. Yang, X. Zhang, Q. Bao, J. Su, and X. Liu. Waste image classification based on transfer learning and convolutional neural network. *Waste Management*, 135, 2021a. ISSN 18792456. doi: 10.1016/j.wasman.2021.08.038.

S. Zhang, D. Hu, T. Lin, W. Li, R. Zhao, H. Yang, Y. Pei, and L. Jiang. Determinants affecting residents' waste classification intention and behavior: A study based on tpb and a-b-c methodology. *Journal of Environmental Management*, 290, 2021b. ISSN 10958630. doi: 10.1016/j.jenvman.2021.112591.