Garbage Classification: A Comparative Analysis of Multi-Class Classification using Feature Descriptors and CNNs

Giovanni Lucia
MSc in Digital Transformation Management
University of Bologna, Cesena Campus
Email: giovanni.lucia@studio.unibo.it

***Abstract*—This work is aimed at analyzing different techniques for the performance of a multi-class classification task, involving the classification of waste into 5 distinct classes: glass, paper, cardboard, plastic, and metal. The project's objective is to address the classification problem using both Machine Learning and Deep Learning methods. Specifically, an initial classification will be performed using Feature Descriptors, and then the task will be performed using two CNNs, both pre-trained on ImageNet, comparing their architecture and metrics. The first CNN would be based on a MobileNet model and the second on VGG19.**

February 06, 2024

1. Introduction

We live in a society that struggles to enact circular economy policies and its 3-R paradigm: reuse, reduce, and recycle. Although an effective waste separation practice is fundamental to correctly address the problem of wastage, the efficiency of the systems and facilities in place to handle and process garbage and the diffusion of the recycling practice is still subdued. This represents a problem both for the environment and the economy.

This project is meant to automate the classification of different types of waste, with the intent to automate garbage recognition and disposal to enhance waste management efficiency. Given a weak awareness of waste separation and recycling among residents, the use of Artificial Intelligence techniques for the automatization of garbage recognition and classification can increase the effectiveness of waste disposal activities, reducing the impact on the environment in terms of pollution, and on the operation of waste management companies.

The proposed solution has been developed using both Machine Learning and Deep Learning techniques, involving ensemble methods for multi-classification tasks, feature extraction, and Convolutional Neural Networks.

2. Related work

The possibility of automating garbage recognition for recycling purposes has been addressed by several studies, that, although the different objectives and methodologies, share the purpose of addressing the inefficiency of garbage collection and disposal. The three papers selected are all based on Convolutional Neural Networks, but cover different applications of the classification information.

The first research paper, titled "Application of MobileNetV2 to waste classification" [1], was aimed at optimizing the model of garbage classification through a lightweight CNN to create a WeChat applet, meant to help final users identify the type of waste and where the latter can be disposed. The up-mentioned applet allows to user to upload the image of the waste and obtain the result of the classification within seconds. The model was trained on a training set of over 80,000 images, classified into 4 main classes: recyclable waste, kitchen waste, hazardous waste, and other waste. These are further subdivided into over 250 waste categories. The model was based on MobileNetV2, a Convolutional Neural Network architecture and achieved an accuracy of 82.92%.

The following research, titled "Fine-Tuning Models Comparisons on Garbage Classification for Recyclability" [2], addresses the optimization of garbage classification models based on Deep Neural Networks, for the identification of non-recyclable waste. The different models employed for this research have been trained on the TrashNet dataset, a GitHub-hosted image dataset, counting 2,527 images, subdivided into 6 classes: glass, paper, cardboard, plastic, metal, and trash. The optimization has involved 5 different models, all pre-trained on ImageNet: AlexNet, VGG-16, GoogLeNet, ResNet, and SqueezeNet. Additionally, the output of the different fine-tuned CNNs has been classified using both Softmax and Support Vector Machines, whose performances were compared in this research. The comparison of the predictions resulting from the different

Model	Accuracy of Fine-tuned Models (%)		Data Aug.	Epoch
	Softmax	SVM	-	200
AlexNet	87.14	97.23	-	200
GoogleNet	88.10	**97.86**	-	200
ResNet	89.38	94.22	-	200
VGG-16	90	97.46	-	200
SqueezeNet	80.43	90.17	-	200

Figure 1. Comparative evaluation of the results of the classification models

combinations of Deep Neural Network and classifier has shown pretty impressive performance, with an accuracy up to 97.86% for the combination of GoogLeNet and SVM, as shown in Figure 1.

The third and last research identified, titled ”An Intelligent Garbage Classification System Using a Lightweight Network MobileNetV2” [3], was meant to develop an intelligent garbage classification system that can recognize and sort different types of waste, to promote awareness of waste classification and ”improve the efficiency and convenience of garbage disposal”. The research identified an extensive, computer-independent solution, that can be deployed on an embedded device (e.g. a recycle bin). The lightweight model, based on MobileNetV2, has been trained on 12,000 images belonging to 5 different classes: harmful waste, kitchen waste, other waste, recyclable waste, and baffle. The trained model achieved an accuracy rate of 98.7% for garbage recognition.

3. Proposed method

The purpose of this project is to identify and compare the results of different classifiers in a multi-class classification task. A multi-class classifier is a machine learning model that, given a set of records as input, classifies the instances into two or more classes. The work has been structured accordingly with the General Paradigm of Machine Learning.

3.1. Data acquisition: identification of publicly available datasets

The first step involved the identification of the two datasets employed for the training of the models.

The reason behind the need for two distinct databases for the classification is due to the fact that the use of a Convolutional Neural Network to pursue the multi-classification task demands a larger amount of data.

The first dataset [4], in fact, counts 2,390 images divided into 5 classes: glass, paper, cardboard, plastic, and metal. It has been used for the training and evaluation of the RandomForest classifier with Feature Descriptor, but the number of records was limited to build a Deep Learning solution based on a Convolutional Neural Network, although pre-trained on ImageNet.

The second dataset [5] contains instead 5.586 items, over two times the number of images contained in the dataset used to train the Machine Learning model based on RandomForest. This dataset has been used for those classification tasks involving Deep Learning techniques. It's necessary to mention that this dataset contains almost all of the images of the first dataset.

The test set, identified as a portion (20%) of the first dataset, was used to evaluate the classification performances of the models using the second dataset as well.

3.2. Data processing

The pre-processing has involved modifications on both the datasets.

The only modification made on the first dataset was the remotion of the images belonging to the ”trash” class from the dataset, for simplification reasons. The ”trash” class was in fact intended as a fictitious class, containing the negative elements that couldn't be classified in any of the other classes.

The second dataset has required multiple modifications: first of all, 7 out of 12 classes not involved in the classification has been either merged (white glass, brown glass, green glass) or removed. As mentioned, this dataset contains almost all of the images of the first dataset, but, since it is crucial to ensure that the training set and the test set are well distinguished from each other to prevent biased results, the images used for the test set has been removed from the dataset through an hash value comparison operation.

3.3. Model

In this section are described the models employed, Random Forest (with feature descriptor), and two CNNs respectively based on MobileNet and VGG19. The performance of the classifiers on the test set will be analyzed and discussed in the next section.

3.3.1. Random Forest with Feature Descriptor. The Random Forest algorithm has been trained on the first dataset, previously split into training, validation, and test sets using the 60-10-30 rule.

On the three data sets has been done the feature extraction. The feature descriptor used was HOG, the Histogram of Oriented Gradients, since it returned a better classification performance with respect to LBP.

The training has involved 500 estimators. The validation has achieved an accuracy score of 54.39%

3.3.2. MobileNet. Once the images used in the test set from the second, wider dataset were removed, the dataset got split into training and validation sets, respectively 80% and 20%, using Keras's pre-build methods.

As already mentioned, the test set is the same as the one used for the classifier based on Random Forest. Since the test set is a list of paths, but the Keras `image_dataset_from_directory` method requires the dataset to be structured in a dedicated tree in the file system, the images pointed by the paths in the test set list has been copied in a new directory in the folder. The test set has been generated starting from this folder.

The model, pre-trained on ImageNet, has been trained using Softmax as classification activator and on 10 epochs. The number of epochs has been determined to balance execution time and performance.

The validation has achieved an accuracy score of 89.54% after 10 epochs.

3.3.3. VGG19. The datasets on which the VGG19 Convolution Neural Network has been trained are the same as the ones on which has been trained MobileNet.

The model has been trained on ImageNet as well, and for this reason, accordingly with the Keras `compile` method documentation, uses Softmax as classification activator. This model has been trained on just 5 epochs, since the training accuracy reached 99.92%. Although this may be a symptom of overfitting, the validation accuracy for the 5th epoch states a slight improvement with respect to the 4th epoch. For this reason, the number of epochs has not been further reduced.

The validation has achieved an accuracy score of 90.36% after 5 epochs.

4. Results

The discussion of the results takes into consideration both the confusion matrices and the classification reports for all the models involved in the project. Both have been generated using SciKit-Learn's library starting from the prediction obtained from the model and the actual labels associated with the test data.

All the classification results here discussed have been obtained by applying the different trained models on the same test set.

The `classification_report` tool builds a text report showing the main classification metrics. This includes overall accuracy of the classification as well as precision, recall, and f-measure metrics per each class identified.

4.0.1. Random Forest with Feature Descriptor. In Figure 2 are reported the classification results for the model based on Random Forest using HOG as feature descriptor. The overall accuracy of the model is quite mediocre as it doesn't reach 60%. From the classification report, we can state the precision and recall per each class involved in the classification.

	precision	recall	f1-score
glass	0.45	0.64	0.53
paper	0.58	0.80	0.67
cardboard	0.79	0.72	0.76
plastic	0.63	0.39	0.48
metal	0.63	0.31	0.42
accuracy			0.58

Figure 2. Classification report for Random Forest with HOG

The recall, which states the ability of the model to recognize relevant (positive) instances among the dataset, is diversified for the different classes: the model returns pretty good results for paper and cardboard, but it isn't able to correctly identify plastic and metal instances. As evinced in the confusion matrix (Figure 3), the latter are more likely to be classified as glass, than in the correct class.

Although the decent recall, the precision of the classification of glass and paper instances is low. This impacts the f-measure and the reliability of the classification for these classes. It means that, although the classifier is good at finding most of the relevant instances, it also tends to include a lot of false positives in its predictions.

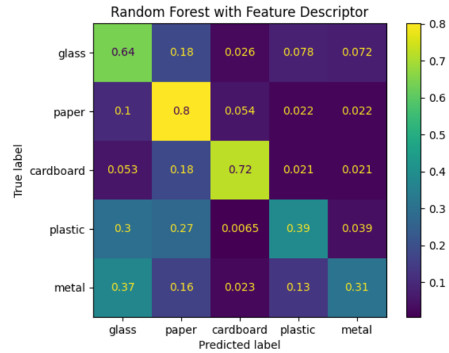

Figure 3. Confusion matrix for Random Forest with HOG

4.0.2. MobileNet. MobileNet has reached pretty good results in the overall accuracy, but also in terms of precision and recall for the single classes, as shown in Figure 4. The accuracy of the trained model on the test images is 81.45%, over 20% higher than the accuracy of the model trained using Random Forest and HOG.

	precision	recall	f1-score
glass	0.75	0.74	0.74
paper	0.92	0.88	0.90
cardboard	0.89	0.94	0.91
plastic	0.73	0.79	0.76
metal	0.79	0.74	0.77
accuracy			0.81

Figure 4. Classification report for MobileNet

The f-measure, the harmonic mean of precision and recall, states that the best performances have been achieved in the classification of paper and cardboard, reporting good scores in both precision and recall. The classification of glass and metal has been more lacking: for instance, as

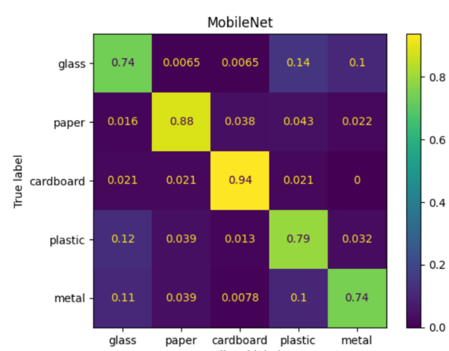

Figure 5. Confusion matrix for MobileNet

highlighted in the confusion matrix (figure 5), a small, although relevant, percentage of glass instances got classified as plastic (14%) and metal (10%).

4.0.3. VGG19. The last trained model to pursue the classification is VGG16, which achieved the highest overall accuracy on the test set of 86.05%. As shown by the classification report (Figure 6), the model has achieved good results also in terms of precision and recall for the majority of the classes involved in the classification with respect to the prediction resulting from MobileNet.

	precision	recall	f1-score
glass	0.77	0.86	0.81
paper	0.90	0.88	0.89
cardboard	0.88	0.91	0.89
plastic	0.88	0.79	0.83
metal	0.87	0.87	0.87
accuracy			0.86

Figure 6. Classification report for VGG19

The classification of paper and cardboard instances has obtained worse results with respect to MobileNet, achieving a slightly lower precision and recall. On the other hand, the model has been able to substantially improve the classification performances in terms of recall for glass and metal of over 10%, with a relevant gain also in precision for the classification of plastic and metal. The confusion matrix (Figure 7) is essentially diagonal, with a single relevant misclassification event, for which 12% of plastic instances have been incorrectly classified as glass.

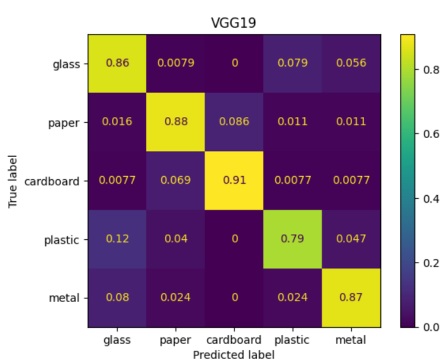

Figure 7. Confusion matrix for VGG19

5. Conclusion

This report has discussed the context, the methodologies and the results of a garbage classification task aimed at automating the recognition of different kinds of waste to promote an effective, sustainable and efficient waste management.

This activity has involved the training of three models over two image datasets. The first model was based on Random Forest, a Machine Learning ensemble method that uses decision trees as base classifiers, and HOG (Histogram of Oriented Gradients) as a feature descriptor of the input images. The second and third models were Convolutional Neural Networks pre-trained on ImageNet: MobileNet, a lightweight deep neural network optimized for the mobile vision application, and VGG19, used for accurate image recognition tasks.

The comparison of the results obtained from the three models highlights an outperformance of the two Convolutional Neural Networks, with respect to the model based on Random Forest and HOG. The best performing model was VGG19 with an overall accuracy of 86.05%, followed by MobileNet with an accuracy of 81.45

These results were somehow expected: Convolutional Neural Networks are meant to address the problem of image classification and recognition, in which a manual feature extraction can be fallible. The fact that VGG19 had better performances with respect to MobileNet, although a lower number of epochs has been used, can be justified by the fact that VGG19 has a deeper architecture with more parameters and that MobileNet is designed to be computationally efficient and lightweight.

References

[1] L. Yong, L. Ma, D. Sun, L. Du (2023), Application of MobileNetV2 to waste classification. PLoS ONE 18(3): e0282336. https://doi.org/10.1371/journal.pone.0282336

[2] U. Ozkaya, L. Seyfi (2019), Fine-Tuning Models Comparisons on Garbage Classification for Recyclability. arXiv. https://doi.org/10.48550/arXiv.1908.04393

[3] H. Ma, Y. Ye, J. Dong and Y. Bo, "An Intelligent Garbage Classification System Using a Lightweight Network MobileNetV2," 2022 7th International Conference on Signal and Image Processing (ICSIP), Suzhou, China, 2022, pp. 531-535, doi: 10.1109/ICSIP55141.2022.9886985.

[4] Garbage classification. (2018, November 24). Kaggle. https://www.kaggle.com/datasets/asdasdasasdas/garbage-classification

[5] Garbage Classification (12 classes). (2021, January 24). Kaggle. https://www.kaggle.com/datasets/mostafaabla/garbage-classification