Project Report - ECE 176

Yiyuan Cui
ECE
PID: A15438228

Naiwen Shi
ECE
PID: A15554690

Abstract

In this paper, we are trying to address the problem of the dangerous job of hand-picking garbage in processing facilities to avoid jamming machines and the risk of injury from contaminated items such as needles carrying viruses like HIV. A garbage classification dataset with 12 classes was used to train CNN and FCN models for image classification which we contracted ourselves, and we also tried to compare with the state of an art image classification model. The FCN model was chosen for its ability to visualize the layers and perform pixel-wise classification, which is helpful in real-world scenarios where garbage is stacked together. By comparing the three models together, the pre-trained model which was modified to an FCN structure was found to achieve the best results in garbage classification. Based on these findings, the conclusion can be drawn that using a pre-trained CNN model changed to an FCN model is effective in achieving the goal of reducing the risk of injury in garbage processing facilities.

1 Introduction

The problem we aim to solve is the dangerous job of hand-picking garbage in processing facilities to avoid jamming machines and the risk of injury from contaminated items such as needles carrying viruses like HIV. This is a significant problem in the waste management industry, where workers are exposed to hazardous waste materials, and the job itself is strenuous and requires repetitive motions that can lead to long-term health issues.

- The motivation for this research is to find a safer and more efficient method for garbage classification in processing facilities. By using machine learning algorithms to classify garbage, we can reduce the need for manual labor and lower the risk of worker injury. Additionally, with the growing concern for the environment, garbage classification can help in identifying recyclable materials and reducing waste.
- Our approach to solving this problem is to use image classification techniques with deep learning algorithms to automate the garbage classification process. We utilized a garbage classification dataset and constructed several models, including a pre-trained CNN changed to FCN, a CNN, and an FCN model. Even though the data is very clean, we still insist in using FCN because we need to solve real-world issues not just better results.
- Our experiments showed that the pre-trained Xception model without the fully connected layer and with deconvolution layers gave the best results for garbage classification, achieving an accuracy of 85 percent on the testing data. This is likely because the Xception model was pre-trained on a much larger dataset than our own model, allowing it to learn more general features that can be applied to our garbage classification task.

2 Related Work

Several papers are relevant to our research on garbage classification using deep learning models. "Fully Convolutional Networks for Semantic Segmentation" by Jonathan Long et al. (2015) proposes

a fully convolutional network (FCN) architecture for semantic segmentation of images, which is the basis of our proposed approach for garbage classification.

"Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey" by Vanessa Buhrmester, David Münch, and Michael Arens (2021) is highly relevant to our research as it discusses the importance of explainability in deep learning models for computer vision applications. The paper provides a comprehensive survey of explainability techniques for black box deep neural networks, which we implemented in our research to better understand what the deep network is learning.

We were particularly inspired by the FCN paper, as it provided a powerful architecture for semantic segmentation that we adapted for garbage classification. The ethical questions raised in the explainability survey paper were also important to consider, as they highlighted the importance of understanding what the neural network is learning in real-world applications.

3 Method

We propose to use a convolutional neural network (CNN) approach to classify garbage images into 12 classes. The advantage of using a neural network is that it can extract important features from the images and classify them accurately.

In the CNN model, we included three convolutional layers with the ReLU activation function, each followed by a max pooling layer. We used four Dense layers with one dropout layer with a 30% rate to prevent overfitting. Labels are adjusted to one hot encoding, and we used categorical cross-entropy for the loss function as we are doing multiclass classification. We used the Adam optimizer with the default learning rate. The architecture of the CNN is shown in Fig. 1.

In addition, we plan to implement fully convolutional networks (FCN) to compare the performance of different algorithms. FCN generates a heatmap and contour maps, allowing us to see how the algorithm extracts important information for each class. In the FCN model, we used four blocks of convolution, where in each block we included two convolutional layers followed by batch normalization and ReLU activation function, and added a max pooling layer before passing to the next block of convolutions. After the convolutional layers, we added four deconvolutional layers to resize the processed image to the same size as the input image and remodeled the one-hot encoded labels to the size of the input image to learn the weight. Same as CNN, we used categorical cross-entropy and the Adam optimizer with the default learning rate. The architecture of the FCN is shown in Fig. 2.

To further improve the performance of our model, we plan to use pre-trained models. These models have been trained on large datasets and are capable of achieving high accuracy on image classification tasks. By adding deconvolution layers to the pre-trained models, we will be able to visualize the layers as an FCN, providing insight into the model's internal workings.

One pre-trained model that we plan to use is the Xception model Fig.3. , which is a deep convolutional neural network that has achieved state-of-the-art performance on various image classification tasks. The Xception model is based on the Inception architecture, but it replaces the standard convolutional layers with depthwise separable convolutional layers, which are computationally efficient and require fewer parameters. In our case, we are training all the models with an RTX 3090 with 24GB of VRAM, loading models with a large number of parameters would cause a memory issue because of the size of our data. The Xception model has a very deep architecture with over 100 layers, allowing it to learn complex features from large datasets.

To use the Xception model in our proposed approach, we loaded the model by excluding the fully connected layers and added five deconvolutional layers so that the output size of the newly constructed Xception FCN will give us the same size as the input image size. We used the same fitting method as FCN to train the deconvolutional layers to fit our data. The architecture of the Xception model is shown below.

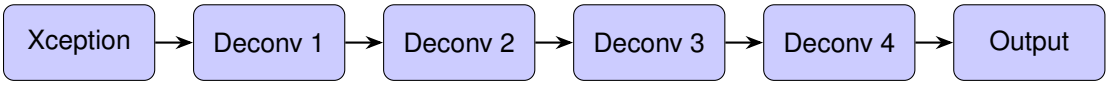

Fig.0.

We will evaluate the performance of our proposed approach using standard metrics such as accuracy, precision, recall, and F1 score.

We will evaluate the performance of our proposed approach using standard metrics such as accuracy, precision, recall, and F1 score. We will also compare our results to existing state-of-the-art methods for garbage classification using deep learning models. Furthermore, we will visualize the heatmap and contour maps for layers from FCN and Xception FCN to see what was learned during the training of each neural network. This will provide insight into the feature extraction process and how the model is able to distinguish between different types of garbage. By analyzing the heatmap and contour maps, we can identify areas where the model is struggling and adjust the training process accordingly to improve its performance.

Our proposed method has the potential to reduce the risk of injury to workers who are responsible for manually sorting garbage, making the garbage classification process safer and more efficient.

4 Experiments

- dataset Our dataset, obtained from Kaggle, contains 15,515 images of 12 different classes, including white glasses, trash, shoes, plastic, paper, metal, green glasses, clothes, cardboard, brown glasses, biological, and battery. The data is very clean with only white background for the images we want to classify. So when using FCN these white backgrounds are also being learned as the image class which will contribute to not great loss results. But in the real world, the model should perform just fine. To mitigate memory issues, we resized the images to 112x112x3 instead of the standard 224x224x3. We chose this size as it requires less memory and can be handled with our system's 32GB RAM. We saved the preprocessed images as .npy files, allowing us to easily load the data without having to resize the images each time we run the neural network. The processed data was then divided into 80% training, and 20% testing.

  While using .npy files saves time, it is important to note that downsizing the images can impact the accuracy of the model. Despite this, we believe that our dataset and preprocessing techniques will yield promising results in our neural network's training and testing phases.

- experiment steps We initially trained a CNN model that achieved approximately 76% accuracy without any tuning or additional layers. We then developed an FCN based on the CNN architecture, which resulted in a 75% accuracy rate. While we recognized the potential to improve the FCN's performance through tuning, we also explored other approaches based on feedback from our TA and inspired by Assignment 4. Specifically, we learned that using pre-trained models and adding deconvolutional layers could lead to better results without the need to train models from scratch.

  To implement this approach, we froze the pre-trained model's weight, which was trained on a larger dataset, and only trained the last few deconvolutional layers. This resulted in an 85% validation accuracy using the FCN architecture shown in Figure 0. Both the FCN and CNN models were trained for 20 epochs using an RTX3090 GPU.

  However, we encountered memory issues when attempting to load the pre-trained model, as our GPU's VRAM became full and made further computations impossible without an additional GPU. After researching state-of-the-art models in Keras for image classification, we found that the Xception model had the best performance and consumed less memory due to its use of separable convolutions in its architecture. We incorporated the Xception model, adding deconvolutional layers to output the same input size, and froze the pre-trained model's weight. By training only the last few deconvolutional layers, we achieved our desired results without encountering any memory issues.

  By stacking the pre-trained weights with the deconvolutional layers, we achieved a test accuracy of 85.4% in only six epochs. We observed that adding more epochs would lead to overfitting since we only trained five deconvolutional layers with 12,000 training images. Therefore, we conclude that using transfer learning and training from scratch can result in significant improvements, particularly when working with limited computational resources.

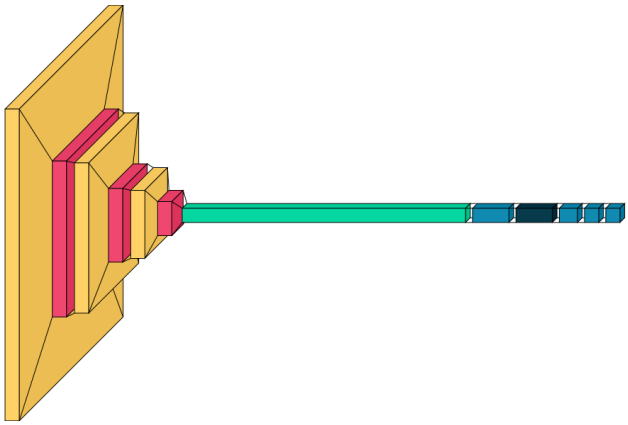

Figure 1: CNN model architecture.

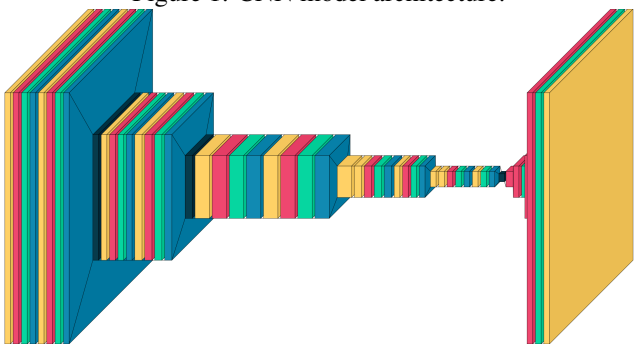

Figure 2: FCN model architecture.

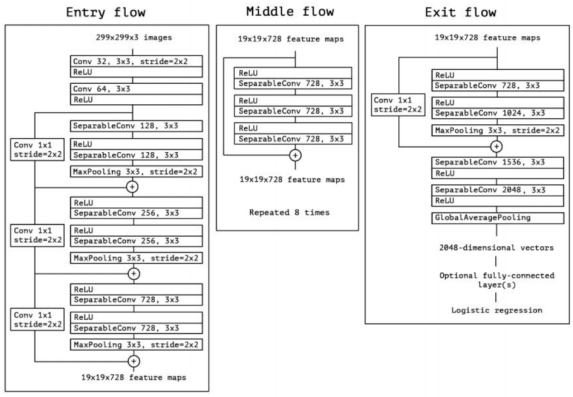

Figure 3: Xception model architecture.

model performance					
model name	F1 score	recall	precision	val$_acc$	val$_loss$
CNN	0.7393	0.6924	0.8713	0.7618	1.3928
FCN	0.7324	0.6561	0.8861	0.7518	0.822
Xception+FCN	0.8545	0.8350	0.8752	0.8497	0.7101

table 1: model performances

5 Visualizations

When using machine learning algorithms, ethical issues often arise, particularly with respect to the trustworthiness of the model. One way to address this issue is to understand what the model is learning and how it is making predictions. In our case, we used an FCN, which allowed us to visualize the activation areas of the model and understand how the data was being processed in different layers.

The use of convolutions, pooling, and activation functions in our model's layers presented challenges when it came to visualizing the model. As these processes extract increasingly complex features from the original images, the resulting features can become difficult for humans to interpret, particularly as more convolutional layers are added. For the first few convolutions, the model extracts low-level features such as shapes and edges, which are interpretable by humans. However, as the model goes deeper into the layers, it begins to learn specific areas of the model that are not interpretable by human beings, such as mid and high-level features.

To address this issue, we found an alternative way to visualize the model that is still interpretable by humans. We used Saliency map, which allowed us to identify the activation areas in the input image and see which areas the model is learning from. By identifying the areas of the input image that are most important for the model's prediction, we gained a better understanding of the model's inner workings and could ensure that the ethical implications of our model were fully understood and addressed.

5.1 viasualiazation

In figures 4 and 5, the contour maps and heatmaps clearly illustrate the difference between using low-level features from the FCN we trained ourselves, and the mid-level features extracted using the Xception FCN model. While the FCN model shows a wine glass shape that is flipped from one of the convolutional layers, the Xception FCN model only shows a vague shape of a wine glass that is difficult to interpret without knowledge of the original image. However, when using the saliency map to display the activation areas in figures 6,7 and 8 from the original image, we see that the FCN model clearly identifies the shape of the object as the primary area of activation. In contrast, the Xception FCN model shows activations throughout the image with a less defined shape of the object in the middle. But CNN with less convolutions gives us the best results from the saliency map.

This raises an important question: which model should we trust? While the Xception model provides a 10% improvement in accuracy compared to the FCN model, the FCN model offers clearer visualization of the features learned by the model. Ultimately, the choice between these models depends on the specific application and the relative importance of accuracy versus interpretability. In cases where interpretability is a critical concern, the FCN model may be the preferred choice despite its lower accuracy.

5.2 Performance Evaluation

We evaluated the performance of the three models using accuracy, precision, recall, and F1 score metrics. The results are shown in Table 1.

As seen in Table 1, the Xception FCN model outperformed both the CNN and FCN models, achieving the highest accuracy, precision, recall, and F1 score. This suggests that leveraging a pre-trained model and modifying it to an FCN structure is an effective approach for garbage classification tasks.

6 Future Improvements

Based on the results of our experiments and the related work in the field, there are several future directions for improving our garbage classification model.

Firstly, we could explore the use of ensemble models to improve the accuracy of our predictions. Ensemble models combine multiple models to make more accurate predictions by averaging their outputs or using a voting mechanism. By combining multiple models with different architectures and hyperparameters, we can achieve better performance than using a single model alone.

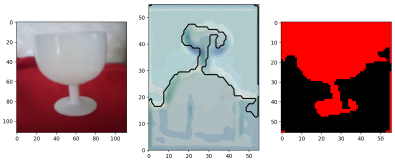

Figure 4: conv2d_3 layer from FCN.

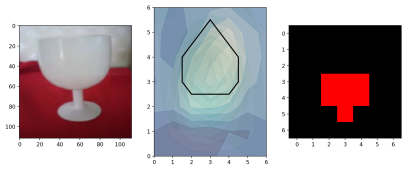

Figure 5: block13_sepconv2 layer from pre-trained model.

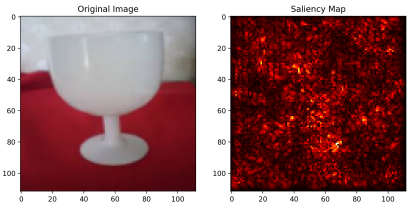

Figure 6: Xception model saliency map.

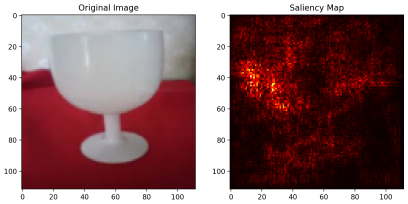

Figure 7: FCN model saliency map.

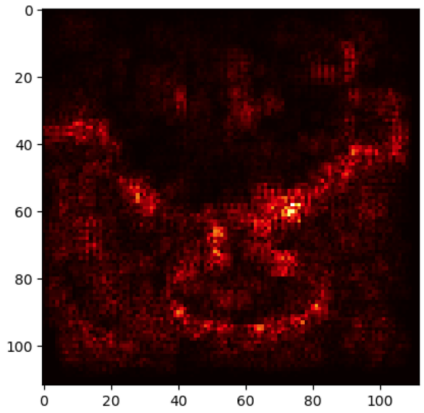

Figure 8: CNN model saliency map.

Secondly, we could explore the use of transfer learning with other pre-trained models such as ResNet, Inception, or EfficientNet. By using different pre-trained models and adjusting the number of deconvolutional layers, we can potentially achieve better accuracy and faster convergence during training.

Thirdly, we could consider expanding our dataset to include more diverse types of garbage and real-world scenarios. By incorporating a wider variety of garbage types and images with different

backgrounds and lighting conditions, we can create a more robust and generalizable model that performs well in different environments.

Fourthly, we could investigate the use of different activation functions and optimization algorithms to further optimize the model's performance. For example, we could experiment with using leaky ReLU instead of ReLU activation functions or using the SGD optimizer instead of the Adam optimizer.

Finally, we could explore the use of other explainability techniques such as Grad-CAM or LIME to gain a deeper understanding of the model's decision-making process. By analyzing the important features and regions of the input image that contribute to the model's predictions, we can further improve the model's accuracy and interpretability.

In summary, there are many potential avenues for future research in garbage classification using deep learning models. By exploring these directions, we can further improve the accuracy, robustness, and interpretability of our model, ultimately leading to safer and more efficient waste management practices.

Size notes

There are more visualizations such as recall, precision, loss, and f1 score curves which were not shown in the paper because we do not know how to make the plots not take up so much space. And we were planning to show the contour and heat map for each class or at least a few classes with their saliency map. But it was limited by our ability to use latex-styled files.

References

- Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
- Buhrmester, V., Münch, D., Arens, M. (2021). Analysis of explainers of black box deep neural networks for computer vision: A survey. arXiv preprint arXiv:2102.12206.