Image Detection

**Paper baseline:** [**https://pdf.sciencedirectassets.com/271837/1-s2.0-S0956053X21X00189/1-s2.0-S0956053X21006474/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEEAaCXVzLWVhc3QtMSJIMEYCIQCAlRttO3vn0t%2B0WcLW1rTpeKso7Ar2L83C8cr9EqvEvwIhAPvRoKJcNxtVGqolHWG4v2WJ%2FYbZoKsdkQm2i0ujJSD0KrMFCEkQBRoMMDU5MDAzNTQ2ODY1IgxTvRNFAmiFrHrRFZUqkAWubRdsnaxry0HQ7rCg4w2wLlUyTPG0ZuPmFFV29QyKCcR7X55wpzG%2FdlpqxpebnMwxZDXlKPy2cJyBCufZm2sfReEAhRmxRKyClKzH5RY8qSuNitr%2BJalhlNWoGrfXJbEu0uDvqtm6eAE3fGilrnR14WmgUpxFftsaSHglJ8vKTb7hMeMx9csR8ZDhlY%2Fv5OhLdWPvOvJkKet6d2CQ0TCcfcRrpUSJX4FqIo3qeya14N%2FoqUc38KGqB%2FvztwdNHUSvQ8DxeydW9XJ%2BQxr29fjoHCEPS6CfRDiRa08XpMO1LxpVSadLDm8iBvFL8qO9qvtt5TU%2BHfBIVO1qGpvD%2BbONRFXi2SHieOHaMv3qBxtmyPAJdhSv2DKiW759VGEPaMk%2B7%2BmxEMM%2FwNP2wWYhZzFFRqioLrmW7GMBlUebPTBRcXCkAXS%2B%2BT4cnCCTzvMymp13Y%2BljtGmihm0IfhHsN7x9cxKX1Xj82X%2Bqm66YIUZKW%2BlfS1WTG%2BfDb%2Fz1b46GdjXNEru0duiiENEfmkP44AhnbpJjUo1hwjnbMp2mrKA3eWefVQkwN9W8jo96Tv9pbU3hRZwNgRP5xJGNu6RHq7v7ulFfDVvMWZqt3dWf2hGjl50SvwO8Bwi7pd6FxXpxRxc7URI4KY%2FYI7v9vsqRuvgsFUvImfzfXcvM0R2Diz4snQrCgDXTCzDLrWu3u9TQTXC2%2FmYTKyVPj3qVLkjbH2Drk96enKMlaa7%2B%2B0FX28S1E%2F%2BOCf2bbdKzt0v1ceYhoLsScU49YwPUrDf77cKrCzFd7B7v%2FxPSyRYPlgn40hrtdN2uA8J8ldywuou4jhWXtbuSBW%2BXp1z0pSTKXj2uE92vyS%2FCTTRBFs3EzDNJrhvcwzD07sGvBjqwAfRc2MY%2FnGdLlb8C20IIHQa0ngCy8dEuSPa%2FMc4QiOtJqCowN6eRY1cIVT5uL9G0UkX1EqyqeNxWT94gSotfvLpfBpAadHFo6bU6I5VM3xv07a1cWOY%2FlnCL5QyqrRTpcis%2BBu0aJRcf4EdaRPYo7QDzWAhbAqbMbTEDAR5ILdonttwyHQWJbS3Df3olzD9aVZ%2F%2FJHWw%2BHZlzG3AagMbluPU1d2kpOmY7ruXEg%2BsWioh&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20240312T162234Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTY2UQVB3OX%2F20240312%2Fus-east-1%2Fs3%2Faws4\_request&X-Amz-Signature=81364473cfdaa2e6fbc68f8b8fbbd0c9e5c102bc7beee22c9ccb2fd53bd3af26&hash=d4957561f76e75ba2419aeb4258f10e89e5e6c8c2f5f60a81aa9acb41e23961a&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S0956053X21006474&tid=spdf-9de6bdf9-422b-4c94-83a5-145a72ac4e87&sid=62ad14a57c3f72413b3b90f-08ccaa62b6d6gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=1d045d5350575306030205&rr=86351fafb93635c5&cc=gb**](https://pdf.sciencedirectassets.com/271837/1-s2.0-S0956053X21X00189/1-s2.0-S0956053X21006474/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEEAaCXVzLWVhc3QtMSJIMEYCIQCAlRttO3vn0t%2B0WcLW1rTpeKso7Ar2L83C8cr9EqvEvwIhAPvRoKJcNxtVGqolHWG4v2WJ%2FYbZoKsdkQm2i0ujJSD0KrMFCEkQBRoMMDU5MDAzNTQ2ODY1IgxTvRNFAmiFrHrRFZUqkAWubRdsnaxry0HQ7rCg4w2wLlUyTPG0ZuPmFFV29QyKCcR7X55wpzG%2FdlpqxpebnMwxZDXlKPy2cJyBCufZm2sfReEAhRmxRKyClKzH5RY8qSuNitr%2BJalhlNWoGrfXJbEu0uDvqtm6eAE3fGilrnR14WmgUpxFftsaSHglJ8vKTb7hMeMx9csR8ZDhlY%2Fv5OhLdWPvOvJkKet6d2CQ0TCcfcRrpUSJX4FqIo3qeya14N%2FoqUc38KGqB%2FvztwdNHUSvQ8DxeydW9XJ%2BQxr29fjoHCEPS6CfRDiRa08XpMO1LxpVSadLDm8iBvFL8qO9qvtt5TU%2BHfBIVO1qGpvD%2BbONRFXi2SHieOHaMv3qBxtmyPAJdhSv2DKiW759VGEPaMk%2B7%2BmxEMM%2FwNP2wWYhZzFFRqioLrmW7GMBlUebPTBRcXCkAXS%2B%2BT4cnCCTzvMymp13Y%2BljtGmihm0IfhHsN7x9cxKX1Xj82X%2Bqm66YIUZKW%2BlfS1WTG%2BfDb%2Fz1b46GdjXNEru0duiiENEfmkP44AhnbpJjUo1hwjnbMp2mrKA3eWefVQkwN9W8jo96Tv9pbU3hRZwNgRP5xJGNu6RHq7v7ulFfDVvMWZqt3dWf2hGjl50SvwO8Bwi7pd6FxXpxRxc7URI4KY%2FYI7v9vsqRuvgsFUvImfzfXcvM0R2Diz4snQrCgDXTCzDLrWu3u9TQTXC2%2FmYTKyVPj3qVLkjbH2Drk96enKMlaa7%2B%2B0FX28S1E%2F%2BOCf2bbdKzt0v1ceYhoLsScU49YwPUrDf77cKrCzFd7B7v%2FxPSyRYPlgn40hrtdN2uA8J8ldywuou4jhWXtbuSBW%2BXp1z0pSTKXj2uE92vyS%2FCTTRBFs3EzDNJrhvcwzD07sGvBjqwAfRc2MY%2FnGdLlb8C20IIHQa0ngCy8dEuSPa%2FMc4QiOtJqCowN6eRY1cIVT5uL9G0UkX1EqyqeNxWT94gSotfvLpfBpAadHFo6bU6I5VM3xv07a1cWOY%2FlnCL5QyqrRTpcis%2BBu0aJRcf4EdaRPYo7QDzWAhbAqbMbTEDAR5ILdonttwyHQWJbS3Df3olzD9aVZ%2F%2FJHWw%2BHZlzG3AagMbluPU1d2kpOmY7ruXEg%2BsWioh&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20240312T162234Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTY2UQVB3OX%2F20240312%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=81364473cfdaa2e6fbc68f8b8fbbd0c9e5c102bc7beee22c9ccb2fd53bd3af26&hash=d4957561f76e75ba2419aeb4258f10e89e5e6c8c2f5f60a81aa9acb41e23961a&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S0956053X21006474&tid=spdf-9de6bdf9-422b-4c94-83a5-145a72ac4e87&sid=62ad14a57c3f72413b3b90f-08ccaa62b6d6gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=1d045d5350575306030205&rr=86351fafb93635c5&cc=gb)

**ViT paper:** [**https://arxiv.org/pdf/2010.11929.pdf**](https://arxiv.org/pdf/2010.11929.pdf)

**Trashbox:** [**https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9770922**](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9770922)

**Cup Dataset:** [**https://www.kaggle.com/datasets/vencerlanz09/plastic-and-paper-cups-synthetic-image-dataset**](https://www.kaggle.com/datasets/vencerlanz09/plastic-and-paper-cups-synthetic-image-dataset)

**Recyclable Material dataset:** [**https://www.kaggle.com/datasets/techsash/waste-classification-data**](https://www.kaggle.com/datasets/techsash/waste-classification-data)

Related Work

When researching a suitable image model for our classification task, we came across a recent paper, attempting to achieve the same goal. The waste management paper [] uses a two-step process to classify and detect waste using deep learning.

Model

The model was split into two steps, similar to the waste management paper but with one alteration. The first step of object detection will be replaced with a unique object classification. This step will classify any items which we have decided must be further inspected by a human to check their recyclability fully. For example, coffee cups may sometimes require manual checking inside of the cup, in case it contains plastic lining. We decided to remove the first step of the waste management paper as we decided that our images would already be centring the object and we would require the user to classify one object at a time, so we would not need any detection algorithms.

[INSERT DIAGRAM OF MODEL ARCHITECTURE]

Waste material Classification

The waste management paper [] tries to classify waste into separate categories by splitting the process into two parts: Detection and Classification. The detection creates a bounding box around the object allowing us to crop the input for the second phase (Classification). We have taken the classification stage of the paper and adjusted the model architecture used in substitution with a ViT model []. This has several advantages over the traditional convolutional deep learning models used for image classification. By using attention heads instead of kernels to get the relationships between parts of an image, we can capture the global attention of a section of an image, instead of being bound by the context of the kernel size. However, since the ‘global attention’ requires an inner dot-product of each vector representing a pixel to each other (num\_pixels^2 time complexity), we are heavily restricted by processing power. We can combat this by splitting the image into patches, say 16 by 16 pixels, and getting the attention between these patches, allowing us to divide the number of vectors by 16.

Unique Object Detection

This step also follows the same architecture (ViT) as before. The only difference is the change in the dataset.

Data

For the waste material classification we fine-tuned a general ViT model with data from the trashbox dataset [], this is a dataset created for a different research paper and classifies waste into multiple categories: cardboard, e-waste, glass, medical, metal, paper, plastic; all materials of waste that could be found in an office space. The dataset also provides us with test and validation folders, with 80, 10, and 10 splits for training, testing and validation respectively. Once we acquired the dataset, we did some preprocessing by checking for any corrupt images, the dataset was already well processed. We did find one or two image files which seemed to be corrupt, but the rest were fine.

The unique object detection was trained on a custom dataset consisting of coffee cup images [] and a recyclable objects image dataset []. The dataset had two classes: Cup and No Cup. In total, there were 30k+ images, around 15000 for each category. The splits were around 80, 10, and 10 for training, validation and testing data respectively.

Results

From Figure [] we can see the confusion matrix of the waste material model when evaluated on the test data. This shows clear signs of success, except for a few outliers and some expected errors. The expected errors were discussed before the training of the model and we realised that some materials such as plastic and glass could be misclassified from each other. Due to the nature of the material, when calculating the inner dot products of the vectors representing the gloss of the object, the results would be very similar for both materials. This is highlighted in the confusion matrix, as we can see misclassifications outside of the range of the general outliers. From table [], we can see the metrics of the testing process, we can see the accuracy is very similar to the F1 score, showing that our model doesn’t have that many false positives or false negatives and this is maintained across all classes.

From Figure [], we can see the confusion matrix for the unique object detection classification when evaluated on the test data. Given we provided some difficult images for testing, we expected some outliers as shown in the figure. However by looking at Table [], we can say this model is a success as we have retained a 95%+ accuracy, and F1 score, reassuring ourselves that our model is well-balanced in terms of performance.

Evaluation

Upon looking at our results and progress up to this point, we deemed the entire model suitable for suggesting waste management in our web application. The table of results in Table [] and Table [], provides us with confidence that we can safely guide our users on what material their waste is, and in turn how it should be recycled.

Our ViT model results:

| Class | Accuracy (%) | Precision | Recall | F1 Score |
| --- | --- | --- | --- | --- |
| Total | 90.57446 | 0.90577 | 0.90574 | 0.90576 |
| Cardboard | 91.73554 | 0.90244 | 0.91736 | 0.90984 |
| E-Waste | 94.70199 | 0.93464 | 0.94702 | 0.94079 |
| Glass | 91.33858 | 0.93173 | 0.91339 | 0.92247 |
| Medical | 91.37056 | 0.91837 | 0.91371 | 0.91603 |
| Metal | 90.38462 | 0.91085 | 0.90385 | 0.90734 |
| Paper | 87.03704 | 0.87361 | 0.87037 | 0.87199 |
| Plastic | 87.31343 | 0.86989 | 0.87313 | 0.87151 |

Total Accuracy Precision Recall F1

90.57446 0.90577 0.90574 0.90576

Cardboard 91.73554 0.90244 0.91736 0.90984

Ewaste 94.70199 0.93464 0.94702 0.94079

Glass 91.33858 0.93173 0.91339 0.92247

Medical 91.37056 0.91837 0.91371 0.91603

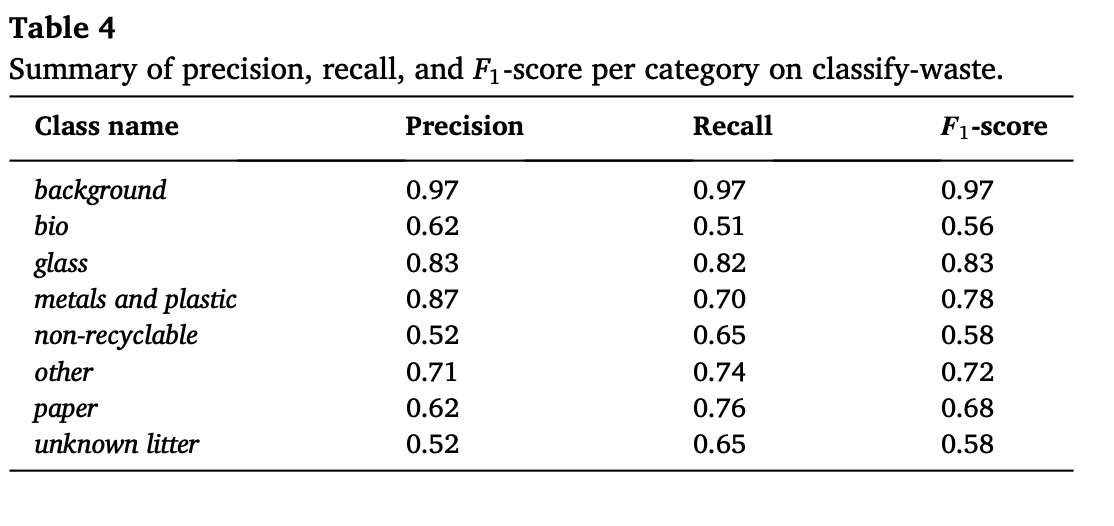
Metal 90.38462 0.91085 0.90385 0.90734

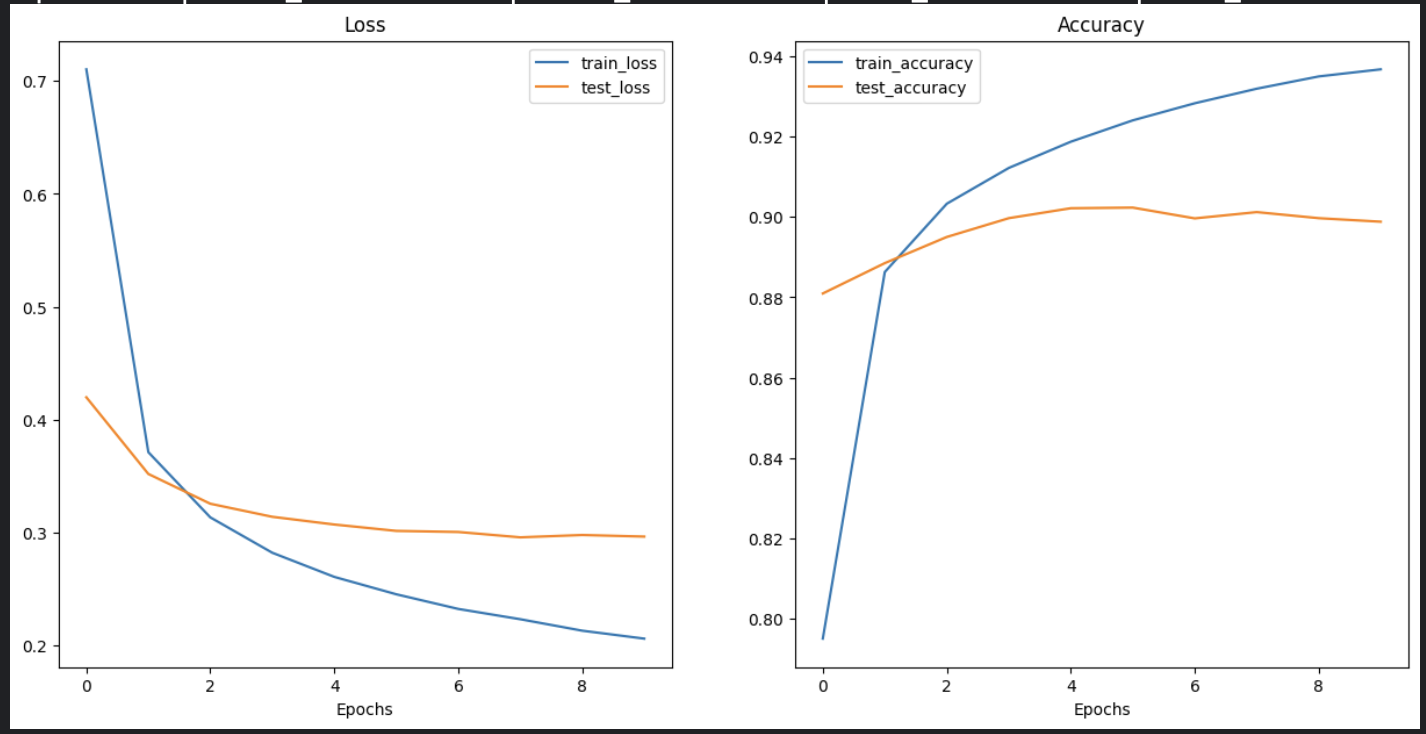
Paper 87.03704 0.87361 0.87037 0.87199

Plastic 87.31343 0.86989 0.87313 0.87151

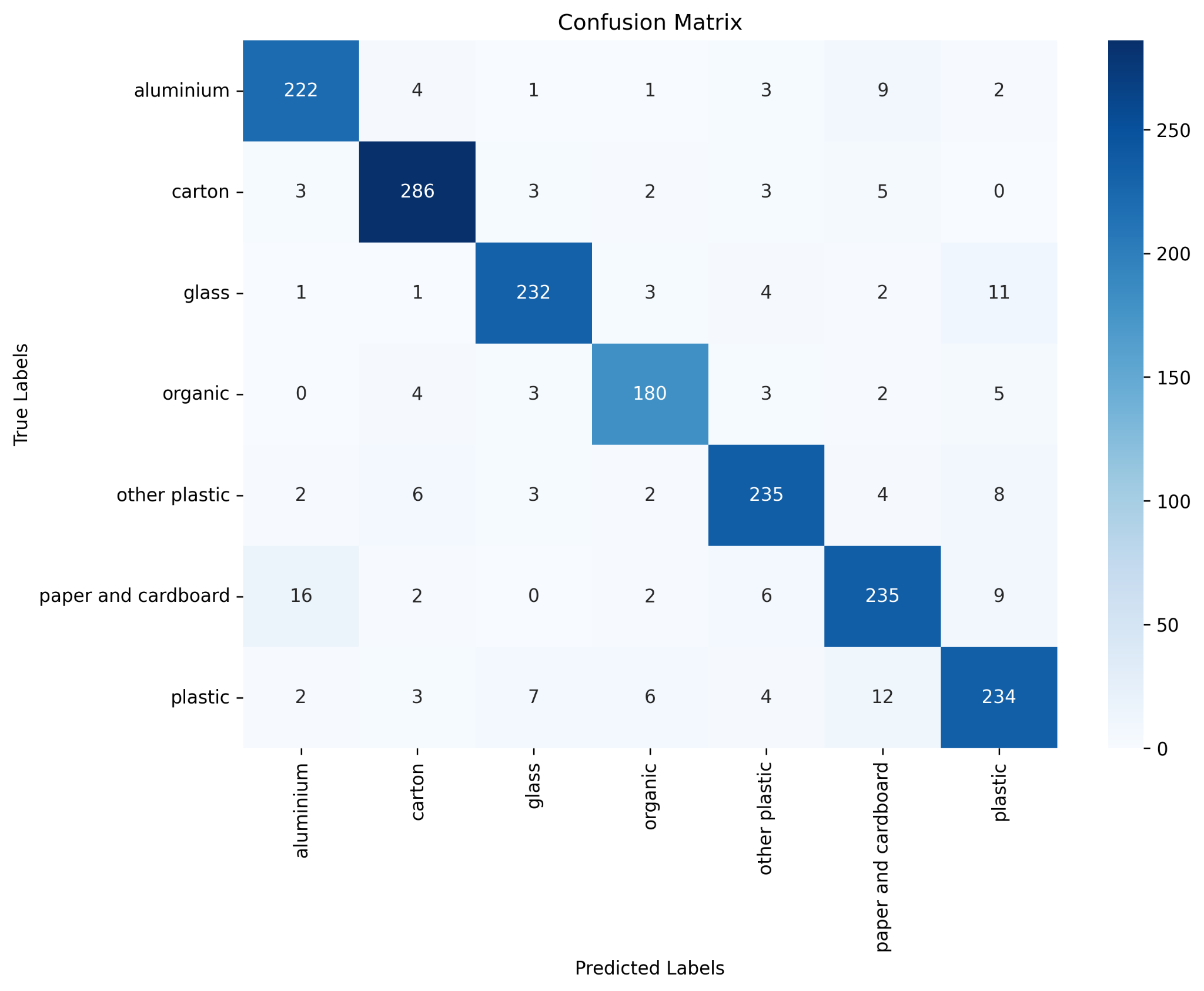
Accuracy of the model on the test images: 90.574456218628%

Benchmark (paper):



Testing results:  
  


Confusion Matrix:

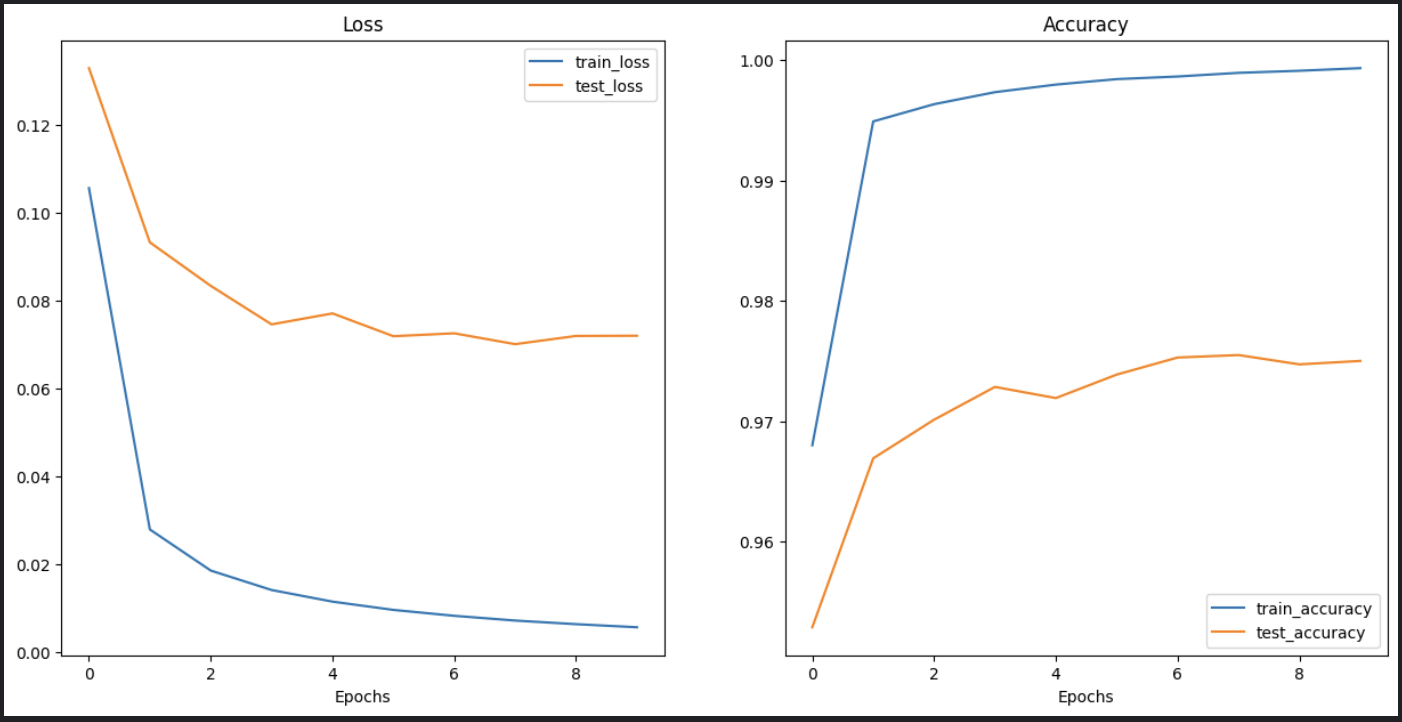


Cup Classification

Metrics:

| Category | Accuracy | Precision | Recall | F1 |
| --- | --- | --- | --- | --- |
| Total | 97.55122 | 0.97639 | 0.97551 | 0.97595 |
| Cup | 95.4 | 0.99687 | 0.954 | 0.97496 |
| No Cup | 99.7003 | 0.95594 | 0.997 | 0.97604 |

Testing results:



Confusion Matrix:

