Software Development Oriented to ML Garbage Classifier

Melchor Lafuente
Iker Jauregui

1 Introduction

The **GarbageClassifier** project is a deep learning approach to automatically classify images of common waste into six categories: cardboard, glass, metal, paper, plastic, and trash. The system leverages **PyTorch Lightning** for structured training and organization of the deep learning pipeline, using **ResNet18** as the backbone model.

The project dataset originates from the Garbage Classification Dataset on Kaggle and has been reduced in size for experimentation purposes. A sample dataset was created for rapid prototyping, containing 100 images per class (90 for training and 10 for testing).

2 Exploratory Data Analysis (EDA)

EDA was performed to understand the dataset characteristics and to guide model development. Key observations include:

- The classes are relatively balanced as shown in figure 1.

- All images are 256×256 pixels.

- Class prototypes, calculated as the mean image per class, revealed distinctive visual characteristics:

  - **Cardboard:** Brown-orange tones.
  - **Plastic:** Blue-gray hues.
  - **Glass:** Greenish tint, with bottle shape visible.
  - **Metal:** Dark gray, with central square-like pattern.
  - **Trash:** Brown-gray tones, diverse textures.
  - **Paper:** Uniform light gray.

- Confusion is likely between metal, plastic, and paper classes due to visual similarity, as highlighted by the prototypes.
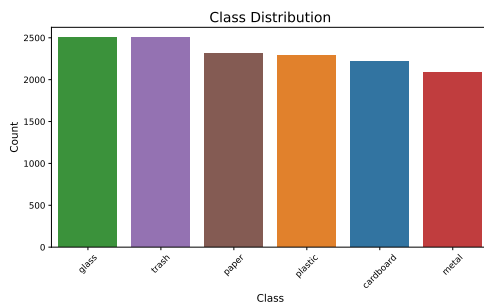
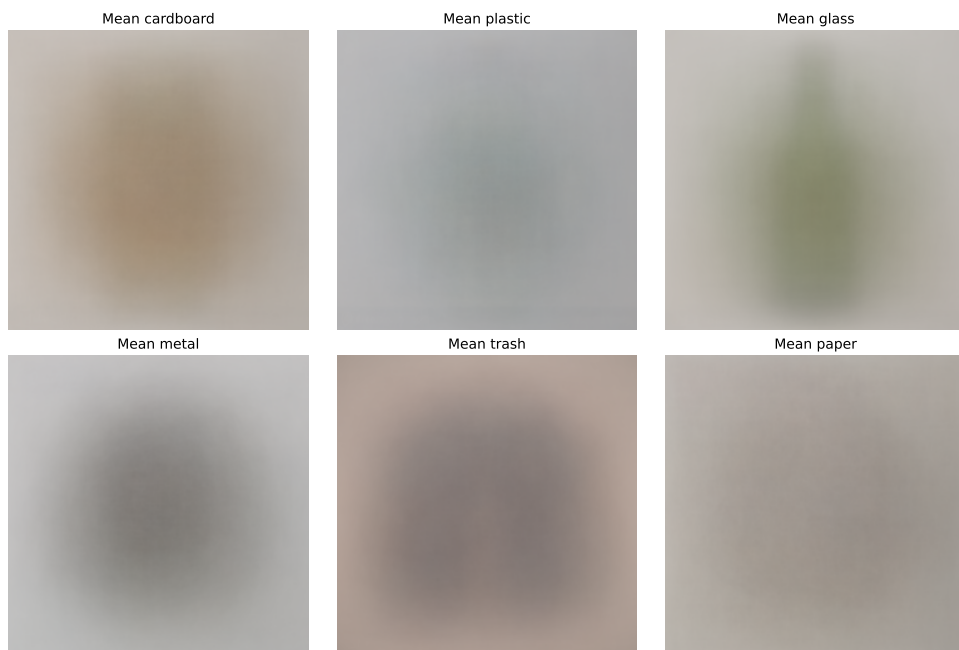

Figure 1: Number of samples per class.

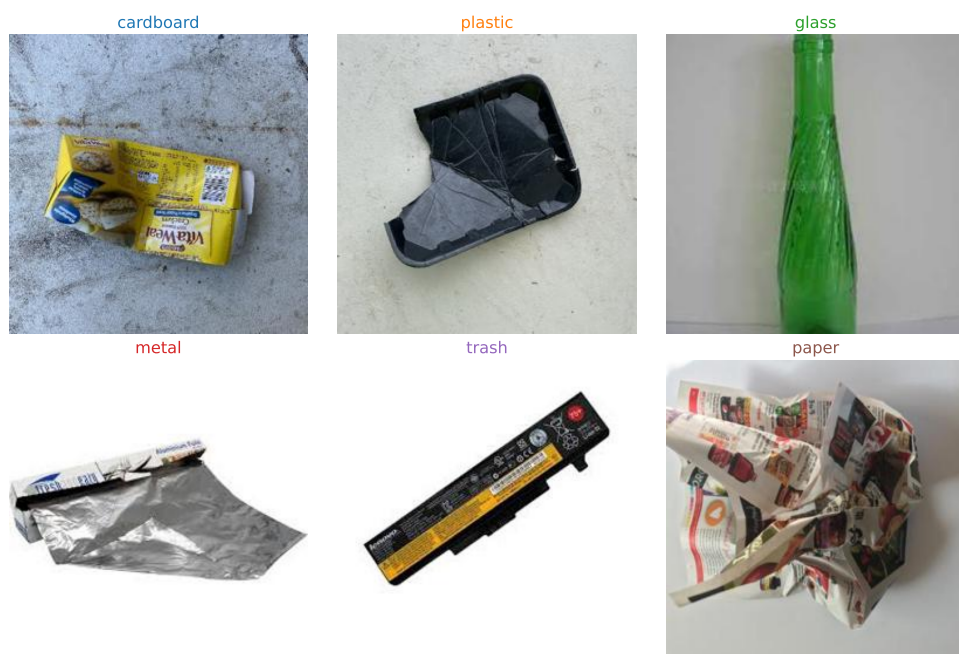Figure 2: Mean images per class highlighting prototypical characteristics.

Figure 3: Random samples per class.

3 Model Training and Performance

The **ResNet18**-based `GarbageClassifier` was trained for 10 epochs on the reduced dataset. Performance metrics were evaluated on training and testing subsets.

3.1 Loss and Accuracy Curves

The training and validation loss curves indicate that the model has not fully converged, suggesting potential improvements with additional epochs. Validation accuracy shows room for enhancement as well.
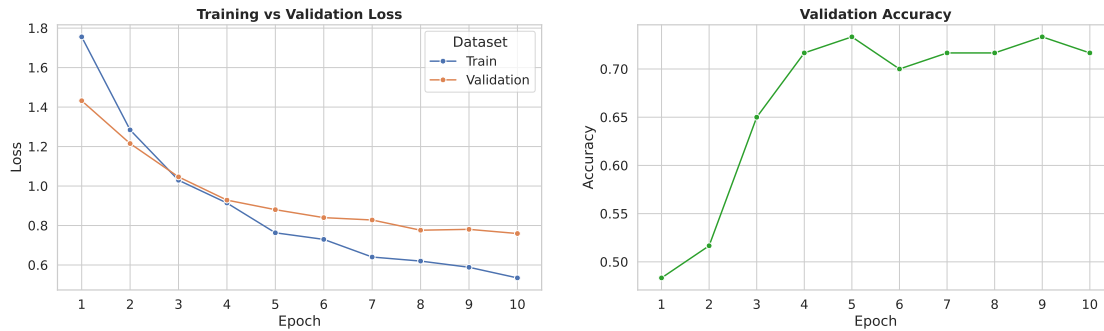

Figure 4: Left: Training vs validation loss. Right: Validation accuracy.

3.2 Confusion Matrices

Confusion matrices reveal that most misclassifications occur among **metal, paper, and plastic**, aligning with the visual similarities observed in the EDA.
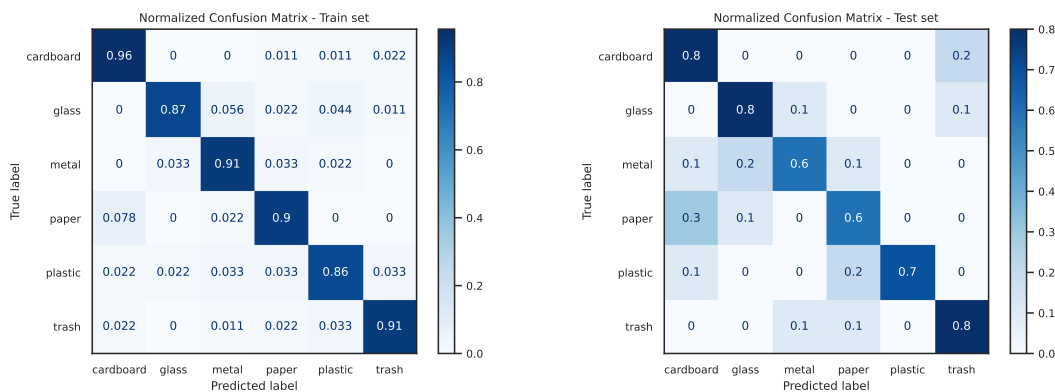

Figure 5: Normalized confusion matrices for training (left) and test (right) sets.

3.3 Calibration Curves

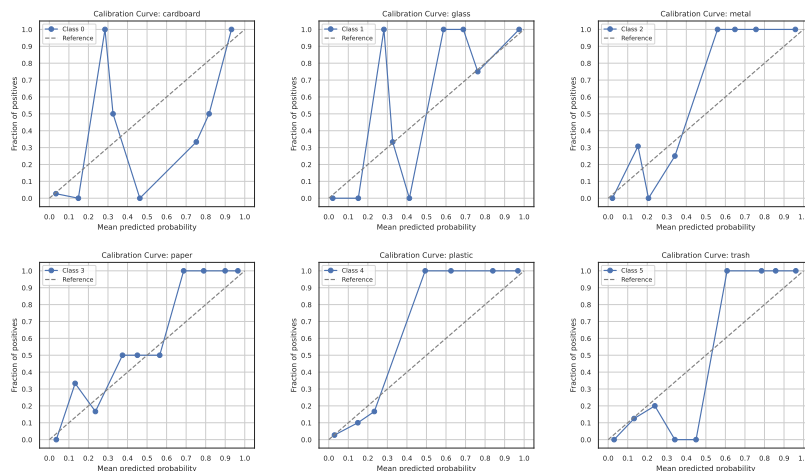Calibration curves suggest that the model is not very well calibrated except for the **paper** class.

Figure 6: Calibration curves per class.

3.4 Error Analysis

Examining the samples with the lowest confidence we can try to think why is the model making bad decisions:

- The model relies too heavily on color.

- The red underwear and blue masks are completely unrelated to each other and to the other trash examples, which confirms that the model struggles because there is no clear pattern.
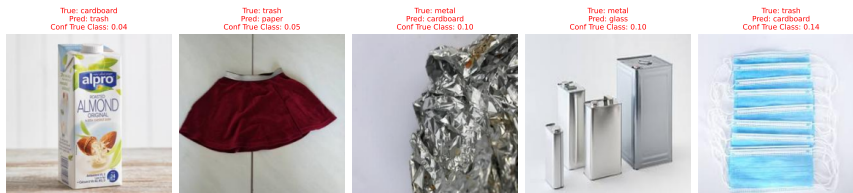

Figure 7: Worst classified test samples, indicating challenging textures and ambiguous objects.

4   Conclusions

The EDA helped identify potential sources of confusion among classes and informed model evaluation. Key takeaways include:

- The ResNet18-based model shows promising initial results but requires more training data and epochs to fully converge.

- Misclassifications mostly involve metal, plastic, and paper, as predicted by EDA prototypes.

- Advanced data augmentation or input transformations could improve feature extraction for challenging textures and colors.

- A rejection-class approach may help address the heterogeneous nature of the trash class.

Future work should focus on scaling the dataset, enhancing augmentation strategies, and exploring more complex architectures or training paradigms to improve robustness and generalization.