Real Waste Image Classification

Alessio Lani - 1857003
Enrico Giordani – 2143412
Lorenzo Arcioni – 1885377
Marta Lombardi – 2156537
Valeria Avino - 1905974

Fundamentals of Data Science
Sapienza University of Rome

Task and motivation

What?

Classification of waste images in real-world scenarios.

Why?

Improper waste disposal significantly threatens the **environment**.

How?

Fine-tuned state-of-the-art **pretrained models** (ResNet, DenseNet, MobileNet, VGG16, EfficientNetV2, Swin Transformer, Vision Transformers) along with a **custom-designed** CNN.

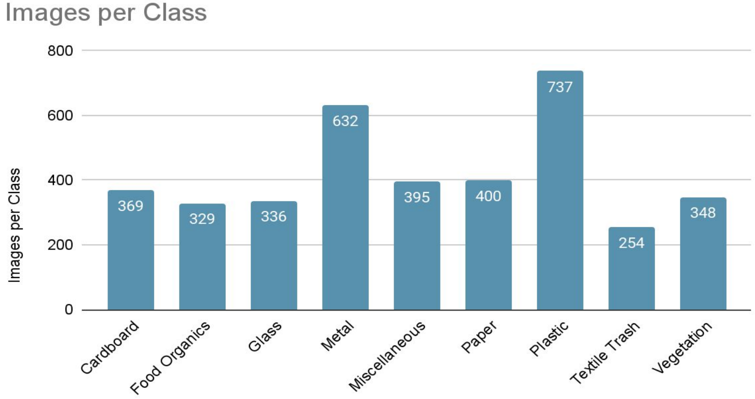
The dataset

4752 images (524 x 524)

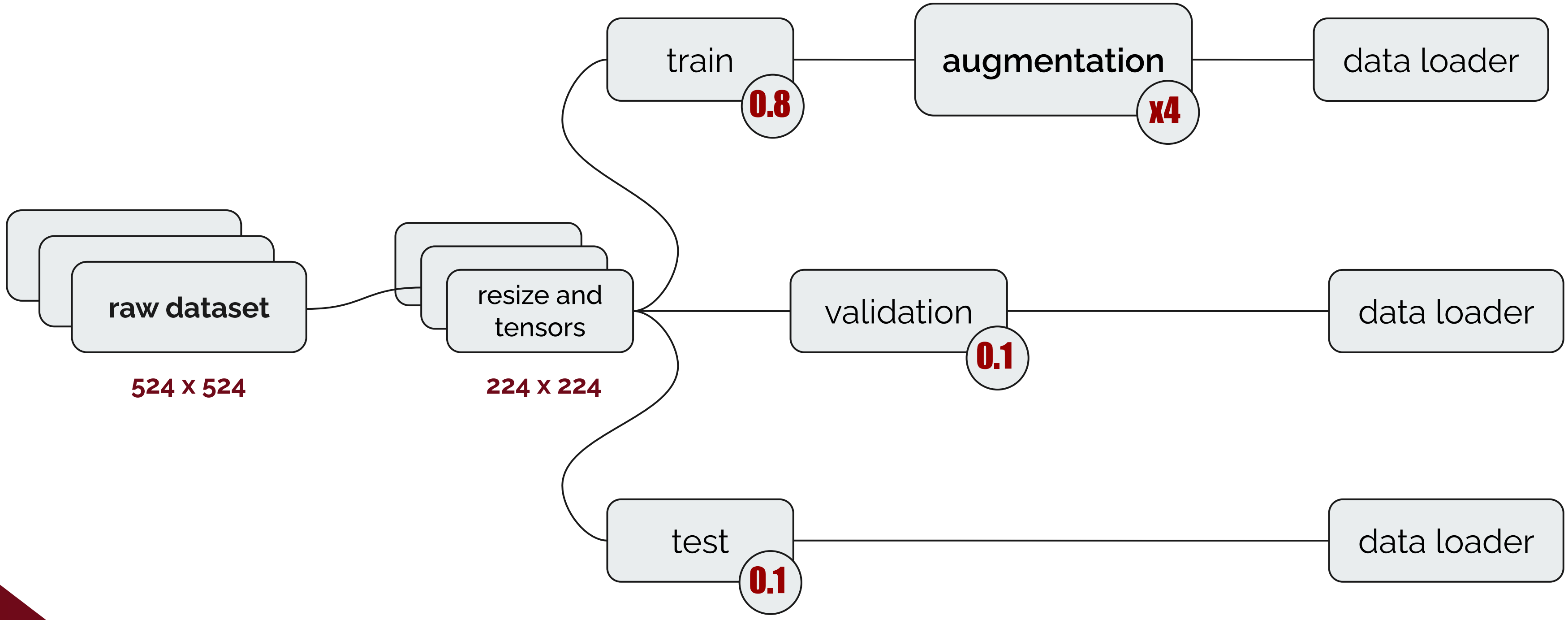
Cardboard
Food organics
Glass
Metal
Miscellaneous Trash
Paper
Plastic
Textile Trash
Vegetation

The dataset is split into

- ❑ **training**: 3800 (**80%**)
- ❑ **validation**: 476 (**10%**)
- ❑ **test**: 476 (**10%**)



Data pipeline

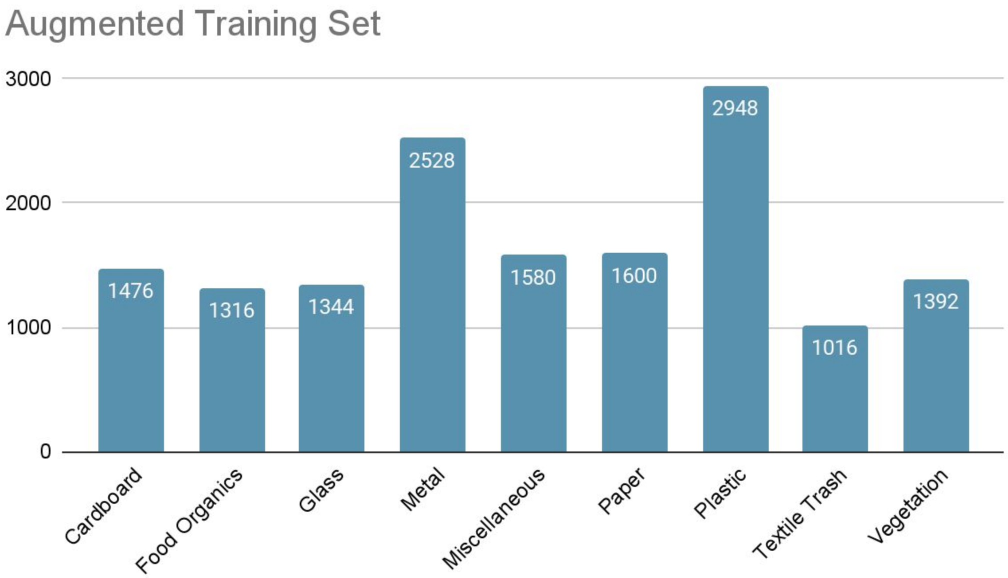


Data Augmentation

Default transformations as:

- **flipping**;
- **rotation**;
- **slight zoom**.

Images are **quadrupled**: the original, plus three randomly transformed augmented versions.



Baseline (custom CNN)

Each (of the 5) convolutional block consists of:

- **Two convolutional layers** (3x3 kernel, same padding) for local feature extraction.
- **Batch normalization** after each convolution for stable training.
- **ReLU activations** for non-linearity.
- **Max-pooling** (2x2 kernel, stride 2) for down-sampling.

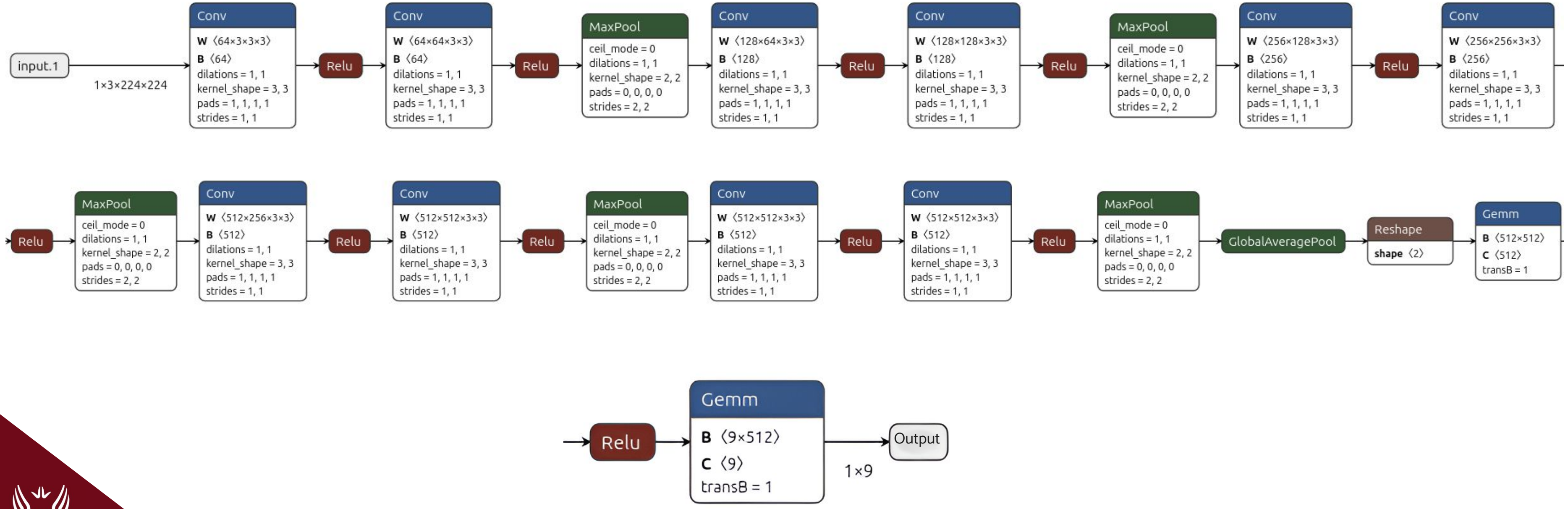
And then a **Global Average Pooling (GAP)** layer:

- A **fully connected** layer with 512 neurons and ReLU activation.
- A **dropout layer** (30%) for regularization.
- A final fully connected layer for classification into 9 classes.

Pretrained Models

ResNet50: *CNN*
Addresses gradient vanishing issues
DenseNet: *CNN*
Dense connections, reuses features to reduce parameters
MobileNet: *CNN*
Depthwise separable convolutions for mobile and resource-constrained devices.
VGG16: sequential *CNN*
3x3 convolutions and pooling layers
EfficientNetV2: *CNN*
*S*peed and accuracy, scales depth, width, and resolution
SwinTransformer:*Transformer*
Shifted windows for local and global relations
Vision Transformers (ViT): *Transformer*
Images as patch sequences to capture local patterns and global relationships

Baseline Model (Custom CNN)



Training Process Summary

1. **Pretrained Models:**
   Fine-tuned by replacing the classification layer to suit the target dataset, optimized using the Adam optimizer.

2. **Class Weights:**
   Calculated to address class imbalance, with **weights inversely proportional** to class **frequencies** in the training set.

3. **Loss Function and Optimizer:**
   Used **CrossEntropyLoss** integrated with class weights for balanced training. The **Adam optimizer** ensures efficient learning and regularization.

4. **Hyper Parameters:**
   Models trained for **15 epochs** to optimize performance. **Learning rate** (between 1e-5 and 1e-4) and **weight decay** (1e-2) tuned based on accuracy and loss.

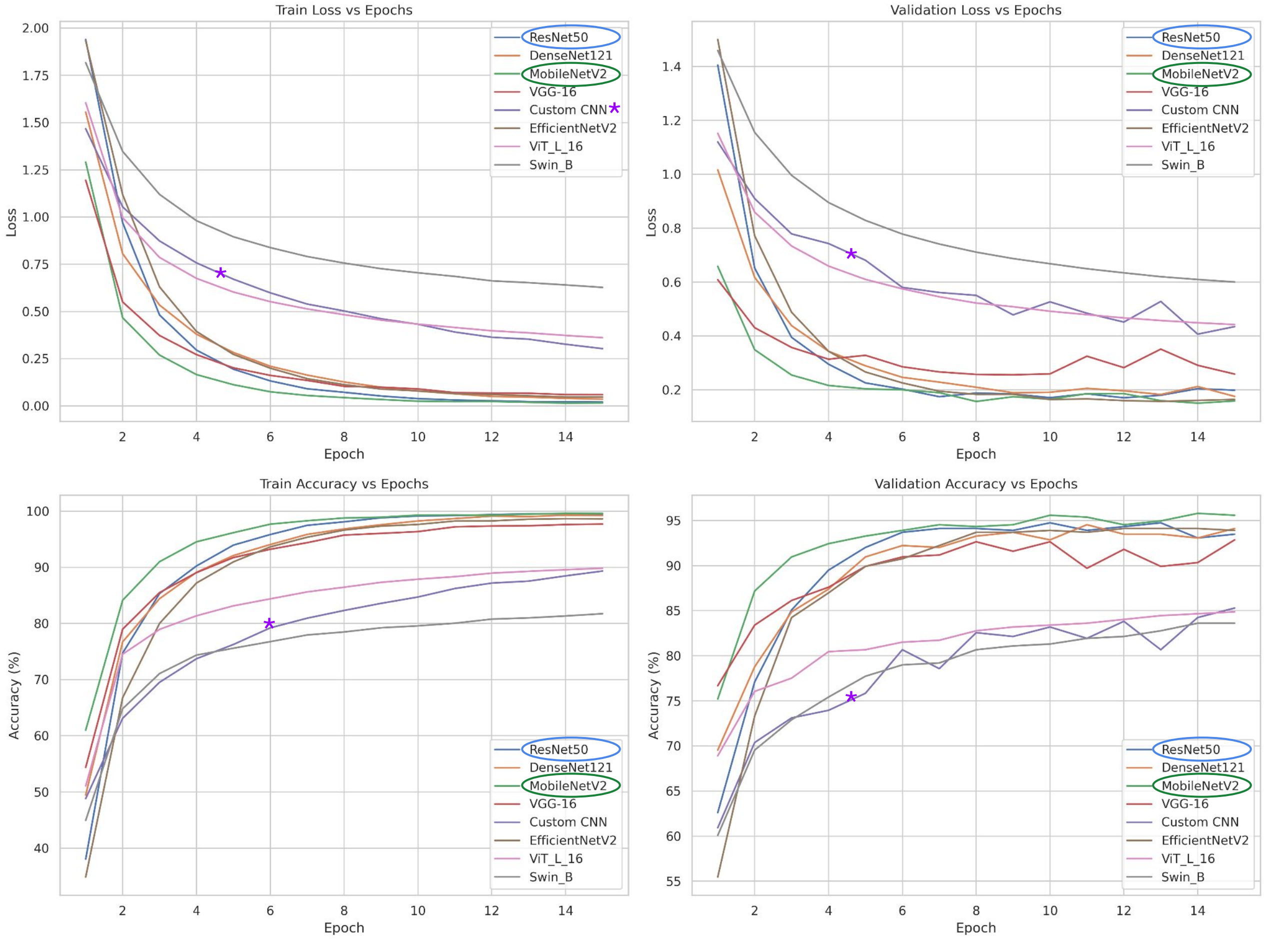
Class Weights

$$c_i = \text{number of samples of class } i \quad \forall i \in \{1, 2, \ldots, K\}$$

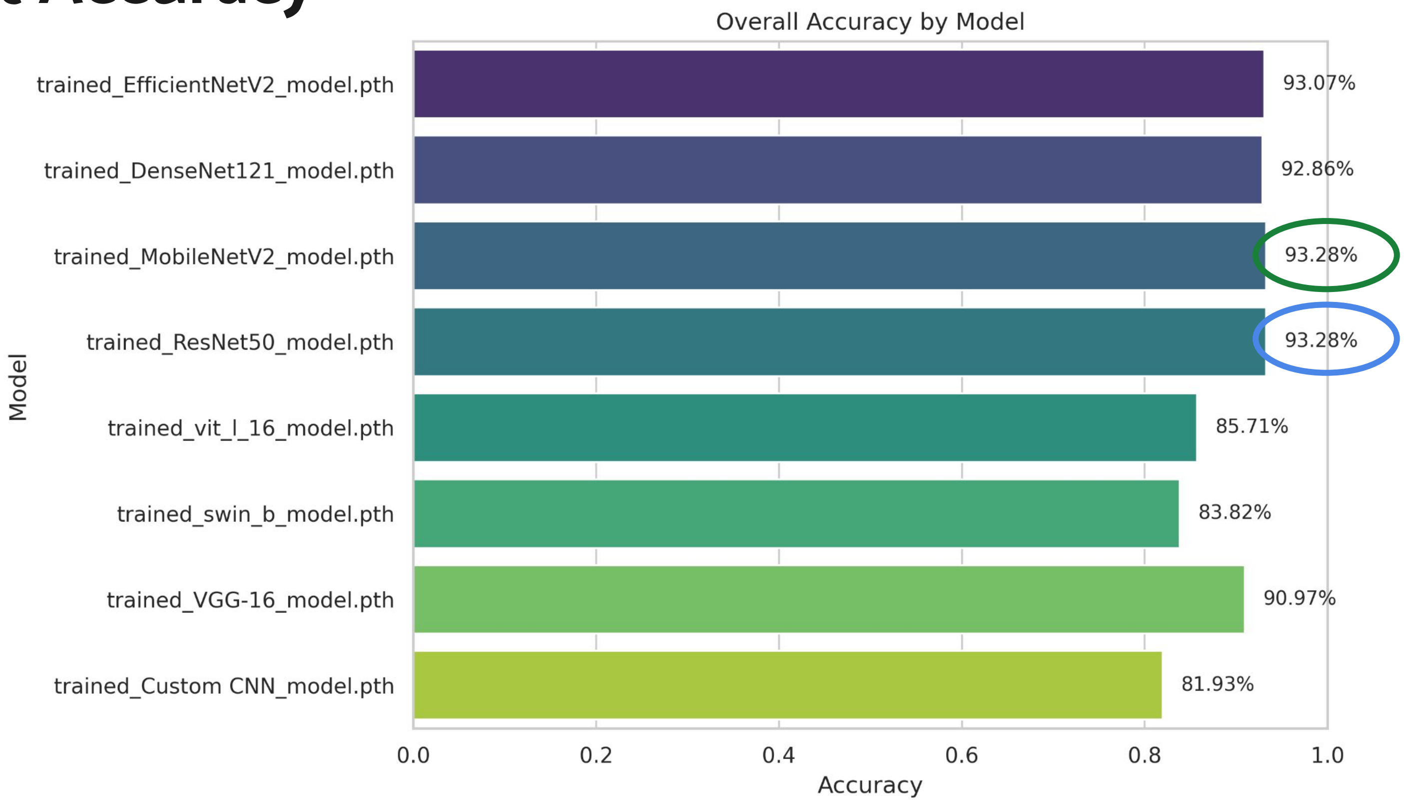
$$\boxed{w_i = \frac{\frac{1}{c_i}}{\sum_{j=1}^{K} \frac{1}{c_j}} \cdot K \quad \forall i \in \{1, 2, \ldots, K\}}$$

where K is the number of classes (9 in our case)

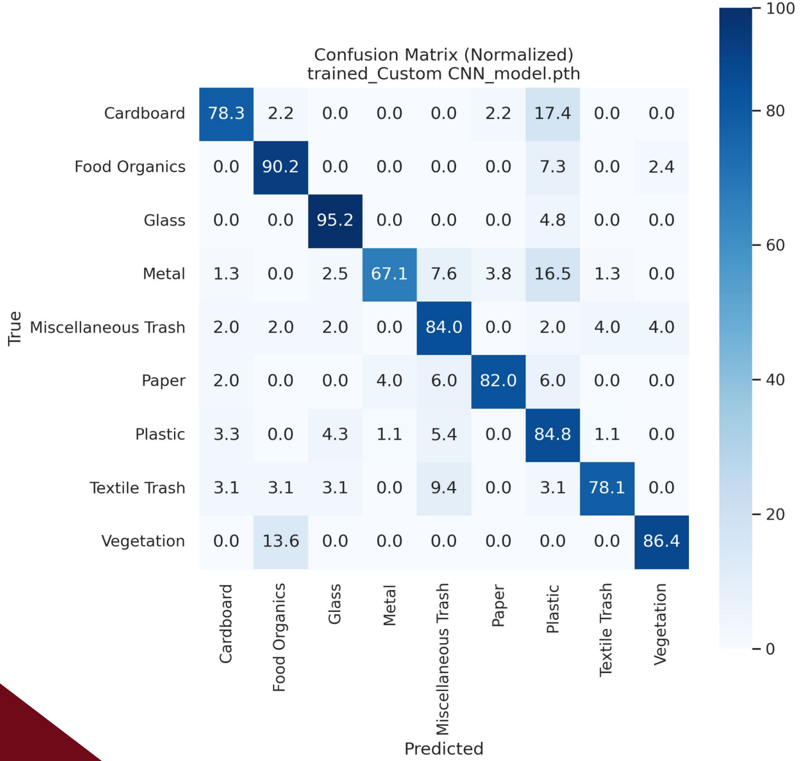
Results (Loss and Accuracy)

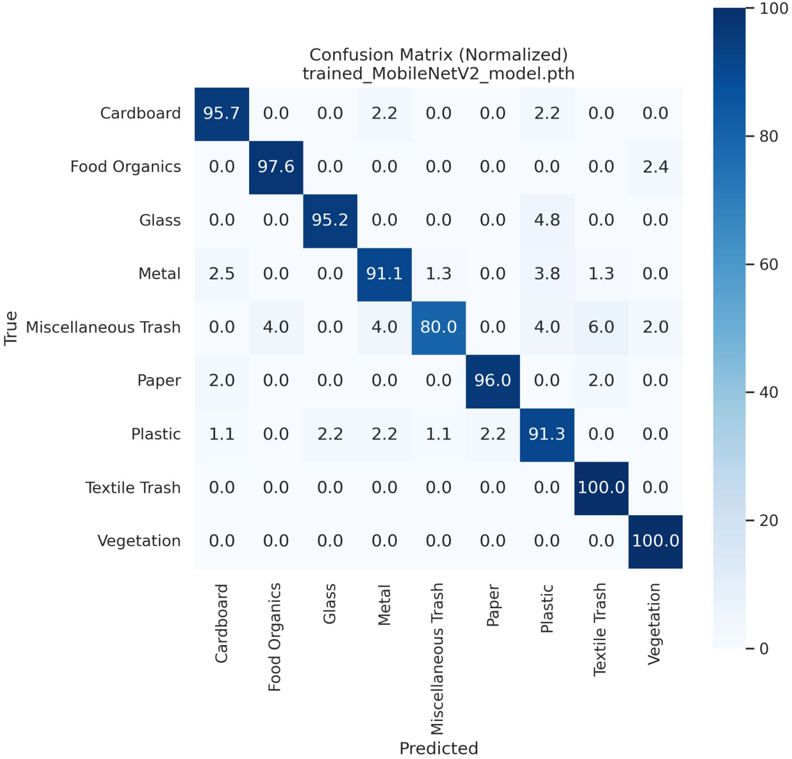


Test Set Accuracy

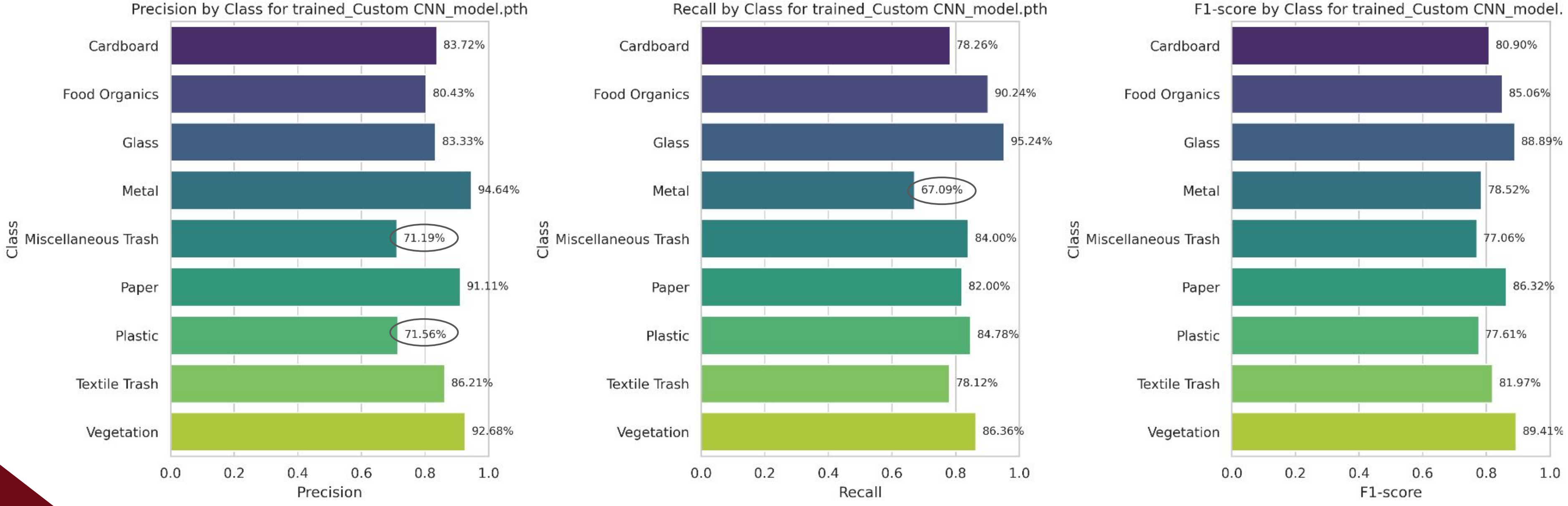


Test Metrics: Baseline vs MobileNetV2



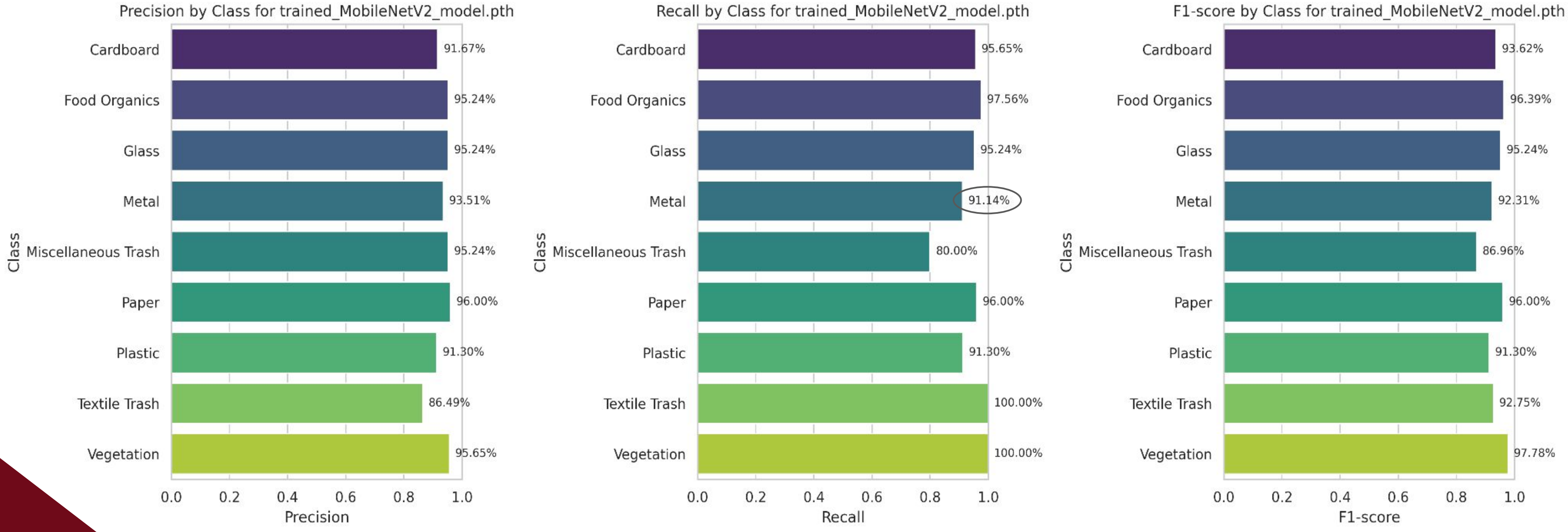


Test Metrics: Baseline (custom CNN)



Precision, Recall, F1-score for the Custom CNN

Test Metrics: MobileNetV2



Precision, Recall, F1-score for the MobileNetV2

Conclusions

- Classifying waste from a single image can be a challenging task (material diversity, varying conditions and lighting).
- **MobileNetV2** and **ResNet50** accuracy: **93.28%**.
- **Inception V3** accuracy: **89.19%** (best performing in [2]).
- **MobileNetV2** requires **less training time** than **ResNet50**.
- As expected, **Transformers**, which perform well on large datasets, do not deliver optimal results in our case study.
- *Miscellaneous Trash* is consistently misclassified, impacting all models; refining or removing it improves accuracy for well-classified categories like *Textile Trash* and *Vegetation*.

Bibliography

1. Dataset source: [RealWaste Image Classification];

2. Single, S., Iranmanesh, S., & Raad, R. (2023). RealWaste: A Novel Real-Life Data Set for Landfill Waste Classification Using Deep Learning. Information, 14(12), 633. https://doi.org/10.3390/info14120633;

3. Younis, H., & Obaid, M. (2024). Performance Comparison of Pretrained Deep Learning Models for Landfill Waste Classification. https://doi.org/10.14569/ijacsa.2024.0151166.

Thank you for the attention!