UNIVERSITY OF SCIENCE AND TECHNOLOGY OF HANOI
BACHELOR DATA SCIENCE (B2)

ICT3.011
MACHINE LEARNING & DATA MINING II

Instructor: Dr. Doan Nhat Quang

FINAL PROJECT REPORT

Clustering Household Waste Images Using Pre-Trained Deep Learning Models and K-Means Clustering

Student	Contact
Nguyễn Đông San	[sannd.22bi13388@usth.edu.vn]
Bùi Nguyễn Ngọc Huyền	**[huyenbnn.22bi13199@usth.edu.vn]**
La Thị Thùy Chi	[chiltt.ba12-26@st.usth.edu.vn]
Nguyễn Bích Thu	[thunb.22bi13418@usth.edu.vn]

Hanoi, June 2024

Table of contents

1. Introduction and Objective

In this report, we explore advanced clustering methods to improve waste image recognition. By combining pre-trained machine learning models with K-means clustering, we extract features and group similar waste images. Our analysis determines the optimal clustering approach for accurate classification, addressing challenges such as visual similarities.

This project evaluates ResNet, VGG16, and VGG19 with K-means clustering for waste image sorting. It aims to identify the most effective model by systematically comparing their ability to capture complex visual elements.

We align cluster labels with original true labels to assess clustering accuracy and analyze model performance, aiming to identify the model that best captures complex visual elements for accurate clustering. Ultimately, our goal is to support scalable and efficient waste sorting solutions.

2. Methodology

2.1. Dataset

The dataset used in this project is an image classification dataset of waste items collected within an authentic landfill environment called "realwaste". It consists of images representing 9 major material types found in household waste. However, for the purpose of this project, 6 common types of household waste were selected. The dataset contains images of waste items such as plastic bottles, paper, glass containers, food waste, carton, plant.

2.2. Framework of Architectural Design

This project combines pre-trained deep learning models (VGG16, VGG19, ResNet50) with K-means clustering to sort household waste images.

- **Feature Extraction:** Pre-trained models are employed to extract features from images, converting them into numerical arrays representing significant visual elements.
- **Shuffling and Label Retention:** The image arrays are shuffled randomly while retaining the original labels. This ensures unbiased model evaluation and comparison.

- **Dimensionality Reduction:** Principal Component Analysis (PCA) is applied to reduce the dimensionality of the extracted features. This enhances computational efficiency while preserving essential information.
- **Clustering with K-means:** K-means clustering is utilized to group the extracted features based on their similarity. This enables the identification of clusters of similar visual patterns within the image dataset.
- **Object Isolation for Performance Analysis:** Object isolation techniques are implemented to analyze how the performance of the clustering algorithm changes when specific objects or features are isolated.

3. Experiments and Results

3.1. Data Preprocessing

- **Image Resizing**: Resize images from 554x554 to 224x224 pixels for compatibility with pre-trained deep learning models.
- **Color Space Conversion**: Convert images from BGR to RGB color space to ensure consistency with image processing libraries.
- **Conversion to NumPy Arrays**: Convert images and labels into NumPy arrays, flattening 3D arrays into 2D arrays for feature extraction and clustering.
- **Image Normalization**: Normalize pixel values to a standardized range to reduce illumination variations and enhance feature learning.

3.2. Feature Extraction

Pre-trained deep learning models (VGG16, VGG19, ResNet50) extract features from waste images by processing them through the models, excluding the classification layers.

Steps

- Model Preparation: Load models with pre-trained weights.
- Image Input: Resize images to (224, 224, 3).
- Feature Extraction:VGG16/VGG19/ResNet50 - Process images to generate 3D feature maps, then flatten these maps into 2D arrays.

By comparing and clustering these feature vectors, the framework can identify similar waste items, enabling effective waste image clustering for household waste management.

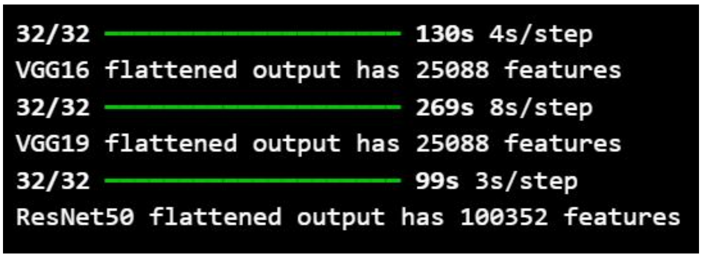


Figure 4.2.1: The training progress

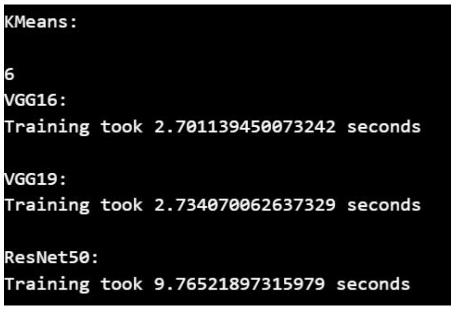


Figure 4.2.2: Time training K-Means clustering

3.3. PCA Analysis

PCA suggests that the number of features can be reduced by transforming original features into a smaller set of principal components that capture the majority of the variance in the data while retaining a significant portion of the variance.

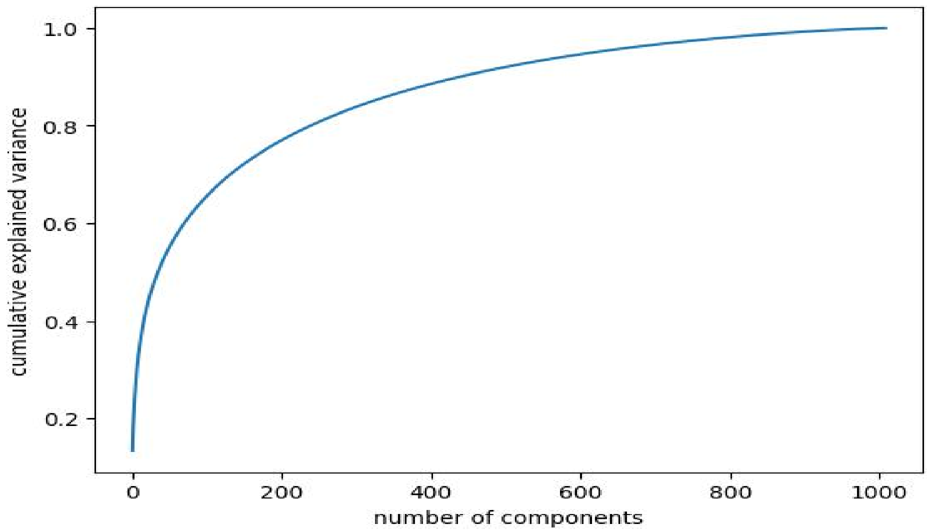
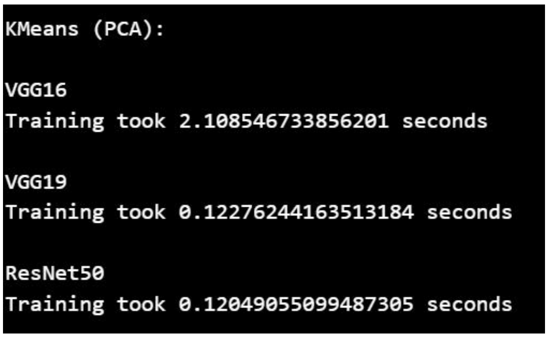


Figure 4.3.1: The cumulative explained variance of model VGG16



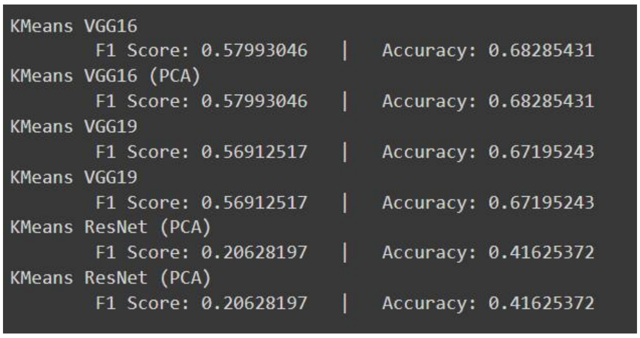
```
KMeans (PCA):

VGG16
Training took 2.108546733856201 seconds

VGG19
Training took 0.12276244163513184 seconds

ResNet50
Training took 0.12049055099487305 seconds
```

Figure 4.3.2: Time taken to train K-Means



```
KMeans VGG16
        F1 Score: 0.57993046   |   Accuracy: 0.68285431
KMeans VGG16 (PCA)
        F1 Score: 0.57993046   |   Accuracy: 0.68285431
KMeans VGG19
        F1 Score: 0.56912517   |   Accuracy: 0.67195243
KMeans VGG19
        F1 Score: 0.56912517   |   Accuracy: 0.67195243
KMeans ResNet (PCA)
        F1 Score: 0.20628197   |   Accuracy: 0.41625372
KMeans ResNet (PCA)
        F1 Score: 0.20628197   |   Accuracy: 0.41625372
```

Figure 4.3.3: F1 score and accuracy score

By reducing the number of dimensions, PCA can significantly speed up training times while remain the performances of models.

3.4. Label assignment and Model Evaluation

- Objective: To align the cluster labels with the true labels for better evaluation of the clustering performance.
- Steps:
    - Count Labels within Clusters: For each cluster, count the number of data points belonging to each true label. This helps understand the distribution of true labels within each cluster.
    - Determine the Most Frequent Label: For each cluster, identify the label that appears the most frequently. Assign this label to the entire cluster, assuming it represents the majority of data points.
    - Evaluate Cluster Performance: We perform model evaluation by comparing assigned labels to original true labels.

3.5. Cluster Evaluation Analysis

The optimal number of clusters is k = 4, offering a good balance of visual separation, quantization error reduction, and relatively higher silhouette score. Although the dataset has 6 labels, the model suggested 4 clusters instead of 6 because the clustering algorithm groups images based on feature similarity, and when different categories have overlapping visual characteristics(for example: carton and paper, plastic bottle and glass bottle) , they are combined into fewer clusters than the actual number of labels. This can be explained by the confusion matrix as belows

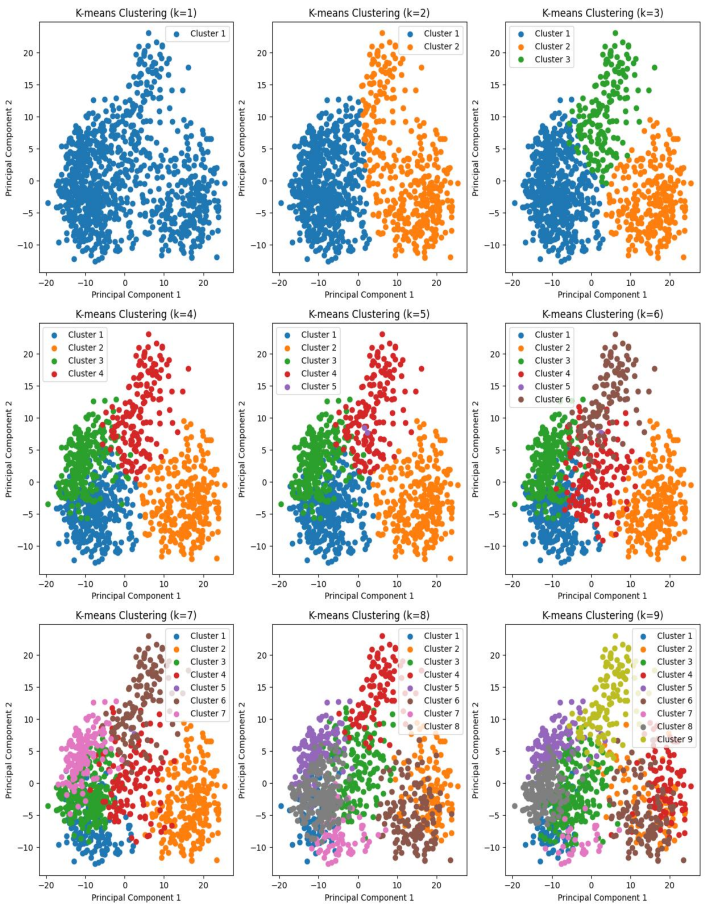


Figure 4.5.1: The result of applying K-means clustering on the dataset with k = 1 to 9 clusters on a 2D PCA-reduced feature space

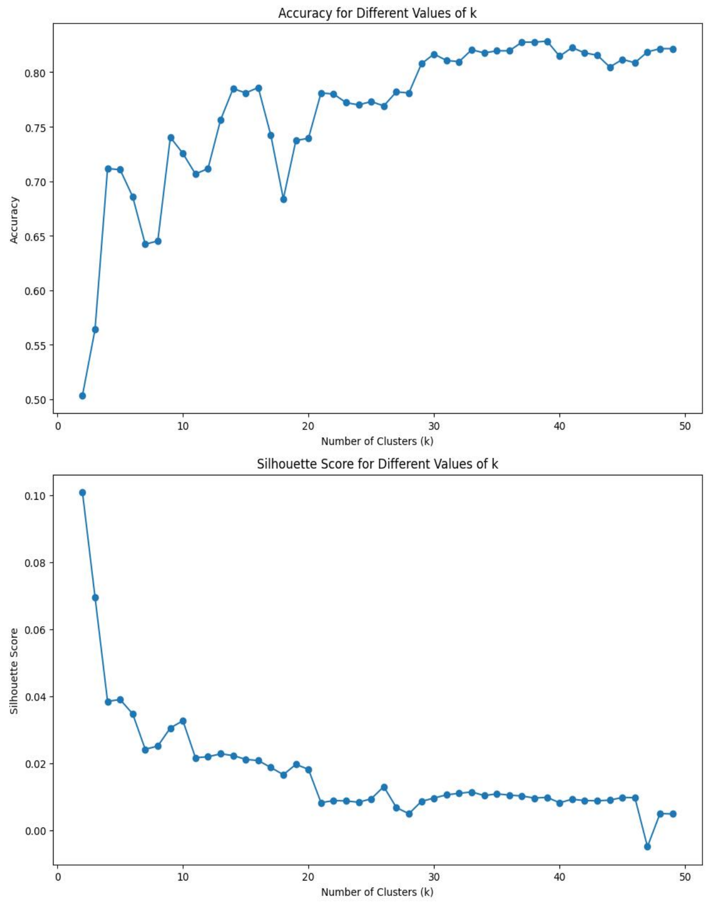


Figure 4.5.2: The accuracy of the clustering (top plot) and the silhouette score (bottom plot)

The confusion matrix provides insights into the waste image clustering model's performance across various waste categories:

Inter-class similarity, such as between Paper and Carton or Glass Bottle and Plastic Bottle, results in significant misclassifications, indicating the model's struggle to distinguish visually similar waste categories.

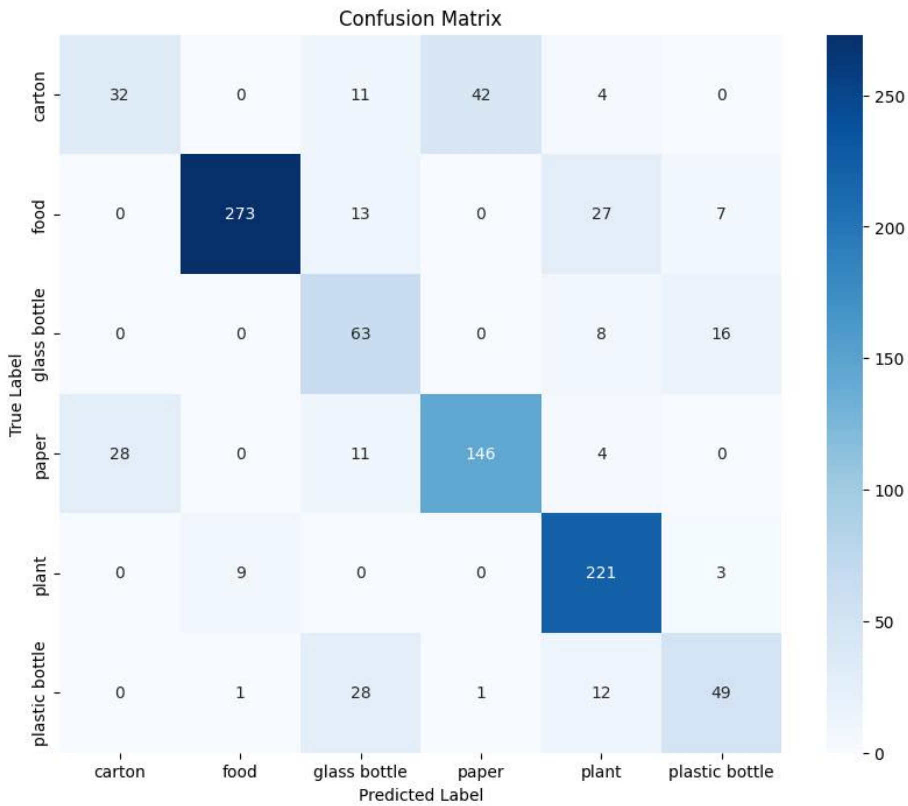


Figure 4.7.1: The confusion matrix of waste clustering

3.6. Object isolation

Object isolation is the process of identifying and separating individual objects within an image or a scene. This process improves dataset quality by providing clean and accurately labeled images. With isolated objects, machine learning models can focus solely on relevant features without being influenced by background clutter, resulting in more robust and accurate predictions.

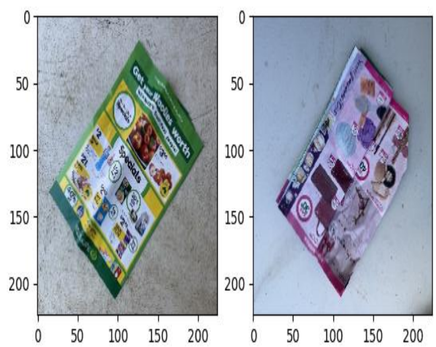


Figure 4.8.1: Two random images before object isolation

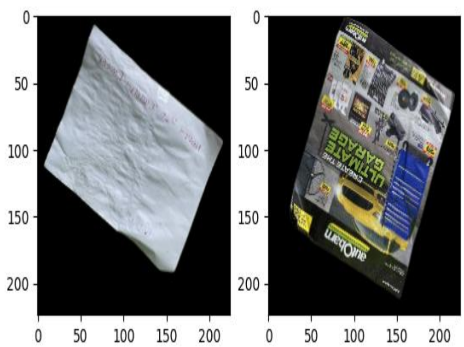


Figure 4.8.2: Two random images after object isolation

Feature Extraction and Model Comparison (After Object Isolation)

```
KMeans VGG16
	F1 Score: 0.57993046   |   Accuracy: 0.68285431
KMeans VGG16 (PCA)
	F1 Score: 0.57993046   |   Accuracy: 0.68285431
KMeans VGG19
	F1 Score: 0.56912517   |   Accuracy: 0.67195243
KMeans VGG19(PCA)
	F1 Score: 0.56912517   |   Accuracy: 0.67195243
KMeans ResNet
	F1 Score: 0.20628197   |   Accuracy: 0.41625372
KMeans ResNet (PCA)
	F1 Score: 0.20628197   |   Accuracy: 0.41625372
```

```
KMeans VGG16
	F1 Score: 0.44973897   |   Accuracy: 0.63545817
KMeans VGG16 (PCA)
	F1 Score: 0.44973897   |   Accuracy: 0.63545817
KMeans VGG19
	F1 Score: 0.46751677   |   Accuracy: 0.65936255
KMeans VGG19 (PCA)
	F1 Score: 0.46751677   |   Accuracy: 0.65936255
KMeans ResNet
	F1 Score: 0.19249090   |   Accuracy: 0.42928287
KMeans ResNet (PCA)
	F1 Score: 0.19249090   |   Accuracy: 0.42928287
```

Figure 4.8.3: The F1 score and accuracy score after Object isolation

The performance slightly dropped after applying object isolation because this technique may have inadvertently removed important contextual information that could aid in clustering. Additionally, simple thresholding methods were often insufficient, and more advanced segmentation techniques were computationally demanding.

4. Challenges and Limitations

- Dataset quality and size: The RealWaste dataset had image variability, class imbalance, and inconsistent resolutions, affecting feature extraction and clustering accuracy.
- Object isolation: Intended to enhance clustering by reducing background noise, but performance slightly dropped due to potential loss of contextual information and computational demands of advanced segmentation.
- Dependence on pre-trained models: VGG16, VGG19, and ResNet may not capture waste-specific features and produce high-dimensional feature vectors with potentially irrelevant information, impacting clustering.
- Overlapping Features between Classes: Certain categories may share similar visual features, causing the model to misclassify them.

5. Conclusion

- Pre-trained models: VGG16 và VGG19 are suitable for this project and VGG16 has slightly better performance.

- Challenges: Computational constraints, dataset variability, object isolation difficulties, and limitations of k-means clustering were encountered. Label alignment for accuracy measurement added complexity.
- Suggested improvements: Create or augment datasets with more diverse and balanced waste images to improve the generalizability and robustness of the models.

6. References

- Source code: https://drive.google.com/drive/u/1/folders/1rg5JmIXI-edM41pYsYJMZt-eNMMTHE3w
- Reference code: GitHub
- Ruiz, V., Sanchez, A., Velez, J. F., & Raducanu, B. (2019). Automatic Image-Based Waste. *Springer Nature Switzerland*, 422–431. https://doi.org/10.1007/978-3-030-19651-6
- Torija, A. J., & Ruiz, D. P. (2015). A general procedure to generate models for urban environmental-noise pollution using feature selection and machine learning methods. *Science of the Total Environment*, *505*, 680–693. https://doi.org/10.1016/j.scitotenv.2014.08.060