Monster Recognition and Classfication

Elia Avanzolini, Alejandro Enrique Barbi, Martina Panini

February 13, 2025

Abstract

This project focuses on automating the recognition of Monster Energy cans using advanced image processing and machine learning techniques. The workflow includes preprocessing images through random transformations to enhance dataset diversity, extracting visual features such as color histograms (RGB and HSV) and texture metrics based on the Gray Level Co-occurrence Matrix (GLCM), and training a classification model using a Multi-Layer Perceptron (MLP).

Can detection within images is performed using the Roboflow service, which identifies and automatically crops cans through bounding boxes. The cropped images are then classified based on color and text present on the cans, with results combined to improve overall recognition accuracy.

The results highlight how the combination of color and textual features can significantly enhance system performance.

1 Introduction

The project aims to develop a classifier for Monster Energy drinks by focusing on finding the best techniques for constructing an accurate bounding box around the can in the image.

We explore a series of computer vision methods aimed at identifying Monster Energy cans in images and determining their flavour. Our approach begins with classical computer vision techniques, such as bounding box detection using OpenCV[Bra00]. Specifically, we employ the template matching function, which identifies the location of cans based on predefined templates. This method is particularly effective in controlled environments where the appearance and positioning of the cans are relatively consistent. The problem we encountered is that our dataset does not only contain images with such characteristics. Therefore, a pre-trained model was needed to extract the bounding boxes from the images. This model delivers more accurate localization and achieves higher classification accuracy compared to the classical methods.

Then, we incorporate text recognition to enhance differentiation between various Monster Energy can types by reading printed labels or logos. This step allows for greater specificity in identifying the exact product variant.

Furthermore, there has been added a classification that takes into account the color of cans.

2 Datasets

The dataset used, taken from kaggle[Kag24], is composed of 19 classes, representing different type of cans immages, divided in training and test set. The data are user-generated. In fact, there was a great diversity in the dataset, such that some photos were totally on the can, while others had broader contexts like individuals or other objects that formed the background, making bounding box dectection and classification even more complicated.

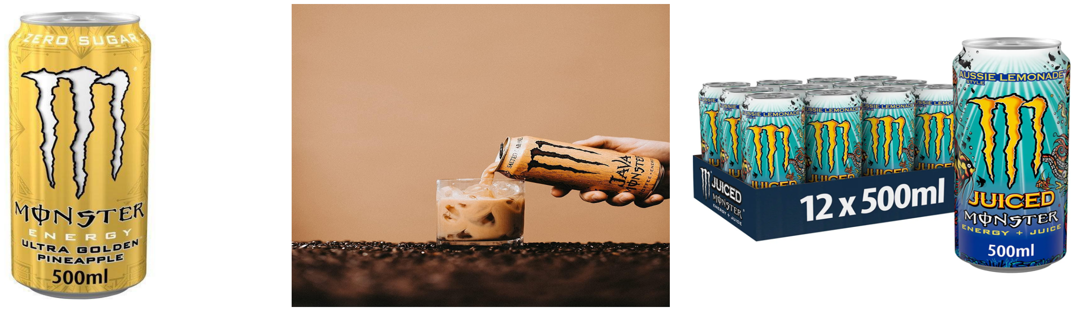

Figure 1: Examples of images taken from Monster Dataset

In order to enrich the dataset, other photos are taken. The number of classes has been increased by this operation to 38.

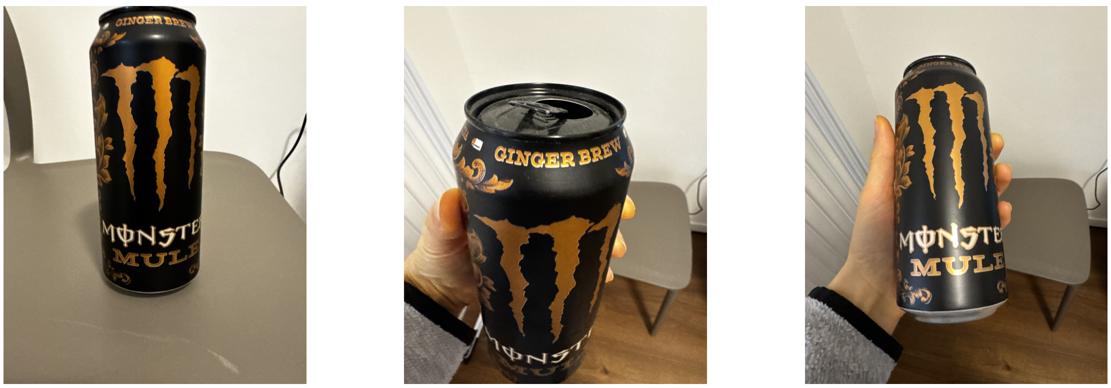

Figure 2: Examples of captured images

3 Methodology

The methodology of this project integrates several stages of template matching and find bounding boxes, image processing, feature extraction, object detection, and machine learning classification to achieve accurate recognition of Monster Energy cans.

This section explains in detail how pre-processing was done for every train image used in the classification task, then present and discuss in detail the work for each method.

3.1 Template Matching

Template matching is employed here as the principal method for segmentation of the regions of interest. One image from the training dataset was taken and settted as a template to find match regions in other images. It is so that Monster Energy cans are be easily distinguished. The function matchTemplate provided by OpenCV was used to compare the template to regions of the scene. The function normally returns a score of similarity as a result of comparing a few different methods among each region in the scene about the template. We can thus work out the possible matches using this technique. To cope with the possible variability in both rotation and scaling, a set of variations of the template was generated thanks to space of rotations and scales. In this procedure, firstly, the template is rotated over angles ranging from 0 up to 360 degrees and then it is scaled from 50% to 150% of its original size. Now, the next step is to find the best match of the template in the image, by itereating the for all possibile combination of parameters until the best match, based on a similarity scored, is found. Once founded a bounding box was drawn around the matched area of the scene, obtaining an output image showing where the can is possible located. However, the bounding boxes drawn are decent only in simpler scenarios, such as detecting a single can, but they become really poor when the scenario becomes more complex, such as the case where multiple cans appear in the same image, or where other subjects-such as a person-or complicate backgrounds are present. In such cases, the accuracy and precision of the bounding box placement are significantly reduced.

Figure 3: An example of good bounding box with Template Matching

3.2 Pre-Trained Model

The detection of bounding boxes around Monster Energy cans is a critical step in the recognition pipeline, ensuring that only the relevant portions of the images are processed for classification. This task is handled by a Roboflow API to perform object detection with high accuracy.

The script utilizes the Roboflow Inference API, a cloud-based service for object detection, to identify and localize Monster Energy cans within the images. The specific model used for detection is identified as `"tin-can-r0yev/1"`, which is trained to recognize the shapes and characteristics of cans. The `JSON` response from Roboflow includes information about each detected object, such as the coordinates of the bounding box, its dimensions, and the confidence score of the detection. To draw and crop bounding boxes around detected cans, the script translates the center-based coordinates and dimensions into pixel coordinates for the top-left and bottom-right corners of the bounding box.

Once the bounding box coordinates are calculated, the script crops the detected cans from the original image based on the bounding box coordinates. The cropped images are stored for further processing and classification.

Through this robust bounding box detection process, the script ensures that only the relevant portions of the images (the Monster Energy cans) are used for feature extraction and classification, significantly improving the accuracy and efficiency of the recognition system.

3.3 Pre-Process

After the extraction of bounding boxes, the new dataset is made only by images showing cans only. The next step involves preprocessing the images to improve model generalization and robustness.

In creating the training dataset, a pre-prossessing pipeline used a sequence of random transformations on every image so as to improve the dataset generalization. Each randomly selected image is applied:

- Horizontal Flipping: Randomly flips images horizontally to introduce variations in orientation.
- Rotation and Scaling: Applies random rotations (between -30° and 30°) and zooms (scaling between 0.8 and 1.2) to simulate different perspectives.

- Translation: Shifts the images up to 10% of their dimensions to vary object positioning.
- Contrast and Brightness Adjustment: Randomly adjusts contrast and brightness to simulate different lighting conditions.
- Gaussian Blur and Denoising: Introduces blur and noise reduction to make the model robust to real-world image imperfections.
- Histogram Equalization: Enhances image contrast by equalizing the histogram on the luminance channel.

At the end of this process the new train dataset is made by original images and processed images.

3.4 Color Classification

The classification of Monster Energy cans based on color is a multi-step process that involves the extraction of detailed color and texture features from images, followed by the application of a machine learning model for classification.

3.4.1 Features Extraction

The first step in color-based classification involves extracting relevant features from the images, which are used to represent the unique color characteristics of each can.

- RGB Histograms: The script computes histograms for each of the Red, Green, and Blue channels, capturing the distribution of primary colors within the image. Each channel is divided into 32 bins, resulting in 96 features (32 bins $\times$ 3 channels).
- HSV Histograms: To capture more nuanced color information, the images are also converted to the HSV (Hue, Saturation, Value) color space. Histograms are computed for each channel, providing an additional 96 features that account for variations in hue and intensity, which are crucial for distinguishing cans with similar RGB compositions but different tones or brightness levels.
- GLCM Texture Features: Beyond color, the script extracts texture features using the Gray Level Co-occurrence Matrix (GLCM). Five key texture properties: contrast, dissimilarity, homogeneity, energy, and correlation—are calculated from the grayscale version of the image. These features help in differentiating cans not only by color but also by the unique textural patterns on their surfaces.

All these features are concatenated into a single feature vector, representing both color distribution and texture, and are saved for model training.

3.4.2 Classification

Once the features are extracted, they are used to train a classification model. The classification process follows these steps:

1. Data Standardization: The extracted features are standardized using StandardScaler to ensure consistent scaling across all features, which is critical for the effective training of neural networks.
2. Label Encoding: Can labels, representing different Monster Energy flavors or types, are encoded into numerical values using LabelEncoder, preparing them for the machine learning model.

3. Model Training: The core of the classification is handled by a Multi-layer Perceptron (MLP) classifier. The MLP is configured with two hidden layers, each containing 50 neurons, and uses the ReLU activation function for non-linearity. The model is trained with an adaptive learning rate and early stopping to prevent overfitting. The training process adjusts the model weights to minimize classification errors based on the color and texture features.

4. Classification: After training, the model is used to classify new images based on their color features. The predictions are decoded back to their original labels using the LabelEncoder, and confidence scores are provided for each classification.

5. Evaluation: The model's performance is evaluated using accuracy metrics and classification reports, providing insight into how well the color features distinguish between different can types. The evaluation results are saved for further analysis.

3.5 Text Recognition

EasyOCR[JM21] was employed to perform Optical Character Recognition (OCR), which is used to detect flavor names and branding details by extracting text, playing an essential role in classifying Monster cans. However, the images first undergo a preprocessing stage to enhance text readability.

The preprocessing begins with converting images to grayscale, followed by binarization to increase contrast and make text clearer. To further refine text visibility, different filtering techniques are applied. Specifically:

- **Gaussian Blur** is used to smooth the image and reduce noise before OCR is performed. This helps in enhancing the clarity of the text.

- **Laplacian Filter** is applied to detect edges in the image, making textual regions stand out against the background.

- **Median Filter** is particularly effective in preserving text structure while reducing noise. It is applied before Sobel edge detection to ensure the text remains intact with minimal artifacts.

- **Sobel Filter (X and Y directions)** is then used on the median-filtered image to further enhance edges, improving the distinction of text from the background.

The OCR process is applied to targeted regions of interest where textual information is most likely to appear. This ensures efficiency and precision. Additionally, a filtration step removes text with low confidence scores or disproportionately large bounding boxes to minimize errors.

Post-processing further refines results by matching OCR outputs against a predefined vocabulary of Monster Energy flavor names. This is done using string-matching techniques to correct inaccuracies.

The results are visualized by overlaying bounding boxes and recognized text onto the original images, demonstrating the effectiveness of this pipeline in accurately extracting and validating textual information.

4 Results

To further enhance classification accuracy, the color-based predictions are combined with text recognition results. This dual approach ensures that cans with similar colors but different labels can be accurately identified, leveraging both visual and textual information for robust can recognition.
The classifications made on a test image (Fig.4) is shown below:

Figure 4: Test image

Image	**Text**	**TextAcc**	**Color**	**ColorAcc**	**Label**
crop2	PACIFIC PUNCH	80.0	java salted caramel	36.67	Pacific Punch
crop3	KHAOTIC	77.14	java salted caramel	31.12	Khaotic
crop1	JAVA TRIPLE SHOT	72.73	java salted caramel	36.10	Java Triple Shot
crop0	ROSA	47.06	ultra red	47.67	Red
crop4	ASSAULT	69.23	ultra black	26.74	Assault
crop5	ROSA	60.0	ultra black	38.03	Rosa
crop7	MONARCH	83.08	java salted caramel	41.58	Monarch
crop6	MANGO LOCO	85.5	ultra blue	60.37	Mango Loco
crop8	RED	68.4	ultra rosa	54.74	Red

Table 1: Color and Text Classification Results

Results of classification are resumed in a `.txt` file:

```
COLOR AND TEXT CLASSIFICATION RESULTS:
                  Image      TextFilter        Text TextAccuracy Color Prediction  Color Accuracy
0  tris2_crop0.png  Original Image     KHAOTIC    77.142857          Khaotic           98.25
1  tris2_crop1.png  Original Image     MONARCH         90.0          Monarch           89.46
2  tris2_crop2.png  Original Image  MANGO LOCO         90.0       Mango Loco           98.31

MONSTER THAT YOU HAVE:
Khaotic
MONARCH
Mango Loco
```

Figure 5: Example of file that show the result for the image tris2

Figure 6: Image tris2

Tests can be performed on all images in the Images folder. Each time you test an image, the predictions are different, never quite accurate. This is because there is a margin of improvement in both color classification and text recognition.

5 Conclusion

In conclusion, this project successfully demonstrated the potential of integrating advanced image processing and machine learning techniques for the automated recognition and classification of Monster Energy cans. By combining color-based features, texture analysis, and text recognition, we were able to create a robust system capable of accurately identifying different can types, even under challenging conditions such as varying lighting, complex backgrounds, and diverse image compositions.

The preprocessing stage, which applied random transformations like rotations, scaling, and contrast adjustments, significantly improved the model's ability to generalize across a diverse dataset. This step was crucial in simulating real-world scenarios, making the system more resilient to variations in image quality and can positioning. Additionally, leveraging the Roboflow API[Rob24] for object detection allowed us to achieve highly accurate bounding box detection, addressing the shortcomings of traditional methods like template matching, which struggled in more complex scenes with multiple objects or non-uniform backgrounds.

The feature extraction process, which included RGB and HSV histograms along with texture metrics derived from the Gray Level Co-occurrence Matrix (GLCM), provided a comprehensive representation of each can's visual characteristics. This rich feature set enabled the Multi-Layer Perceptron (MLP)[PVG+11] classifier to achieve respectable levels of accuracy, particularly when combined with the results from the text recognition pipeline powered by EasyOCR. The dual approach of color and text classification proved to be a powerful strategy, especially for distinguishing cans with similar color schemes but distinct branding or flavor labels.

Despite these successes, the project also highlighted several areas for improvement. The classification accuracy, while promising, showed inconsistencies when tested across different images, indicating the need for further optimization. The color classification model, in particular, could benefit from fine-tuning the feature extraction process or exploring more sophisticated deep learning architectures.

Similarly, the text recognition pipeline, although effective, sometimes struggled with low-contrast text or distorted fonts, suggesting the potential for enhancements through more advanced OCR techniques or post-processing methods.

Future work could focus on expanding the dataset to include a wider variety of can images under different environmental conditions, which would help improve model generalization. Additionally, exploring ensemble methods or hybrid models that combine convolutional neural networks (CNNs) with traditional feature-based approaches could further boost classification performance. Finally, integrating real-time detection capabilities or deploying the model in mobile or embedded systems could open up practical applications for automated product recognition in retail or inventory management contexts.

Overall, this project lays a solid foundation for future research and development in the field of object recognition, demonstrating the value of combining multiple machine learning techniques to tackle complex classification tasks.

References

[Bra00] Gary Bradski. The opencv library. In *Dr. Dobb's Journal of Software Tools*, 2000.

[JM21] S. Jaeger and M. A. Mermet. Easyocr: Scene text detection and recognition, 2021. Retrieved from https://github.com/JaidedAI/EasyOCR.

[Kag24] Kaggle. Monster energy cans dataset, 2024. Retrieved from https://www.kaggle.com.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[Rob24] Roboflow. Roboflow api documentation, 2024. Retrieved from https://roboflow.com.