UNIVERSITÀ DI PAVIA

Machine Learning Course

Cake Classification: a CNN + MLP approach

Andrea Alberti

Department of Computer Engineering - Data Science

University of Pavia, Italy
Email: andrea.alberti01@universitadipavia.it

February 25, 2024

Abstract

This project aimed to develop accurate classification models for 15 different cake images using two approaches: Multi Layer Perceptron (MLP) and a pre-trained Convolutional Neural Network (CNN) based on PVMLNet, a simplified version of AlexNet. The models were evaluated using various feature types, including low-level features such as color histogram and co-occurrence matrix, as well as combinations of these features. Additionally, neural features extracted from different layers of PVMLNet were explored. The results demonstrate that the best accuracy of 90% was achieved by utilizing neural features, showcasing the power of leveraging pre-trained CNNs for image classification tasks. In contrast, the utilization of low-level features yielded a comparatively lower accuracy of 31%. These findings highlight the importance of utilizing deeper representations learned by neural networks, which can capture complex patterns and nuances in the cake images.

Contents

1 Introduction

This project focuses on implementing a Neural Network for classifying images of cakes. To tackle this problem, two neural networks are utilized: a Convolutional Neural Network (CNN) and a Multi Layer Perceptron (MLP). Additionally, the project compares and evaluates different types of features employed by each network. The primary library employed for the implementation is "pvml."

1.1 Available Data

The dataset consists of 15 categories of cakes, with each category containing 120 images. Within each category, 100 images are allocated for the training set, while the remaining 20 images form the test set. All the images in the dataset have been resized to a uniform size of 224 x 224 pixels.

1.2 Goal

The objective of this project is to identify the most effective method for extracting features from the cake images. The selected feature extraction technique will then be utilized to train a Neural Network capable of accurately classifying new, unseen data into the 15 different cake classes.

1.3 CNN and MLP

A Multi-Layer Perceptron (MLP) is a type of Feed-Forward Neural Network. It comprises an input layer, an output layer, and zero or more hidden layers consisting of interconnected neurons. In an MLP, each neuron in a layer receives input signals from all neurons in the previous layer. It then applies an activation function to the weighted sum of its inputs and passes the result to the next layer. During training, the parameters of the MLP, which include weights and biases, are learned by optimizing a chosen Loss Function, typically Cross Entropy. This optimization process is performed using Backpropagation, where the derivatives are propagated backward through the network, adjusting the parameters iteratively. On the other hand, a Convolutional Neural Network (CNN) is specifically designed for processing and analyzing visual data, such as images. CNNs consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers apply filters to the input data, capturing spatial patterns and features. The pooling layers downsample the feature maps, reducing their dimensionality. Finally, the fully connected layers combine the extracted features to make predictions. The activations of different layers can be used as input features for a classification network such as an MLP.

2 Features Extraction

Before proceeding with the model construction, it is fundamental to pre-process the data to make them suitable for the algorithmic processing. This preprocessing step, known as "Feature Extraction," is essential in the model creation process. In this project, two primary categories of feature extraction methods are compared: "Low-Level Features" and "Neural Features." The first category involves extracting specific features directly from the images in an 'handcrafted' way. On the other hand, the second category utilizes a pre-trained neural network to extract features from the images. In this case the used CNN is the *PVMLNet*, designed as a slight simplification over the *AlexNet* architecture. These two approaches are compared to determine the most effective feature extraction method for classifying cake images.

2.1 Low Level Features

There are many types of low level features and in this project different combinations of the following ones are considered: Color Histogram, Edge Direction Histogram, and Co-occurrence Matrix.

2.1.1 Color Histogram

A color histogram is a frequency representation of the color distribution in an image and can be used as a quantitative representation of the image. In Figure 1, the test and train accuracies of a multilayer perceptron trained on color histogram features are depicted. The model was trained for 5000 epochs, achieving a test accuracy of approximately 21% while exhibiting a slight overfitting tendency. The growth of the test accuracy appears to plateau after 1000 epochs, suggesting that the model struggles to extract further information from the data. This limitation can be attributed to the simplicity of color histogram features, which may not adequately capture the intricate complexities present in the images.

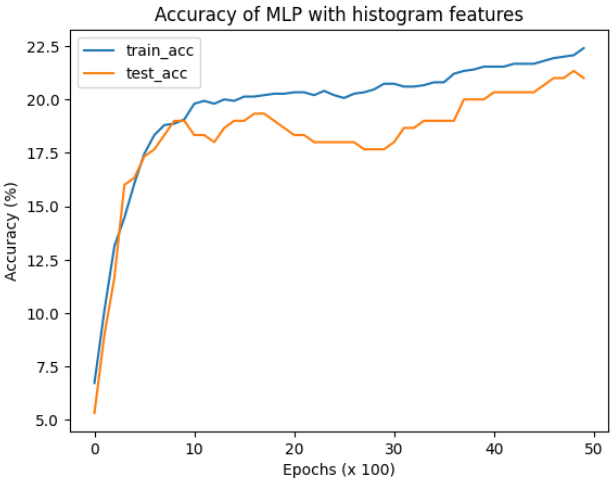

Figure 1: Color Histogram

2.1.2 Combined Features

To try enhancing the model performance, the color histogram features were combined with other low-level features, listed in Table 1. To merge these features, they were normalized using three different techniques: "Mean-var," "Min-max," and "Max-abs."

Feature Extraction Combinations
Color Histogram + Edge Direction Histogram
Color Histogram + Co-occurrence Matrix
Edge Direction Histogram + Co-occurrence Matrix
Color Histogram + Edge Direction Histogram + Co-occurrence Matrix

Table 1: Feature Extraction Combinations

The results are presented in Figure 2, showing the train and test accuracies achieved for each feature type using various normalization techniques and without normalization. Notably, the combination of Color Histogram and Edge Direction Histogram yielded similar results to the combination of all three feature extraction techniques, while demonstrating a reduced tendency for overfitting. The introduction of normalization techniques improved the test accuracy, with the "Minmax" and "Max-abs" methods outperforming "Mean-var" and achieving an approximate accuracy level of 31%. However, despite the enhanced performance from the feature combination, the results remain unsatisfactory, as the model still struggles to accurately classify the images. As a potential attempt for improvement, the next step is to consider neural features.

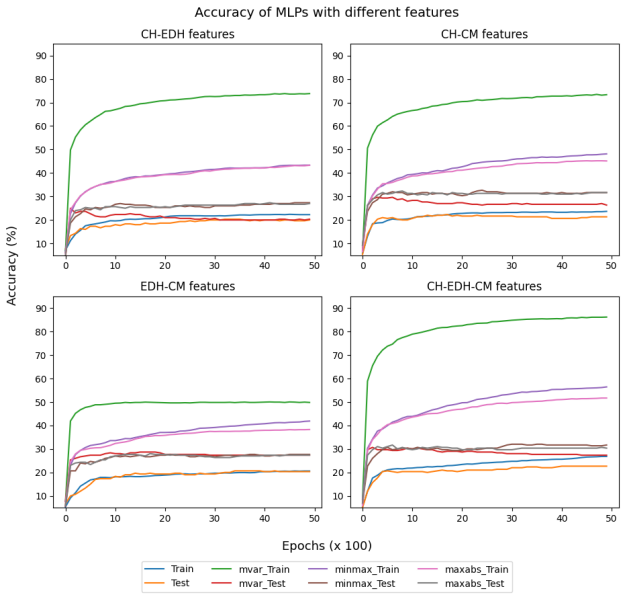

Figure 2: Feature Extraction Combinations Results

2.2 Neural Features

In contrast to low-level features, neural features are obtained by utilizing a pre-trained neural network. These features are derived from the activations of different layers within the network, which can serve as valuable inputs for the classification task. In this project, the CNN layers considered for feature extraction are denoted as -1, -2, -3, -4, -5, -6, and -7, where the numbers represent the layer indices in Python, with -1 representing the last layer.

Different Activation Layers

In Figure 3 are shown the results of the accuracies achieved by the MLP trained with different activation layers as features. Two ways of extracting the features have been considered. the first involved making a spatial mean across the values, the second one instead just considered all the values, flattening them. The best approach in terms of results were this latter and it was chosen for the following analysis. It is worth nothing that this approach requires an higher computational cost since the number of features is much higher. The layers -5 and -6 were observed to be the most effective, achieving a test accuracy of approximately 90%. This value is significantly larger than the one obtained using low-level features, proving the effectiveness of the neural features. Another proof of their bounty is the extremely reduced number of epochs it was necessary to reach a plateau in the accuracy growth.

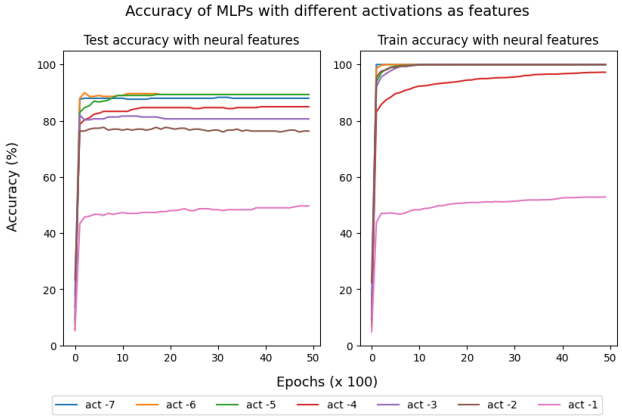

Figure 3: Different Activation Layers

3 Transfer Learning

In the context of this project, transfer learning is applied by replacing the last layer of PVMLNet, with the weights of a trained MLP. By doing so, the PVMLNet can benefit from the learned weights of the perceptron, potentially improving its performance and enabling it to classify cake images more accurately. This transfer of knowledge allows for efficient utilization of the previously learned features and can expedite the training process for this specific classification task. The reached test accuracy is around 80%, largely smaller than the 90% achieved before.

4 Model Analysis

To analyze the behavior of the chosen model ('-5' layer neural features) was used the confusion matrix in figure 8 showing the number of correct and incorrect predictions for each class. The darker the color, the higher the number of samples of class i (row index) classified as class j (column index). As expected, the diagonal is the most populated part of the matrix, meaning that the model correctly classifies the images most of times.

4.1 Most Exchanged Classes

The most exchanged classes are those in which the model has more difficulties to distinguish between. In figure 4 are reported for each cake the class to which the model classifies it (excluded correct classifications) most of times. An investigation on the most exchanged classes could be useful to understand the reason behind the misclassifications and to improve the model performance. The factors affecting the choice are a lots, including the shape of the cake, the colors and even the texture.

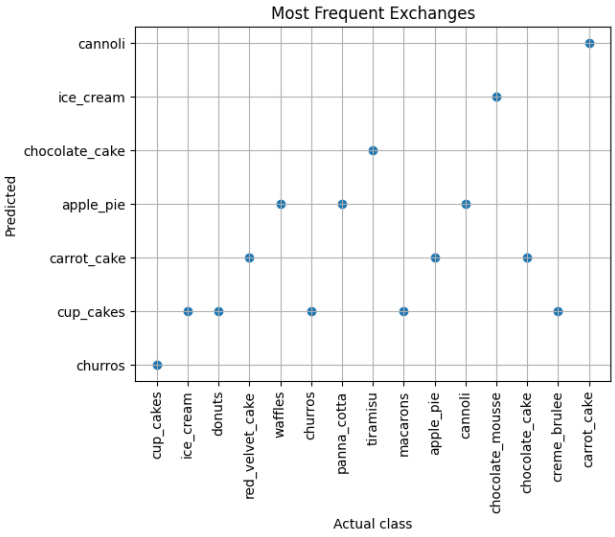

Figure 4: Most exchanged

4.2 Most Misclassified Samples

Some cakes are more difficult to be classified than others and in figure 5 are reported the most misclassified samples. *Chocolate-mousse*, *Apple-pie* and *Tiramisu* are the most misclassified cakes, and they are exchanged respectively with *Ice-cream*, *Carrot-cake* and *Chocolate-cake*. To understand the reason behind these misclassifications, further investigations are needed.

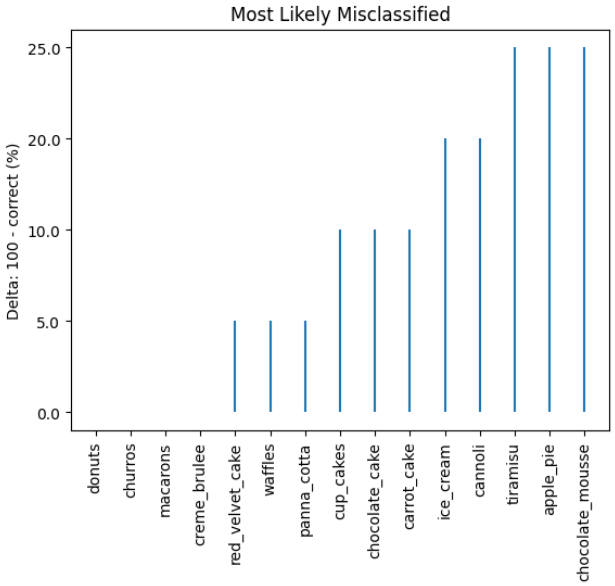

Figure 5: Most misclassified

4.3 Misclassified Images

A last interesting and useful analysis is to look at the images that the model misclassifies and to their predicted classes. In figure 6 are reported 4 images misclassified by the model. The title indicate the actual class, while the bars indicate the probability of the 5 most likely classes predicted by the model. Understanding the reason behind these misclassifications could be not so easy, but achieving this goal could be useful to improve the model performance. To provide an enhanced vision on the misclassified cakes, in Figure 7 are reported their actual images.

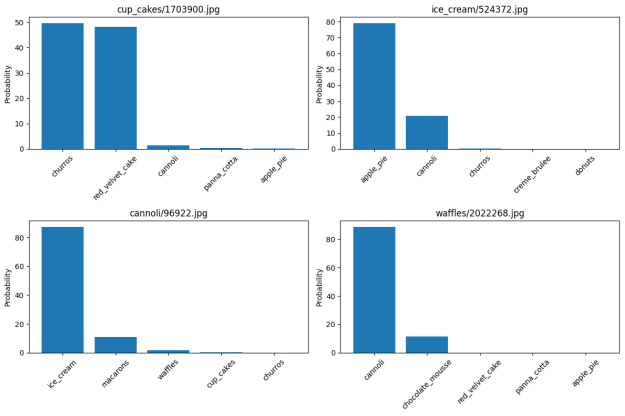

Figure 6: Predictions for the images

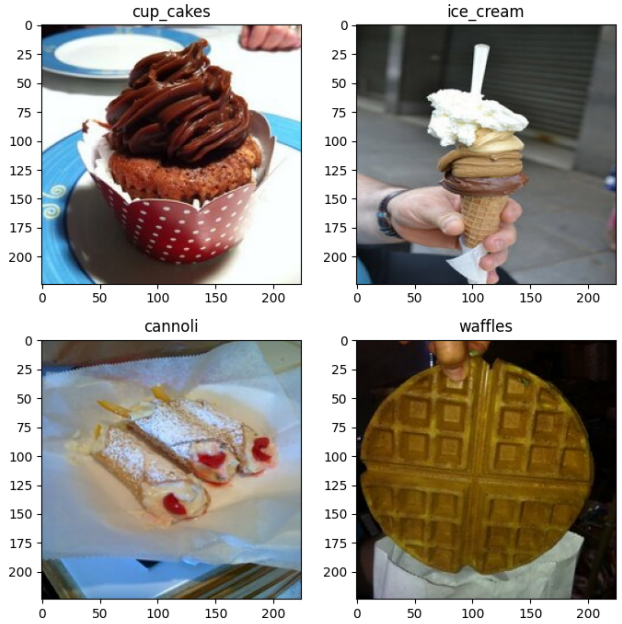

Figure 7: Actual images

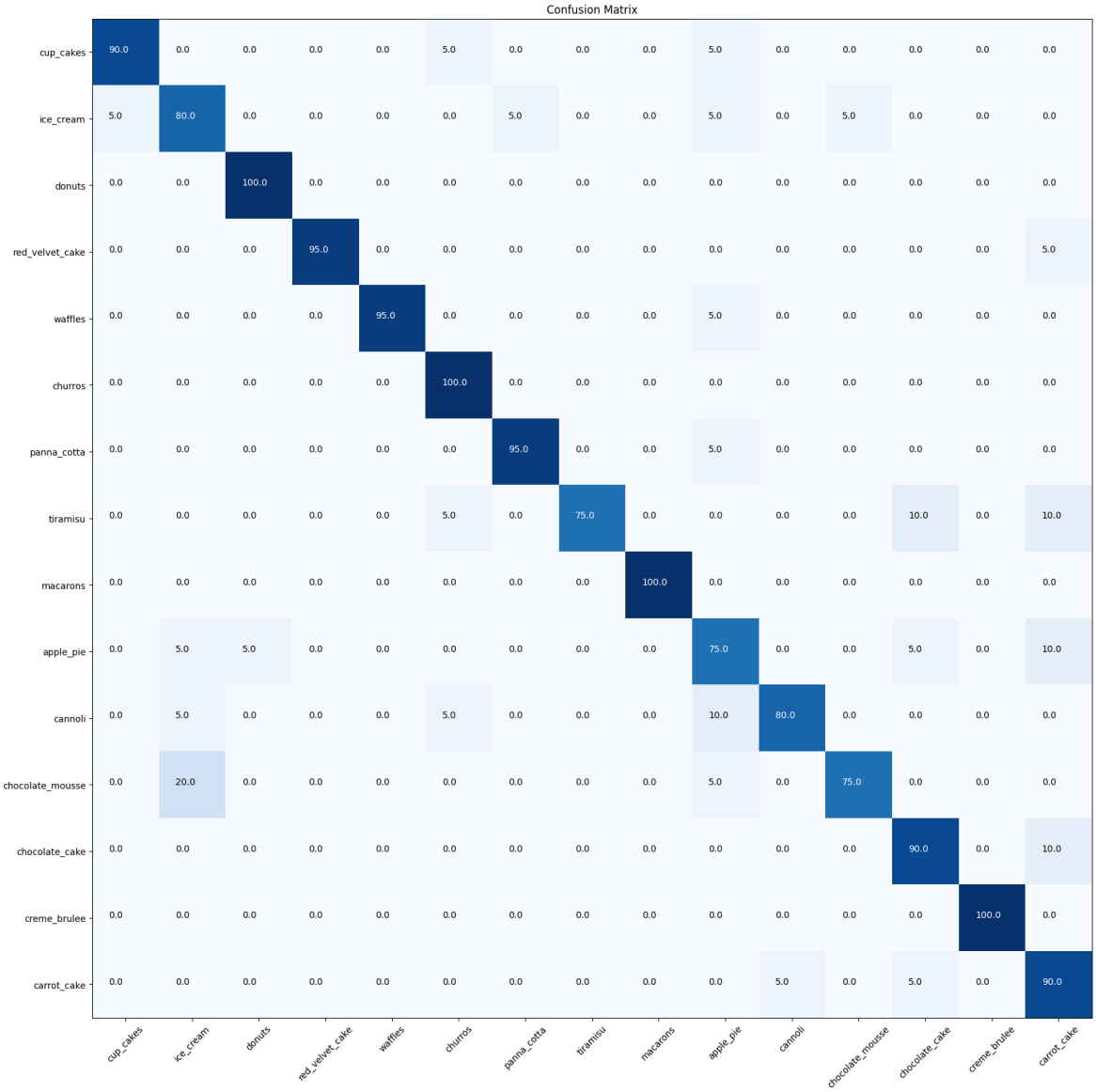

Figure 8: Test accuracies comparison