Image Classification with Multiple Models and Features

Yi-Ting Lan

March 24, 2024

Abstract

This report presents a comparative analysis of different image classification techniques by employing a variety of feature extraction methods and classification models. The study focuses on evaluating the performance of three feature extraction techniques: Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and color histograms, across three classification models: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest. The effectiveness of these techniques is assessed based on F1-Score and Accuracy metrics.

1 Introduction

Image classification stands as a pivotal component in numerous computer vision applications. The choice of feature extraction methods and classification models significantly impacts the performance of such systems. This report delves into the comparison of three distinct feature extraction techniques — HOG, SIFT, and color histograms — combined with three classification models — KNN, SVM, and Random Forest. Our investigation aims to unearth the synergy between different features and models to optimize classification accuracy and F1 scores.

2 Methodology

2.1 Feature Extraction

- **Color Histograms** serve as a fundamental and powerful method for image feature extraction by quantifying the distribution of colors within an image to capture global information. Their unique advantage lies in their insensitivity to spatial changes in the image, effectively reflecting the color composition of the image, making them crucial in applications such as content-based retrieval and scene understanding.

- **Histogram of Oriented Gradients (HOG)** effectively captures texture features and local shape information of an image by statistically analyzing the distribution of gradients or edge orientations within local regions of the image. HOG feature extraction is based on the local gradient direction and intensity of the image, allowing it to describe the contours and structural information of objects in the image.

- **Scale-Invariant Feature Transform (SIFT) + Bag Of Words** The SIFT algorithm is widely recognized in the field of image analysis because it can extract keypoints and their descriptors from images that are invariant to rotation, scale, and changes in brightness. These keypoint descriptors encode detailed information about local regions in the image, making them crucial for matching and recognition across images. However, the SIFT algorithm itself does not provide a direct way to convert these local features into a unified format usable by machine learning models.

  This is where the bag-of-words (BoW) model comes into play. By quantizing the feature descriptors extracted by SIFT into BoW vectors, we can construct a "visual dictionary" that represents each image as a histogram of visual words from the dictionary. This representation effectively converts the image into a fixed-length feature vector that can be directly used for training machine learning models.

  The main reason for choosing to combine SIFT with the BoW model is that this approach combines the strengths of both: the efficient keypoint extraction capability of SIFT with the powerful feature representation ability of BoW. Although the bag-of-words model ignores spatial relationships between features, this simplification actually allows the model to focus more on the content of the images, reducing the complexity of model processing and thereby improving the efficiency and accuracy of image classification tasks.

2.2 Classification Models

- **K-Nearest Neighbors (KNN)** algorithm is an intuitive pattern recognition method based on a simple principle: the class of a sample is most likely to be the same as the class of its K nearest neighbors in the feature space. The advantage of this method lies in its simplicity and ease of implementation, but when dealing with large datasets or high-dimensional features, the computational cost of finding the nearest neighbors can be very high.

- **Support Vector Machine (SVM)** is a powerful supervised learning algorithm designed to find an optimal hyperplane that maximizes the margin between different class samples. SVM handles non-linearly separable datasets by introducing kernel functions, enabling it to find classification decision boundaries in high-dimensional spaces.

- **Random Forest** is an ensemble learning technique based on decision trees, which improves model accuracy and stability by constructing multiple decision trees and combining their prediction results. Each tree uses randomly selected features and samples during the training process, which helps enhance the model's generalization ability to new data and effectively avoids overfitting.

3 Experiments

Our experiment aimed to evaluate the performance of nine feature-model combinations. The dataset comprised 99,600 images in the training set and 200 images in the test set, covering 200 categories.

Taking Color Histogram as an example, each image was transformed into a color histogram using the function, with bins=(8, 8, 8). This means that for each channel of the HSV color space, the color values were divided into 8 bins, resulting in a total histogram bins quantity of $8 * 8 * 8 = 512$. This histogram was normalized and flattened into a one-dimensional array. Therefore, each image was converted into a 512-dimensional vector. Then, SVM, KNN, and

RF models were used for image classification tasks, with the relevant hyper-parameters as shown in Table 1. Each model was trained and tested on the entire image dataset, and performance metrics were recorded. Table 1 below summarizes the test accuracy and F1 scores for each combination.

Table 1: Performance of Feature-Model Combinations

Feature + Model	**Testing acc**	**F1 Score**	**Model Parameters**
Color Histogram + SVM	0.06	0.0281	C=0.8 dual=False
Color Histogram + KNN	0.045	0.0364	n_neighbors=5 algorithm='auto'
Color Histogram + Random Forest	0.115	0.0752	random_state=42 n_jobs=-1
HOG + SVM	0.05	0.0294	C=0.8 dual=False
HOG + KNN	0.02	0.0083	n_neighbors=5 algorithm='auto'
HOG + Random Forest	0.055	0.0337	random_state=42 n_jobs=-1
SIFT + SVM	0.11	0.0802	C=1.0 kernel='rbf'
SIFT + KNN	0.02	0.0175	n_neighbors=5 algorithm='auto'
SIFT + Random Forest	0.045	0.0293	random_state=42 n_jobs=-1

Table 2: Runtime of Feature-Model Combinations

Feature + Model	**Total Time**
Color Histogram + SVM	36min 57s
Color Histogram + KNN	19min 21s
Color Histogram + Random Forest	19min 21s
HOG + SVM	1h 47min 10s
HOG + KNN	2.27 s
HOG + Random Forest	30min 26s
SIFT + SVM	37min 46s
SIFT + KNN	580 ms
SIFT + Random Forest	13min 19s

4 Result

The experimental results show variations in performance among different feature-model combinations. Notably, the combination of SIFT feature extraction method and SVM model achieved the highest F1 score, indicating a strong compatibility between SIFT's detailed local feature representation and SVM's effective high-dimensional space classification. Conversely, the performance of the HOG and KNN combination was subpar, highlighting potential limitations in using edge orientation features with simplistic models like KNN for complex image classification tasks.

Table 2 displays the runtime of various feature extraction techniques and their combined models in this study. We observed that among all feature extraction methods, the combination of Color Histogram and KNN exhibited the lowest runtime. As mentioned earlier, the computational complexity of color distribution as a feature is relatively low. In contrast, the SIFT technique incurred more time during the feature extraction stage due to its involvement in dense local feature computations and subsequent Kmeans clustering. Regarding model training, SVM's training time was significantly higher than other models, highlighting the computational intensity of SVM. Of particular note is that the combination of SIFT and SVM exceeded an hour in overall runtime, indicating the importance of considering their comprehensive impact on computational efficiency when selecting feature extraction techniques and models.

5 Conclusion

This study provides insight into the effectiveness of different combinations of feature extraction methods and classification models for image classification. Our findings suggest that the choice of feature extraction technique plays a crucial role in model performance, with SIFT and SVM standing out as a particularly potent combination. Future work could explore the integration of feature selection methods to further enhance classification accuracy and computational efficiency.