Image Classification with Multiple Models and Features

Chen-Yang Yu
AI Robotics, National Cheng Kung University

March 26, 2024

Abstract

In this work, we investigate the performance of various image classification models and feature extraction techniques on the given dataset(TinyImageNet). The primary objective is to analyze how different feature representations and classification algorithms impact the accuracy and effectiveness of image classification tasks.

We extract image features using three different methods: Histogram of Oriented Gradients (HOG), color histograms, and Scale-Invariant Feature Transform (SIFT). Additionally, We explore three different classification models: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest. Through a series of experiments, we evaluate the performance of each feature-model combination using the F1-score and accuracy metrics. Our findings provide insights into the strengths and weaknesses of different approaches and highlight the importance of selecting appropriate feature representations and classification algorithms for optimal performance in image classification tasks.

1 Introduction

Image classification is a fundamental task in computer vision and machine learning, with applications ranging from object recognition and scene understanding to content-based image retrieval and automated image annotation. The ability to accurately classify images into predefined categories is crucial for many real-world applications, such as automated surveillance, medical imaging analysis, and content-based image search engines.

Traditional image classification approaches relied heavily on hand-crafted feature extraction techniques, which aimed to capture relevant visual information from images. These features were then fed into various classification models, such as Support Vector Machines (SVMs) or K-Nearest Neighbors (KNN), to perform the classification task. However, the selection of appropriate feature extraction methods and classification algorithms played a critical role in the overall performance of the system.

In recent years, with the advent of deep learning techniques, end-to-end trainable models have gained significant popularity in image classification tasks. These models can automatically learn relevant features from raw pixel data, potentially alleviating the need for hand-crafted feature extraction. Nevertheless, traditional feature extraction methods and classical machine learning models remain valuable tools, particularly in scenarios where computational resources are limited or when interpretability and explainability are crucial.

In this study, we explore the performance of three different classification models: KNN, SVM, and Random Forest. We extract image features using three different techniques: Histogram of Oriented Gradients (HOG), color histograms,

Scale-Invariant Feature Transform (SIFT). By comparing the performance of these feature-model combinations, we aim to gain insights into the strengths and weaknesses of each approach and provide practical guidelines for selecting the most appropriate methods for image classification tasks.

2 Methodology

2.1 Dataset

For our experiments, we utilized the TinyImageNet dataset, which consists of a total of 10000 images across 200 classes and a diverse collection of natural images.

The original images in the TinyImageNet dataset have a resolution of **64x64x3** pixels. To prepare the data for our experiments, we performed some preprocessing steps as metioned at the following section.

2.2 Preprocess

2.2.1 Train Test Split

To evaluate the performance of our models and ensure reliable results, we employed train-test split strategy. Specifically, we divided the dataset into training and testing subsets, with 80% of the data allocated for training and the remaining 20% for testing.

- Training: 80,000 images across 200 classes
- Testing: 20,000 images across 200 classes

2.2.2 Resize

The original images in the TinyImageNet dataset have a resolution of **64x64x3** pixels. We downsize every image at resolution of **32x32x3** pixels for speed consideration during feature extraction and model training.

2.3 Feature Extraction Methods

In this study, we explored three different feature extraction techniques to represent the image data effectively. These techniques aimed to capture relevant visual information from the images, which could then be used by the classification models to perform the image classification task.

2.3.1 Histogram of Oriented Gradients (HOG)

The Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005) is a feature descriptor that captures the distribution of intensity gradients or edge directions within localized regions of an image.

2.3.2 Color Histogram

Color histograms are a simple yet effective way to represent the color distribution within an image. To compute the color histogram, we first quantized the color space (RGB to HSV) into a fixed number of bins (8, 8, 8). Then, for each pixel in the image, we incremented the corresponding bin in the histogram based on the pixel's color value. The resulting histogram represented the relative frequencies of different color values within the image, providing a compact yet informative representation of the image's color content.

2.3.3 Scale-Invariant Feature Transform (SIFT)

The Scale-Invariant Feature Transform (SIFT)(Lindeberg, 2012) is a popular local feature descriptor that is widely used in various computer vision tasks, including image classification. SIFT features are designed to be invariant to image scaling, rotation, and partially invariant to changes in illumination and viewpoint. To extract SIFT features, we first identified keypoints in the image using a difference-of-Gaussians approach. Then, for each keypoint, we computed a SIFT descriptor,

which captured the local gradient information around the keypoint's neighborhood.

2.3.4 Bag of Words (BoW)

Since the number of SIFT features extracted from an image can vary, we employed the Bag of Words (BoW) model to obtain a fixed-length feature vector for each image. The BoW approach involves clustering the SIFT descriptors from all images into a predefined number of visual words or codewords, effectively quantizing the feature space. For each image, we then constructed a histogram of codeword occurrences, representing the frequency of each visual word within the image. This histogram served as the final feature vector for the image, allowing us to maintain a consistent feature dimension across all images while capturing the essence of the SIFT descriptors.

2.4 Classification Models

In our study, we explored three different classification models to evaluate their performance on the image classification task using the extracted features. These models were chosen for their widespread use, effectiveness, and ability to handle diverse types of data.

2.4.1 Support Vector Machine (SVM)

Support Vector Machines (SVMs)(Hearst, Dumais, Osuna, Platt, & Scholkopf, 1998) are a powerful class of supervised learning algorithms that have been widely used for classification and regression tasks. SVMs operate by finding the optimal hyperplane that maximally separates the different classes in the feature space.

2.4.2 K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm is a non-parametric method that classifies instances based on their similarity to the nearest neighbors in the feature space. In our implementation, we used the Euclidean distance metric to measure the similarity between feature vectors. During classification, the algorithm assigns a label to a new instance based on the majority vote of its k nearest neighbors in the training set.

2.4.3 Random Forest

Random Forest is an ensemble learning method that combines multiple decision trees to improve the overall classification performance. Each individual decision tree in the ensemble is trained on a random subset of the training data and a random subset of features, promoting diversity among the trees.

3 Experiments

To evaluate the performance of the different feature extraction methods and classification models, we conducted a series of experiments and analyzed the results in terms of accuracy, F1 score, and computational time. The experimental results are summarized in the following tables:

3.1 Method 1. HOG

When using the Histogram of Oriented Gradients (HOG) as the feature extraction method, the Support Vector Machine (SVM) model achieved the highest accuracy of 0.17 and an F1 score of 0.12. However, it took a significant amount of time (2032.11 seconds) to train the SVM model. The K-Nearest Neighbors (KNN) and Random Forest models performed relatively poorly, with lower accuracy and F1 scores, but they were significantly faster. See table 1

Model	Accuracy	F1 Score	Time(s)
SVM	*0.17	*0.12	2032.12
KNN	0.04	0.02	4.01
RF	0.05	0.04	1647.76

Table 1: HOG with different models

3.2 Method 2. Color Histogram

When using the Color Histogram as the feature extraction method, the Random Forest model achieved the highest accuracy of 0.12 and an F1 score of 0.07. The SVM model performed slightly better than the KNN model in terms of accuracy and F1 score, but it took significantly more time to train (3196.23 seconds) compared to the other two models. See table 2

Model	Accuracy	F1 Score	Time(s)
SVM	0.08	0.05	3196.23
KNN	0.02	0.01	7.27
RF	*0.12	0.07	423.65

Table 2: Color Histogram with different models

3.3 Method 3. SIFT

When using the Scale-Invariant Feature Transform (SIFT) with the Bag of Words (BoW) model to handle variable feature dimensions, all three classification models (SVM, KNN, and Random Forest) achieved an accuracy of 0.01. However, the F1 scores were significantly higher, with KNN having the highest F1 score of 5.13. The SVM model took the longest time to train (1543.12 seconds), while the KNN and Random Forest models were relatively faster. See table 3

Model	Accuracy	F1 Score	Time(s)
SVM	0.01	5.12	1543.12
KNN	0.01	5.13	415.17
RF	0.01	5.12	417.43

Table 3: SIFT with different models

4 Results and Discussion

Based on these results, we can make the following observations:

1. The choice of feature extraction method and classification model significantly impacts the overall performance of the image classification system. Different combinations of features and models yield varying levels of accuracy, F1 scores, and computational times.

2. The SIFT features with the Bag of Words (BoW) model achieved the highest F1 scores across all three classification models, suggesting that this combination of feature extraction and feature representation methods effectively captured relevant information for the image classification task.

3. The SVM model generally took the longest time to train, potentially due to the complexity of finding the optimal hyperplane in high-dimensional feature spaces. However, in some cases (e.g., HOG features), the SVM achieved the highest accuracy.

4. The Random Forest model consistently performed well across different feature extraction methods, striking a balance between accuracy, F1 score, and computational time.

5. The KNN model was generally the fastest to train but often exhibited lower accuracy and F1 scores compared to the other models.

5 Conclusion

In this study, we explored the performance of various image classification models and feature extraction techniques on the TinyImageNet dataset, which consists of 200 classes and a diverse collection of natural images. We implemented three different classification models: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest. Additionally, we extracted image features using four different methods: Histogram of Oriented Gradients (HOG), color histograms, Scale-Invariant Feature Transform (SIFT) with Bag of Words (BoW) representation, and raw pixel values.

Through a series of experiments, we evaluated the performance of each feature-model combination using accuracy, F1 score, and computational time as metrics. Our results demonstrated that the choice of feature extraction method and classification model significantly impacts the overall performance of the image classification system.

References

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 ieee computer society conference on computer vision and pattern recognition (cvpr'05)* (Vol. 1, pp. 886–893).

Hearst, M., Dumais, S., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, *13*(4), 18-28. doi: 10.1109/5254.708428

Lindeberg, T. (2012). Scale invariant feature transform.