National University of Sciences and Technology
School of Electrical Engineering and Computer Science
Department of Computing

CS893: Advanced Computer Vision Spring 2022

Assignment 1

Image Classification

Instructor: Dr. Muhammad Moazam Fraz

Submitted by: Muhammad Salman Akhtar
Registration No: 398895

1 Introduction

Recognition of traffic signs is one of the fundamental tasks for a smart transportation system. While correctly identifying traffic signs are extremely essential for autonomous vehicles, it can also be helpful for human drivers by providing useful cues in the form of warnings, prohibitions and suggestions thus improving the road safety. This task is comprised of two steps, the detection of traffic signs followed by the classification of those detected signs. This assignment is concerned with the second part of the process that is the classification of traffics signs.

2 Dataset

The dataset is a subset of BelgiumTSC comprising 10 classes of traffic signs. The dataset can be downloaded from this link. A sample of each class is depicted in figure 1. For training purposes 800 images (80 from each class) have been provided while a total of 200 images (20 images from each class) have been designated for testing dataset. Images in dataset comprises of varying dimensions and have been taken under different lighting conditions with changing viewing angles.

(a) Hump

(b) Speed Limit 70

(c) Speed Limit 50

(d) Children crossing

(e) Forbidden direction

(f) Give way

(g) No entry

(h) Bicycle way

(i) Bicycle and children way

(j) No parking

Figure 1: Traffic Sign Dataset

3 Motivation for the chosen method

Traffic signs classification is a challenging problem because of multiple factors caused by occlusions, different lighting conditions, different viewing angles, out of focus due to motion and many more. Thus, only a robust algorithm can accurately classify the traffic sign under these challenging conditions. Convolutional Neural Networks (CNN) are considered to be excellent at classifying objects among computer vision algorithms. They have been known to outperform human test subjects in German Traffic Sign Recognition Benchmark held at IJCNN in 2011 (1). However, the deep CNNs require large training dataset for model training in order to develop a predictor model that gives accurate results on general inputs.
Since, the available dataset is of a small size, therefore CNN-based classifier would not be a suitable choice for this problem. As an alternative, Machine Learning (ML)-based classifiers; Support Vector Machine (SVM) and Random Forest were considered. SVMs have been regarded as robust and accurate classifier among ML-based image classifiers (2). A good ML classifier requires equally good features as input to give satisfactory results. A good feature detector should exhibit a significant degree of invariance to photometric and geometric transformations. Some of the available choices of feature detectors/feature descriptors include SIFT, SURF, BRISK, ORB and HOG. SIFT is known to be a very robust feature detector which exhibits invariance with respect to scale, rotation and photometric transformations. Thus, it was selected for extracting features for implementing Approach 1 that is key point-based feature formulation. Moreover, Histogram of Gradients (HOG) was selected for implementing Approach 2 taking inspiration from (3).

4 The Approach

Two separate approaches were implemented for extracting features which were to be passed to a classifier. Both approaches are summarized in the following paragraphs:-

4.1 Approach 1

The image training dataset is read and converted to grayscale and resized to 128x128 dimensions. These basic image processing steps are followed by key-point detection and building of feature vectors. For this purpose, Scale Invariant Feature Transform (SIFT) was used. SIFT is a patented algorithm and its commercial usage is restricted, however, it can be used freely for academic purposes. The extracted feature vectors are passed to Kmeans clustering which assigns them to the respective clusters. Similar features vectors are assigned to the same cluster. Thus, each cluster represents a particular a particular feature or a visual word in the image. A histogram of these features is formed where each cluster acts as a bin which can as a whole be looked as a bag of visual words. This bag of visual word is used for training an SVM-based classifier. In the end, the test dataset is fed to the trained SVM for making predictions. Figure 2 shows the algorithm flow for Approach 1 while table 1 depicts the hyperparameters.

Table 1: Hyperparameters of Approach 1

Hyperparameter	**Value**
no of clusters	500
C	20
gamma	0.001
kernel	rbf

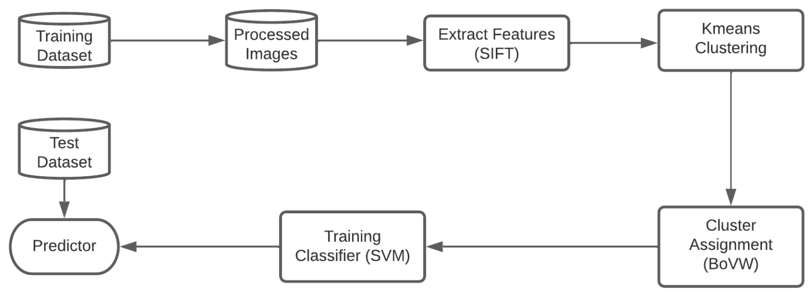


Figure 2: Traffic Sign Classifier Model (Approach 1)

4.2 Approach 2

The image training dataset is read and converted to grayscale and resized to 128x128 dimensions. These basic image processing steps are followed by taking Histogram of Gradients (HOG). Owing to the properties of gradients, the extracted features have photometric as well as geometric invariance. In contrast to SIFT, HOG is calculated on a uniformly spaced dense grid. The HOG features are normalized and their dimensions are reduced through Principal component analysis. The tailored features are then used for training an SVM-based classifier. In the end, the test dataset is fed to the trained SVM for making predictions. Figure 3 depicts the algorithm flow for Approach 2 while table 2 depicts the hyperparameters.

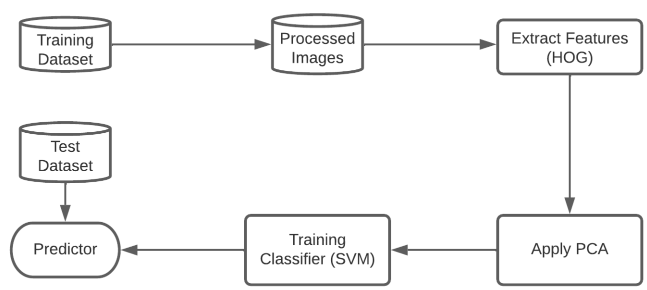


Figure 3: Traffic Sign Classifier Model (Approach 2)

Table 2: Hyperparameters of Approach 1

Hyperparameter	**Value**
C	10
gamma	0.001
kernel	rbf

5 How to run the code

The algorithm has been implemented using python programming language in Jupyter notebook and corresponding "TrafficSignClassifierCV.ipynb" file can be accessed from github link. The code expects that image database folder "CS893 Sp2022 A1 Dataset" is placed in same directory as the "TrafficSignClassifierCV.ipynb". All the cells should be ran sequentially from top to bottom. The notebook also includes all the supporting functions that are required for training and testing the model. Before running "Execute

Model Training" cell, all the cells before it should be executed. To select "Approach 1", edit (featureExtractorType = "SIFT") and for "Approach 2", edit (featureExtractorType = "HOG") in "Execute Model Training" and "Execute Model Testing".

6 Results

The results were predicted for both Approach 1 and Approach 2 with dataset of 200 images. Among 200 images each class contained 20 image.

6.1 Quantitative Results

Following paragraphs discuss quantitative results for both approaches:-

6.1.1 Approach 1

Overall accuracy for Approach 1 was 91 percent. Figure 4 shows the confusion matrix and classification report for Approach 1. Most number of False Positives (8) appear for "No Parking" sign while most number of False Negatives (07) appear for "No entry" sign. Least number of True positives (13) were detected for "Give way" sign. This modeled performed poorly for "Give way", "No entry" and "No Parking" signs as depicted by their f1-scores of 0.74, 0.77 and 0.81 respectively.

6.1.2 Approach 2

Overall accuracy for Approach 2 was 99 percent. Figure 5 shows the confusion matrix and classification report for Approach1. Only 01 False Positives were observed for "Bicycle way" and "No Parking" signs. Moreover, only 01 False Negatives were observed for "Speed Limit 50" and "No entry" signs.

6.2 Qualitative Results

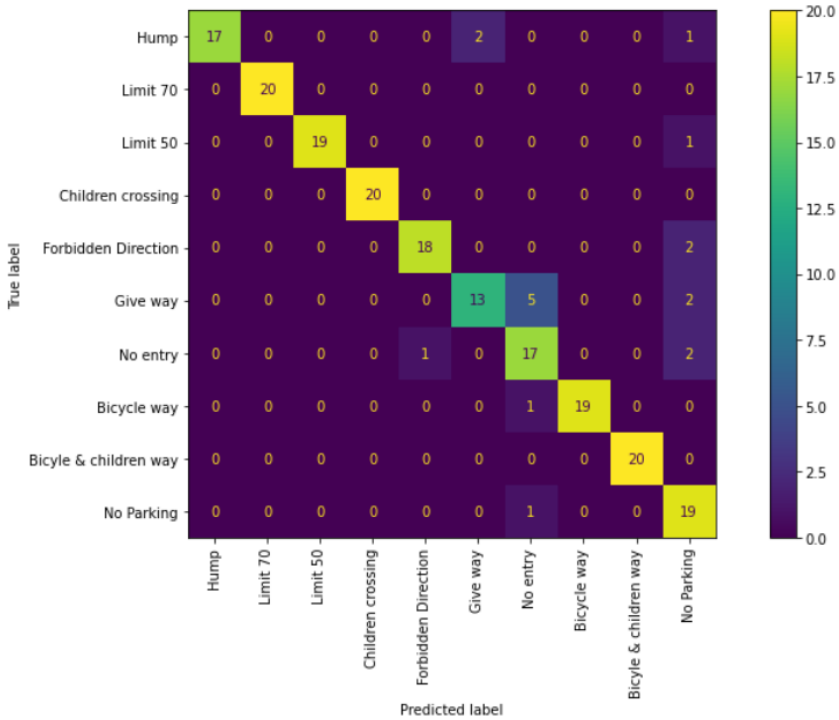
Following paragraphs discuss qualitative results for both approaches:-

6.2.1 Approach 1

Approach 1 made a lot of errors particularly while predicting "Give way" traffic signs. Confusion matrix for Approach 1 reveals that the model confused 5 "Give way" signs with "No entry". One reason could be the similarity of both these traffic signs as can be visually seen that could have resulted in enhanced similarity of extracted features as well. Some of the correctly classified images are shown in figure 6 while incorrectly classified images by approach 1 are shown in figure 7.

6.2.2 Approach 2

Approach 2 predicted almost all the results correctly. Some of the correctly classified are shown in 8 while a couple of misclassified images by approach 2 are shown in figure 9. Only 01 "Speed Limit 50" sign was



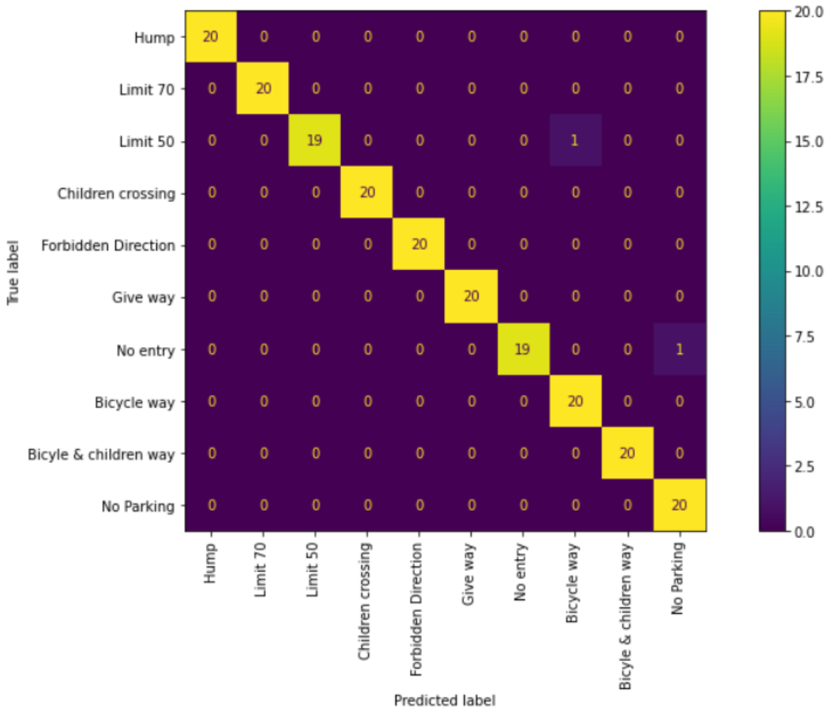
(a) Confusion Matrix (Approach 1)

	precision	recall	f1-score	support
Hump	1.00	0.85	0.92	20
Limit 70	1.00	1.00	1.00	20
Limit 50	1.00	0.95	0.97	20
Children crossing	1.00	1.00	1.00	20
Forbidden Direction	0.95	0.90	0.92	20
Give way	0.87	0.65	0.74	20
No entry	0.71	0.85	0.77	20
Bicycle way	1.00	0.95	0.97	20
Bicyle & children way	1.00	1.00	1.00	20
No Parking	0.70	0.95	0.81	20
accuracy			0.91	200
macro avg	0.92	0.91	0.91	200
weighted avg	0.92	0.91	0.91	200

(b) Classification Report (Approach 1)

Figure 4: Approach 1 Results

misclassified as "Bicycle way" sign while 01 "No entry" sign was misclassified as "No Parking" sign. It is pertinent to mention here that the test image number 134 ("No entry") was misclassified as "No Parking" sign by both approaches.



(a) Confusion Matrix (Approach 2)

	precision	recall	f1-score	support
Hump	1.00	1.00	1.00	20
Limit 70	1.00	1.00	1.00	20
Limit 50	1.00	0.95	0.97	20
Children crossing	1.00	1.00	1.00	20
Forbidden Direction	1.00	1.00	1.00	20
Give way	1.00	1.00	1.00	20
No entry	1.00	0.95	0.97	20
Bicycle way	0.95	1.00	0.98	20
Bicyle & children way	1.00	1.00	1.00	20
No Parking	0.95	1.00	0.98	20
accuracy			0.99	200
macro avg	0.99	0.99	0.99	200
weighted avg	0.99	0.99	0.99	200

(b) Classification Report (Approach 2)

Figure 5: Approach 2 Results

7 Conclusion

In this assignment, two separate approaches were implemented for feature extraction of images to be classified into 10 traffic sign categories. Approach 1 used feature extraction using key point detector (SIFT) and subsequently formed bag of visual words through K-means clustering. The bag of visual words was then used for training the ML-based SVM classier which was finally used for predicting the test dataset. This approach achieved an accuracy of 91 percent while performing rather poorly in at least two classes of traffic signs ("Forbidden direction" and "No entry"). Approach 2 used feature extraction through HOG which was followed by feature scaling and dimensionality reduction through PCA. The squeezed feature dataset is fed to ML-based SVM classier for training and the resultant trained model is used for predicting the test dataset.

(a) Hump

(b) Speed Limit 70

(c) Speed Limit 50

(d) Children crossing

(e) Forbidden direction

(f) Give way

(g) No entry

(h) Bicycle way

(i) Bicycle and children way

(j) No parking

Figure 6: Correctly Classified Images (Approach 1)

(a) Wrongly classified as Give way

(b) Wrongly classified as No parking

(c) Wrongly classified as No parking

(d) Wrongly classified as No entry

(e) Wrongly classified as No parking

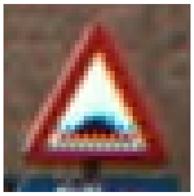
(f) Wrongly classified as No parking

(g) Wrongly classified as No entry

(h) Wrongly classified as No entry

Figure 7: Wrongly Classified Images (Approach 1)

This approach surprising achieved a very high accuracy of 99 percent. The resulting better performance of second approach could be due to the presence of very simple shaped objects (cropped Traffic signs) in the dataset. The HOG efficiently extracted these shapes in the form of gradient features for subsequent classifier and resultantly achieved high accuracy on given test set.



(a) Hump

(b) Speed Limit 70

(c) Speed Limit 50

(d) Children crossing

(e) Forbidden direction

(f) Give way

(g) No entry

(h) Bicycle way

(i) Bicycle and children way

(j) No parking

Figure 8: Correctly Classified Images (Approach 2)

(a) Wrongly classified as Bicycle way

(b) Wrongly classified as No parking

Figure 9: Wrongly Classified Images (Approach 2)

References

[1] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural networks*, vol. 32, pp. 323–332, 2012.

[2] A. E. Mohamed, "Comparative study of four supervised machine learning techniques for classification," *International Journal of Applied*, vol. 7, no. 2, 2017.

[3] N. Ahmed, S. Rabbi, T. Rahman, R. Mia, and M. Rahman, "Traffic sign detection and recognition model using support vector machine and histogram of oriented gradient," *International Journal of Information Technology and Computer Science*, vol. 13, no. 3, pp. 61–73, 2021.