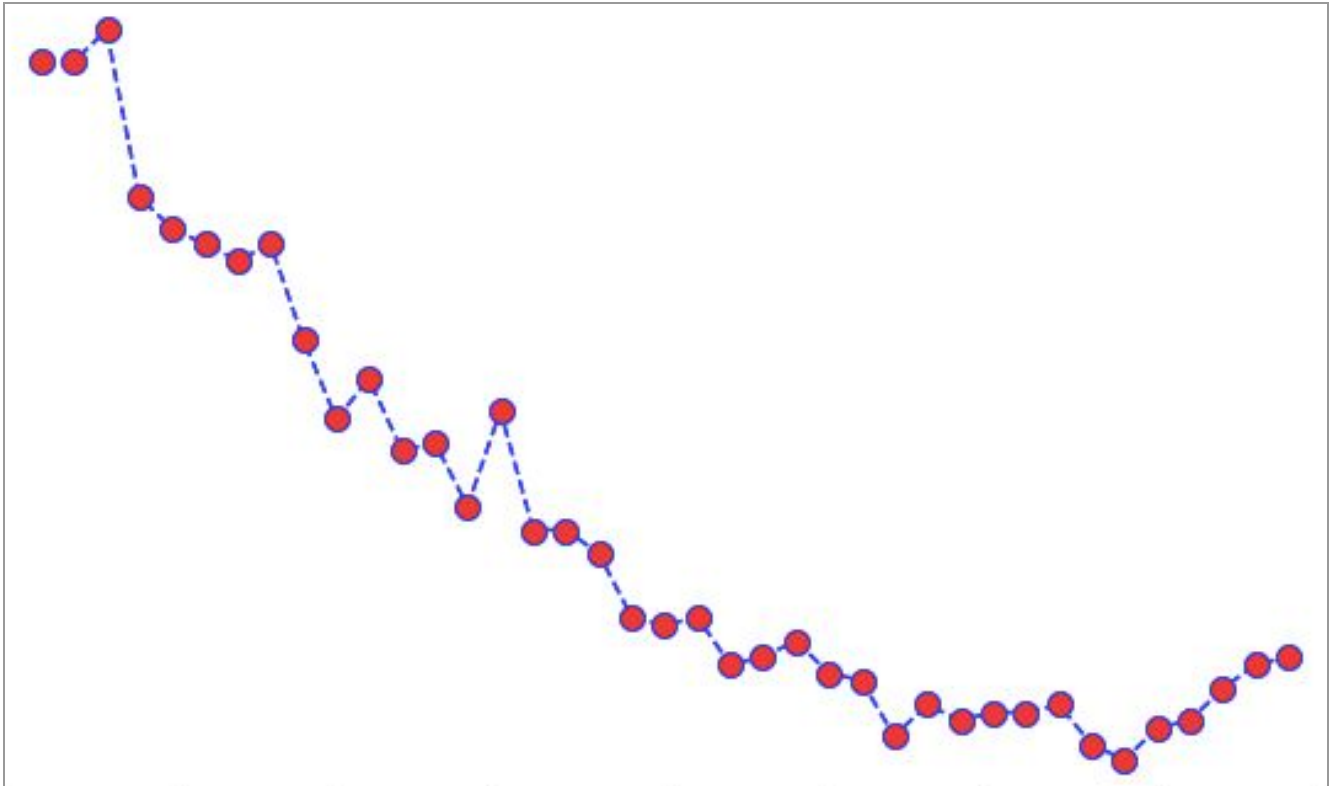


PROGRAM 2 REPORT

CMPE 255 - Data Mining



DEVASHISH NYATI

013732342

SAN JOSE STATE UNIVERSITY

03.07.2019

INTRODUCTION

The aim of this assignment was to develop a predictive model that can determine, given an image, which one of 14 classes it is. Classes: car, SUV, small_truck, medium_truck, large_truck, pedestrian, bus, van, people, bicycle, motorcycle, signal_green, signal_yellow, signal_red.

APPROACH

1. Experiment with different image feature extraction techniques and extract features
2. Apply dimensionality reduction techniques
3. Apply classification models and evaluate the models

PROCEDURE

1. Experiment with different image feature extraction techniques and extract features:

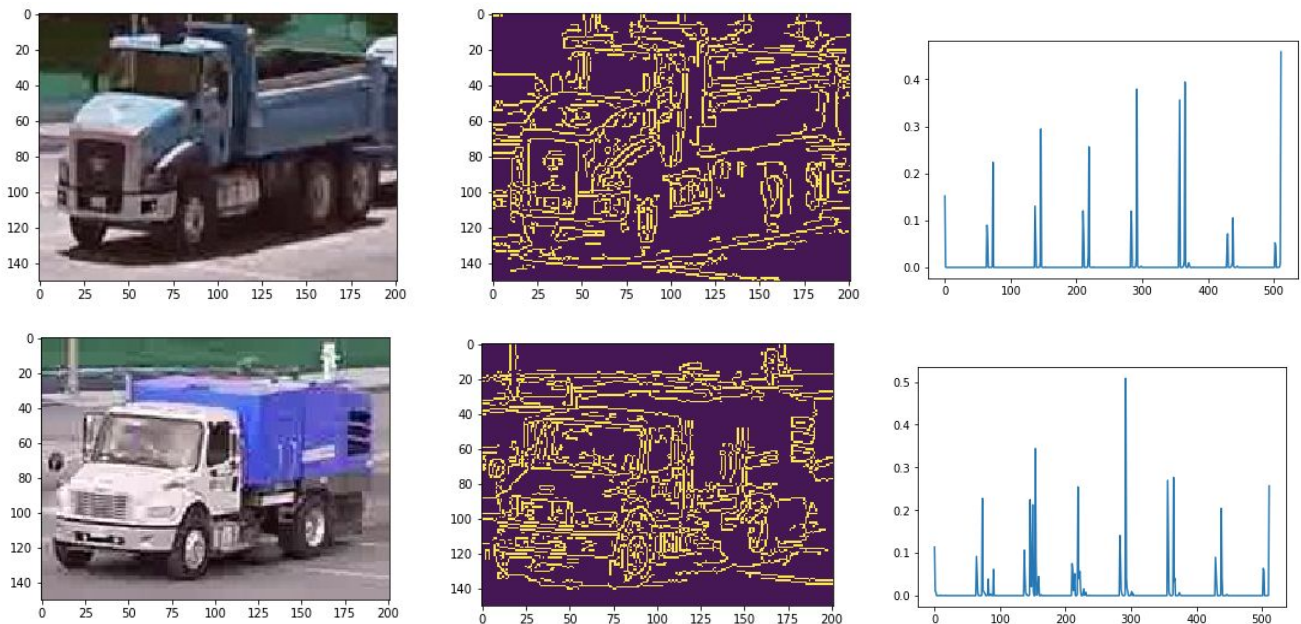
I extracted a few features like the edged image, Histogram of Oriented Gradients (HOG), Color Histogram and compared the images.

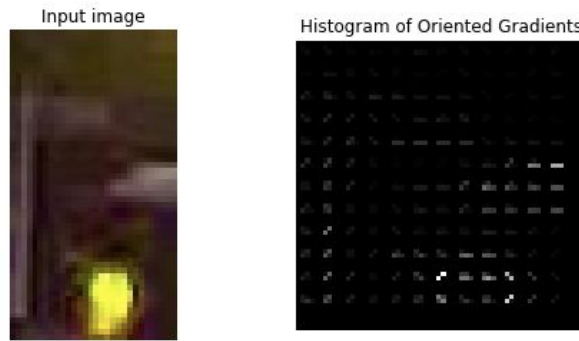
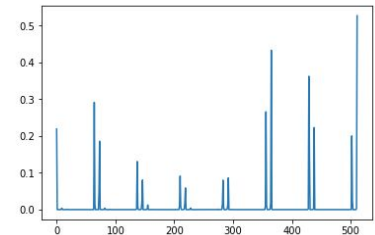
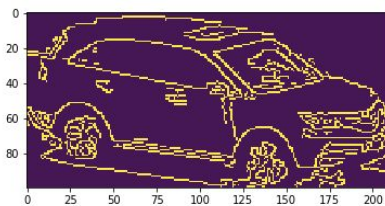
a. Histogram of Oriented Gradient:

I calculated HOG by resizing the image, then converting the image to grayscale and then using skimage's feature, I calculated HOG.

b. Color Histogram:

I calculated color histogram using cv2 calcHist(). I normalized the histogram to reduce the number of values.





After calculating and experimenting with different features, I created the feature matrix with the color histogram values and histogram of oriented gradient.

2. Apply dimensionality reduction techniques:

My feature matrix became very long since I added color histogram extracted features and histogram of oriented gradient extracted features. I used PCA to reduce the number of features. I reduced 1256 features to 300 components.

3. Apply classification models and evaluate the models:

I used different classification models like random forest and KNN and played with the parameters. My best result came with using KNN with 10 nearest neighbors. I evaluated the small data by calculating f1 score.

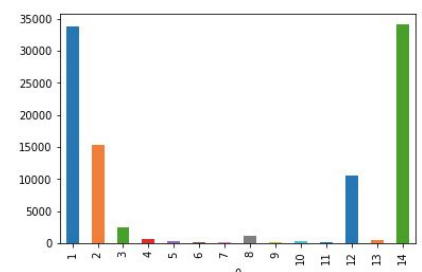
4. Imbalanced Data:

I used different parameters in my classification models to deal with imbalanced data like setting `class_weight='balanced'`. I even plotted the result to make sure all the classes are present.

RESULTS

After submission, the result was as follows:

- Rank: 3
- F1 (on 50%): 0.9082



CONCLUSION

I was successfully able to implement image classification. This assignment has helped me learn the concepts of feature extraction from images, dimensionality reduction and classification models.