Classification of single images with and without the presence of pedestrians

Chirag Panikkasseril Unni
School of Computer Science and Engineering,
UNSW Sydney, NSW 2052, Australia
z5241855@unsw.edu.au

***Abstract*—When it comes to Computer vision and related study, one of the most important topics is detection of pedestrians. Numerous scientific fields like automated cars, public surveillance, traffic monitoring and robotics technologies make use of pedestrian detection for various applications. This paper aims to provide various feature extraction techniques in computer vision and use classification model for detecting the presence of pedestrian in single images. The feature extraction and techniques and various Machine learning models do not aim to produce the best model for detection as selection of best model depends on the usage, requirement and the environment of the system but the experiment and evaluation is expected to produce one or more model which can be used to detect the presence of pedestrian in single images.**

I. Introduction

Pedestrian detection is one of the most important application in the field of computer vision and it's been used for developing systems like driver assistance , automated vehicles, surveillance traffic monitoring, etc. The Pedestrian detection problem involve mainly two steps, first step is to extract the relevant features in image dataset and further to train a classification model to detect the presence of pedestrians in Images.

The data provided by CSIRO uses NICTA images. The dataset contain large number of positive and negative images including 3 positive(with pedestrian) and 1 negative(without pedestrian) set for training and separate set for testing.

Here we will try to implement various feature extract technique in the field of computer vision and use it to train and test a classification model based on various machine learning algorithms used for binary classification. In the experiment we will be using Histogram of Oriented Gradients (HOG) feature extraction [1] and also be discussion feature extraction techniques like haar wavelet features, Scale Invariant Feature Transform (SIFT), Speeded-Up Robust Feature (SURF) features for pedestrian classification. Most of the research on pedestrian classification involves some kind of supervised or semi-supervised learning using Machine learning algorithms on classification models. In this experiment we will try to implement models using Linear Support Vector Machine classifier, Decision Tree classifier, Stochastic Gradient Descent classifier as a few techniques for our evaluation. The accuracy and efficiency of the model will be verified using various evaluation techniques like Receiver operating characteristic (ROC) Area Under the Curve (AUC) score, accuracy, recall , f1-score, etc.

II. Overview of Feature Extraction Techniques

A. Scale Invariant Feature Transform (SIFT)

SIFT proposed by Lowe solves the image rotation, affine transformations, intensity, and viewpoint change in matching features. It uses the following steps in feature extraction using the Difference of Gaussian method (DoG), a key point localization by eliminating low contrast points, a key point orientation assignment based on local image gradient and by computing the local image descriptor for each key point based on image gradient magnitude and orientation [2].

B. Speed up Robust Feature (SURF)

Speed up Robust Feature (SURF) technique, which is an approximation of SIFT [3]. It can perform much faster than SIFT even without compromising the quality of detected points. It uses Difference of Gaussian DoG with box filter rather than averaging and also by computing this parallelly for different scales. For description purpose it also use haar wavelet responses by finding localized region and for each region computing the SURF feature and thus produced matches are compared with similar contrast blobs.

C. Histogram of Oriented Gradients (HOG)

HOG, or Histogram of Oriented Gradients, is often used in computer vision research for object detection. HOG is a feature descriptor that is widely used to extract features from image datasets. When we try to use HOG feature extraction technique the function, unlike, other edge detection method which only produce edge detection, HOG extract edges as well as focus on the shape and the structure of the object in the image. Further the edge detection is achieved by extracting gradient and the orientation

The calculation of orientation is done by focusing on localised portions. That is, it takes a full image and break it down to smaller images regions and each region calculation of gradient and orientation is computed. Separate Histograms will be computed for each of these localised regions in image i.e., for each such regions 1-D histogram of gradient directions or edge orientations will be computed over the pixels of the region using gradients and orientations technique and hence the name or Histogram of Oriented Gradients.

Implementing HOG in the experiment is done with the help of predefined function called hog in the skimage.feature library [4]. This can also be achieved by using tools like OpenCV. Computations of features can be easily achieved by these functions.

III. Classification Model

A. Linear Support Vector Machine Classifier

Linear Support Vector machine (SVM) classifier is a very common classifier used for pedestrian classification experiments [5]. The algorithm creates a line or a hyperplane which separates the data into classes. But usually with a dataset there is no unique line separating these tasks.

SVM helps in identifying the point closest to both classes and these points are support vectors and the distance between these are called margin. The Hyperplane which gives the maximum margin is the optimal hyperplane. Thus, the optimal hyperplane will gives you the decision boundary between the classes with maximum separation.

As in our experiment we use Linear SVM find the line separating the data into pedestrian or no pedestrian from the feature data obtained from feature extraction. The hyperparameter which is considered in tuning are C value which controls the trade-off between classifying training points correctly with high C and smooth decision boundary with low C and the Gamma values which defines how far the influence of a single training reaches as in low value means far reach and high means close reach.

B. Decision Tree Classifier

Decision Tree classifiers are used widely in data mining and statistics. Main advantage is that there is no assumption about data distribution and are very fast to compute and to interpret as the structure formed are tree with leaves forming the classification based on if-then rule[6]. It's a supervised learning model , as it classifies instances into one of predefined sets of classes. Common metrics used over the algorithm are Gini impurity measure ,Information gain based on concept of entropy and Variance Reduction.

C. Stochastic Gradient Descent Classifier

Stochastic Gradient Descent (SGD) algorithms employ a model in which it use a single training sample in each iteration and the information will be process in a stream of line model [7]. Especially in object detection where the image information is highly symmetrical and not much change observed with the background data SGD computes it much faster way thus it can run for smaller data set and can be scaled linearly for large data set and still converge optimal results. Thus its been used extensively in image classification experiments especially for pedestrian detection problems.

IV. Method

A. Data

The experiment focus on the National ICT Australia Limited (NICTA) dataset for developing a classification model for pedestrian detection. The dataset contains 72114 images for training and 9457 for testing with images with size 80x64. The dataset includes both negative as well as positive classification of pedestrian classification.

B. Feature Extraction

The images are read using OpenCV function and then converted to grayscale images. The image set is read by iteratively calling each of the files under testing and training folder and each of the subset folders. The images are then called by the function to extract the feature using sklearn.features hog function. The extracted fd variable and corresponding classes are stored in numpy arrays for the classification model.

Hog function uses parameters given below:

Orientations = 9, pixels_per_cell = (8, 8),cells_per_block = (2, 2)

where orientation is the number of orientation bins, pixels_per_cell is the size (in pixels) of a cell and cells_per_block is the number of cells in each block. In addition to this by default it uses L2 norm for normalisation of the data

V. Experimental setup

A. Model Training

The extracted features and corresponding classes are stored in below numpy arrays. *X_train, y_train,X_test,y_test* .The test variables are then passed to the various training models to fit the model and the observed prediction of classes is stored in the variable *y_pred*. Here we use various training models which were previously discussed for binary classifications.

The data is fitted with Linear SVM model using sklearn.svm SVC function . The kernel used is linear , probability is enabled and gamma is set to auto.

Further the data is fitted with SGDClassifier function from sklearn.linear_model. And then the data is fitted with DecisionTreeClassifier function from sklearn.tree. In each of the case *y_ pred* is calculated and stored for evaluation.

B. Model Evaluation

The model evaluation is done by calculating various metrics. For each model accuracy_score , precision, recall and f1-score is calculated from sklearn.metrics functions.

Further for the Linear SVC models the corresponding False Positive Rate or (1 - Specifity) and True Positive Rate or (Sensitivity) is obtained. Using this information a plot of Receiver Operating Characteristics is plotted for evaluating and observing the Area under the curve.

VI. Results and discussion

The following result were obtained after model training. The evaluation shows that the Linear model obtained a higher accuracy score among all three models with value 0.9860420852278735 as in “Fig. 1”, which followed by SGD classifier with value 0.9819181558633816 as in “Fig. 2”, and then Decision Tree Classifier with value 0.9240774029819182. as in “Fig. 3”.

Weighted average and macro average metrics for precision, recall and f1-score is also obtained for each of the model which produced similar evaluation results.

The area under the curve produced by the linear SVM model obtained a value of 0.998 given by “Fig. 4”.

Accuracy: 0.9860420852278735

	precision	recall	f1-score	support
0	0.99	0.99	0.99	6000
1	0.98	0.98	0.98	3457
accuracy			0.99	9457
macro avg	0.99	0.98	0.98	9457
weighted avg	0.99	0.99	0.99	9457

Fig. 1 *Classification report for Linear SVM model*

Accuracy: 0.9819181558633816

	precision	recall	f1-score	support
0	0.98	1.00	0.99	6000
1	0.99	0.96	0.97	3457
accuracy			0.98	9457
macro avg	0.98	0.98	0.98	9457
weighted avg	0.98	0.98	0.98	9457

Fig. 2 *Classification report for Stochastic Gradient Descent model*

Accuracy: 0.9240774029819182

	precision	recall	f1-score	support
0	0.94	0.94	0.94	6000
1	0.90	0.90	0.90	3457
accuracy			0.92	9457
macro avg	0.92	0.92	0.92	9457
weighted avg	0.92	0.92	0.92	9457

Fig. 3 *Classification report for Decision Tree model*

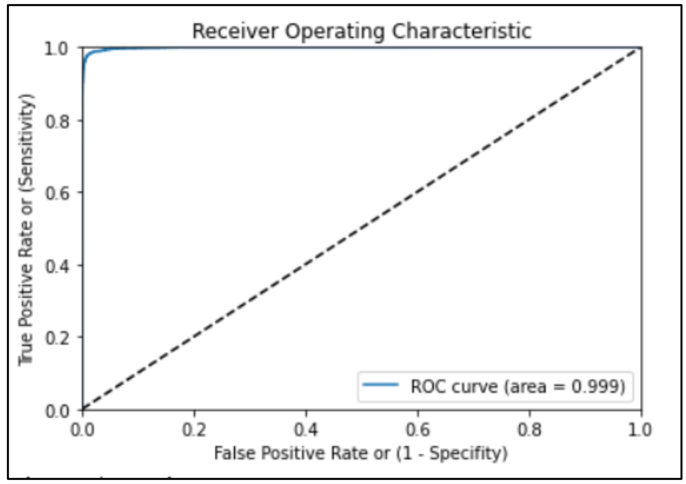

Fig. 4 *Plot for Reciever Operating Charesteristics for Linear SVC model*

VII. Conclusion

The paper thus proposes a method to detect the presence of pedestrian in single images. The proposed method is built upon the feature extraction based on Histogram oriented gradient method. The HOG features are subsequently coded and aggregated for training on a Machine learning algorithm based on Linear Support vector machine. The proposed method can be seen producing superior results on classification to identify pedestrian in the single images.

References

[1] X. Yuan, L. Cai-nian, X. Xiao-liang, J. Mei and Z. Jian-guo, "A two-stage hog feature extraction processor embedded with SVM for pedestrian detection," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, 2015, pp. 3452-3455

[2] David G Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, vol.50, No. 2, 2004, pp.91-110.

[3] Herbert Bay, Tinne Tuytelaars and Luc Van Gool, "Speeded-up robust features (SURF)," Computer vision and image understanding, vol.110, No.3, 2008, pp. 346-359.

[4] "Module: feature—skimage v0.17.dev0 docs", Scikit-image.org.[Online]. Available:https://scikit-image.org/docs/dev/api/skimage.feature.html#skimage.feature.hog

[5] Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; San Diego, CA, USA. 20–25 June 2005; pp. 886–893.

[6] Witten, I. H., Frank, E., Hall, M. A., Pal, C. J.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann (2016).

[7] R. G. J. Wijnhoven and P. H. N. de With, "Fast Training of Object Detection Using Stochastic Gradient Descent," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010, pp. 424-427.