Pedestrian detection using Retinex based image enhancement algorithm and Faster R-CNN

MSc Research Project
Data Analytics

Vaibhav Aher
Student ID: x18104215

School of Computing
National College of Ireland

Supervisor: Dr. Paul Stynes

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Vaibhav Aher
Student ID:	x18104215
Programme:	MSc Data Analytics
Year:	2019
Module:	MSc Research Project
Supervisor:	Dr. Paul Stynes
Submission Due Date:	12/08/2019
Project Title:	Pedestrian detection using Retinex based image enhancement algorithm and Faster R-CNN
Word Count:	
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	15th September 2019

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Pedestrian detection using Retinex based image enhancement algorithm and Faster R-CNN

Vaibhav Aher
x18104215

Abstract

There are many practical applications where pedestrian detection is playing a key role. The pedestrian detection can be useful in smart automated driving vehicles, Intelligent surveillance system, Heavy Industrial-Worker safety systems. Nowadays pedestrian detection has been achieved by various machine learning methods however the main requirement for further research is computational speed and low error rate. The factors like an illusion, complex human poses and occlusion of the pedestrian make this topic challenging in the computational vision field. Deep learning has achieved good accuracy in pedestrian detection, models like YOLO and SSD are remarkable methods with great accuracy for pedestrian detection but due to pooling layers in this model causes the loss of valuable information. Changing Illumination of the day causes difficulties in pedestrian detection. For this problem, Retinex helped to enhance the image with an increase of colour intensity and this helped for improvement in accuracy. This research aims to implement Faster RCNN with non-maximum suppression algorithm for pedestrian detection. Non-maximum suppression algorithm helped for deciding one bounding box among multiple Faster RCNN model is applied on Caltech-USA dataset and evolution of the model is checked based on accuracy which results in 75.74%. The use of Retinex helped to improve the accuracy by 2.8% for FRCNN model. The transfer learning model based on Histogram oriented gradient is also implemented with an accuracy of 56%.

Keywords Pedestrian detection, SSD, YOLO, Faster RCNN, Retinex, Sliding window.

1 Introduction

Most of the shops, malls and houses are under cover of CCTV for security purpose but nowadays an intelligent surveillance system is on-demand. It is important to detect the suspicious activity of a human, the first stage in an intelligent surveillance system is pedestrian detection. It is challenging for a human to see and catch suspicious activity of every moment so in this case, it's better to train the machine with some algorithm. Along with advanced security systems pedestrian detection also area of interest for an automated driving vehicle. Most of the accident on the road are due to human error so smart vehicle system can control the accidents Hongquan, Tongyang, Zhiyong and Yuan (2018).

Object detection has a long history in the computer vision field. During early research, the main challenge to the researcher is low-quality image data in addition to that the hardware configuration was also low at the early time. This causes more computational cost and less accuracy for object detection. Today deep learning is one of the popular machine learning technique which is applicable in many fields. Computer vision researcher focusing on improving the precision, performance with low computational cost. Pedestrian detection can be done by three techniques background modelling, template matching and statistical learning. The earliest research in the field of object detection has done by Viola and Jones algorithm Fahim et al. (1999). Many methods have been implemented by researchers for pedestrian detection. Some are tried by considering specials characters for pedestrian detection. Chebrolu et al. (2019) has developed the model for pedestrian detection using night vision camera and he has used SVM (Support vector machine) for the same task.

Cheng et al. (2018) Have applied deep learning neural network approach for detection and in addition to that SVM and AdaBoost for feature extraction. They have treated each body part of the pedestrian as a single separate segment. In this work, In this research the Retinex is implemented for image enhancement purpose and which also reduce the noise in the image along with that the model has been trained on Faster RCNN (FRCNN). The purpose of use of the non-maximum suppression algorithm is to select the best bounding box or best coordinate which gives the information of the presence of a pedestrian in the image

1.1 Motivation and Background

Pedestrian detection is one of the important branches of computer vision study. Though the research on pedestrian detection has started since the early 80s but nowadays due to a different application, pedestrian detection is today's need. Surveillance system nowadays becoming more smarter than earlier. It's difficult for a normal human to observe CCTV footage all time and without error. The intelligent surveillance system is the system trained with a machine learning algorithm which contains trained model to find pedestrians with complex poses and suspicious activities. Training and implementation of the intelligent Surveillance system are comparatively cheap and error-free. Along with intelligent Surveillance system human detection is famous in heavy industrial machines. In most of the forging industry with automated forging machines, it is important to check the human presence in the forging area. Due to high temperature and risky handling, it is important to set up an automatic human detection system which can abort machine operation and turn warning alarm on.

The smart driving vehicle is also another application in which pedestrian detection research is needed most thoroughly. For the smart vehicle, the pedestrian detection algorithm should be quicker with less computational time and very high accuracy. Ford is developing a new system which will warn the driver after detection of human at a specific distance and if the driver ignores warning then it will automatically control the speed for safety. pedestrian detection research has the potential to use it in an automatic driving system or vehicle security system only important thing is that It should come up with a minimal error and quick processing time.

There are many challenges in pedestrian detection, first and most important challenge is complex human poses. Due to different human poses, the model needs to train with huge data which increases the training cost of the model. Another challenge is the different background and texture. Hence, the model must be trained with the dataset which contains the images with a variety of background and different texture. Pedestrian detection is very sensitive as it directly related to applications which are useful in human life, hence it is important to localize the pedestrian correctly in the image. Due today and night, the changes in illumination take place in recording data hence in all type of illusion condition the model should work with the same efficiency. There is various dataset available which are subset by morning, evening and night dataset or dataset with a different background. There are many techniques available by which pedestrian detection has been achieved.

A decade ago there were systems with very low configuration as compare to the hardware available today hence the computational time and the cost for training was very high, but still, the accuracy was comparatively low. The image quality of the image was also poor but the challenges in todays phase are different. Now pedestrian detection is accuracy and time-based. There are various ways by which pedestrian detection can be achieved out of which the models with deep learning showed remarkable accuracy. CNN is better for deep learning due to its back-propagation. The back-propagation causes the trained weight with a better value which helps in improving the result.

Although these models are good for accuracy in case of the pedestrian with fewer occlusion areas. In the case of occlusion due to multiple bounding boxes of region proposal affect the overall accuracy. Using non-maximum suppression the bounding box with the highest values can be selected as a major region proposal for pedestrian detection. The model with Faster RCNN can be implemented with less computational time and multiple bounding boxes problem can be resolved by combining it with non-maximum suppression. This research is based on the same model FRCNN with non-maximum suppression and its motive is to check the accuracy or an improvement in the model.

1.2 Research Question

"To what extent the accuracy of pedestrian detection model can be increase by using Retinex based image enhancement with Faster RCNN ? "

The aim of this research project is to analyse accuracy of pedestrian detection model by applying Retinex as an image enhancement algorithm.

1.3 Research Objectives and Contribution

This research cover the following objectives to answer this research question

- **Objective 1:** Data Pre-processing and application of retinex on pre-processed data
- **Objective 2:** Development and implementation of FRCNN model with Non maximum suppression for pedestrian detection for enhanced images by retinex and for original images separately.
- **Objective 3:** Building Histogram oriented gradient model for the reference purpose.
- **Objective 4:** Comparison and evolution of FRCNN model with and without Retinex based enhanced images, Histogram oriented gradient model.

In this research project. We have implement the Deep learning based FRCNN model for pedestrian detection. The Non maximum suppression is implemented for improvement of localizing the pedestrian using region proposal bounding box. The second model is based on the knowledge of Histogram oriented gradient. This research contributed to the field of pedestrian detection by following value addition.

- The implementation of Retinex for image enhancement with FRCNN model.
- Region proposal bounding boxes using Non maximum suppression in the case of pedestrian occlusion situation.
- Analysis of the results of developed models with existing model.

The structure of the report is as follows 2 presents the critical study of different pedestrian detection researches and methodologies.In this section different machine learning algorithms and deep learning techniques are critically analysed. 3 present the methodology which is further followed by 4.3 in which results are discussed.

2 Related Work

2.1 Introduction

In object detection pedestrian detection is also a key challenge because of different background, texture and changing poses of the pedestrians. Occlusion of the pedestrian also makes it more complex. Pedestrian detection is one of the essential parts of the video surveillance system. Many car manufacturers are working on automatic pedestrian detection systems. Since early 90s pedestrian detection has been always active research area in machine learning. There are many real-life applications of the object detection system like safe driving systems in the car, Automatic driving vehicle, advanced video surveillance, Industrial robotic vision systems *Deep Learning based on CNN for Pedestrian Detection: An Overview and Analysis.* (2018a).Many researchers have tried to detect and classify fast-moving objects in the video. There are various techniques to detect the object and aim of this section is to review different methodologies and techniques of object detection. This review includes step by step progress of object detection from the early stage of researches to the latest research in the same domain.

2.1.1 A critical review of Pedestrian detection

2.1.2 Introduction

In the above section different challenges and the critical overview of the pedestrian detection are explained.In addition to that the factors which are important to pedestrian detection are explained with the help of previous research.

2.1.3 A critical review

The main need in the field of pedestrian detection field is accuracy and speed. In the initial phase in the research of computer vision field, the main challenge was low configuration hardware. It takes more training and processing time in image processing if the hardware is low configuration. Also, Initially the image quality was low as compared to now hence the training with noisy images resulted in low accuracy model Liu and T. (2017). Pedestrian detection is more challenging due to background interference. The factors like illumination, brightness, darkness and different background texture viz (mountains, beaches, streets) make this topic more challenging. Melo and D. (2014) proposed a sliding window approach for pedestrian detection. This algorithm developed in term of scanning the different region of the pedestrian in the image, it scans the part of the body or covers part of a person but the problem with this model is the detection of the pedestrian body part as a single bounding box. However, the model proposed by L et al. (2018) proposed a deep learning-based model and it has more accuracy for pedestrian detection. This CNN classified the object as a pedestrian and non-pedestrian. There is various public dataset available for pedestrian detection. Caltech is one of the richest datasets. This research also used the same dataset. The occlusion is the main problem in the detection of the pedestrian. The demand for smart surveillance system is increasing day by day hence the pedestrian detection is selected as a research topic. The next section covers the previous research done on pedestrian detection and identification of the research gap in the literature.

2.1.4 Conclusion

From this section we come to know about the different challenges of pedestrian detection. The factors like illumination, texture are the key challenges in pedestrian detection.

2.2 Critical review of Deep learning and machine learning for pedestrian detection and identification of gaps

2.2.1 Introduction

This section is about the critical analysis and overview of the different deep learning and machine learning techniques for pedestrian and as well as object detection. The techniques discussed in above section are HOG with SVM as a classifier, SSD, YOLO, Multi DNN model and MFSNN (Multi-DNN with fusion siamense Neural Network)

2.2.2 Critical review of Deep learning Techniques

The accuracy and computation speed are two important factors in the computer vision system. The machine learning algorithms are used to classify the images into an image that contains a pedestrian and the image that does not contain the pedestrian. Moreover, machine learning helps to extract the feature from the complex dataset. The earlier research on pedestrian detection was using histogram oriented gradient (HOG) model. Fahim et al. (1999) implemented histogram oriented gradient with (SVM) support vector machine as a classifier. Though the accuracy of this model is 82% the computation time is very high. This approach follows template matching as a first layer for the human detection then the output of template matching passed through histogram analysis model.

Single-shot detection (SSD) was implemented by Wenjing et al. (2018) for pedestrian detection in a crowded situation. This research follows the implementation of 5*1 convolutional filters whereas generally, the standard is of 3*3 kernels. This customised filter is to avoid square shape which caused a reduction in redundant information. However, this causes the horizontal direction for searching the pedestrian due to which this model took more computational power and hence it is slow in comparison with YOLO model which is discussed further. The implemented SSD model was single-stage detection whereas FRCNN is a two-stage model. This model has achieved the accuracy of 88% for pedestrian with occlusion more than 75%. The model was tested on INRIA dataset.

Another innovative approach implemented by Chuang et al. (2019) which was a region-based algorithm for pedestrian detection. In this model, the original image is sampled into two different ratios to form a pyramid structure for better learning of the model. The developed pyramid input was given to YOLO model and it resulted in the improvement of normal YOLO model by 2%. However division of image resulted into more computational cost as well as time.

Multi-DNN with fusion siamense Neural Network (MFSNN) was advocated as a novel approach for object detection based on features by Chao et al. (2017). The model implemented in this research is for vehicle detection which was done by using the two-step model. In the first step, the colour identification was done by using Alex Net and for the model, identification was based on VGG net. In this model, the output of Alex net was given to VGG net model for object identification. The MFSNN used here for pattern matching for a small and large object. Multiple branches were developed to extract the different feature like a colour, model which helped to identify the background and the object separately. The evolution of this model was done based on Missing rate which was 30%.

Amir and M (2018) implemented a convolutional neural network-based model for pedestrian detection but in this research, it was focused on an aerial view of the pedestrian. This research is different from others because the object that localized in the image is smaller but in another case, the pedestrian occupied more than 25% of the space from image. This model is based on two-stage sub-models in which in first stage frame is applied to the pedestrian in the image by ignoring the confidence score less than threshold average score and this is done by single-shot detection technique. For the second VGG16 is connected to the first model. For this research total, 77365 images were used. In the second model, the human actions were classified in three classes viz interacting (hug, handshake), non-interacting (eating, sleeping, yawning), the interaction of human with objects (talking on the phone, playing) etc. This two-stage multiple labelled framework showed the accuracy of 72.70%. Although this model is good enough accurate to consider for further research the cases of occlusion are not considered in this research.

Cheng et al. (2018) in his research presented the model for pedestrian detection in occlusion cases. High occlusion of the pedestrian is one of the important problems in pedestrian detection. This research proposed Part and Context Network (PCN) based model in which it has used two features for pedestrian detection, first one was body part segmentation and another one was contextual information. Long short-term memory (LSTM) was used to synchronise the contextual information of different body parts. RoI-Pooling layer was used to extract the synchronised features. The evolution of this model was based on the missing rate which was 16% for Caltech dataset and this evolution is better of the models that discussed earlier. The body segmentation made this model complex and due to the RoI pooling layer, the loss of valuable information from the image may occur. Moreover, the two-stage model caused more computational time for training and testing.

2.2.3 Conclusion

The different deep learning and machine learning techniques has been reviewed in this section. The most focused techniques for this research are YOLO D model with Retinex based image enhancement and Fast RCNN model.

2.3 Critique for implementation of pedestrian detection and image enhancement techniques

2.3.1 Introduction

This section introduce the actual implementation of different deep learning techniques with application of different image enhancement algorithms. The image enhancement techniques are critically analysed with the positive and negative effects in the perspective of data loss.

2.3.2 Critique for Image enhancement techniques

At the early stage of the research in computer vision along with low configuration hardware, another problem is low-quality image dataset. Colour image consists of high information for object detection and classification techniques. Mostly the problem occurred in the low-quality images that contain blur or low contrast. Due to the presence of noise in the image dataset caused the formation of low accuracy models. Qina et al. (2018) implemented image enhancement algorithm named Laplace operator image enhancement algorithm. The main requirement for Faster RCNN was high-quality image dataset. The high-quality dataset is better for better performance. The performance of pedestrian detection was good in case of image enhancement as compare to the model with original images used. But the problem with Laplace algorithm is the longer processing time. Also, Laplace algorithm is mainly used for the reduction of blur effect in the image but it does not perform at the edges of the object.

Hongquan, Meihan, Changnian and Yun (2018) proposed CNN model with image enhancement by ACE (Automated colour Enhancement). The research has done on tunnel pedestrian dataset with 43954 images. The pedestrian were sampled as per different body part by using AdaBoost algorithm. The ACE used in this research it automatically decide the contrast ratio in the image as per colour contrast in the present image. The colour formulation was not similar for all images, it varied from image to image. This was done by a mathematical formula which run time calculate the brightness parameter for the image. As the name suggested automatic colour enhancement hence this automation is slow as compare to Qina et al. (2018). But the result of this model is better than other model and is improvement of 3% in accuracy by using ACE for image enhancement.

Another model with image enhancement for pedestrian detection was by using Retinex and this model was proposed by Hongquan, Tongyang, Zhiyong and Yuan (2018). Retinext used in this model was a simple algorithm used for image enhancement by providing some colour contrast parameters in a JSON file. Retinex is comparatively fast. In this model YOLO-D (You Look Only Once) model was proposed in which object detection was carried by horizontal frames. This model has achieved remarkable accuracy of 94% for pedestrian detection. But this research was performed on a very small dataset which contained 200 pedestrians only. The same image enhancement technique we have used in this research project with Faster RCNN model. The contrast format which gave remarkable accuracy for YOLO-D model, the same format we have included in this research for pedestrian detection. The image enhancement is the part of data pre-processing which is done separately before passing the data to the model. Retinex is an algorithm which fixes the light in the image if required hence it is applied to the image when there is a requirement of light fixing only. This makes Retinex faster comparative to other image enhancement method discussed earlier.

2.3.3 Conclusion

This section explained the different image enhancement techniques and its effect on the evolution of the models. The best suitable technique was selected for this research.

2.4 Comparison of Reviewed Models and Techniques for pedestrian detection

2.4.1 Introduction

Comparison of different model on basis of their accuracy is summarised in the Table 1. The table is based on the different models discussed in section 2.

Table 1: Comparison of Machine learning Model and accuracy.

ML Model	Dataset	Accuracy	Author
YOLO-D	INRIA	86.77%	Zhong et al. (2018)
Single Shot Detection	Caltech	77%	Wenjing et al. (2018)
Fast RCNN	Caltech	81.33%	Mahmoud and Ali (2018)
Multi-stage Contextual DNN	Caltech	55.60%	Zeng et al. (2013)
Joint Deep Learning	Caltech	60.70%	Ouyang and Wang (2013)
Deep Learning Sementic task	Caltech	71.20%	Tian et al. (2015)

2.4.2 Conclusion

This section gives the information about different machine learning and deep learning techniques on different dataset with the accuracy achieved.

3 Methodology

3.1 Introduction

There are various methodologies have been used to implement the data mining techniques. It depends on the flow of the project like gathering the data than processing it and execute the suitable model and finally evaluate the result etc. Among all methodologies, CRISP-DM and KDD are the two most used methods. After analysing project requirement and studying different methodologies it was concluded that it is suitable to use Knowledge Discovery Databases (KDD) methodology to implement machine learning algorithm for pedestrian detection.

3.2 Knowledge Discovery Database method

3.3 Introduction

KDD methodology is suitable where the requirement is of deep analysis of the data and extract the knowledge from the data. The KDD methodology is used in this project as per following the existing steps.

3.3.1 KDD Method

(i) Collect the data with positive and negative annotation of the pedestrian. (ii) Convert the video file which is present in the .seq (sequential) format to a jpeg file and link the respective annotations to the respective image by using LSTM (long short term memory) algorithm. (iii) feature extracted by using the convolutional neural network (CNN) (iv) Faster RCNN with Non-maximum suppression and Histogram oriented gradient with SVM as a classifier. (iv) The result of both Machine learning techniques is evaluated and compared concerning their accuracy.

3.3.2 Conclusion

KDD methodology was selected for this research project.The project followed by the different KDD steps mentioned in above section.

3.4 Process Flow Diagram

3.4.1 Introduction

In the above section the process flow diagram is explained in the detailed. The process flow diagram is divided in to three stages and all the stages are explained in above section.

3.4.2 Process flow diagram of pedestrian detection Model

Figure 1 Indicates the process flow of this research project. It is divided into three parts which are Pre-processing, Pedestrian detection and Evolution and Results. All the processes in Figure 1 are thoroughly explained in this section. This diagram help us to understand the flow of the project. The first section processes are explained in 3.5, 3.5.4 and 3.5.5. The second stage which is pedestrian detection it is the model that is implemented in this research project and it is explained in 4. The last step in the process flow diagram is Evolution and Result which is explained in 4.3

3.4.3 Conclusion

From process flow diagram it is easier way to understand the flow of the project. This is diagrammatic representation and the detailed part is explained step by step.

3.5 Data Extraction and pre-processing

3.5.1 Introduction

Caltech USA pedestrian dataset provides 10 hours of 640x480 30Hz video which is publicly available dataset for pedestrian detection purpose. The dataset consists of annotations file in which annotations of the pedestrian region of different frames are provided with the same dataset. Pedestrian detection is useful in many application like smart vehicles and robotics hence California technological university also developing the models for pedestrian detection and they have made this dataset publicly available for research purpose.

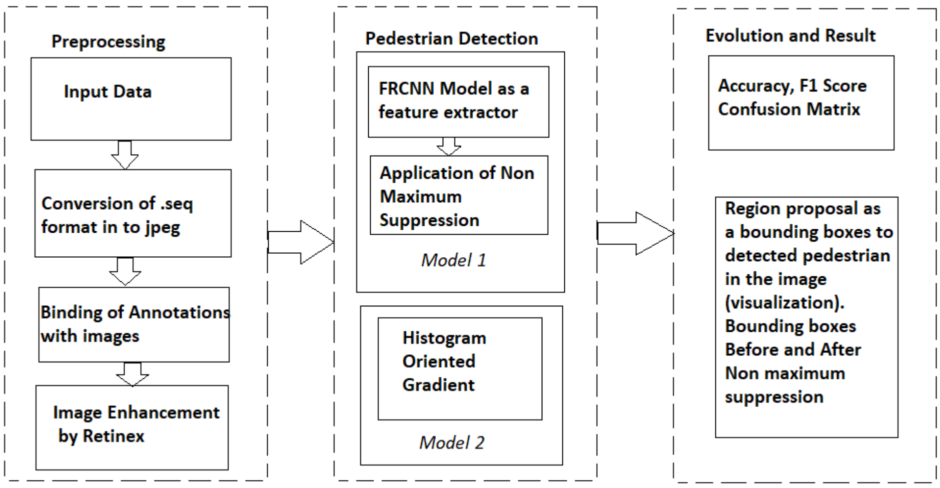

Figure 1: Process flow diagram

3.5.2 Data Extraction

The data can be directly downloaded from the website which contains the video recording of 10 hours. Along with the video file, the annotations of the recorded video file is also available. Annotations consist of the file path, x1, y1, x2, y2 coordinates and the file is available in JSON format. There are total 2,50,000 frames in total and 3,50,000 bounding boxes. But due to the limitation of hardware configuration only first three sequential files are used for this model.

3.5.3 Data pre-processing

As the data is available in sequential format (.seq) files, hence for this research first task is to convert it into jpeg files. for this conversion purpose python version, 3.7.2 is used. To convert sequential files into jpeg following libraries were used 1) cpickle 2) skipy 3)struct. For our second model, we have divided training and testing data as positive and negative in terms of the presence of a pedestrian in the image.

3.5.4 Exploratory Data Analysis

After completion of downloading data, it is important to check whether the file is corrupt or not because if it is corrupt then during conversion it throws the error and the possibility of repetitive data with wrong annotation bounding. There are approximately 2300 unique pedestrian detected. About 50% of data have no pedestrian, whereas 30% have two or more. Figure 2 shows the height distribution in terms of pixels far scale is under 30 pixel and near is above 80 pixels and around 69% data are medium scale. The video is available into 640 X 480 resolution. In Table 2 description of dataset is provided.The Caltech USA pedestrian dataset and its detailed description is available on caltech usa website.

3.5.5 Image Enhancement

Noise in the image means blur or imbalance brightness. There are various techniques used for noise reduction in the image. Retinex is one of the most popular techniques used for

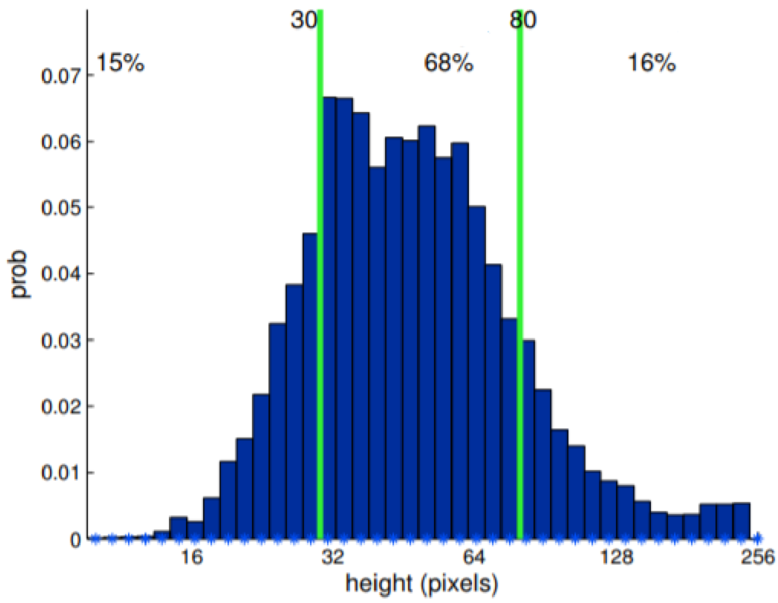

Figure 2: Height distribution

Table 2: Dataset Summary

Attributes	Numbers
Labeled Frames	1000k
Bounding boxes	350K
Occluded Bounding boxes	126K
Unique Pedestrian	2300
Avg Pedestrian duration	5 seconds

image enhancement purpose. For this research, the testing data images are enhanced by using the Retinex algorithm. The script for retinex is developed in python using numpy and cv2 libraries. Retinex simply improves the brightness in the picture which helps to decrease the noise in the image. For object detection image with less noise help to increase the object localization probability. *Deep Learning based on CNN for Pedestrian Detection: An Overview and Analysis.* (2018b)

3.5.6 Feature Extraction

Deep learning, features are automatically extracted within different layers whereas in machine learning features are need to be extracted separately. This is how deep learning is different from machine learning, no need to extract feature manually or by using different algorithms. The power of deep learning is the different layers present act as a feature extractor. There are input layer, an output layer and in between then, there are several hidden layers which act as a feature extractor. The input layer is used to feed the data in CNN whereas output layer gives the output and in between hidden layers helps CNN to optimise the result by calculating the precise weight for the model. The weights are

Figure 3: Image before Retinex

Figure 4: Image After Retinex

updated continuously throughout the process while training the dataset.

3.5.7 Conclusion

The implementation of Deep learning model and its results are discussed in the above section.The above section explained the data pre-processing part in the detailed manner. In addition to that the exploratory analysis is also explained in the same section. This is the first step in the implementation of the research project.This project follows the KDD approach in which Caltech pedestrian dataset is used which is publicly available. The extracted dataset is then pre-processed and image enhancement is achieved by the Retinex algorithm.

4 Implementation and results

4.1 Introduction

This section discussed the implementation of deep learning model, its result and evaluation. Along with implementation pre-processing of data, image enhancement steps in Retinex and transfer learning of Histogram oriented approach is also discussed. For evaluation of model Accuracy as a measure is used along with accuracy F-Score and confusion matrices also compared. The implementation is done by using advanced Tensorflow library. Tensorflow has many inbuild function which makes it easier to implement deep learning models.

4.2 Data pre-processing

4.2.1 Introduction

This is the initial stage for the implementation of FRCNN model for pedestrian detection. Some part of the data pre processing is explained in the section 4.2.

4.2.2 Data Pre-processing

As discussed in section 3.5.3 the data pre-processing is one of the important steps in object detection modelling. The raw data available for this project was in a sequential format which can not feed directly to the model. The first thing was to convert the sequential data into image format which is jpeg format. Along with with dataset the JSON file with the metadata of the sequential file was downloaded. In annotation files the various attributes like position of the pedestrian in the image are given in terms of coordinate, along with coordinate the more information like the presence of occlusion , one or more pedestrian this information is also given, The available information is for a particular frame of the video file hence using python script the annotations are linked with respective frame. The next step is noise reduction in the image by enhancing the feature of the image. The reduction in the image noise gives better features in the layers. As discussed the output of the previous layer is input for next layer hence it is important to pre-process the data to avoid misleading results if this step went wrong then it can affect the performance of the whole model. For noise reduction and image enhancement purpose, Retinex is used as it fixes the brightness in the picture which helps in object detection Hongquan, Tongyang, Zhiyong and Yuan (2018). The packages like numpy, pillow, cv2 were used for image enhancement.

4.2.3 Implementation of the models

The input data to this model are images linked with annotations which contain the region proposal information in the format of coordinates is added to the input layer of the model. The training dataset is from seq1 to seq 5 and the testing dataset is set 6 to seq 9 but for hardware limitation purpose this model is trained and tested on every 10th frame instead of the whole dataset. Tensorflow API is used to implement Faster RCNN model. The first layer after the input layer in the model is to covert the input image dimension which is 600x800x3 into 37x50x256 dimension.

Step1 : Conversion of seq. files in to jpeg

Step2 : Exploratory analysis and image enhancement by Retinex.

Step3 : Developed and implemented the model for the original data (images without Retinex)

Step4 : Developed and implemented the model for the enhanced images data (images with Retinex)

Step5 : Developed HOG (Histogram Orieneted Gradienet Model) for comparison purpose.

Step6 : Compare and analysed the results of each model individually.

Faster RCNN uses the selective search algorithm which detects the region of interest and passes it to the convolutional layer. Generally, the ROI pooling layer is popular for deep learning models but In this model deep region proposal network RPN is used as a pooling layer in CNN. The specific ratio for the image is defined as 2:1 which reduce the size of the input image, this layer saves computational time for CNN. The output of this layer is connected to the input for two different layer one by one, the first one is the classification layer and after that regression layer. The additional layer of Non-maximum suppression is added before the classification layer which finds the bounding box around the detected pedestrian in the testing data. Non-maximum suppression (NMS) helps the model to reduce the bounding boxes in case of occlusion images of the pedestrian. In the final stage, softmax function is used to classify between pedestrian or non-pedestrian. The classification layer is the final layer in this model, It predicts the class name for the given input in testing data. In the end, accuracy is measured as a metric.

Another model is trained by transfer learning. The model is based on Histogram oriented gradient in which the image is scanned slide by a slide from left most coordinate to bottom side right-most coordinate. For implementation purpose, the libraries cv2, numpy, sclearn, argparse are used. Initially, in one function the size of the slide have to be declared and with increment direction of x and y-axis, the image is parsed step by step. Each slide has the probability as a measure and the slide with the highest probability contains the desired object which is pedestrian here. This model is comparatively simple but it takes more computational time. The model is tested on the same dataset.

4.2.4 Conclusion

The data pre-processing part is explained in the above section. The image enhancement was also the part of data pre-processing. The well prepared data was then passed through the model for further implementation. The outputs were recorded in the form of confusion matrix. The evolution section contains the details of the result of each model and explanation in next section.

4.3 Evaluation and Result

4.3.1 Introduction

In this section the output of each model is explained individually and compared between. The analysis of the result also been discussed in this section. The results are also expressed in graphical format for easy understanding.

4.3.2 Evaluation and analysis of Result

Accuracy is used as a performance metric for Faster RCNN model. From confusion matrix, it can be seen the true as well as false prediction. The model was run twice. On the first iteration, it was run on the original dataset that is the dataset without Retinex. Figure 5 was the resulted accuracy of the model which run on original dataset where as Figure 6 is the resukt of Retinex based model.The Accuracy obtained on the original dataset is 72.94% with a sensitivity of 0.66 and specificity of 0.79 whereas in second iteration model was tested on the dataset which was enhanced by using Retinex and the model gives the accuracy of 75.74 %. Hence it can be seen that Retinex improves the accuracy of FRCNN model by 2.8%.

The F1 score for the improved model is 0.74. and the precision score is 0.7582. In this model 11249 images were classified as a pedestrian whereas 6350 images are classified as non-pedestrian.

Figure 7 is about the result of region proposal without bounding boxes, as for each slide the part of pedestrian form one bounding box and multiple slides causes the number

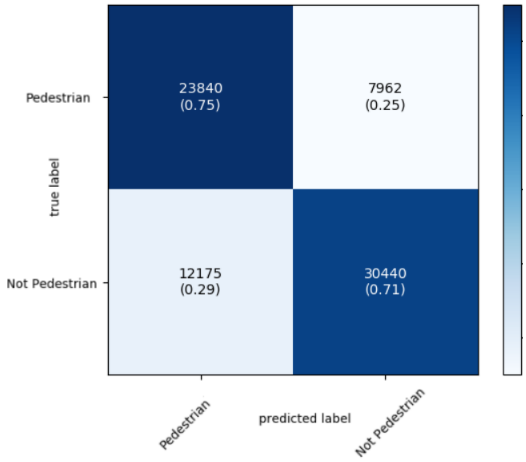

Figure 5: Confusion matrix of CNN model on original dataset

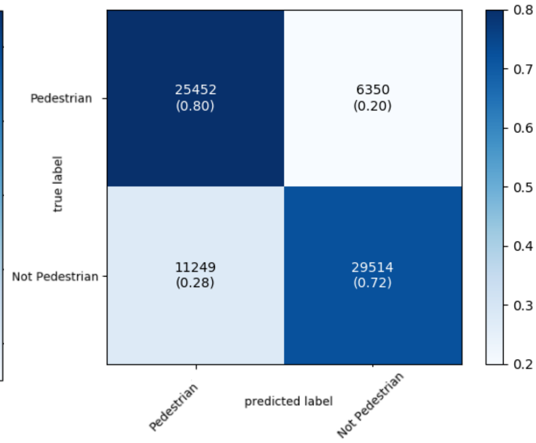

Figure 6: Confusion matrix of CNN model on Retinex based enhanced image dataset

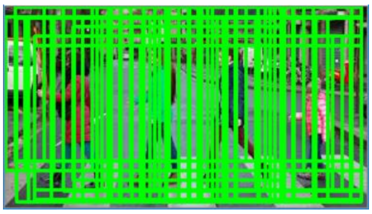

Figure 7: Bounding boxes before Non maximum suppression

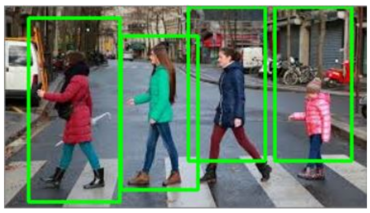

Figure 8: Bounding boxes after Non maximum suppression

of bounding boxes whereas after applying non maximum suppression the bounding box with the highest probability of the pedestrian is the main bounding box which causes reduction in the bounding boxes. Figure 8 shows the result of the bounding box around a pedestrian after Non-maximum suppression.

In the second model which is based on Histogram oriented gradient approach, Figure 9 was the output of HOG model which resulted in to the low accuracy as compare to FRCNN model. The accuracy of this model is around 58% and the F1 score value is 56.14%. The reason behind the low accuracy of this model is sliding window require huge dataset for training and also unlike there is forward and backpropagation in FRCNN which help the model to calculate the weight.

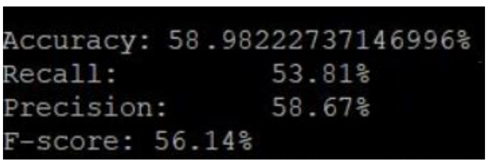
```
Accuracy: 58.98222737146996%
Recall:          53.81%
Precision:       58.67%
F-score: 56.14%
```

Figure 9: Accuracy and F1 score of Histogram Oriented Gradient

5 Discussion

Various image enhancement techniques are discussed in 2 section. From the result, it comes to know that the images with good quality and less noise improve the performance of the model. In some images, pedestrians were not detected due to contrast or brightness issue in the image which can be fixed by different noise reduction techniques. The increase in accuracy is 2.8%. As this model is not tested and trained on the whole dataset because of hardware limitations. The number of data we trained the better will be the performance of the model. The testing and training of the whole dataset may also result in improved accuracy for this model.

The second model considered for this research is the Histogram oriented gradient which is quite an old technique but the result obtained from this model is much lower than Faster RCNN model. The one reason behind the higher accuracy of FRCNN is trained weights. The FRCNN model was trained on around 2k iterations and the best suitable weight is considered for testing purpose. in FRCNN there is forward as well as backpropagation which trained the weight with better accuracy. Another important thing in this project is non-maximum suppression. In general approach for object localization, the multiple bounding boxes is a prime problem but NMS made this work easy by considering the pixel values greater than threshold values which helped to draw less bounding boxes for the region proposal. The bounding boxes can be used in much different application hence the pedestrian localization part is also involved in this research project.

6 Conclusion and Future Work

As research question framed in section 4.1 (To what extent the accuracy of pedestrian detection model can be increase by using Retinex based image enhancement with Faster RCNN ?) The FRCNN model has successfully implemented for pedestrian detection. The image enhancement using Retinex also reflected in an improved result with an accuracy of 75.82%. The pedestrian also localised in the image using Non-maximum suppression based bounding boxes. The data processing task for this project was quite time-consuming because of the low configuration of the GPU. For training and testing of CNN models, high-end hardware with good GPU makes work faster and with better performance. The testing of this model on whole data may result in better accuracy. The dataset was very huge as mentioned in section 4.2.

The limitation in this research is thermal vision-based images were not included in this research. The same model can be tested on infrared images for pedestrian detection. This can help us in an intelligent surveillance system at night. Also for the automated driven vehicle along with computer vision system the use of sensors based model may also useful. The combination of a computer vision system and sensor-based model can also be considered as the next part of this project.

7 Acknowledgement

I am thankful to Dr Paul Stynes and Sachin Sharma for their guidance for this project.

References

Amir, S. and M, N. N. (2018). Convolutional neural networks for aerial multi-label pedestrian detection., *2018 21st International Conference on Information Fusion (FUSION), Information Fusion (FUSION), 2018 21st International Conference on* p. 1005.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Chao, C., Nong, S., Changxin, G. and Lei, Z. (2017). Vehicle re-identification by fusing multiple deep neural networks., *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Image Processing Theory, Tools and Applications (IPTA), 2017 Seventh International Conference on* p. 1.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Chebrolu, R., Naga, K. and P.N., K. (2019). Deep learning based pedestrian detection at all light conditions., *2019 International Conference on Communication and Signal Processing (ICCSP), Communication and Signal Processing (ICCSP), 2019 International Conference on* p. 0838.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Cheng, W., H, L., F, W. and H, Z. (2018). Pedestrian detection via body part semantic and contextual information with dnn., *IEEE Transactions on Multimedia, Multimedia, IEEE Transactions on, IEEE Trans. Multimedia* (11): 3148.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Chuang, X., Pin, L. and Yurong, S. (2019). Pedestrian detection and location algorithm based on deep learning., *2019 International Conference on Intelligent Transportation, Big Data Smart City (ICITBS), Intelligent Transportation, Big Data Smart City (ICITBS), 2019 International Conference on, ICITBS* p. 582.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Deep Learning based on CNN for Pedestrian Detection: An Overview and Analysis. (2018a). *2018 9th International Symposium on Telecommunications (IST), Telecommunications (IST), 2018 9th International Symposium on* p. 108.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Deep Learning based on CNN for Pedestrian Detection: An Overview and Analysis. (2018b). *2018 9th International Symposium on Telecommunications (IST), Telecommunications (IST), 2018 9th International Symposium on* p. 108.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Fahim, A., Ahmed, T. B. and Mohidul, I. S. (1999). Hog and gabor filter based pedestrian detection using convolutional neural networks., *1999 International Conference on Electrical, Computer and Communication Engineering (ECCE), Electrical, Computer and Communication Engineering (ECCE), 2019 International Conference on* p. 1.

URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds
livescope=sitecustid=ncirlib*

Hongquan, Q., Meihan, W., Changnian, Z. and Yun, W. (2018). A study on faster r-cnn-based subway pedestrian detection with ace enhancement., *Algorithms* **11**(12): 192.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=a9hAN=13407
livescope=sitecustid=ncirlib*

Hongquan, Q., Tongyang, Y., Zhiyong, S. and Yuan, Z. (2018). A pedestrian detection method based on yolov3 model and image enhanced by retinex., *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2018 11th International Congress on* p. 1.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds
livescope=sitecustid=ncirlib*

L, M. K., N, H. A. and Satyakama, P. (2018). Pedestrian detection using r-cnn object detector., *2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Computational Intelligence (LA-CCI), 2018 IEEE Latin American Conference on* p. 1.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds
livescope=sitecustid=ncirlib*

Liu, T. and T., S. (2017). Enhanced pedestrian detection using deep learning based semantic image segmentation., *2017 22nd International Conference on Digital Signal Processing (DSP), Digital Signal Processing (DSP), 2017 22nd International Conference on* p. 1.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds
livescope=sitecustid=ncirlib*

Mahmoud, S. and Ali, A. (2018). Pedestrian detection using an extended fast rcnn based on a secure margin in roi feature maps., *2018 9th International Symposium on Telecommunications (IST), Telecommunications (IST), 2018 9th International Symposium on* p. 155.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds
livescope=sitecustid=ncirlib*

Melo, C. D. and D., L. S. M. (2014). An optimized sliding window approach to pedestrian detection., *2014 22nd International Conference on Pattern Recognition, Pattern Recognition (ICPR), 2014 22nd International Conference on, Pattern Recognition, International Conference on* p. 4346.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds
livescope=sitecustid=ncirlib*

Ouyang, W. and Wang, X. (2013). Joint deep learning for pedestrian detection., *2013 IEEE International Conference on Computer Vision, Computer Vision (ICCV), 2013 IEEE International Conference on, Computer Vision, IEEE International Conference on* p. 2056.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds
livescope=sitecustid=ncirlib*

Qina, T., GuangDa, X., YanPing, W. and Yuan, Z. (2018). Pedestrian detection based on laplace operator image enhancement algorithm and faster r-cnn., *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2018 11th International Congress on* p. 1.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Tian, Y., Luo, P., Wang, X. and Tang, X. (2015). Pedestrian detection aided by deep learning semantic tasks., *2015 7th International Conference on Games Virtual Worlds for Serious Applications (VS-Games)* p. 5079.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edbAN=11112 livescope=sitecustid=ncirlib*

Wenjing, Z., Lihua, T., Chen, L. and Haojia, L. (2018). A ssd-based crowded pedestrian detection method., *2018 International Conference on Control, Automation and Information Sciences (ICCAIS), Control, Automation and Information Sciences (ICCAIS), 2018 International Conference on* p. 222.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Zeng, X., Ouyang, W. and Wang, X. (2013). Multi-stage contextual deep learning for pedestrian detection., *2013 IEEE International Conference on Computer Vision, Computer Vision (ICCV), 2013 IEEE International Conference on, Computer Vision, IEEE International Conference on* p. 121.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*

Zhong, H., Lei, Z. and Pengfei, W. (2018). Pedestrian detection based on yolo-d network., *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Software Engineering and Service Science (ICSESS), 2018 IEEE 9th International Conference on* p. 802.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=edseeeAN=eds livescope=sitecustid=ncirlib*