Bachelor of Science in Computer Science & Engineering

Zebra Crossing Detection and Recognition Using Grab Cut Algorithm, Radon Transform and Improved Mask R-CNN.

by

J.M. Imtinan Uddin

ID: 1504039

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

April, 2021

Zebra Crossing Detection and Recognition Using Grab Cut Algorithm, Radon Transform and Improved Mask R-CNN.

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

J.M. Imtinan Uddin
ID: 1504039

Supervised by

Dr. Pranab Kumar Dhar
Professor
Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)
Chattogram-4349, Bangladesh.

The thesis titled '**Zebra Crossing Detection and Recognition Using Grab Cut Algorithm, Radon Transform and Improved Mask R-CNN.'** submitted by ID: 1504039, Session 2019-2020 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

Board of Examiners

______________________________ Chairman

Dr. Pranab Kumar Dhar
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

______________________________ Member (Ex-Officio)

Dr. Asaduzzaman
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

______________________________ Member (External)

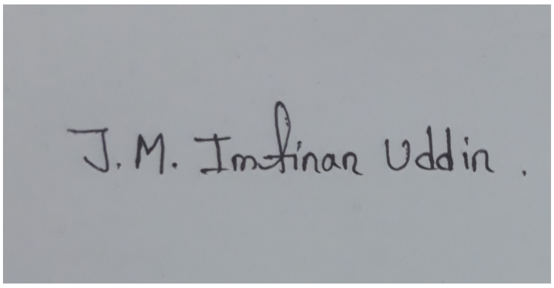
Dr. Muhammad Ibrahim Khan
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.



Signature of the candidate

Date: 19-04-2021

Acknowledgements

The journey of writing this thesis towards its completion has only been possible though all the support and supervision from many individuals. This has been a fulfilling experience, both personally and professionally. I would like to express my heartfelt gratitude towards my supervisor Dr. Pranab Kumar Dhar ,Professor of Computer Science and Engineering Department at Chittagong University of Engineering and Technology. By whose guidance this thesis becomes a reality. I am thankful for his continuous guidance and critical questions that has encouraged me to think out of my comfort zone and has levered my capabilities to take challenges that I never thought could be done by me. I am thankful to the panel members for providing the approval of my work. I am grateful to my parents, sister and brother for their patience in bearing with me thoroughout this whole time. Their unconditional love and encouragement has provided support in every aspect of my life.

Abstract

Zebra crossing is an important traffic sign in our modern traffic system. Zebra crossing helps modern traffic systems to operate smoothly and ensures pedestrian's security and safety. So, detecting and recognizing zebra crossing is a colossal task in this modern era. In this paper, we have come up with a method where handcrafted features of zebra crossing region are combined with deep features of zebra crossing extracted by improved Mask RCNN to detect and recognize zebra crossing. Handcrafted features of zebra crossing region are extracted through image processing techniques such as histogram equalization, radon transform, automatic grab cut algorithm, and uniform local binary pattern. For this, the contrast and sharpness of the zebra crossing images are enhanced by the CLAHE method if the image's intensity value is less than a threshold value. After that images are converted into binary images using the OTSU method and then opening was performed to remove noise. Then we determined the contours of all the objects that are present in the images and kept those objects which have an area that falls in a particular range, have a shape of rectangular, and within a particular aspect ratio. After that we have performed canny edge detection technique to detect all the edges in the images and performed radon transform to detect all the potential parallel edges. Then these potential parallel edges are justified as zebra crossing lines and then through the automatic grab cut algorithm which uses a binary mask of this region we extracted zebra crossing ROI. After detecting ROI of zebra crossing, uniform local binary pattern is used to extract features of this ROI and concatenated this extracted handcrafted features with the deep features extracted by improved Mask RCNN. The result of the proposed method shows that it can effectively detect and recognize the zebra crossing region from various images. Moreover,it shows better performance compared to other state-of-the-art recognition methods which yields an accuracy of 99.3 %.

Keywords: zebra crossing, improved Mask RCNN , radon transform, automatic grab cut, handcrafted features, deep features.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Zebra crossing is being used to regulate traffic systems. Zebra crossing can be of many types, generally zebra crossing is collection of white stripes on a black surface road. By using zebra crossing, pedestrians cross the road safely and at the time of crossing vehicle remain stopped in front of zebra crossing. But it is a matter of sorrow that many people are dying in the world for not obeying zebra crossing rules properly. Not only that but also drivers are not willing to obey the zebra crossing rule. Again, it is very unfortunate that blind people are unable recognize zebra crossing on a road. So, automatically detecting and recognizing zebra crossing to help visually impaired people to cross the road using zebra crossing is a necessary task. Again, in modern age with the help of technological advancement automated vehicle is increasing day by day. So automated vehicles also need to detect and recognize zebra crossing efficiently and in real time in order to obey the zebra crossing rules. In order to detect traffic rules violation at zebra crossing, we also need to detect and recognize zebra crossing at first. For continuously monitoring situation at zebra crossing area, we also need to detect and recognize zebra crossing. For this reason many methods had been developed by researchers to efficiently detect and recognize zebra crossing.

1.2 Framework of detection and recognition

Object detection and recognition is a subject of image processing techniques. For detection and recognition at first it is mandatory to extract features. For features extraction there are mainly two approaches handcrafted features and deep features. After extracting the features, classification or recognition algorithm is performed on those extracted features. Our proposed method mainly consists of following steps:

1. Extracting handcrafted features through image processing techniques such as histogram equalization,radon transform,automatic GrabCut algorithm and uniform local binary pattern.

2. Extracting deep features through improved Mask RCNN.

3. Normalize both extracted feature vectors before combining.

4. Combining handcrafted and deep feature vectors.

4. Using this combined feature vector we have detected and recognized zebra crossing region and instance segmentation.

1.3 Difficulties

Many difficulties may arise while developing a object detection and recognition system. We also have faced difficulties during the development of this framework. Few difficulties that occur frequently is given below:

- **Orientation:** Region of interest may be oriented in different ways in an image.
- **Illumination:** Images may be of different illuminations.
- **Noise:** There may be noise in image.
- **Size of ROI:** Sizes of ROI is may be of different sizes. So finding out appropriate method for segmenting ROI is quite cumbersome.

1.4 Applications

There are enormous applications of zebra crossing detection and recognition framework. Some are given below:

- Detecting zebra crossing for automated vehicle.

- Monitoring traffic situation at zebra crossing areas.
- To help visually impaired people.

1.5 Motivation

Object detection and recognition is an important task in this modern world. In the world of automation zebra crossing detection and recognition has a vital role. Many methods have been developed to detect and recognize zebra crossing. But Not every method is perfect. So we have proposed a zebra crossing detection and recognition method that can efficiently detect zebra crossing with higher speed and accuracy and also improve the previously developed method. So our main motivation of our work is given below:

- To detect zebra crossing under various complex situations and illuminations.
- To detect and recognize zebra crossing in real time.
- To increase accuracy of detecting and recognizing zebra crossing.

1.6 Contribution of the thesis

Research work is conducted to achieve a set of goals whether it is to define a new methodology or to improve the existing ones.In this project, the main focus was given to improve the existing zebra crossing detection and recognition system for obtaining better accuracy and speed.The primary contribution of this thesis is the following:

- We have combined handcrafted features extracted by uniform local binary pattern and deep features extracted by improved Mask RCNN algorithm to detect and recognize zebra crossing.
- We have added an extra convolutional layer in improved mask RCNN algorithm so that it gives 56*56 resolution of mask.
- We have increased the accuracy of detecting and recognizing zebra crossing.

1.7 Thesis Organization

The rest of this report is organized as follows:

- Chapter 2 gives a brief summary of previous research works in the field of zebra crossing detection and recognition.
- Chapter 3 gives description of proposed methodology.
- Chapter 4 provides the description of the working data set and analysis of the performance measure for the proposed framework.
- Chapter 5 contains the overall summary of this thesis work and provides some future recommendations as well.

1.8 Conclusion

In this chapter, an overview of our proposed zebra crossing detection and recognition algorithm is given. Along with the difficulties, the summary of the recognition system framework is described in this chapter. The motivation behind this work and contributions are also described here. In the next chapter, background and present state of the problem will be provided.

Chapter 2

Literature Review

2.1 Introduction

In recent years there have been done several works for developing ideal zebra crossing detection and recognition system by overcoming all the challenges. This section provides the brief discussion related to the existing and previous works of zebra crossing recognition and detection method.

2.2 Related Literature Review

Many researchers have worked on zebra crossing recognition and detection system. Still because of some extreme and complex situations researchers are facing problem to detect zebra crossing more accurately. For example, in [1] authors have used block-based Hough transform to detected crosswalk in natural scene images. But due to lack of in-depth information of images it sometimes detect stairways as zebra crossing. In [2] authors have developed a zebra crossing detection algorithm but this method can not detect zebra crossing that are faded. In [3] a RGB-D image based method it cannot detect zebra crossing for lacking of proper in-depth information in outdoor scenes. The method described in [4] can not detect crosswalk under shadow and noise. In [5] authors used Mask RCNN to detect zebra crossing region as well as instance segmentation of that region. In [6] they have detected and recognized crosswalk using flood-fill operation and uniform local binary pattern. In [7] authors have developed a framework for zebra crossing detection and recognition which has two stage, called coarse stage and fine stage. At coarse stage identifying vanishing points and straight lines associated with the stripes of zebra crossing, rough candidate regions of interest (ROIs) are determined. After that in fine stage by exploring their prior constraint information they have determined whether these candidate ROIs are indeed zebra crossing. A zebra crossing detection method was proposed in [8] where ROI are segmented using a unique geometric features such as horizontal edge segments are arranged in a sorted manner. After that ROI is justified using vertical vanishing point.

In recent time it has been observed that deep learning is being used to detect objects from video and images [9]. Deep learning is a method where it can automatically extract features from images with large amount of training data. So many deep learning based model is being used to detect and recognize objects as it reduces complexity. There are many deep learning based model to detect and recognize zebra crossing. In [10] authors have developed a framework using convolutional neural network to recognize zebra crossing. But the dataset used in this framework is minimally labelled. In [11] proposed method can detect zebra crossing from a specific direction. In [12] authors have developed a CNN architecture using a dataset created by themselves. But in terms of crosswalk detection accuracy it is not very promising method. In [13] they have used deep learning approach based on VGG architecture with relatively large dataset developed by their own. Also they evaluated their model with cross database.In [14] they have used traditional image crossing technique for zebra crossing classification and embed the system in a waerable device. Also various types of zebra crossing detection system is described in [15] - [16]

Many deep learning architectures have been used such as Resnet101, VGG-16 to detect and recognize pedestrian crosswalk region.

However deep learning based model does not always outperforms other model in terms of accuracy [17]. But researchers found out that combination of handcrafted features with the automatically extracted features of deep learning architecture yields better performance than the other state-of-the art method. Various model has been developed using this type approach to detect object from images [18]-[19].

In this paper, we have proposed a method to detect and recognize zebra crossing combining handcrafted features of crosswalk region (ROI) with the deep features extracted by improved Mask RCNN[20]. For extracting handcrafted features, we have used various image processing techniques such as clip adaptive histogram equalization, binarization using Otsu method, canny edge operator for detecting edge, radon transform for justifying crosswalk's parallel lines and automatic grab cut algorithm [21] for ROI segmentation, uniform local binary pattern to extract features of ROI. After extracting the handcrafted features,we have combined theses features with the deep features extracted by improved Mask RCNN. Our proposed model outperforms other state-of-the art model for detecting and recognizing zebra crossing.

2.3 Conclusion

In this chapter we have discussed previously developed method of zebra crossing detection and recognition. We have come to know about various approaches to detect zebra crossing. We have also found out some method's advantage and flaws.

Chapter 3

Methodology

3.1 Introduction

Zebra crossing detection and recognition is a well recognized problem in the arena of Artificial intelligence and computer vision. Many methods have been developed to solve this problem. But not every method is not flawless. To overcome the problems of previous zebra crossing detection framework we have proposed a method to detect and recognize zebra crossing. Our proposed method combines handcrafted features extracted by uniform local binary pattern with the deep features extracted by improved Mask RCNN model. In this section we provide the brief description about our proposed methodology.

3.2 Diagram/Overview of Framework

The primary objective of our proposed method is to detect and recognize zebra crossing and as well as giving zebra crossing's mask as output. There are many methods that detects zebra crossing, though all of them are not robust and accurate. Sometimes, these methods fail to detect crosswalk region at complex environments and situations. In this paper, we are proposing a method to detect crosswalk. This method has two phases. In first phase we have extracted the hand-crafted features using rotation invariant uniform local binary pattern and deep features using improved Mask RCNN deep learning method. In the second phase, we have combined these two types of extracted features and then passed this combined feature through improved Mask RCNN to train our framework. Figure 3.1 shows the Flow chart of our proposed method.

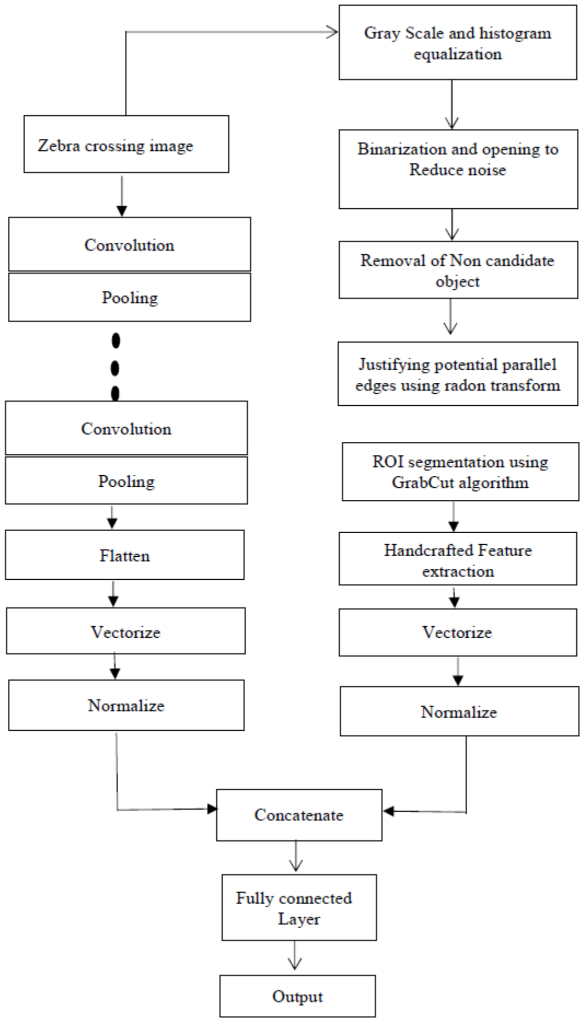


Figure 3.1: Flow chart of the proposed method

3.3 Detailed Explanation

3.3.1 Hand Crafted feature extraction

For the extraction of handcrafted features, we have performed some image preprocessing techniques on the input images. The performed image preprocessing steps are: 1) Converting to gray scale and enhancement of sharpness and contrast, 2) Conversion to binary image and performing opening operation to remove noise, 3) eliminating non candidate objects, 4) detecting edges using canny edge operator, 5) justifying parallel edges of crosswalk region using radon transform 6) segment the ROI using automatic grab cut algorithm. 7) extract features of ROI In the following subsection we have discussed the above steps.

3.3.1.1 Converting to gray scale and enhancement of sharpness and contrast

We have converted the images into gray scale to reduce the operational complexity. Since our input images may have different contrast and sharpness, we have enhanced the contrast and sharpness of images using contrast limited adaptive histogram equalization (CLAHE) method for those images which have contrast and sharpness less than a threshold value. We have set the threshold value to 90. By examining different threshold value, we have observed that it gives the better contrast and sharpness. If any image's average intensity value is less than 90 then we have performed the clip adaptive histogram equalization where contrast limit is set equal to 2 and tile grid size is (8,8). This histogram equalization technique gives us a high intensity image.

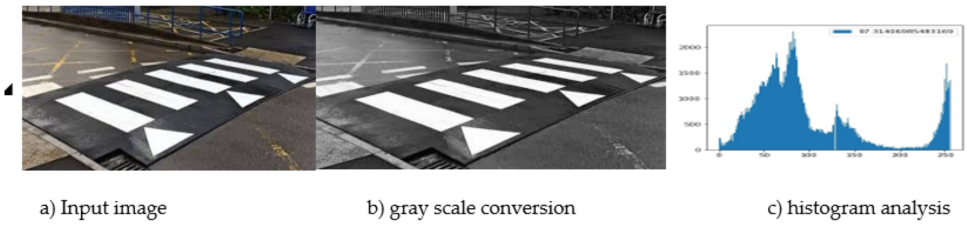


Figure 3.2: Sample for the gray scale conversion and histogram analysis

Image shown in figure 3.2(a) is the input image and image 3.2(b) and 3.2(c) is the gray scale conversion and histogram analysis of this input image respectively. From the histogram analysis of the input image, we see that average intensity value is 97.31 which is greater than the threshold intensity value 90. So, we have not performed histogram equalization on that image.

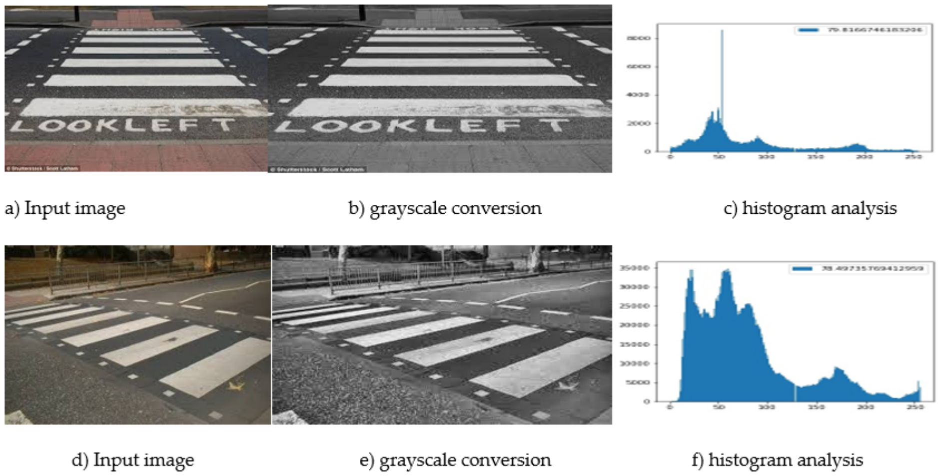


Figure 3.3: Sample for the gray scale conversion and histogram analysis

From the histogram analysis of the sample shown in figure 3.3 we see that their average intensity value is less than threshold value which is 79.81 and 78.49 respectively. So, we have performed histogram equalization on these images to enhance sharpness and contrast.

Figure 3.4: After applying histogram equalization on the input image 3.3(b) and 3.3(e) respectively

3.3.1.2 Performing binarization and opening to remove noise

In this step, we have performed gaussian blur technique and after then binarization technique on the pre-processed images through Otsu method. After performing binarization there are some noises that should be reduced. For reducing noises, we have performed opening operation on the images. In opening operation, the shape of kernel is (3,3).

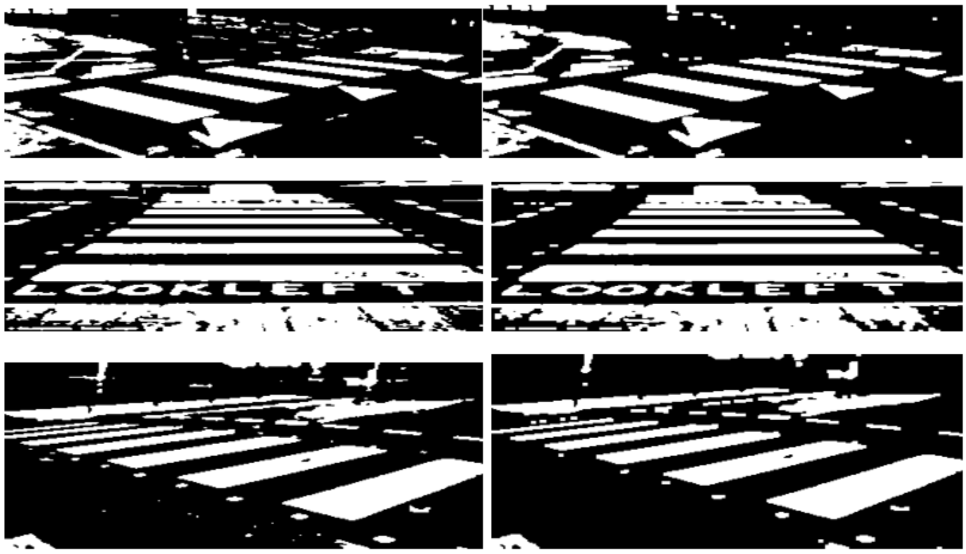


Figure 3.5: samples after performing binarization and opening operation

3.3.1.3 Removal of non-candidate objects

Our primary goal is to extract zebra crossing region from the input images. We know that crosswalk is a parallel collection of black and white stripes. Generally, these stripes resemble the shape of rectangular shaped objects. Using the properties of rectangular shaped objects, we have removed the all unwanted shaped objects from the images keeping only the zebra crossing stripes. For keeping only, the zebra crossing stripes, we have used three conditions. For this, let us take three variables P, Q, R. We have estimated that the area of an object should be between 220-250, the rectangular objects should have a contour length of 4 and the width to height ratio should be between 0.2-0.3. Assigning this condition to variables P, Q, R we get,

$$P = \textit{Area of objects should be between 220-250} \tag{3.1}$$

$$Q = \textit{the contour length of each rectangular objects is equal to 4} \tag{3.2}$$

$$R = \textit{width to height ratio should be between 0.2-0.3} \tag{3.3}$$

From 3.1,3.2,3.3 we get the conditions of all the candidate objects

$$\textit{Candidate objects} = PQR \tag{3.4}$$

For keeping only, the candidate objects, we have calculated contours of all objects present in the image. Then we calculated contours area, length and aspect ratio. After calculating these values and using equation 3.4 we have kept all the candidate objects of the pedestrian crosswalk region and removed non-candidate objects.

Figure 3.6: Samples after eliminating all the non-candidate objects

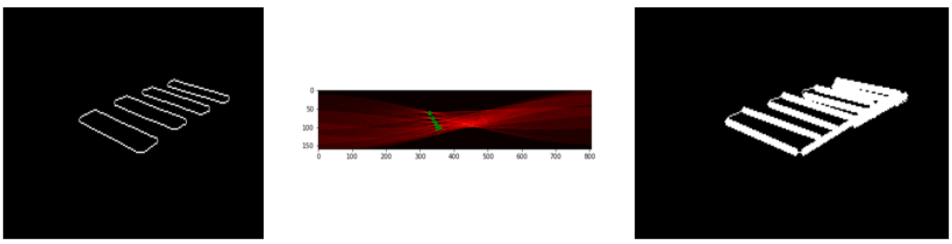
3.3.1.4 Justifying potential parallel edges

In this section we have used radon transform to justify potential crosswalk edges and keep these edges. Before applying radon transform, we have performed canny edge operation on the images to extract all the edges. After that we have performed radon transform. Radon transform can be used to detect lines. From the radon space of the lines we get (r, θ) .

The parameter r denotes the distance between starting and ending point of edge lines and (θ) represents the angle of the edge lines. So, the equation of the edge lines in polar form can be represented by equation (3.5)

$$r = xcos\theta + ysin\theta \tag{3.5}$$

From the radon transform of the canny images we get some peak that corresponds to each edge lines. Interpreting these peak's location using equation (3.5), we can get all the parallel lines of crosswalk.



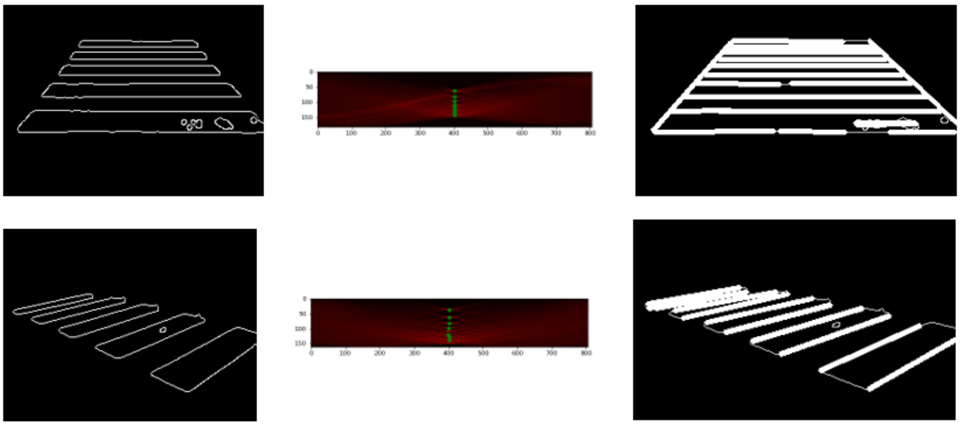


Figure 3.7: Samples for justifying potential parallel edges

3.3.1.5 Extracting ROI

After justifying crosswalk edges, we have used automatic grab cut algorithm. This segmentation algorithm works using gaussian matrix by drawing boundary box around the crosswalk region. Using the boundary box position this algorithm segments the ROI.

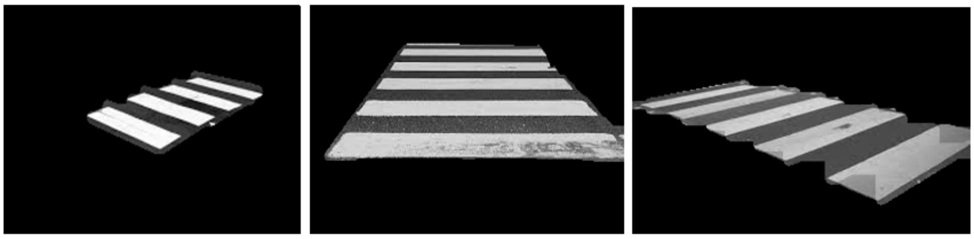


Figure 3.8: Samples of extracted ROI

3.3.1.6 Extracting features

After segmenting ROI, we have used rotation invariant uniform local binary pattern to extract the features of the ROI. We have resized all the ROI to 128*128. For extracting features, we have divided the image matrix by 16*16 window cell. So, the total number of cells in the image is 8. We have set the neighbors P size of 8 because if we set larger neighbor size, we will lose local information. Rotational invariant uniform local binary pattern returns a feature vector of 1 -by-N dimension. Where N represents the number of features. N depend on number of cells * number of bins each cell.

Number of cells = size of image / cell size

Number of bins = P+2 (for rotation invariance LBP)

So rotational invariance uniform LBP returns a feature vector of 80.

Figure 3.9: Feature's vector extracted by LBP

3.3.2 Deep features extraction by improved mask RCNN

After extracting handcrafted features by uniform local binary pattern discussed in section 3.3.1, we have extracted deep features of crosswalk region by improved mask RCNN. In deep learning method deep features are extracted by automatically. For extracting deep features, we have used improved mask rcnn based on resnet101 deep learning architecture. In deep learning architecture there are mainly two parts convolutional layer and fully connected layer. Convolutional layer is the most important building block of deep learning architecture. In convolutional layer, features of images are learned by repetitive convolutional layer by applying filter. Each convolutional layer gives a feature map as output that are fed into following convolutional layer. In our deep learning architecture, each convolutional layer is followed by a pooling layer. Pooling layer functions without dependency. The main task of pooling layer is to reduce the dimensionality of feature map by keeping only the main features. The final layer of a deep learning architecture is fully connected layer. This layer takes the final feature map from last convolutional layer and makes prediction. But in this phase while extracting deep features we have not used this fully connected layer since we only need deep features vector. We get a 2-dimensional feature vector. For each image it returns 4000 features. Now all the hyper-parameter setting is given below:

- **Backbone**: resnet101
- **Batch Size**: 1
- **Learning Momentum**: 0.9

- **Learning Rate:** 0.001
- **Number of Epoch:** 30
- **Steps per Epoch:** 100
- **Validation Steps:** 50
- **Weight decay:** 0.0001
- **Loss Function:** smooth l1 loss

Figure 3.10: Deep features extracted by improved Mask RCNN

3.3.3 Concatenation of handcrafted features and deep features.

In section 3.3.1 and 3.3.2, we have discussed how we have extracted handcrafted features and deep features. In this step we have combined these two features to make a hybrid feature map that will be used for further training for crosswalk detection and recognition. The deep features we have extracted is a two-dimensional tensor. On the other hand, handcrafted features are a Two-dimensional vector. For the sake of combination of two features vector, we have transformed the deep features into one dimensional feature vector. Converting the features into one dimensional vector we have normalized both handcrafted and deep features vector using z-score method. After that we have combined these feature vector which yields a hybrid feature vector.

Figure 3.11: Feature vector after concatenation

3.3.4 Combined feature vectors to classify

After combination of two features vector, we have passed these combined features to fully connected layer.In most popular deep learning architecture, the last few layers are full connected layers which compiles the data extracted by previous layers to form the final output.After feature extraction we need to classify the data into various classes.we generally end up adding FC layers to make the model end-to-end trainable. The fully connected layers learn a (possibly non-linear) function between the high-level features given as an output from the convolutional layers. We have used improved masked RCNN model based on resnet101 architecture. In the training phase, we have tried pre-trained weights of COCO dataset and ImageNet dataset. Finding the best features is done by the convolutional and pooling layer of improved mask RCNN model. In the last layer of this architecture which is fully connected layer, the obtained feature map is passed to predict and classify.

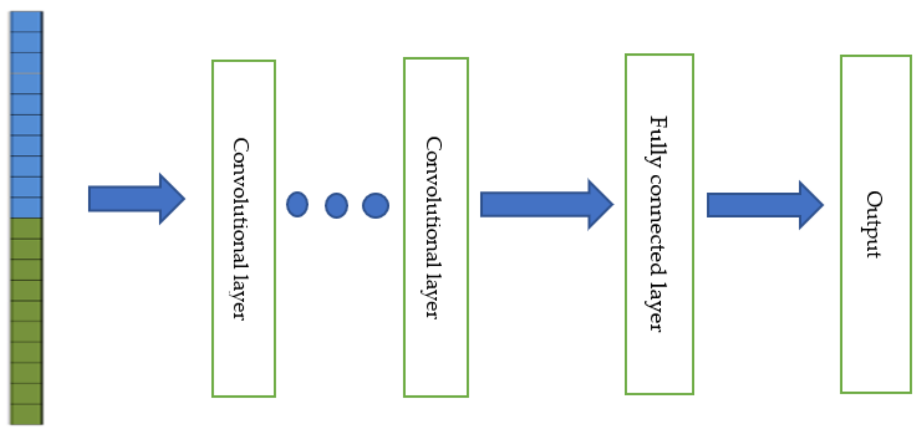


Figure 3.12: Block diagram of training phase after concatenating features

3.3.5 Implementation

To implement our proposed model we took the help of various python image processing libraries. That is given below:

- OpenCV.
- Sci-kit image.

- Tensorflow.
- Keras.
- Tensorboard.

3.4 Conclusion

This chapter gives a brief overview of our proposed method for zebra crossing recognition. Our proposed method was tested using two different CNN architecture. The experimental analysis of the proposed framework is discussed in the following chapter.

Chapter 4

Results and Discussions

4.1 Introduction

In the previous chapter, a detailed explanation of the proposed framework for zebra crossing detection and recognition was given. This chapter examines the performance of the proposed framework. All the experiments were accomplished in the Intel(HQ) core(TM) i7 CPU@2.55 GHZ processor with 16 GB ram with Nvidia GTX 960M graphics card. In this paper, the experiments were processed in python environment.

4.2 Dataset Description

A total of 1800 images was gathered from various environments and illumination conditions. Out of 1800 images ,1500 images is for training purpose and rest 300 is for validation. Some negative examples were also taken for analyzing the robustness of the proposed system. This paper considered the images from the different view of the zebra crossing only. The height and angular displacement were different in every image to cope with various views of the pedestrian. Some sample images are shown in figure.

Figure 4.1: zebra crossing sample images: a) normal illumination b) noisy images c) d) uneven illumination e) night images

4.3 Impact Analysis

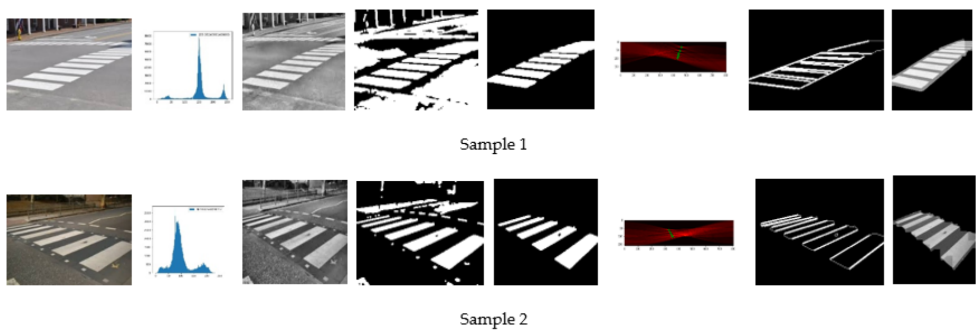
Our proposed method has huge impact. In terms of detection and recognition it has higher accuracy and speed. Also its robustness has made it possible to detect zebra crossing under various complex situations and illuminations.

4.3.1 Social and Environmental Impact

Our proposed method has immense social impact. Our model can be used in automated vehicle. It can also help visually impaired people not only that but also monitoring traffic situations at zebra crossing areas. Our method has no environmental issues.

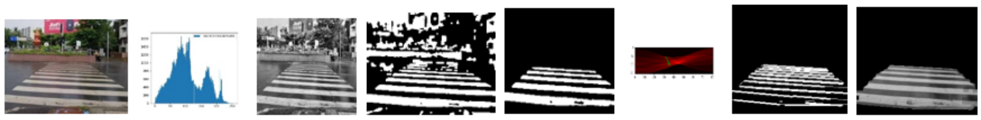
4.4 Evaluation of Framework

The results of different hand-crafted feature extraction samples are presented in figure 4.2. In this figure the input zebra crossing image, gray scale conversion and clip adaptive histogram equalization, binarization, elimination of non candidate object, justifying potential horizontal edges using radon transform, extracted ROI are given.

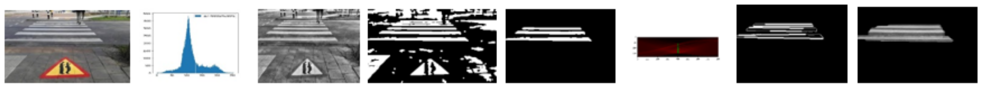


Sample 3

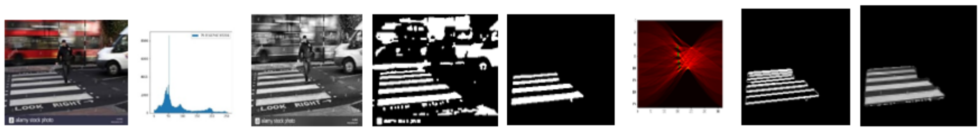
Sample 4



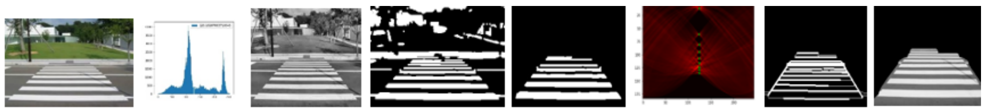
Sample 5



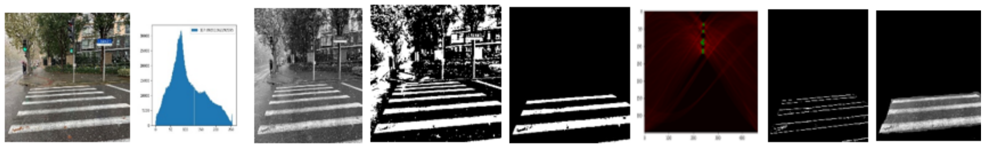
Sample 6



Sample 7



Sample 8



Sample 9

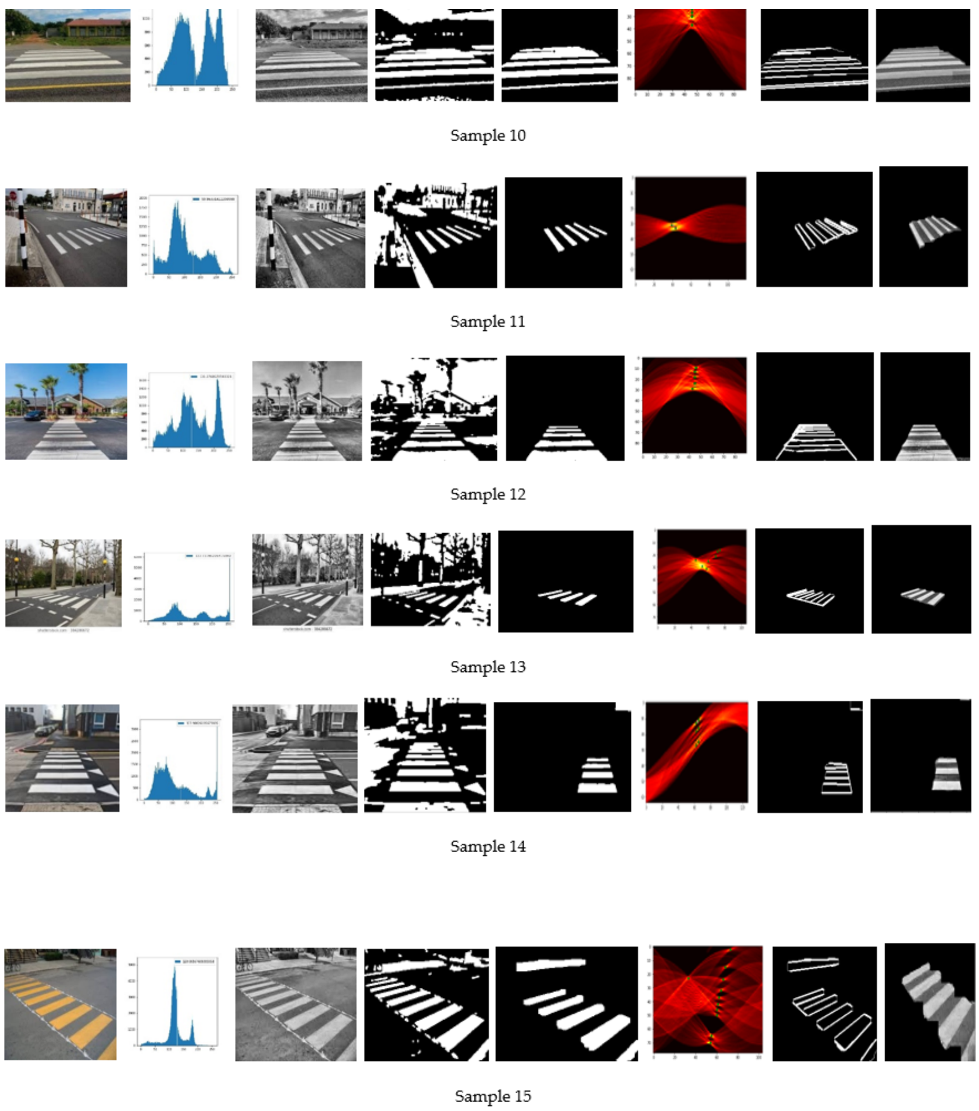


Figure 4.2: processing example of zebra crossing region detection: a) input image b) histogram analysis c) histogram equalization d) binarization and opening e) candidate region f) peak detection of line g) justifying potential parallel edges h) extracted ROI

In figure 4.2 sample images are taken from the different environment and illumination conditions. In it sample 1,2,3 are taken in normal illuminations. But in sample 2 average intensity value is less than threshold value. So, contrast limited adaptive histogram equalization is performed to enhance the sharpness and contrast. The outcome of the histogram equalization is shown is figure 4.2(c) iii. Sample 4,5 is captured during rainy day. Sample 6,7 contains non candidate object like traffic signs, pedestrian, bus etc hence removed using morphological operation. Sample 9 was taken while it was raining. We had been able to extract ROI successfully except a small portion was cut in the vertical end of the

ROI. Sample 10,11,12 is of normal illumination for this reason ROI was extracted without any complications. But in case of sample 14,15 we have observed that still there was unwanted object in spite of eliminating non candidate object. However, we have removed those objects while justifying potential parallel edges. After extracting ROI from these sample features are extracted by uniform binary pattern and combined with deep features extracted by improved mask RCNN for classification purpose.

4.4.1 Results of proposed approach

After training of our proposed model, we have tested many images under different situations. Our proposed approach successfully detected zebra crossing in all the images along with boundary box and instance segmentation.

Figure 4.3

4.5 Evaluation of Performance

After training our proposed approach, we have evaluated our model with around 312 images of different conditions. For the purpose of evaluation, we have trained our model also without combining handcrafted features and deep features. Then we have compared our proposed model with this model. In the following subsection we have discussed experimental results of our proposed method with pretrained weights of COCO dataset and ImageNet dataset.

4.5.1 Training without combining handcrafted features

In this step, we have trained our model with improved Mask RCNN without combining handcrafted features with pre-trained weights of ImageNet dataset and COCO dataset. We have trained this model with 30 epoch and learning rate was 0.001. The training and validation dataset consisted of 1500 and 300 images respectively. We have annotated training images with VIA image annotator.

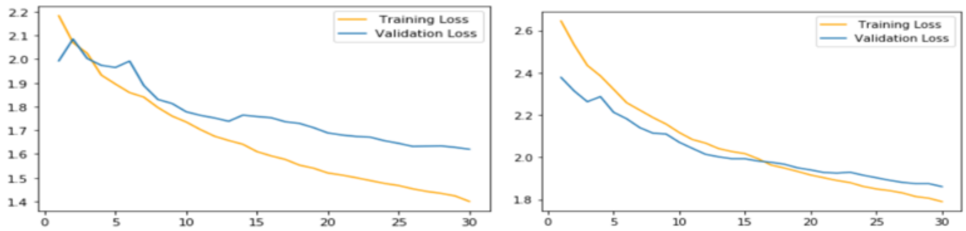


Figure 4.4: a) Loss with pre-trained weights of ImageNet dataset b) Loss with pre-trained weights of COCO dataset

4.5.2 Training after combining handcrafted features

After combining deep features with handcrafted features discussed in section 3.3.3, we have trained our combined feature vector with pre-trained weights of ImageNet dataset and COCO dataset. For training we have set the learning rate .001 and epoch was 30. After training we have observed that proposed approach yields better performance compared to other approach.

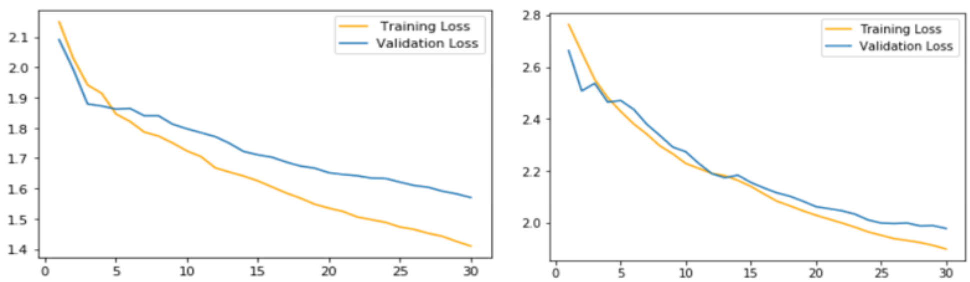


Figure 4.5: a)Loss with pretrained weights of ImageNet dataset b) Validation Loss with pre-trained weights of COCO dataset

From the above optimization learning curve based on loss metric we can observe that combination of hand-crafted feature with deep feature gives better model. From the above figure we can say that optimization curve of figure 4.5(b) Which is combined feature vector loss with pretrained weights of COCO data set is good fitted and it has minimum gap between training loss and validation loss.

We have evaluated our proposed method under different environmental situations and illuminations. The performance of our proposed framework was evaluated with respect to various evaluation metrics. That is discussed below:

- **TP:** A true positive is an output where the predictive model accurately predicts the positive class.
- **TN:** A true negative is an output where the predictive model correctly predicts the negative class.
- **FP:** A false positive is an output where the predictive model incorrectly predicts the positive class.
- **FN:** a false negative is an output where the predictive model incorrectly predicts the negative class.
- **Accuracy:** Accuracy refers how many of each class did the model accurately predict. This considers both TP (True Positive) and TN (True Negative). Accuracy can be defined by the following formula in 4.1

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

- **Precision:** Precision refers how many of the positive groups expected by the model are actually positive. Precision can be defined by the following formula in 4.2

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

- **Recall:** Recall refers how much of each positive class did the model correctly predict. Recall can be defined by the following formula in 4.3

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

- **F1 score:** It is cumbersome to make comparison between two models that have low precision but high recall, or the other way around. But we use F1 score to compare them. The F1 score is a method for measuring both recall and precision at the same time. F1 score can be defined by the following formula in 4.4

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4.4}$$

4.5.3 Performance of our proposed framework

After developing our proposed framework we have tested this framework with 312 images. This images is taken at different illumination and environmental situations. Table 4.1 shows the testing accuracy of our proposed framework.

Table 4.1: performance on our test dataset

Conditions	*Precision* **(%)**	*Recall* **(%)**	*Accuracy* **(%)**
Normal Illumination	100	100	100
Uneven Illumination	100	100	100
Noisy Background	97.5	98.8	98

4.5.4 Comparison with other state of the art Framework

We also have compared our proposed method with other existing zebra crossing detection method. In table 4.2 we have shown the comparison of our proposed model with methods discussed in [6,10,13,14]. The comparisons were performed using their respective dataset. The table 4.2 shows the comparison in terms of precision, recall and F1 score, accuracy metrics. Our proposed model shows significantly better performance than the previously proposed method because in our method we have combined hand crafted features with deep features that has not been performed previously in any zebra crossing detection method.

Table 4.2: Comparison with other proposed method

Method	Instance Segmentation	Precision (%)	Recall(%)	F1(%)	Accuracy(%)
[6]	No	95.10	97.00	96.04	96.47
[10]	No	98.00	96.00	97.00	97.10
[13]	No	92.82	93.33	93.07	93.21
[14]	No	94.00	96.00	95.00	95.41
Our proposed model	Yes	99.17	99.60	99.13	99.30

4.6 Conclusion

In this chapter experimental result and evaluation of our proposed method is discussed. Experiments on the adopted CNN architecture is described. We have also evaluated our proposed framework with other state-of-the art zebra crossing detection framework. In the next chapter we have concluded our thesis with future recommendations.

Chapter 5

Conclusion

5.1 Conclusion

Zebra crossing recognition system has enormous effect in our society. In our proposed methodology we have developed a zebra crossing recognition framework combining handcrafted and deep features. For this thesis work we have developed our own dataset. We have also added convolutional layer in the improved Mask RCNN model to generate better resolution of mask.In the methodology we have demonstrated every steps that how we have extracted handcrafted features and deep features and all the hyper parameter settings. We have also described how we combined the both feature vectors. Finally we showed output our proposed framework.

5.2 Future Work

In this paper, we have proposed a method to detect and recognize zebra crossing in which we combines the hand-crafted features and deep features of crosswalk ROI. We have evaluated our proposed method's performance based on our dataset. Our proposed method shows better performance in terms of detection, recognition and instance segmentation. This proposed method is very efficient than previous crosswalk detection methods. It can also successfully detect unusual shapes of crosswalk. However, future work may be done to improve accuracy of the instance segmentation of the ROI. We will train our method on bigger dataset to make it more robust. The training dataset also may contain more complex and difficult crosswalk situation so that it can detect and recognize crosswalk in every possible situation. After the combination of feature vector we may use PCA ,LDA for dimensitionality reduction.

References

[1] X.-H. Wu, R. Hu and Y.-Q. Bao, 'Block-based hough transform for recognition of zebra crossing in natural scene images,' *IEEE Access*, vol. 7, pp. 59 895–59 902, 2019 (cit. on p. 5).

[2] D. F. Zhang *et al.*, 'An effective detection algorithm of zebra-crossing,' in *International Conference on Electrical and Information Technologies for Rail Transportation*, Springer, 2017, pp. 809–816 (cit. on p. 5).

[3] S. Wang, H. Pan, C. Zhang and Y. Tian, 'Rgb-d image-based detection of stairs, pedestrian crosswalks and traffic signs,' *Journal of Visual Communication and Image Representation*, vol. 25, no. 2, pp. 263–272, 2014 (cit. on p. 5).

[4] B. Qin, W. Liu, X. Shen, Z. J. Chong, T. Bandyopadhyay, M. Ang, E. Frazzoli and D. Rus, 'A general framework for road marking detection and analysis,' in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, IEEE, 2013, pp. 619–625 (cit. on p. 5).

[5] M. A. Malbog, 'Mask r-cnn for pedestrian crosswalk detection and instance segmentation,' in *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, IEEE, 2019, pp. 1–5 (cit. on p. 5).

[6] M. I. Meem, P. K. Dhar, M. Khaliluzzaman and T. Shimamura, 'Zebra-crossing detection and recognition based on flood fill operation and uniform local binary pattern,' in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, 2019, pp. 1–6 (cit. on p. 5).

[7] Y. Fan, Z. Sun and G. Zhao, 'A coarse-to-fine framework for multiple pedestrian crossing detection,' *Sensors*, vol. 20, no. 15, p. 4144, 2020 (cit. on p. 5).

[8] M. Khaliluzzaman and K. Deb, 'Zebra-crossing detection based on geometric feature and vertical vanishing point,' in *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, IEEE, 2016, pp. 1–6 (cit. on p. 6).

[9] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu and M. Pietikäinen, 'Deep learning for generic object detection: A survey,' *International journal of computer vision*, vol. 128, no. 2, pp. 261–318, 2020 (cit. on p. 6).

[10] S. Yu, H. Lee and J. Kim, 'Lytnet: A convolutional neural network for real-time pedestrian traffic lights and zebra crossing recognition for the visually

impaired,' in *International Conference on Computer Analysis of Images and Patterns*, Springer, 2019, pp. 259–270 (cit. on p. 6).

[11] C.-R. Dow, H.-H. Ngo, L.-H. Lee, P.-Y. Lai, K.-C. Wang and V.-T. Bui, 'A crosswalk pedestrian recognition system by using deep learning and zebra-crossing recognition techniques,' *Software: Practice and Experience*, vol. 50, no. 5, pp. 630–644, 2020 (cit. on p. 6).

[12] V. Tümen and B. Ergen, 'Intersections and crosswalk detection using deep learning and image processing techniques,' *Physica A: Statistical Mechanics and its Applications*, vol. 543, p. 123 510, 2020 (cit. on p. 6).

[13] R. F. Berriel, F. S. Rossi, A. F. de Souza and T. Oliveira-Santos, 'Automatic large-scale data acquisition via crowdsourcing for crosswalk classification: A deep learning approach,' *Computers & Graphics*, vol. 68, pp. 32–42, 2017 (cit. on p. 6).

[14] E. T. Silva Jr, F. Sampaio, L. C. da Silva, D. S. Medeiros and G. P. Correia, 'A method for embedding a computer vision application into a wearable device,' *Microprocessors and Microsystems*, vol. 76, p. 103 086, 2020 (cit. on p. 6).

[15] H. Wang, H. Zhang, M. Gu and C. Wang, 'Comparison of the pedestrian crossing intention parameters and research on intention recognition model under different road conditions,' *IEEE Access*, vol. 8, pp. 196 675–196 689, 2020 (cit. on p. 6).

[16] W.-J. Chang, L.-B. Chen, C.-Y. Sie and C.-H. Yang, 'An artificial intelligence edge computing-based assistive system for visually impaired pedestrian safety at zebra crossings,' *IEEE Transactions on Consumer Electronics*, 2020 (cit. on p. 6).

[17] F. Shaheen, B. Verma and M. Asafuddoula, 'Impact of automatic feature extraction in deep learning architecture,' in *2016 International conference on digital image computing: techniques and applications (DICTA)*, IEEE, 2016, pp. 1–8 (cit. on p. 6).

[18] D. T. Nguyen, T. D. Pham, N. R. Baek and K. R. Park, 'Combining deep and handcrafted image features for presentation attack detection in face recognition systems using visible-light camera sensors,' *Sensors*, vol. 18, no. 3, p. 699, 2018 (cit. on p. 6).

[19] M.-I. Georgescu, R. T. Ionescu and M. Popescu, 'Local learning with deep and handcrafted features for facial expression recognition,' *IEEE Access*, vol. 7, pp. 64 827–64 836, 2019 (cit. on p. 6).

[20] X. Wu, S. Wen and Y.-a. Xie, 'Improvement of mask-rcnn object segmentation algorithm,' in *International Conference on Intelligent Robotics and Applications*, Springer, 2019, pp. 582–591 (cit. on p. 7).

[21] Y. He and Y. Sun, ‘An automatic image segmentation algorithm based on grabcut,’ in *6th International Conference on Wireless, Mobile and Multi-Media (ICWMMN 2015)*, IET, 2015, pp. 185–190 (cit. on p. 7).