“A Bangladeshi Road Sign Detection & Recognition System Based on Template matching & Convolutional Neural Network”

<u>Submitted By</u>

Md. Ziaul Karim

ID: 171-35-1883

<u>Supervised By</u>

Md. Shohel Arman

Lecturer (Senior Scale)

Department of Software Engineering

Daffodil International University

This Project report has been submitted in fulfillment of the requirements for the Degree of

Bachelor of Science in Software Engineering.

APPROVAL

This thesis/project/internship titled on "**A Bangladeshi Road Sign Detection & Recognition System Based on Template matching & Convolutional Neural Network**", submitted by, **Md. Ziaul Karim, ID: 171-35-1883** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

-- Chairman

Dr. Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Daffodil International University

-- Internal Examiner 1

Md. Shohel Arman
Senior Lecturer
Department of Software Engineering
Daffodil International University

-- Internal Examiner 2

Farhan Anan Himu
Lecturer
Department of Software Engineering
Daffodil International University

-- External Examiner

Professor Mohammad Abu Yousuf, PhD
Institute of Information Technology
Jahangirnagar University

DECLARATION

I hereby declare that I have done this thesis: **"A Bangladeshi Road Sign Detection & Recognition System Based on Template matching & Convolutional Neural Network"** under the supervision of **Md. Shohel Arman,** Senior Lecturer, Department of Software Engineering, Daffodil International University. I also declare that this thesis or any part of this has been done by myself and has not been submitted anywhere else for the award of any degree.

Md. Ziaul Karim
ID: 171-35-1883
Batch: 22nd
Department of Software Engineering
Faculty of Science & Information Technology
Daffodil International University.

Certified by

Md. Shohel Arman
Senior Lecturer

Department of Software Engineering
Faculty of Science & Information Technology
Daffodil International University.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to Almighty Allah who has provided me with knowledge & ability to carry out this thesis and a journey of about four years of studying in this wonderful vicinity.

Also I would like to thank my honorable supervisor **Md. Shohel Arman,** Senior Lecturer, for his guidance, support, encouragement and co-operation. Without whom the thesis would not have been complete.

Md. Ziaul Karim

ABSTRACT

Road Signs are very important but a neglected topic in the discipline of driving vehicles in Bangladesh. As a highly populous country chaos on the road has taken a disastrous shape over the past decade as the number of vehicles on the road had went up and road safety became a crying need as the lives lost in road accidents are constantly on the rise. In case of Road Signs in Bangladesh the descriptive text for it is given more emphasis than the sign itself. Whereas, in the developed countries the signs are given more importance, because Signs are supposed to be universally accepted in case of communication. Hence, the Road Signs in Bangladesh are put quite arbitrarily, especially in the metropolitan areas the signs are credited by Metropolitan Police, or sponsored by companies, they look more like posters or banners than a legitimate Road Sign and it becomes very difficult to locate them.

As we are witnessing the development of Autonomous Driving Systems all over the world by companies like Tesla, it is becoming more and more popular in demand. The detection & recognition of Road Signs play a crucial role in it. In Bangladesh Road Signs are ignored most of the time & text based road signs are given the priority. While human brain is capable enough to detect text based Road Signs, a smart system can struggle. Nonetheless, two types of measures could be taken into account in this case, one is through text recognition based and the other is detecting the sign in the poster-like signs and then run it through a neural network based classification to identify it. The later solution was taken into account for this paper, as all the Road Signs (both poster-like & basics) contain the basic signs for directions, they may or may not contain descriptive text along with them, but the road signs are always present in larger or smaller scales. The solution was built around Deep Learning and Image Processing techniques for object detection, segmentation and classification. Supervised Learning and Transfer Learning were used in terms of technical development. Text based recognition systems are not ideal for metropolitan areas because there are so many posters and banners attached to the poles and around the Road Signs.

Key words : Road Sign, Object Detection, Region of Interest, Artificial Intelligence, CNN – Convolutional Neural Network, Deep Learning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABREVIATIONS

(i) RS – Road Signs
(ii) RoI – Region of Interest
(iii) CNN – Convolutional Neural Network
(iv) YOLO – You only look once.

CHAPTER 1

INTRODUCTION

1.1. Thesis Overview

This study is composed of two parts, road signs detection and road signs recognition. A Road Sign detection and recognition solution is offered in this study. For detecting Road signs a model based on computer vision techniques like image segmentation and shape features matching has been used. For classification & recognition purpose a LeNet-5 architecture based Convolutional Neural Network has been used. The solution was created combining two parts together serving separate purposes of both detection & recognition. But we discuss unique instances regarding Road Signs used inside the metropolitan areas of Bangladesh and try to conduct tests with the developed solutions.

1.2. Problem Definition

Road Signs in Bangladesh are quite difficult to detect and classify since our roads have plenty of posters and banners on the road and also there are road signs that are placed in such ways that it becomes quite difficult to detect them with even human viewpoint. The objective here is to detect and classify road signs against those odds and contribute to the future challenges about to be faced in the matter, regarding smart solutions. The problem that is covered in this paper is to detect road signs from 'poster like' road signs placements in metropolitan areas usually they are sponsored by certain companies with the courtesy of metropolitan police and they are unlike any other RS that are used all over the world except for the fact that they do contain the RS in the corner but in a very small scale, which can be difficult to spot by a camera. An example would be just as below –

Fig. 1.1. Examples of Road Signs that follow out of the norm designs than traditional ones.

1.3. Motivation

This study was carried out keeping 'Safer Road, Safer Future' in mind. Bangladesh needs safer Traffic Control Systems and strict laws to make our roads safer. Recent death tolls of students on the roads of Dhaka Metropolitan City in 2018 leading to "We want safe road" movement have been a major driving force for this study. Smart solutions can play vital a role in terms of saving lives on the road everyday. This study will work on contributing to unique instances of RS in use and also enable future researchers to develop a system that can detect them from live video feed since this study was carried out on input images only.

1.4. Thesis Contribution

This thesis has been carried out keeping a research gap in mind which adds a new dimension to the field of Detection & Recognition systems of RS. This article takes knowledge gained from several authors into account and enables others to do likewise. More over the data collected for the purpose of the thesis is self-collected and preprocessed, which can also be of great use to future researchers.

1.5. Scope

The scope of this study is going to be helpful for:

- Detection & recognition of RS with unique features

- Contribution to the dataset of the Bangladeshi RS.

- Add a unique perspective to the whole concept of RS based smart solutions in Bangladesh.

CHAPTER 2

LITERATURE REVIEW

2.1. Introduction

A literature review is a review article which includes current knowledge, including firm based findings and as well as theoretical and methodological contributions to a particular topic. The review consists of summarization, critical evaluations of previously published work. It can refer to a full scholarly article, book or journal or specific portions of it. The main purpose is to provide a researcher or a general reader to provide with knowledge on the matter.

2.2. A Comparative study on the detection and recognition of RS in Computer Vision with Machine Learning & Deep Learning.

If we look at [12] SURF, Bag of Visual words and K-means methods were used as backbone for real time detection of traffic signs, bag of visual words method, which uses key-point features of an image matching with a reference image to detect and recognize an object, K-means algorithm plays a vital role in the recognition by creating a histogram of the nearest neighbors. In [5] & [8] we can see Adaboost, SURF, SIFT were used as inspiration for key-feature matching to detect RS and extract RoI, however in [5] Artificial Neural Network was used which was based on MLP for shape classification, an additional SLP network also developed for content recognition in Triangular and Circular Road sign design. While [8] uses a two step algorithm that prioritizes shape features, but also eliminates the possibility of false positives by using ratio of the object area, however the accuracy is not too high (Lasota et. al. 2016) . In [3] Matching pursuit has been used, in this paper a pixel by pixel matching of template image and complemented image is performed which is similar to the detection method used in this paper, but the template matching technique is used for recognition of the RS as well, which is good for smaller scale dataset, but not for large scales, the accuracy here is not significantly high, SVM & HOG descriptor based models has been known to have shown higher accuracy rate (up to 95%). [2] Uses deep learning & also a completely innovative concept which is semantic web ontologies, it uses interpreting message sent by an RS to a deep learning model to recognize a traffic sign. This paper implemented RS ontologies, SPARQL query requests are used to detect a traffic a sign based on ontologies created and feed forward this data to a trained Neural Network to classify the sign. [4] and [3] both are based on similar datasets, but here [4] the use of DtBs vector can be seen along with neural networks. [4] uses image segmentation based on color and then uses aspect ratio and area to filter properties, then by calculating Distance to Borders (DtBs) to recognize the shape of an RS, candidate region is extracted and fed to an ANN network for recognition of a particular RS. [6] uses color based segmentation just [3] to better extract RS, then converts it to binary image and uses Fast Radial Symmetry algorithm to recognize the shape of candidate regions and Harries corner detector the other two basic shapes triangle & rectangle is recognized. Overall the detector has shown promising results (Horak et. al. 2016). [11] uses Hough Transformation and CNN to classify RS in pictures, in preprocessing the HSV color space has been used like in [3] & [6] , however this experiment was only conducted on circular RS. However the accuracy rate is much higher (98.2%).

2.3. Conclusion

HSV color is very close to the way human perceive an image [3], for segmentation of red an blue color used the majority of traffic signs it is worth considerable. From a large image [7] feature matching is the simplest way to detect RS and extract RoI. Template library is created for the purpose. [11] CNN is a better choice for recognition purposes since there are large datasets like German Traffic Sign Recognition Benchmark available, pretrained deep neural network architectures such as VGG-Net, Mobile-Net, Le-Net are available to train a robust model for the purpose. The trained model that has a higher accuracy rate can be used in the development of real-time detection and recognition for the future work as well.

CHAPTER 3

METHODOLOGY

We have figured out by now assessing previous works done on the matter from Chapter - 2, is that RS comes with some distinct basic features such as shape & colors. Aside from that we have area of content which dictates the classification & recognition. In terms of detection from a large scale field of view, the distinct features are widely used to extract the Region of Interest. The challenge here is to make the detection method as less time costly as possible but also given the limitations in existing unique instances of which we are solely going to focus on in our later experiments. The classification and recognition part is powered by the training of CNN models, the more it gets to train the more accuracy we can get out of it.

3.1. Collection of Dataset

Keeping the unique RS detection and recognition in mind the collected the data is based on RS found in metropolitan areas of Chittagong & Dhaka. The source is very limited since the instances we are looking for are unique and very small in number. About 102 images (both unique & commonly used instances) of RS has been collected with a 16 megapixel camera & from google image search. For additional data we are using GTSRB database images that match with the ones in Bangladesh. We are also using the dataset layout from it. Some images were also created as synthetic data with illustrators for testing purposes. Examples of images in the Road Signs dataset are shown in Fig. 2.1.

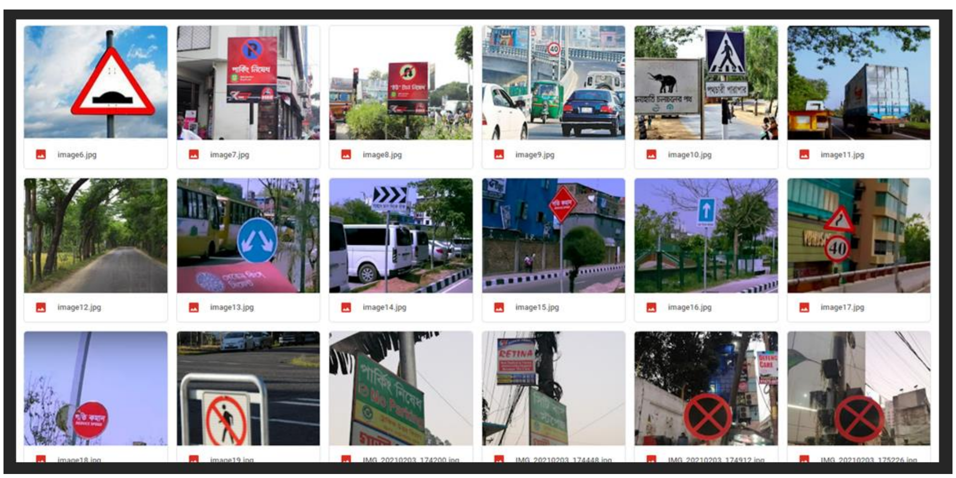

Fig 3.1. Examples of collected image data

3.2. Data Preprocessing

Since image data is being used in the experiment, image pre-processing is a very important step in this regard. It is analogous to the mathematical normalization of a data set, which is a common step in various feature descriptor methods. It may have a positive impact on the quality of feature extraction and the results of image analysis. It is the first step after the collection of training dataset for a machine learning or a deep learning model.

3.2.1. For Detection

For detection purposes, template images were created using illustrators with square image dimensions and also taken from real images by cropping into square dimensions as well. Then threshold binary images were produced and stored in a folder. Then the images were annotated using Microsoft Excel Spreadsheet with headings – 'template_folder, shape_class, sign_type' shown as below:

template_folder	shape_class	sign_type
00	0	prohibitory
01	1	cautionary
02	2	informational
03	3	prohibitory
04	4	prohibitory
....	....	....

Table 3.1. Template Data annotations for RS detection.

Here shape_class 0= circular, 1=triangular, 2 = rectangular, 3=octagonal and 4= circular not allowed.

3.2.2. For Recognition

For recognition purposes, An image with certain RS was cropped to square dimension with the RS in it, then converted into 32x32 dimensions, converted to RGB color format and .ppm file type. Similar signs were stored in a folder named with a distinct number. Images were annotated with the folder number as their class_ID and recognition name with RS_name, shown as below:

class_ID	RS_Name
00	Speed Limit 30 km/h
01	No Passing
....	....

Table 3.2. Image Data annotations for recognition.

3.3. Detection method outline

For RS detection purpose edge detection, HSV color space has been used. The detection model is divided into following steps.

1. Blue and Red color segmentation using HSV color space.
2. Eliminating noise with gaussian blur.
3. Edge Detection with Canny operator.
4. Drawing contours based on occupied pixel area to eliminate small unimportant areas.
5. Template matching with Template gallery for RoI bounding box generation.
6. Cropping the bounding box content.

3.4. Recognition method outline

For RS recognition purpose, LeNet-5 has been used, which uses 32x32 pixel RGB image as input. Following steps are followed:

1. The detected cropped image is scaled down to 32x32 dimension.
2. Converted into RGB format.
3. Fed into the trained CNN network.
4. Recognition Results are produced in text output.

3.5 Requirements for Implementation

Because of the technology involved like usage of OpenCV and a high level CNN architecture training, along with Tensorflow framework, the following setup of the following are required:

Hardwares/Software requirements:

1. NVDIA GPU (CUDA 5.0 or above)
2. RAM (At least 8 GB)
3. Any Operating system that runs Tensorflow framework.

CHAPTER 4

DETECTION METHOD & ARCHITECTURE

4.1. Introduction: Object Detection

Object detection is one of the majorly discussed and researched topic in modern day field of Machine learning and Deep learning. Object detection is a key technology behind driver assistance systems, that enable cars to detect driving lanes or perform pedestrian detection to improve road safety. Object detection is also useful in applications such as video surveillance or image retrieval systems.

4.2. The literature in detection of Road Signs

HOG features were used initially to detect and recognize RS. Fast Radial Symmetry and Harris corner detector [7] has also been effective in detection of RS apart from. By using Hough, the global detection problem that is not easy to solve can be transformed into the local peak detection problem that is easy to solve, making the transformed result easy to detect and recognize. Its advantage is that noise and curve discontinuity have relatively small influence (Sun et. al. 2019). In [1] Single Shot Detector(SSD) has been used to detect RS. YOLO v3 has also been found effective in detection and classification of RS. In most of the studies carried out before the popularity of Deep Neural Network based object detectors, image segmentation, shape classification and template matching has been popularly used [3]. Transfer learning based detection is best used for detection when there is availability of a vast amount of data. But when working with a small amount of data, it is best to use Computer Vision techniques [7].

4.3. Outline of Bangladeshi Road Signs

4.3.1. Commonly used Road Signs in Bangladesh

Road Signs used in Bangladesh are fairly common and they are mostly implemented in highways. But one of the major portion of it also plays a very important role in the traffic control systems Metropolitan areas. The vast population of drivers here are mostly unaware of the importance of them and more or less carefree, hence the metropolitan traffic police departments comes forward to provide RS that are better visible and poster like. That raises the questions below:

- Does smart solutions exist to identify poster-like RS within a field of view?
- Can existing solutions detect road signs in these circumstances?

RS that align with the officially approved law can be categorized between red and blue colors. Shape features like Triangular, Circular, Rectangular, Rhomboidal and Octagonal.

Prohibitory RS	Cautionary RS	Informative RS
		

Fig. 4.1. Outline of Bangladeshi Road Signs

Shape features are very important in terms of Detection of RS. Some commonly used shapes in RS are shown in Fig. 4.2.

Fig. 4.2. Commonly used shapes for Road Signs.

During the data collection some unique RS were discovered that are shown in Fig. 4.3.

Fig. 4.3. Unique Road Signs

4.3.2. Road Signs in Metropolitan Areas.

RS in metropolitan areas are quite different. Since in metro areas there are a lot of distractions on the side of the roads, poster like RS are used where the text description is given priority than the RS, as show in Fig. 1.1 in Chapter 1. Basically the structure of the RS in Metropolitan areas are as follows:

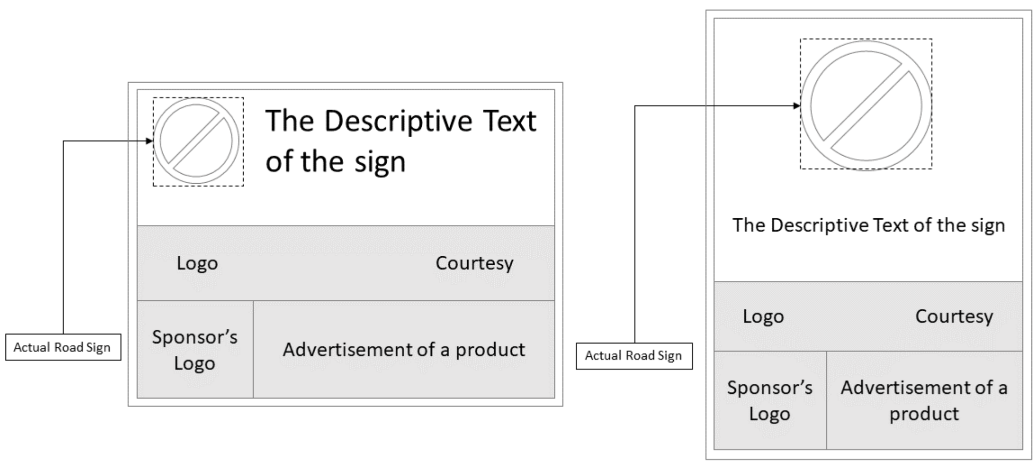

Fig. 4.4. Structures of Road Signs in Metropolitan Areas.

The problem lies here is that each metropolitan area has their own unique background colors of their own. Which makes it harder to detect. Only handful of locations have them and we don't have access to massive amount of data of this kind to support training and testing for a Neural Network based detection system. So, a solution was chosen based on Computer Vision techniques.

4.4. Proposed Detection Method

The proposed detection method is based on color segmentation, edge detection, color segmentation and template matching technique [7]. Overall architecture is shown in a diagram that can be seen in Fig. 4.5.

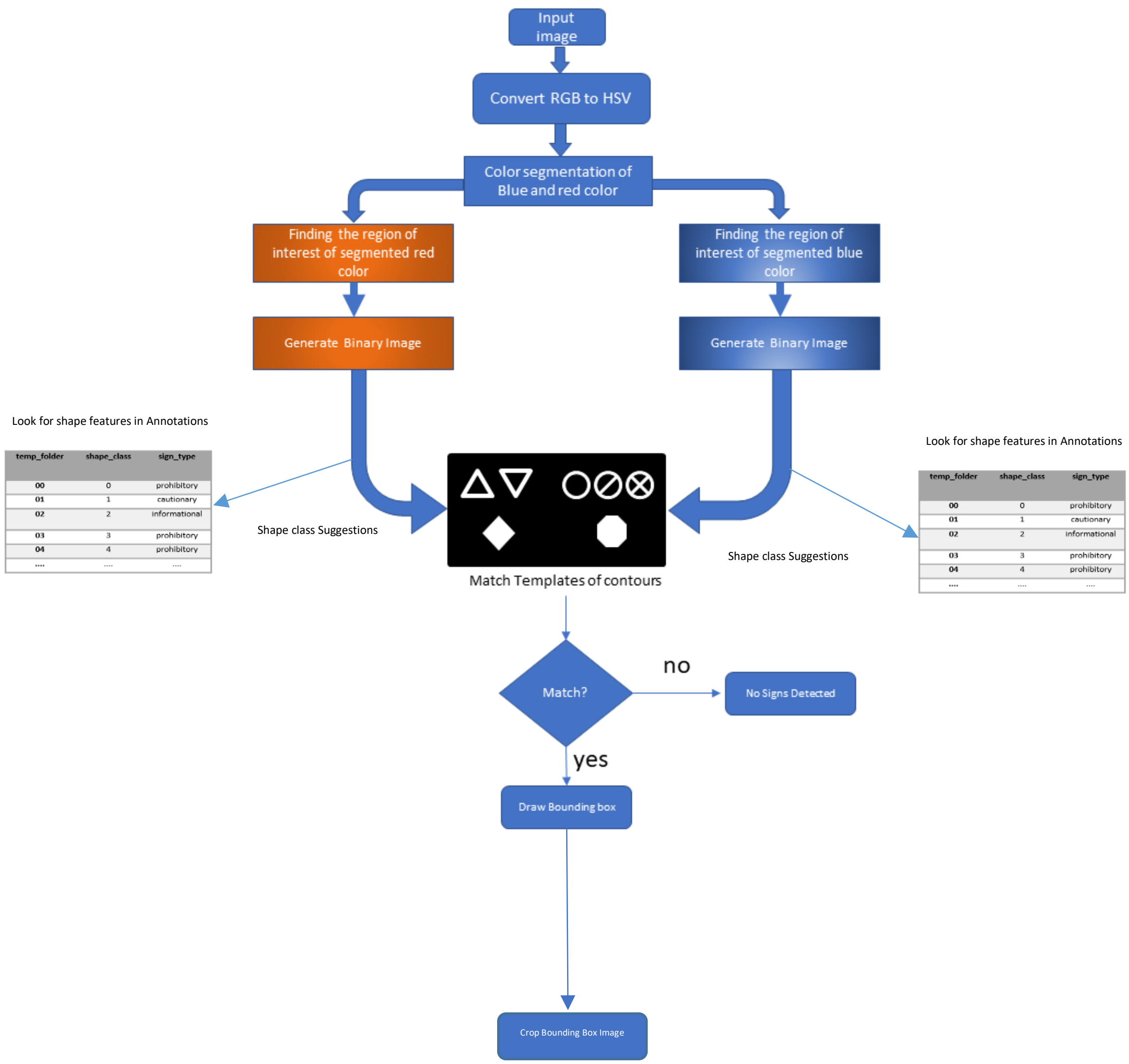

Fig. 4.5. Detection model architecture

4.4.1. Color Segmentation

Usually, RS are placed on the side of the road and often isn't placed in plain sight. There are backgrounds that are tree, posters and buildings etc. But all the RS have distinct colors that make them a bit more visible. Images taken by camera are digitized in RGB color format. But the correlation between the colors **R**ed, **G**reen and **B**lue components makes image segmentation quite challenging for lit up situations. Studies done for this paper have shown that HSV is more suitable for almost any kind of light intensity. HSV color space (or HSB color space) is composed of Hue H, Saturation S and Value V. It is a nonlinear transform of RGB space.

$$R' = R / 255$$
$$G' = G / 255$$
$$B' = B / 255$$

$$Cmax = max(R', G', B')$$
$$Cmin = min(R', G', B)$$
$$\Delta = Cmax - Cmin$$

$$H = \begin{cases} 0°, & \Delta = 0 \\ 60° \times \left(\frac{G'-B'}{\Delta} (mod\ 6)\right), & Cmax = R' \\ 60° \times \left(\frac{B'-R'}{\Delta} + 2\right), & Cmax = G' \\ 60° \times \left(\frac{R'-G'}{\Delta} + 4\right), & Cmax = B' \end{cases}$$

$$S = \begin{cases} 0, & Cmax = 0 \\ \frac{\Delta}{Cmax}, & Cmax \neq 0 \end{cases}$$

$$V = Cmax$$

Fig. 4.6. The formula for conversion of RGB color space into HSV space

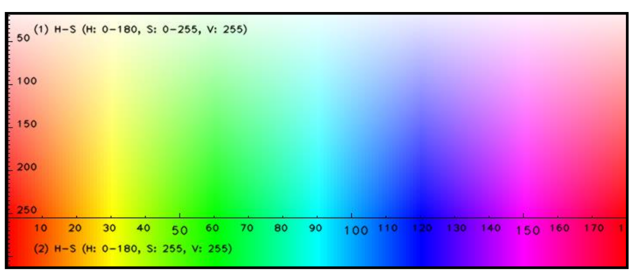

Fig. 4.7. HSV color space

Then segmentation of red and blue color is done by performing following tasks:

(i) Set lower red range to HSV(0, 50, 50) and upper red range to (10, 255, 255) and mask 0 is created. Similarly another red color mask 1 created with lower (170, 50, 50) and upper (180, 255, 255) ranges.

(ii) Set lower blue range to HSV (100, 50,50) and upper blue range (110, 255, 255) and mask 2 is created. Similarly another blue color mask 3 is created with lower (110, 50, 50) and upper (120, 255, 255) ranges.

(iii) All the masks are combined by mask = mask0+ mask1 + mask2 + mask3

(iv) A copy output image is made by setting every pixel value except for ‘mask region’ to 0.

Results of the operations are shown in Fig. 4.5.

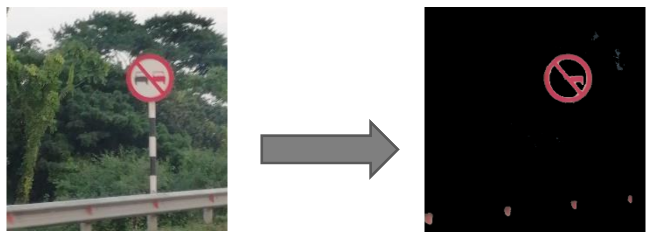

Fig. 4.9. Performing color segmentation with HSV color space.

4.4.2. Eliminating noise with gaussian blur.

The filtered image is applied with gaussian blur to remove left over noises and small red or blue noises picked up in the image.

$$G(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Fig. 4.10. Gaussian function & result

4.4.3. Edge Detection with Canny operator.

Canny Edge detector is used on the gaussian blurred image. But before that grey scale filter is used on the image.

Edge Detection operation is shown in Fig. 4.11.

Fig. 4.11. Canny edge detection & result.

4.4.4. Drawing contours based on occupied pixel area to eliminate small unimportant areas.

After finding the optimum edges by Canny operation done in 4.4.3. contours are found and drawn around the edges using thresh binary methods. Process are shown in Fig. 4.

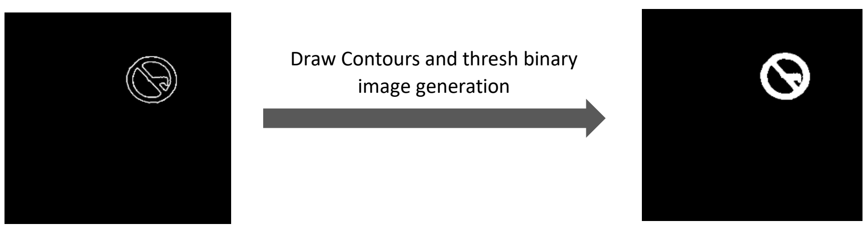

Fig. 4.12. Drawing Contours on occupied Pixel area of Road Sign.

4.4.5. Detecting shapes

After drawing the contours we fill the bounding contour area we detect which shape classes it belongs to based on table 3.1. from Chapter 3. Fig. 4.13. shows filled up contour area of the sign.

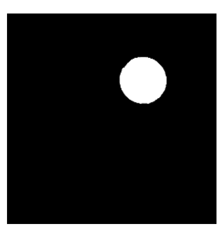

Fig. 4.13. Filling up the area of contour

Based on OpenCV approximation function we detect whether the filled area could belong to which shape classes & give a list suggestion. In this case, class – 0, 3, 4 are suggested.

4.4.6. Template matching with Template gallery for RoI bounding box generation.

4.4.6.1. Template Gallery

Since template matching is used for identification, we tend to identify a specific shape contour that exists within the picture that belongs to the norm of RS. The pictures in the template gallery are all common outlines and shapes used in RS in binary images. That means only shape features are going to matter in this part. Therefore, template gallery uses binary image format. 208 template images of 9 different types have been added to the gallery, all with performed image augmentations using shear, zoom and rotation.

The template folders look as figure 4.14 -

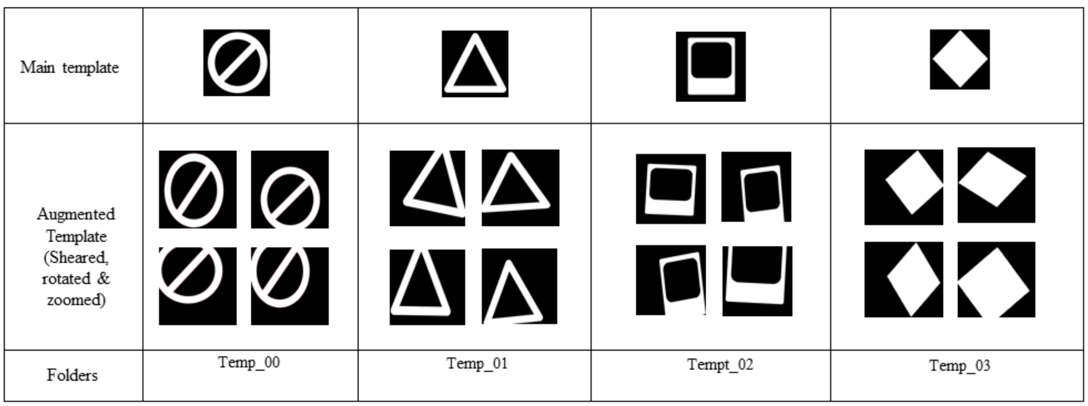

Fig. 4.14. Template Gallery

4.4.6.2. Template Matching

Template matching is a process. Algorithm in [7] is being used in this case. First the template in the gallery is judged for the similarity with each part of the image. Secondly, it is determined whether it has the similar part as the template image. In 4.4.5. detection of shapes cuts down the process of looking through all the folders for templates that could match, based on the suggestions it had made. Finally, the specific location of the RoI is detected in the picture.

First, we can place the detection object template *t(x, y)* on the image. Secondly, we focus on detecting the similarity between the template in the suggested template folders (suggested by the shape detection) and the input image. All the pixels in the image have such operations. Finally we need to determine whether or not there is an object in the image according to the maximum similarity or it exceeds a certain amount of threshold, and find the specific location of the object which here is the RS.

Matching measures include:

$$\max |f - t| \quad (1)$$

$$\iint_s |f - t| dxdy \quad (2)$$

$$\iint_s (f - t)^2 dxdy \quad (3)$$

The smaller the measurement values of equation (1), (2) and (3), the better the matching result; on the contrary, the larger the measurement values below, the better the matching result. (Jia et. al. 2020)

$$m(u, v) = \iint_s t(x, y) f(x + u, y + v) dxdy \quad (4)$$

$$m(u, v) = \frac{m(u,v)}{\sqrt{\iint_S f(x+u,y+v)^2 dxdy}} \quad (5)$$

4.4.7. Drawing Bounding Box and Cropping the Region of Interest

After finding the maximum matching pixels the object is then localized with bounding box. Then using the bounding box co-ordinates (height, width, x, y) the content within is cropped to extract the RS. The operations performed are shown in Fig. 4.15.

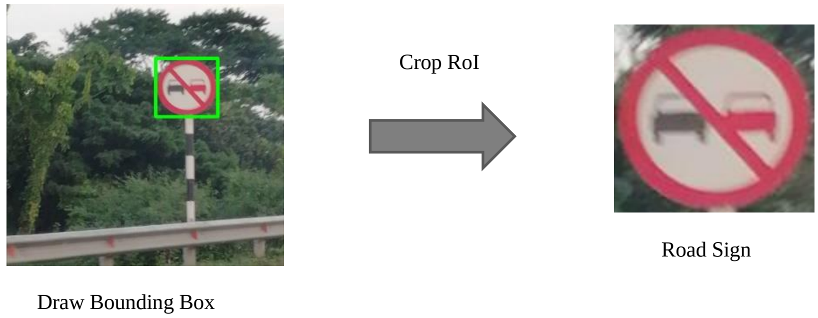

Fig. 4.15. Cropping into RoI.

4.4.7.1. Testing on other instances

Experiments on poster-like RS has been conducted under the same methodology discussed in previous sections. They are shown in Fig. 4.16.

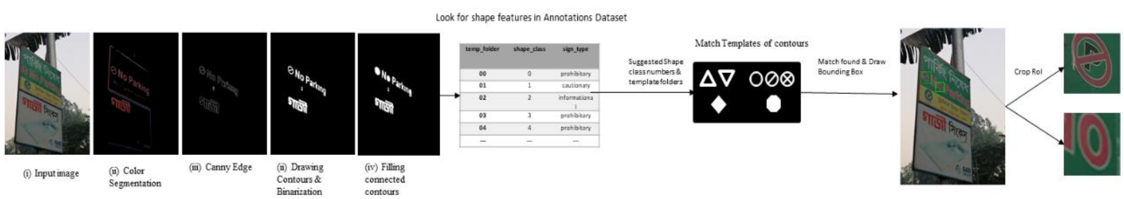

Fig. 4.16. Example of detecting RS in Poster-Like Road Sign

4.5. Results & Discussion

RS detection is a very complex problem and there are a lot of dependencies involved in the matter. A few things are notable as follows:

(i) It becomes difficult to detect RS from Poster-Like circumstances if the background is matching with the RS border color. Then the color segmentation becomes too difficult to isolate the RS. As shown in Fig. 4.17.

Fig. 4.17. Road Sign background matching the sign color.

(ii) If the text description is written in either red or blue color that matches the RS color, then it gives false positive detection of RS. As seen in Fig. 4.16. the English letter ‘o’ in “No Parking” text is also detected as an RS.

(iii) Detection time is relatively fast for almost all the RS shapes. The highest time recorded being 40 milliseconds (for poster-like RS) and lowest being 4.9 milliseconds(for circle shaped RS) . Fig. 4.17. shows time records for detection purposes.

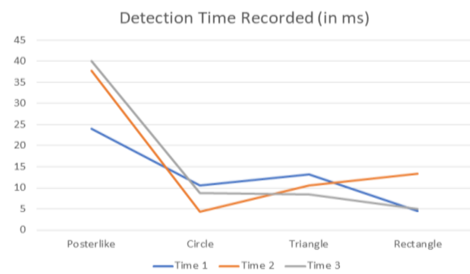

Fig. 4.18. Detection time records

4.6. Conclusion

The detection method used here is not quite robust and can give false positive at times. As we are not using any trained model to detect RS it is safe to say this model could be improved with more accuracy if more data were accumulated. But this has been developed keeping unique instances in mind and the fact that a trained Machine Learning model or Neural Network based model would be able to detect RS more accurately. Then again, it all comes down to the access of a lot of data and the instances here are unique and rarely seen. Some false-positive detection are shown in Fig. 4.18.

Fig. 4.19. Some wrong detections.

CHAPTER 5

RECOGNITION METHOD & ARCHITECTURE

5.1. Introduction: Object Recognition

Object recognition is a technology enabled by computer that is related to computer vision and digital image processing. It deals with both detecting instances of an object to classify and recognize it based on Machine Learning & Deep Learning based algorithms, backed by thousands of training image data. Supervised Learning, Unsupervised Learning and Transfer Learning is widely used in Object Recognition.

5.2. Literature of Road Signs Recognition

Use of visual words [12] has been seen effective in unsupervised learning of RS recognition. Support Vector Machine Based classifications [3] have been used in the past and also several feature matching algorithms [8] like SIFT, SURF, Adaboost Classifier & Template matching has been used in earlier days of RS recognition technologies. In modern days the widespread use of Deep Learning and advancement of Artificial Intelligence, Neural Network based architectures are more suitable of such tasks. ANN architectures have been used paired with Image Processing techniques like DtBs vectors[4] and shape & content classifications [5]. Modern day ease of access in data and impressive architectures for transfer learning for CNN [11] like AlexNet, MS COCO dataset & YOLO[14] architectures enables us to easily train our data into deep networks and get desired accuracy. In this study we have used LeNet-5 architecture for our transfer learning & training purposes and used Convolutional Neural Network (CNN) as the main training model.

5.3. Overview of the used Architecture

Convolutional Neural Networks is the standard form of neural network architecture for solving tasks associated with images. It is solely used for tasks such as object detection recognition purposes. It is a form of feed-forward neural network which has artificial neurons that answers a locality of the encircling cells within the coverage vary. It performs very well in large-scale image processing.

There are several different architectures for CNN. The one that is used in this paper is LeNet-5. Here's a discussion about the methodology architecture –

5.3.1. LeNet-5 Features

The features of LeNet-5 are –

(i) Every convolutional layer consists of three different parts – convolution, pooling and nonlinear activation functions.
(ii) Spatial features are extracted by convolution.
(iii) In average pooling layer subsampling are done.
(iv) tanh activation function is used after (iii)
(v) Multilayer Perceptron (MLP) is used as the last classifier.
(vi) Sparse connection between layers are used to reduce the computation complexity.

5.3.2. Architecture of LeNet-5

LeNet-5 network is composed of 5 layers with learnable parameters provided. It has three sets of convolution layers which consists of three different parts as discussed in 5.3.1. After the convolution and average pooling layers are through, we have two fully connected layers. At the very last, a Softmax classifier which classifies the images into respective class.

Input

$32 \times 32 \times 1$

Fig. 5.1 Input Image

The input to this model is a 32×32 grayscale image.

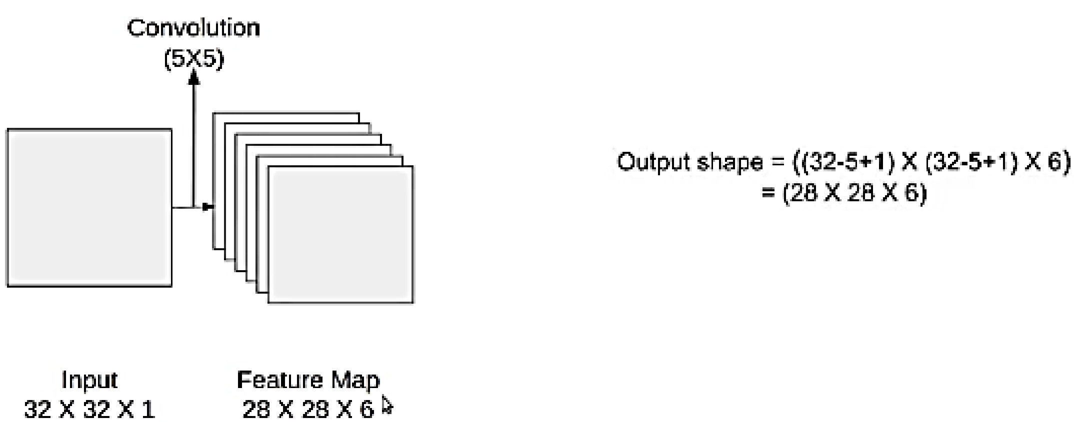

Fig. 5.2 Feature Mapping

The first convolution operation with the filter size 5×5 and we have 6 such filters. As a result, we get a feature map of size $28 \times 28 \times 6$. Here the number of channels is equal to the number of filters that have been applied.

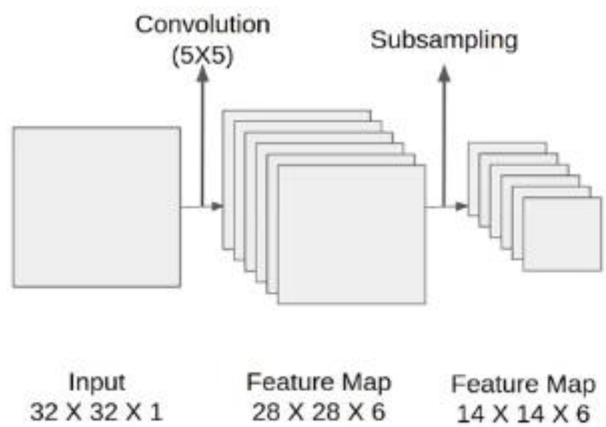

Fig. 5.3 Subsampling

After the first pooling operation, we apply the average pooling and the size of the feature map is reduced by half. Note that, the number of channels is intact.

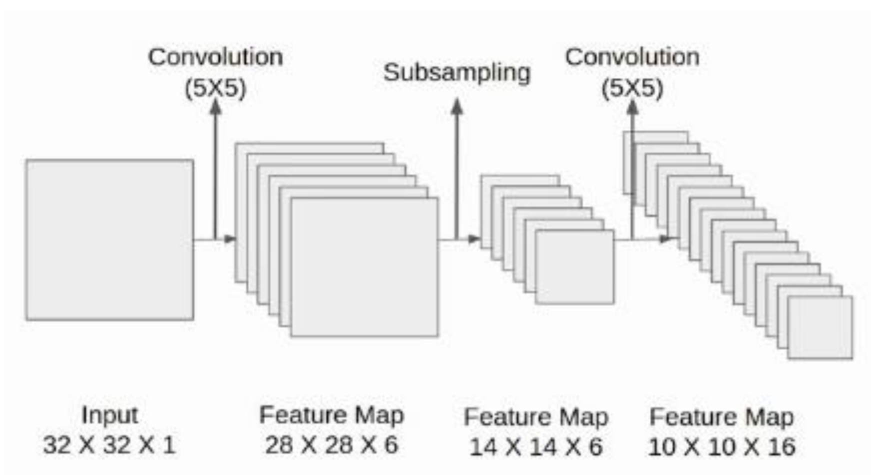

Fig. 5.4. Convolution

Next, we have a convolution layer with sixteen filters of size 5×5. Again the feature map changed it is 10×10×16. The output size is calculated in a similar manner. After this, we again applied an average pooling or subsampling layer, which again reduce the size of the feature map by half i.e. 5×5×16.

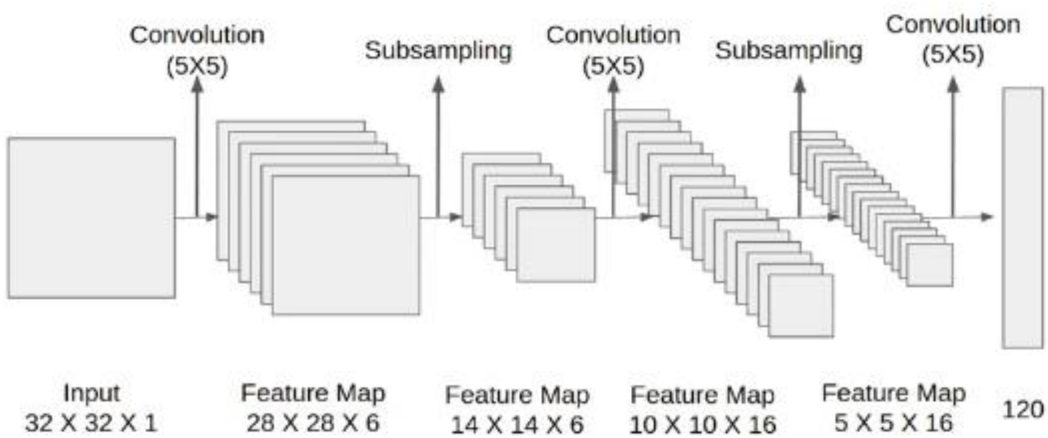

Fig. 5.5. Final convolutional layer.

Then we have a final convolution layer of size 5×5 with 120 filters. As shown in the above image. Leaving the feature map size 1×1×120. After which flatten result is 120 values.

After these convolution layers, we have a fully connected layer with eighty-four neurons. At last, we have an output layer with ten neurons since the data have ten classes.

Here is the final architecture of the Lenet-5 model.

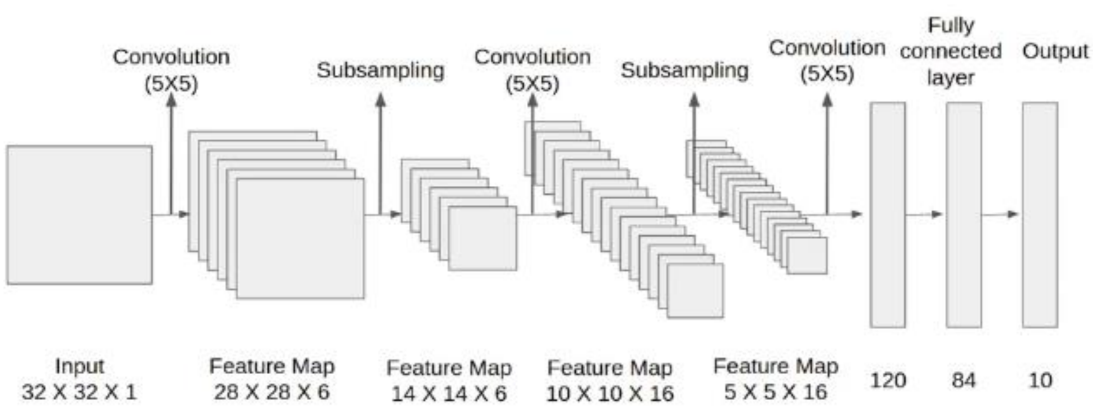

Fig. 5.6. Fully connected layer

Here are the operation performed in LeNet-5 in a nutshell shown in Fig. 5.7.

Layer	# filters / neurons	Filter size	Stride	Size of feature map	Activation function
Input	-	-	-	32 X 32 X 1	
Conv 1	6	5 * 5	1	28 X 28 X 6	tanh
Avg. pooling 1		2 * 2	2	14 X 14 X 6	
Conv 2	16	5 * 5	1	10 X 10 X 16	tanh
Avg. pooling 2		2 * 2	2	5 X 5 X 16	
Conv 3	120	5 * 5	1	120	tanh
Fully Connected 1	-	-	-	84	tanh
Fully Connected 2	-	-	-	10	Softmax

Fig. 5.7. LeNet-5 Architecture

5.4. Training

As we have mentioned in Chapter 3 section 3.1. we are using the format and data from German Traffic Sign Recognition dataset which has over 40 classes and more than 50000 images in total. We are using about 44 class with extra 4 being added classes for signs unique to Bangladesh. Here is the distribution of images plotted in a histogram in Fig. 5.8.

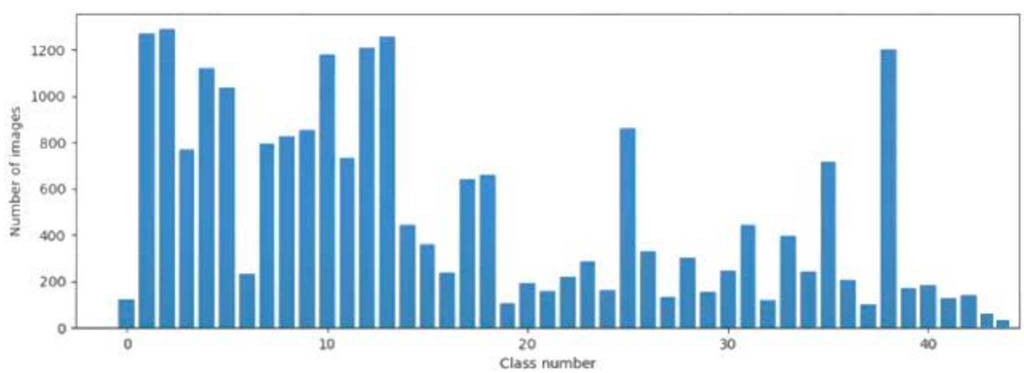

Fig. 5.8. Distribution of training Dataset.

The training was run up to 30 epochs. The training to test data ratio was 0.2, validation ratio was 0.2 as well.

5.5. Testing and Results

5.5.1. Accuracy

We have trained our model to it's limits. The accuracy went up to 98.2%. A model evaluation diagram is plotted in Fig. 5.9.

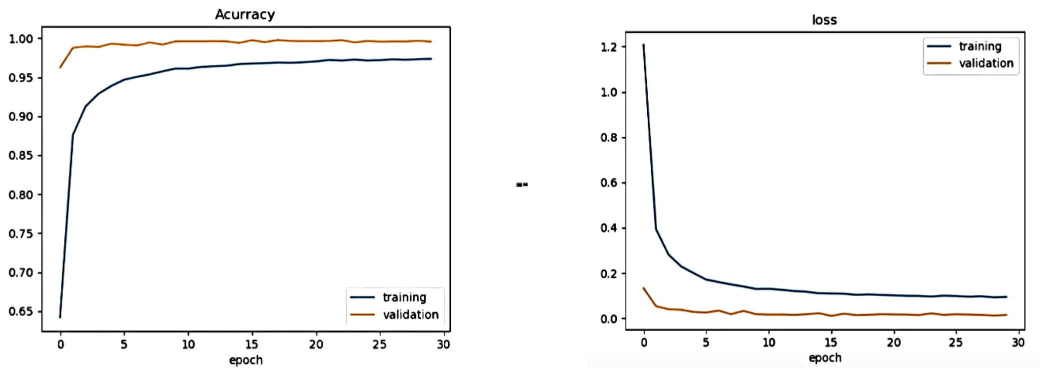

Fig 5.9. Accuracy and loss evaluation diagram

5.5.2. Test Results

After the detection and extraction of RS in Chapter 4, we pick up right after section 4.4.7. in this section.

Some tests were run on the collected data and here are the results shown in Fig. 5.10.

Fig 5.10. Road signs recognition in images

5.6. Conclusion

The accuracy rate is satisfactory in this experiment. Despite the recognition confidence being low for RS that have low amount of data. In this section, a traffic sign recognition method on account of LeNet-5 architecture is proposed, Test result displays that the accuracy of this method is 98.2%. A more well balanced collection of data can solve the lower confidence rate in classes that contain small number of images.

CHAPTER 6

CONCLUSION, LIMITATIONS & FUTURE SCOPE

6.1. Conclusion

In this study two parts of research were carried out. One being RS detection and the other being RS recognition. The first part was solely image processing, based on template matching, color segmentation, shape detection and template matching techniques. The second part was implemented using LeNet-5 architecture and transfer learning process, the accuracy were found up to 98.2% which is quite satisfactory.

6.2. Limitations

(i) Due to lack of Data collection the detection method still isn't that robust.
(ii) Recognition Model needs a more balanced collection of data for certain instances.

6.3. Future Scope

In the future the unique instances of traffic signs could be detected with a more robust algorithm and collections of those instances could escalade the detection process more easily because it will be eligible for using advanced detection algorithms like YOLO v5 or SSD.

References

[1] Ahsan, Sk. Md. M., Das, S., Kumar, S., & La Tasriba, Z. (2019). A Detailed Study on Bangladeshi Road Sign Detection and Recognition. *IEEE Xplore*, 1–6. doi: 10.1109/EICT48899.2019.9068760

[2] Amrani, N. E. A., Abra, O. E. K., Youssfi, M., & Bouattane, O. (2019, October 1). *A new interpretation technique of traffic signs, based on Deep Learning and Semantic Web*. IEEE Xplore. doi:10.1109/ICDS47004.2019.8942319

[3] Asad, M. B., Sharna, F. J., Alam, N., & Islam, M. J. (2016). HSV and Template Matching based Bengali road sign recognition technique. *IEEE Xplore*, 1–4.doi: 10.1109/ICISET.2016.7856527

[4] Chakraborty, S., Uddin, M. N., & Deb, K. (2017). Bangladeshi road sign recognition based on DtBs vector and artificial neural network. *IEEE Xplore*, 599–603. doi: 10.1109/ECACE.2017.7912975

[5] Hamdi, S., Faiedh, H., Souani, C., & Besbes, K. (2017, January 1). *Road signs classification by ANN for real-time implementation*. IEEE Xplore; IEEE. doi: 10.1109/CADIAG.2017.8075679

[6] Horak, Karel & Cip, Pavel & Davídek, Daniel. (2016). Automatic Traffic Sign Detection and Recognition Using Colour Segmentation and Shape Identification. *doi:* 10.1051/matecconf/20166817002 (n.d.). MATEC Web of Conferences 6 , ICIEA 2016 8 17002 (2016).

[7] Jia, L., Shi, X., & Wei, L. (2020, October 1). Design of Traffic Sign Detection and Recognition Algorithm Based on Template Matching. doi: 10.1109/ICCASIT50869.2020.9368750

[8] Lasota, M., & Skoczylas, M. (2016, October 1). *Recognition of multiple traffic signs using keypoints feature detectors*. IEEE Xplore; IEEE. doi: 10.1109/ICEPE.2016.7781397

[9] Ryo, M., Karungaru, S., & Terada, K. (2017, July 1). *Character Recognition in Road Signs Using a Smartphone*. IEEE Xplore. doi: 10.1109/IIAI-AAI.2017.200

[10] Segin, A., Yatskiv, V., & Davletova, A. (2017, September 1). *Specialized computer based real time road signs recognition system for vehicles*. IEEE Xplore. doi: 10.1109/IDAACS.2017.8095120

[11] Sun, Y., Ge, P., & Liu, D. (2019, November 1). Traffic Sign Detection and Recognition Based on Convolutional Neural Network. https://doi.org/10.1109/CAC48633.2019.8997240

[12] Supriyanto, C., Luthfiarta, A., & Zeniarja, J. (2016, October 1). *An unsupervised approach for traffic sign recognition based on bag-of-visual-words*. IEEE Xplore; IEEE. doi: 10.1109/ICITEED.2016.7863253

[13] Zeng, J., Liu, M., Fu, X., Gu, R., & Leng, L. (2019). Curvature Bag of Words Model for Shape Recognition. *IEEE Access*, *7*(18648803), 57163–57171. doi: 10.1109/ACCESS.2019.2913688

[14] Mohd-Isa, W.-N., Abdullah, M.-S., Sarzil, M., Abdullah, J., Ali, A., & Hashim, N. (2020, October 1). Detection of Malaysian Traffic Signs via Modified YOLOv3 Algorithm. https://doi.org/10.1109/ICDABI51230.2020.9325690

Turnitin Originality Report

Processed on: 03-Oct-2021 15:53 +06
ID: 1663724490
Word Count: 6978
Submitted: 1

171-35-1883 By Md. Ziaul Karim

Similarity Index	Similarity by Source	
16%	Internet Sources:	11%
	Publications:	8%
	Student Papers:	8%

3% match (Internet from 17-Jun-2021)
https://www.analyticsvidhya.com/blog/2021/03/the-architecture-of-lenet-5/

2% match (publications)
Liwei Jia, Xiaoming Shi, Li Wei. "Design of Traffic Sign Detection and Recognition Algorithm Based on Template Matching", 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT, 2020

2% match (student papers from 16-Apr-2018)
Class: April 2018 Project Report
Assignment: Student Project
Paper ID: 947633998

1% match (Internet from 05-Jun-2020)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/3927/P148856%20%2828_%29.pdf?isAllowed=y&sequence=1

1% match (Internet from 15-Mar-2020)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/3553/P13659%20%2829%25%29.pdf isAllowed=y&sequence=1

1% match (publications)
Ying Sun, Pingshu Ge, Dequan Liu. "Traffic Sign Detection and Recognition Based on Convolutional Neural Network", 2019 Chinese Automation Congress (CAC), 2019

1% match (Internet from 22-Dec-2015)
http://ojs.academypublisher.com/index.php/jcp/article/download/jcp080513661370/6821

1% match (Internet from 15-Dec-2020)
https://www.mathworks.com/discovery/object-detection.html

1% match (publications)
Scott Krig. "Computer Vision Metrics", Springer Science and Business Media LLC, 2016

< 1% match (Internet from 26-Mar-2021)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/2088/P13003%20%2821%25%29.pdf isAllowed=y&sequence=1

< 1% match (Internet from 22-Oct-2020)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/4028/P14367%20%2821_%29.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 07-Apr-2021)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5145/152-15-6025%20%2818_%29.pdf?isAllowed=y&sequence=1

< 1% match (student papers from 29-Jun-2014)
Submitted to University of Newcastle upon Tyne on 2014-06-29

< 1% match (student papers from 24-Aug-2021)

Submitted to Engineers Australia on 2021-08-24

< 1% match (publications)
Z Akbar. "Community learning centre in improving disasters awareness through environmental education", IOP Conference Series: Earth and Environmental Science, 2021

< 1% match (Internet from 23-May-2019)
https://upcommons.upc.edu/bitstream/handle/2117/133070/138115.pdf?isAllowed=y&sequence=1

< 1% match (student papers from 19-Dec-2020)
Submitted to Florida Atlantic University on 2020-12-19

< 1% match ()
"Power System Stability Improvement via STATCOM and Supercapacitor"

< 1% match (Internet from 05-Apr-2021)
http://dl.lib.mrt.ac.lk/bitstream/handle/123/15050/Full%20thesis.pdf?isAllowed=n&sequence=3

< 1% match (publications)
Mohsina Binte Asad, Fatema Jannat Sharna, Nashid Alam, M. Jahirul Islam. "HSV and Template Matching based Bengali road sign recognition technique", 2016 International Conference on Innovations in Science, Engineering and Technology (ICISET), 2016

< 1% match (Internet from 07-Aug-2020)
https://mafiadoc.com/basic-aerodynamic-study-of-joined-wings_5a99aca31723dde3776dd700.html

< 1% match ()
http://innovexpo.itee.uq.edu.au/1999/thesis/kennedy/thesis.pdf

< 1% match (Internet from 01-Apr-2020)
https://www.slideshare.net/RaihanMahmud5/remote-doctor-project-report

"A Bangladeshi Road Sign Detection & Recognition System Based on Template matching & Convolutional Neural Network" By Md. Ziaul Karim ID: 171-35-1883 Supervised By Md. Shohel Arman Lecturer (Senior Scale) Department of SWE Daffodil International University DAFFODIL INTERNATIONAL UNIVERSITY DHAKA, BANGLADESH SEPTEMBER 2021 This Thesis report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering. APPROVAL OF THESIS This thesis titled "A Bangladeshi Road Sign Detection and Recognition System Based on Template Matching & Convolutional Neural Network.", submitted by Md. Ziaul Karim, ID: 171-35-1883 to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Software Engineering and approved as to it's style and contents. i DECLARATION I hereby declare that I have this thesis under the supervision of Md. Shohel Arman, Senior Lecturer, Department of Software Engineering, Daffodil International University. I also declare that this thesis or any part of this has been done by myself and has not been submitted anywhere else for the award of any degree. _____________________ Md. Ziaul Karim 171-35-1883 Department of Software Engineering Faculty of Science & Information Technology Daffodil International University. Certified by: _____________________ Md. Shohel Arman Senior Lecturer Department of Software Engineering Faculty of Science & Information Technology Daffodil International University. ii ACKNOWLEDGEMENT I would like to express my heartfelt gratitude to Almighty Allah who has provided me with knowledge & ability to carry out this thesis and a journey of about four years of studying in this wonderful vicinity. Also I would like to thank my honorable supervisor Md. Shohel Arman, Senior Lecturer, for his guidance, support, encouragement and co-operation. Without whom the thesis would not have been complete. iii ABSTRACT Road Signs are very important but a neglected topic in the discipline of driving vehicles in Bangladesh. As a highly populous country chaos on the road has taken a disastrous shape over the past decade as the number of vehicles on the road had went up and road safety became a crying need as the lives lost in road accidents are constantly on the rise. In case of Road Signs in Bangladesh the descriptive text for it is given more emphasis than the sign itself. Whereas, in the developed countries the signs are given more importance, because Signs are supposed to be universally accepted in case of communication. Hence, the Road Signs in Bangladesh are put quite arbitrarily, especially in the metropolitan areas the signs are credited by Metropolitan Police, or sponsored by companies, they look more like posters or banners than a legitimate Road Sign and it becomes very difficult to locate them. As we are witnessing the development of Autonomous Driving Systems all over the world by companies like Tesla, it is becoming more and more popular in demand. The detection & recognition of Road Signs play a crucial role in it. In Bangladesh Road Signs are ignored most of the time & text based road signs are given the priority. While human brain is capable enough to detect text based Road Signs, a smart system can struggle. Nonetheless, two types of measures could be taken into account in this case, one is through text recognition based and the other is detecting the sign in the poster-like signs and then run it through a neural network based classification to identify it. The later solution was taken into account for this paper, as all the Road Signs (both poster-like & basics) contain the basic signs for directions, they may or may not contain descriptive text along with them, but the road signs are always present in larger or smaller scales. The solution was built around Deep Learning and Image Processing techniques for object detection, segmentation and classification. Supervised Learning and Transfer Learning were used in terms of technical development. Text based recognition systems are not ideal for metropolitan areas because there are so many posters and banners attached to the poles and around the Road Signs. Key words : Road Sign, Object Detection, Region of Interest, Artificial Intelligence, CNN – Convolutional Neural Network, Deep Learning. iv TABLE OF CONTENTS Content Pg. No. vi LIST OF FIGURES Figure Title Pg. No. vii LIST OF TABLES Table Name Pg. No. viii ABREVIATIONS (i) RS – Road Signs (ii) RoI – Region of Interest (iii) CNN – Convolutional Neural Network (iv) YOLO – You only look once. ix CHAPTER 1 INTRODUCTION 1.1. Thesis Overview This study is composed of two parts, road signs detection and road signs recognition. A Road Sign detection and recognition solution is offered in this study. For detecting Road signs a model based on computer vision techniques like image segmentation and shape features matching has been used. For classification & recognition purpose a LeNet-5 architecture based Convolutional Neural Network has been used. The solution was created combining two parts together serving separate purposes of both detection & recognition. But we discuss unique instances regarding Road Signs used inside the metropolitan areas of Bangladesh and try to conduct tests with the developed solutions. 1.2. Problem Definition Road Signs in Bangladesh are quite difficult to detect and classify since our roads have plenty of posters and banners on the road and also there are road signs that are placed in such ways that it becomes quite difficult to detect them with even human viewpoint. The objective here is to detect and classify road signs against those odds and contribute to the future challenges about to be faced in the matter, regarding smart solutions. The problem that is covered in this paper is to detect road signs from 'poster like' road signs placements in metropolitan areas usually they are sponsored by certain companies with the courtesy of metropolitan police and they are unlike any other RS that are used all over the world except for the fact that they do contain the RS in the corner but in a very small scale, which can be difficult to spot by a camera. An example would be just as below – Fig. 1.1. Examples of Road Signs that follow out of the norm designs than traditional ones. 1.3. Motivation This study was carried out keeping 'Safer Road, Safer Future' in mind. Bangladesh needs safer Traffic Control Systems and strict laws to make our roads safer. Recent death tolls of students on the roads of Dhaka Metropolitan City in 2018 leading to "We want safe road" movement have been a major driving force for this study. Smart solutions can play vital a role in terms of saving lives on the road everyday. This study will work on contributing to unique instances of RS in use and also enable future researchers to develop a system that can detect them from live video feed since this study was carried out on input images only. 1.4. Thesis Contribution This thesis has been carried out keeping a research gap in mind which adds a new dimension to the field of Detection & Recognition systems of RS. This article takes knowledge gained from several authors into account and enables others to do likewise. More over the data collected for the purpose of the thesis is self-collected and preprocessed, which can also be of great use to future researchers. 1.5. Scope The scope of this study is going to be helpful for: - Detection & recognition of RS with unique features - Contribution to the dataset of the Bangladeshi RS. - Add a unique perspective to the whole concept of RS based smart solutions in Bangladesh. CHAPTER 2 LITERATURE REVIEW 2.1. Introduction A literature review is a review article which includes current knowledge, including firm based findings and as well as theoretical and methodological contributions to a particular topic. The review consists of summarization, critical evaluations of previously published work. It can refer to a full scholarly article, book or journal or specific portions of it. The main purpose is to provide a researcher or a general reader to provide with knowledge on the matter. 2.2. A Comparative study on the detection and recognition of RS in Computer Vision with Machine Learning & Deep Learning. If we look at [12] SURF, Bag of Visual words and K-means methods were used as backbone for real time detection of traffic signs, bag of visual words method, which uses key-point features of an image matching with a reference image to detect and recognize an object, K-means algorithm plays a vital role in the recognition by creating a histogram of the nearest neighbors. In [5] & [8] we can see Adaboost, SURF, SIFT were used as inspiration for key-feature matching to detect RS and extract RoI, however in [5] Artificial Neural Network was used which was based on MLP for shape classification, an additional SLP network also developed for content recognition in Triangular and Circular Road sign design. While [8] uses a two step algorithm that prioritizes shape features, but also eliminates the possibility of false positives by using ratio of the object area, however the accuracy is not too high (Lasota et. al. 2016) . In [3] Matching pursuit has been used, in this paper a pixel by pixel matching of template image and complemented image is performed which is similar to the detection method used in this paper, but the template matching technique is used for recognition of the RS as well, which is good for smaller scale dataset, but not for large scales, the accuracy here is not significantly high, SVM & HOG descriptor based models has been known to have shown higher accuracy rate (up to 95%). [2] Uses deep learning & also a completely innovative concept which is semantic web ontologies, it uses interpreting message sent by an RS to a deep learning model to recognize a traffic sign. This paper implemented RS ontologies, SPARQL query requests are used to detect a traffic a sign based on ontologies created and feed forward this data to a trained Neural Network to classify the sign. [4] and [3] both are based on similar datasets, but here [4] the use of DtBs vector can be seen along with neural networks. [4] uses image segmentation based on color and then uses aspect ratio and area to filter properties, then by calculating Distance to Borders (DtBs) to recognize the shape of an RS, candidate region is extracted and fed to an ANN network for recognition of a particular RS. [6] uses color based segmentation just [3] to better extract RS, then converts it to binary image and uses Fast Radial Symmetry algorithm to recognize the shape of candidate regions and Harries corner detector the other two basic shapes triangle & rectangle is recognized. Overall the detector has shown promising results (Horak et. al. 2016). [11] uses Haugh Transformation and CNN to classify RS in pictures, in preprocessing the HSV color space has been used like in [3] & [6] , however this experiment was only conducted on circular RS. However the accuracy rate is much higher (98.2%). 2.3. Conclusion HSV color is very close to the way human perceive an image [3], for segmentation of red an blue color used the majority of traffic signs it is worth considerable. From a large image [7] feature matching is the simplest way to detect RS and extract RoI. Template library is created for the purpose. [11] CNN is a better choice for recognition purposes since there are large datasets like German Traffic Sign Recognition Benchmark available, pretrained deep neural network architectures such as VGG-Net, Mobile-Net, Le-Net are available to train a robust model for the purpose. The trained model that has a higher accuracy rate can be used in the development of real-time detection and recognition for the future work as well. CHAPTER 3 METHODOLOGY We have figured out by now assessing previous works done on the matter from Chapter - 2, is that RS comes with some distinct basic features such as shape & colors. Aside from that we have area of content which dictates the classification & recognition. In terms of detection from a large scale field of view, the distinct features are widely used to extract the Region of Interest. The challenge here is to make the detection method as less time costly as possible but also given the limitations in existing unique instances of which we are solely going to focus on in our later experiments. The classification and recognition part is powered by the training of CNN models, the more it gets to train the more accuracy we can get out of it. 3.1. Collection of Dataset Keeping the unique RS detection and recognition in mind the collected the data is based on RS found in metropolitan areas of Chittagong & Dhaka. The source is very limited since the instances we are looking for are unique and very small in number. About 102 images (both unique & commonly used instances) of RS has been collected with a 16 megapixel camera & from google image search. For additional data we are using GTSRB database images that match with the ones in Bangladesh. We are also using the dataset layout from it. Some images were also created as synthetic data with illustrators for testing purposes. Examples of images in the Road Signs dataset are shown in Fig. 2.1. Fig 3.1. Examples of collected image data 3.2. Data Preprocessing Since image data is being used in the experiment, image pre-processing is a very important step in this regard. It is analogous to the mathematical normalization of a data set, which is a common step in various feature descriptor methods. It may have a positive impact on the quality of feature extraction and the results of image analysis. It is the first step after the collection of training dataset for a machine learning or a deep learning model. 3.2.1. For Detection For detection purposes, template images were created using illustrators with square image dimensions and also taken from real images by cropping into square dimensions as well. Then threshold binary images were produced and stored in a folder. Then the images were annotated using Microsoft Excel Spreadsheet with headings – 'template_folder, shape_class, sign_type' shown as below: template_folder shape_class sign_type 00 0 prohibitory 01 1 cautionary 02 2 informational 03 3 prohibitory 04 4 prohibitory .... Table 3.1. Template Data annotations for RS detection. Here shape_class 0= circular, 1=triangular, 2 = rectangular, 3=octagonal and 4= circular not allowed. 3.2.2. For Recognition For recognition purposes, An image with certain RS was cropped to square dimension with the RS in it, then converted into 32x32 dimensions, converted to RGB color format and .ppm file type. Similar signs were stored in a folder named with a distinct number. Images were annotated with the folder number as their class_ID and recognition name with RS_name, shown as below: class_ID RS_Name 00 Speed Limit 30 km/h 01 No Passing .... Table 3.2. Image Data annotations for recognition. 3.3. Detection method outline For RS detection purpose edge detection, HSV color space has been used. The detection model is divided into following steps. 1. Blue and Red color segmentation using HSV color space. 2. Eliminating noise with gaussian blur. 3. Edge Detection with Canny operator. 4. Drawing contours based on occupied pixel area to eliminate small unimportant areas. 5. Template matching with Template gallery for RoI bounding box generation. 6. Cropping the bounding box content. 3.4. Recognition method outline For RS recognition purpose, LeNet-5 has been used, which uses 32x32 pixel RGB image as input. Following steps are followed: 1. The detected cropped image is scaled down to 32x32 dimension. 2. Converted into RGB format. 3. Fed into the trained CNN network. 4. Recognition Results are produced in text output. 3.5 Requirements for Implementation Because of the technology involved like usage of OpenCV and a high level CNN architecture training, along with Tensorflow framework, the following setup of the following are required: Hardwares/Software requirements: 1. NVDIA GPU (CUDA 5.0 or above) 2. RAM (At least 8 GB) 3. Any Operating system that runs Tensorflow framework. CHAPTER 4 DETECTION METHOD & ARCHITECTURE 4.1. Introduction: Object Detection Object detection is one of the majorly discussed and researched topic in modern day field of Machine learning and Deep learning. Object detection is a key technology behind driver assistance systems, that enable cars to detect driving lanes or perform pedestrian detection to improve road safety. Object detection is also useful in applications such as video surveillance or image retrieval systems. 4.2. The literature in detection of Road Signs HOG features were used initially to detect and recognize RS. Fast Radial Symmetry and Harris corner detector [7] has also been effective in detection of RS apart from. By using Hough, the global detection problem that is not easy to solve can be transformed into the local peak detection problem that is easy to solve, making the transformed result easy to detect and recognize. Its advantage is that noise and curve discontinuity have relatively small influence (Sun et. al. 2019). In [1] Single Shot Detector(SSD) has been used to detect RS. YOLO v3 has also been found effective in detection and classification of RS. In most of the studies carried out before the popularity of Deep Neural Network based object detectors, image segmentation, shape classification and template matching has been popularly used [3]. Transfer learning based detection is best used for detection when there is availability of a vast amount of data. But when working with a small amount of data, it is best to use Computer Vision techniques [7]. 4.3. Outline of Bangladeshi Road Signs 4.3.1. Commonly used Road Signs in Bangladesh Road Signs used in Bangladesh are fairly common and they are mostly implemented in highways. But one of the major portion of it also plays a very important role in the traffic control systems Metropolitan areas. The vast population of drivers here are mostly unaware of the importance of them and more or less carefree, hence the metropolitan traffic police departments comes forward to provide RS that are better visible and poster like. That raises the questions below: - Does smart solutions exist to identify poster-like RS within a field of view? - Can existing solutions detect road signs in these circumstances? RS that align with the officially approved law can be categorized between red and blue colors. Shape features like Triangular, Circular, Rectangular, Rhomboidal and Octagonal. Prohibitory RS Cautionary RS Informative RS Fig. 4.1. Outline of Bangladeshi Road Signs Shape features are very important in terms of Detection of RS. Some commonly used shapes in RS are shown in Fig. 4.2. Fig. 4.2. Commonly used shapes for Road Signs. During the data collection some unique RS were discovered that are shown in Fig. 4.3. Fig. 4.3. Unique Road Signs 4.3.2. Road Signs in Metropolitan Areas. RS in metropolitan areas are quite different. Since in metro areas there are a lot of distractions on the side of the roads, poster like RS are used where the test descrption is given priority than the RS, as show in Fig. 1.1 in Chapter 1. Basically the structure of the RS in Metropolitan areas are as follows: Fig. 4.4. Structures of Road Signs in Metropolitan Areas. The problem lies here is that each metropolitan area has their own unique background colors of their own. Which makes it harder to detect. Only handful of locations have them and we don't have access to massive amount of data of this kind to support training and testing for a Neural Network based detection system. So, a solution was chosen based on Computer Vision techniques. 4.4. Proposed Detection Method The proposed detection method is based on color segmentation, edge detection, color segmentation and template matching technique [7]. Overall architecture is shown in a diagram that can be seen in Fig. 4.5. Look for shape features in Annotations Look for shape features in Annotations Shape class Suggestions Shape class Suggestions Crop Bounding Box Image Fig. 4.5. Detection model architecture 4.4.1. Color Segmentation Usually, RS are placed on the side of the road and often isn't placed in plain sight. There are backgrounds that are tree, posters and buildings etc. But all the RS have distinct colors that make them a bit more visible. Images taken by camera are digitized in RGB color format. But the correlation between the colors Red, Green and Blue components makes image segmentation quite challenging for lit up situations. Studies done for this paper have shown that HSV is more suitable for almost any kind of light intensity. HSV color space (or HSB color space) is composed of Hue H, Saturation S and Value V. It is a nonlinear transform of RGB space. Fig. 4.6. The formula for conversion of RGB color space into HSV space Fig. 4.7. HSV color space Then segmentation of red and blue color is done by performing following tasks: (i) Set lower red range to HSV(0, 50, 50) and upper red range to (10, 255, 255) and mask 0 is created. Similarly another red color mask 1 created with lower (170, 50, 50) and upper (180, 255, 255) ranges. (ii) Set lower blue range to HSV (100, 50,50) and upper blue range (110, 255, 255) and mask 2 is created. Similarly another blue color mask 3 is created with lower (110, 50, 50) and upper (120, 255, 255) ranges. (iii) All the masks are combined by mask = mask0+ mask1 + mask2 + mask3 (iv) A copy output image is made by setting every pixel value except for 'mask region' to 0. Results of the operations are shown in Fig. 4.5. Fig. 4.9. Performing color segmentation with HSV color space. 4.4.2. Eliminating noise with gaussian blur. The filtered image is applied with gaussian blur to remove left over noises and small red or blue noises picked up in the image. Fig. 4.10. Gaussian function & result 4.4.3. Edge Detection with Canny operator. Canny Edge detector is used on the gaussian blurred image. But before that grey scale filter is used on the image. Edge Detection operation is shown in Fig. 4.11. Gray Scale Cann Fig. 4.11. Canny edge detection & result. 4.4.4. Drawing contours based on occupied pixel area to eliminate small unimportant areas. After finding the optimum edges by Canny operation done in 4.4.3. contours are found and drawn around the edges using thresh binary methods. Process are shown in Fig. 4. Draw Contours and thresh binary image generation Fig. 4.12. Drawing Contours on occupied Pixel area of Road Sign. 4.4.5. Detecting shapes After drawing the contours we fill the bounding contour area we detect which shape classes it belongs to based on table 3.1. from Chapter 3. Fig. 4.13. shows filled up contour area of the sign. Fig. 4.13. Filling up the area of contour Based on OpenCV approximation function we detect whether the filled area could belong to which shape classes & give a list suggestion. In this case, class – 0, 3, 4 are suggested. 4.4.6. Template matching with Template gallery for RoI bounding box generation. 4.4.6.1. Template Gallery Since template matching is used for identification, we tend to identify a specific shape contour that exists within the picture that belongs to the norm of RS. The pictures in the template gallery are all common outlines and shapes used in RS in binary images. That means only shape features are going to matter in this part. Therefore, template gallery uses binary image format. 208 template images of 9 different types have been added to the gallery, all with performed image augmentations using shear, zoom and rotation. The template folders look as figure 4.14 - Fig. 4.14. Template Gallery 4.4.6.2. Template Matching Template matching is a process. Algorithm in [7] is being used in this case. First the template in the gallery is judged for the similarity with each part of the image. Secondly, it is determined whether it has the similar part as the template image. In 4.4.5. detection of shapes cuts down the process of looking through all the folders for templates that could match, based on the suggestions it had made. Finally, the specific location of the RoI is detected in the picture. First, we can place the detection object template t(x, y) on the image. Secondly, we focus on detecting the similarity between the template in the suggested template folders (suggested by the shape detection) and the input image. All the pixels in the image have such operations. Finally we need to determine whether or not there is an object in the image according to the maximum similarity or it exceeds a certain amount of threshold, and find the specific location of the object which here is the RS. Matching measures include: max | f -t | (1) $\iint | f - t \; | dxdx \; s \iint (f - t)2dxdx \; s$ (2) (3) The smaller the measurement values of equation (1), (2) and (3), the better the matching result; on the contrary, the larger the measurement values below, the better the matching result. (Jia et. al. 2020) m(u, v)= $\iint s \; t(x, x) f(x + t, x + t) dxdx$ (4) m(u, v)= $m(u,u)$ (5) $\sqrt{\iint s} \; f(x+u,x+u)2dxdx$ 4.4.7. Drawing Bounding Box and Cropping the Region of Interest After finding the maximum matching pixels the object is then localized with bounding box. Then using the bounding box co-ordinates (height, width, x, y) the content within is cropped to extract the RS. The operations performed are shown in Fig. 4.15. Draw Bounding Box Crop RoI Road Sign Fig. 4.15. Cropping into RoI. 4.4.7.1. Testing on other instances Experiments on poster-like RS has been conducted under the same methodology discussed in previous sections. They are shown in Fig. 4.16. Fig. 4.16. Example of detecting RS in Poster-Like Road Sign 4.5. Results & Discussion RS detection is a very complex problem and there are a lot of dependencies involved in the matter. A few things are notable as follows: (i) It becomes difficult to detect RS from Poster-Like circumstances if the background is matching with the RS border color. Then the color segmentation becomes too difficult to isolate the RS. As shown in Fig. 4.17. Fig. 4.17. Road Sign background matching the sign color. (ii) If the text description is written in either red or blue color that matches the RS color, then it gives false positive detection of RS. As seen in Fig. 4.16. the English letter 'o' in "No Parking" text is also detected as an RS. (iii) Detection time is relatively fast for almost all the RS shapes. The highest time recorded being 40 milliseconds (for poster-like RS) and lowest being 4.9 milliseconds(for circle shaped RS) . Fig. 4.17. shows time records for detection purposes. Fig. 4.18. Detection time records 4.6. Conclusion The detection method used here is not quite robust and can give false positive at times. As we are not using any trained model to detect RS it is safe to say this model could be improved with more accuracy if more data were accumulated. But this has been developed keeping unique instances in mind and the fact that a trained Machine Learning model or Neural Network based model would be able to detect RS more accurately. Then again, it all comes down to the access of a lot of data and the instances here are unique and rarely seen. Some false-positive detection are shown in Fig. 4.18. Fig. 4.19. Some wrong detections. CHAPTER 5 RECOGNITION METHOD & ARCHITECTURE 5.1. Introduction: Object Recognition Object recognition is a technology enabled by computer that is related to computer vision and digital image processing. It deals with both detecting instances of an object to classify and recognize it based on Machine Learning & Deep Learning based algorithms, backed by thousands of training image data. Supervised Learning, Unsupervised Learning and Transfer Learning is widely used in Object Recognition. 5.2. Literature of Road Signs Recognition Use of visual words [12] has been seen effective in unsupervised learning of RS recognition. Support Vector Machine Based classifications [3] have been used in the past and also several feature matching algorithms [8] like SIFT, SURF, Adaboost Classifier & Template matching has been used in earlier days of RS recognition technologies. In modern days the widespread use of Deep Learning and advancement of Artificial Intelligence, Neural Network based architectures are more suitable of such tasks. ANN architectures have been used paired with Image Processing techniques like DtBs vectors[4] and shape & content classifications [5]. Modern day ease of access in data and impressive architectures for transfer learning for CNN [11] like AlexNet, MS COCO dataset & YOLO[14] architectures enables us to easily train our data into deep networks and get desired accuracy. In this study we have used LeNet-5 architecture for our transfer learning & training purposes and used Convolutional Neural Network (CNN) as the main training model. 5.3. Overview of the used Architecture Convolutional Neural Networks is the standard form of neural network architecture for solving tasks associated with images. It is solely used for tasks such as object detection recognition purposes. It is a form of feed-forward neural network which has artificial neurons that answers a locality of the encircling cells within the coverage vary. It performs very well in large-scale image processing. There are several different architectures for CNN. The one that is used in this paper is LeNet-5. Here's a discussion about the methodology architecture – 5.3.1. LeNet-5 Features The features of LeNet-5 are – (i) Every convolutional layer consists of three different parts – convolution, pooling and nonlinear activation functions. (ii) Spatial features are extracted by convolution. (iii) In average pooling layer subsampling are done. (iv) tanh activation function is used after (iii) (v) Multilayer Perceptron (MLP) is used as the last classifier. (vi) Sparse connection between layers are used to reduce the computation complexity. 5.3.2. Architecture of LeNet-5 LeNet-5 network is composed of 5 layers with learnable parameters provided. It has three sets of convolution layers which consists of three different parts as discussed in 5.3.1. After the convolution and average pooling layers are through, we have two fully connected layers. At the very last, a Softmax classifier which classifies the images into respective class. Input 32 × 32 × 1 Fig. 5.1 Input Image The input to this model is a 32 × 32 grayscale image. Fig. 5.2 Feature Mapping The first convolution operation with the filter size 5×5 and we have 6 such filters. As a result, we get a feature map of size 28 × 28 × 6. Here the number of channels is equal to the number of filters that have been applied. Fig. 5.3 Subsampling After the first pooling operation, we apply the average pooling and the size of the feature map is reduced by half. Note that, the number of channels is intact. Fig. 5.4. Convolution Next, we have a convolution layer with sixteen filters of size 5×5. Again the feature map changed it is 10×10×16. The output size is calculated in a similar manner. After this, we again applied an average pooling or subsampling layer, which again reduce the size of the feature map by half i.e. 5×5×16. Fig. 5.5. Final convolutional layer. Then we have a final convolution layer of size 5×5 with 120 filters. As shown in the above image. Leaving the feature map size 1×1×120. After which flatten result is 120 values. After these convolution layers, we have a fully connected layer with eighty-four neurons. At last, we have an output layer with ten neurons since the data have ten classes. Here is the final architecture of the Lenet-5 model. Fig. 5.6. Fully connected layer Here are the operation performed in LeNet-5 in a nutshell shown in Fig. 5.7. Fig. 5.7. LeNet-5 Architecture 5.4. Training As we have mentioned in Chapter 3 section 3.1. we are using the format and data from German Traffic Sign Recognition dataset which has over 40 classes and more than 50000 images in total. We are using about 44 class with extra 4 being added classes for signs unique to Bangladesh. Here is the distribution of images plotted in a histogram in Fig. 5.8. Fig. 5.8. Distribution of training Dataset. The training was run up to 30 epochs. The training to test data ratio was 0.2, validation ratio was 0.2 as well. 5.5. Testing and Results 5.5.1. Accuracy We have trained our model to it's limits. The accuracy went up to 98.2%. A model evaluation diagram is plotted in Fig. 5.9. Fig 5.9. Accuracy and loss evaluation diagram 5.5.2. Test Results After the detection and extraction of RS in Chapter 4, we pick up right after section 4.4.7. in this section. Some tests were run on the collected data and here are the results shown in Fig. 5.10. Fig 5.10. Road signs recognition in images 5.6. Conclusion The accuracy rate is satisfactory in this experiment. Despite the recognition confidence being low for RS that have low amount of data. In this section, a traffic sign recognition method on account of LeNet-5 architecture is proposed, Test result displays that the accuracy of this method is 98.2%. A more well balanced collection of data can solve the lower confidence rate in classes that contain small number of images. CHAPTER 6 CONCLUSION, LIMITATIONS &FUTURE SCOPE 6.1. Conclusion In this study two parts of research were carried out. One being RS detection and the other being RS recognition. The first part was solely image processing, based on template matching, color segmentation, shape detection and template matching techniques. The second part was implemented using LeNet-5 architecture and transfer learning process, the accuracy were found up to 98.2% which is quite satisfactory. 6.2. Limitations (i) (ii) Due to lack of Data collection the detection method still isn't that robust. Recognition Model needs a more balanced collection of data for certain instances. 6.3. Future Scope In the future the unique instances of traffic signs could be detected with a more robust algorithm and collections of those instances could escalade the detection process more easily because it will be eligible for using advanced detection algorithms like YOLO v5 or SSD. References [1] Ahsan, Sk. Md. M., Das, S., Kumar, S., & La Tasriba, Z. (2019). A Detailed Study on Bangladeshi Road Sign Detection and Recognition. IEEE Xplore, 1–6. doi: 10.1109/EICT48899.2019.9068760 [2] Amrani, N. E. A., Abra, O. E. K., Youssfi, M., & Bouattane, O. (2019, October 1). A new interpretation technique of traffic signs, based on Deep Learning and Semantic Web. IEEE Xplore. doi:10.1109/ICDS47004.2019.8942319 [3] Asad, M. B., Sharna, F. J., Alam, N., & Islam, M. J. (2016). HSV and Template Matching based Bengali road sign recognition technique. IEEE Xplore, 1–4.doi: 10.1109/ICISET.2016.7856527 [4] Chakraborty, S., Uddin, M. N., & Deb, K. (2017). Bangladeshi road sign recognition based on DtBs vector and artificial neural network. IEEE Xplore, 599–603. doi: 10.1109/ECACE.2017.7912975 [5] Hamdi, S., Faiedh, H., Souani, C., & Besbes, K. (2017, January 1). Road signs classification by ANN for real-time implementation. IEEE Xplore; IEEE. doi: 10.1109/CADIAG.2017.8075679 [6] Horak, Karel & Cip, Pavel & Davídek, Daniel. (2016). Automatic Traffic Sign Detection and Recognition Using Colour Segmentation and Shape Identification. doi: 10.1051/matecconf/20166817002 (n.d.). MATEC Web of Conferences 6 , ICIEA 2016 8 17002 (2016). [7] Jia, L., Shi, X., & Wei, L. (2020, October 1). Design of Traffic Sign Detection and Recognition Algorithm Based on Template Matching. doi: 10.1109/ICCASIT50869.2020.9368750 [8] Lasota, M., & Skoczylas, M. (2016, October 1). Recognition of multiple traffic signs using keypoints feature detectors. IEEE Xplore; IEEE. doi: 10.1109/ICEPE.2016.7781397 [9] Ryo, M., Karungaru, S., & Terada, K. (2017, July 1). Character Recognition in Road Signs Using a Smartphone. IEEE Xplore. doi: 10.1109/IIAI-AAI.2017.200 [10] Segin, A., Yatskiv, V., & Davletova, A. (2017, September 1). Specialized computer based real time road signs recognition system for vehicles. IEEE Xplore. doi: 10.1109/IDAACS.2017.8095120 [11] Sun, Y., Ge, P., & Liu, D. (2019, November 1). Traffic Sign Detection and Recognition Based on Convolutional Neural Network. https://doi.org/10.1109/CAC48633.2019.8997240 [12] Supriyanto, C., Luthfiarta, A., & Zeniarja, J. (2016, October 1). An unsupervised approach for traffic sign recognition based on bag-of-visual-words. IEEE Xplore; IEEE. doi: 10.1109/ICITEED.2016.7863253 [13] Zeng, J., Liu, M., Fu, X., Gu, R., & Leng, L. (2019). Curvature Bag of Words Model for Shape Recognition. IEEE Access, 7(18648803), 57163–57171. doi: 10.1109/ACCESS.2019.2913688 [14] Mohd-Isa, W.-N., Abdullah, M.-S., Sarzil, M., Abdullah, J., Ali, A., & Hashim, N. (2020, October 1). Detection of Malaysian Traffic Signs via Modified YOLOv3 Algorithm. https://doi.org/10.1109/ICDABI51230.2020.9325690