Traffic sign detection and recognition based on convolutional neural networks

Md Tarequl Islam

Department of Electrical and Electronic Engineering

Student of Engineering, Ahsanullah University of Science and Technology

zahin178@gmail.com

Abstract--**Traffic sign detection and recognition topic are one of the most popular topics of computer vision and image processing in recent years, as they play an important role in autonomous driving and traffic safety. This paper proposes a system that will detect and classify different types of traffic signs from images. This paper differs from other papers as it uses signs that are globally recognized and isn't limited to very few signs like many other papers. The number of signs used in this paper for classification is 28, which are used all around the globe. Two separate neural networks have been used for detection and recognition purpose; one classifies the sign and other the shape. Image augmentation has been used to create the training and validation dataset. 40,000 images have been used to train the first classifier with 28000 positive images (images that contain the traffic signs) and 12000 negative images(images that do not contain any traffic signs) and 3600 images were used to train the second classifier with 2400 positive images and 1200 negative images. The images are processed to find the region of interest, which is then fed to two CNN classifiers for classification.**

Keywords—Traffic sign recognition, traffic sign detection, image augmentation, image processing, convolutional neural network

I. INTRODUCTION

Traffic sign detection and recognition have become a popular field in the computer vision sector, as it plays a vital role in assisting the driver with understanding the signs and following traffic rules as well as creating an autonomous driving system. Traffic signs consist various unique shapes and a range of colors, which is why it is easily noticeable by the human eyes. The same features can be used for the computers to distinguish and classify the traffic signs.

The color and shape of each traffic sign make it unique than the rest of the signs. The colors that are commonly associated with the traffic signs are red, blue, green, yellow and the shapes are mostly circular, rectangular or triangular.

The signs chosen in this paper have a background color of red and blue and the shapes of the signs are circular and triangular. HSV color space has been used to detect the red and blue objects in the image. Masking is also used to extract these objects from the rest of the background. Further image processing is also applied to get the proper ROI(region of interest). Then the ROI in converted to 64x64 and 128x128 image patches, which are then fed to sign classifier and shape classifier respectively. The shape classifier tries to classify the shape of the object, while the sign classifier tries to classify the sign. The results from both the classifiers finally decide if the image patch contains a traffic sign or not. Convolutional neural network has been chosen for this objective as it performs better than svm [8]. The cnn used for sign classification is based on the LeNet architecture, which consists of two convolutional layers and two fully connected layers while the cnn used for shape classification consists of one convolutional layer and two fully connected layers.

II. RELATED WORK

Finding the right region is the first step of classifying any object. There are many approaches to finding the optimum ROI. Majid Mirmehdi *et al.* [3] used maximally stable extremal regions (MSERs) to find regions that could consist a traffic sign. Each frame is binarized at a different number of threshold levels, and the connected components at each level are found. These components that maintain their shape throughout several threshold levels are selected as MSERs. For the traffic signs to maintain their shapes throughout the threshold levels, different color enhancement technique has been used [3], [5]. As traffic signs have different unique shapes, the edges and eigenvalues can also be a medium of detecting the traffic signs [6], [7]. Jung-Guk Park *et al.* [4] proposed different machine learning algorithms to find the features directly without depending on color or shape, which is computationally more expensive than the previous methods.

For classification purpose, different machine learning algorithms and neural networks are commonly used. Among the machine learning algorithms, support vector machine (SVM) is the most popular one. SVM has been used with histogram of oriented gradient (HOG) features [3] and with binary images [9]. Neural network is also becoming more robust and efficient in this field. Ioan Cristian Schuszter showed in his paper that the neural network is performing better than SVM [8]. As a result neural network is applied in various papers [9], [10], [11], [12].

III. SYSTEM OVERVIEW

A. Selecting the ROI

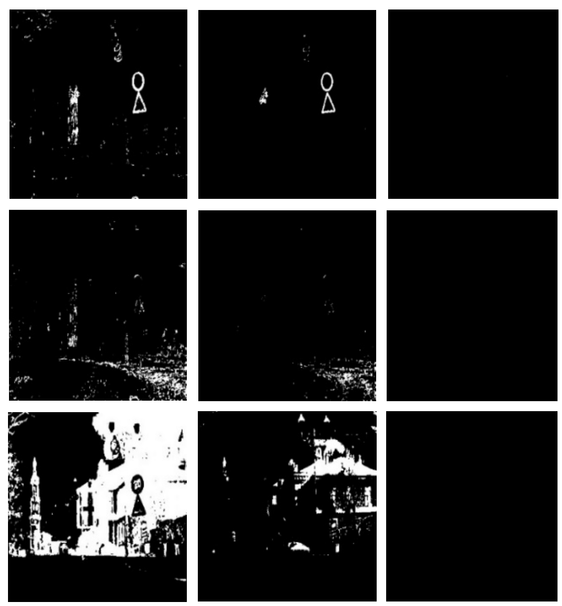
Selecting the ROI is very important in order to classify the signs properly. For this purpose, HSV color space has been used [1]. There are various color spaces available like RGB, HSV, GRAYSCALE, YCbCr etc. Normally when an image is taken, it is taken in RGB format meaning every pixel contains three information, the intensity of red, green and blue color. These three colors are superimposed on each other to get other colors. In the case of RGB, all three of the information is used just to identify the color or the Chroma while not concerning the luminance or Luma. But unlike RGB, HSV separates luma, or the image intensity, from chroma or the color information. As a result, the system is more robust to lighting changes. Thus the accuracy of selecting the correct color gets better regardless of the environment. So the image is converted from RGB to HSV.

At first, the image is blurred to reduce the noise, then it is converted to HSV color space. There are various types of blurring process but Gaussian blur has been chosen for the purpose. Now the task is to extract red and blue color from the image. Therefore the masking process is used to get the blue and red regions. Image masking is a process of graphics software to hide some portions of an image and to reveal some portions. It is based on hue, saturation and value. The masking is a non-destructive process of image editing. A range is needed for each color and then masking is applied according to that range. There was a predefined range for the hue feature, saturation feature and the value feature for both red and blue color. For red, two separate hue ranges had to be defined to cover the whole red color section.

TABLE 1. RANGES USED FOR THE MASKING

Color	Hue	Saturation	Value	Mask Id
Red-I	160 - 180	10 - 255	10 - 255	1
		50 - 255	50 - 255	2
		110 - 255	110 - 255	3
Red-II	0 - 10	10 - 255	10 - 255	4
		50 - 255	50 - 255	5
		110 - 255	110 - 255	6
Blue	90 - 120	10 - 255	10 - 255	7
		50 - 255	50 - 255	8
		110 - 255	110 - 255	9

(a) Original image

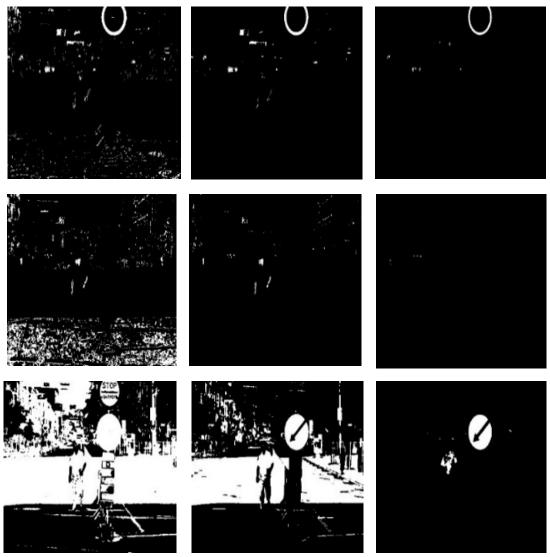


(b) Masked images

Figure 1.

Table 1 shows the range that had been used for the process and figure 1 and 2 show the result. The top left image is the result of mask id 1 and the bottom right image is the result of mask id 9 in both of the figures.

(a) Original image



(b) Masked images

Figure 2.

Figure 1 and 2 express the necessity of using different ranges of saturation and value, as it reduces overlapping and the chances of finding the proper ROI gets better. In the real world, the lighting conditions vary from time to time, so depending on a single predefined range of saturation and value is very inefficient. Some ranges do better in a bright environment, while other ranges do better in the dark environment. So picking a single range is highly inefficient. One can also think of taking the entire range of 0 to 255 but it is also not effective as it would select almost all of the connected objects of the image and there is a high possibility of overlapping. One other technique used by Prashengit Dhar *et al.* [15] is to combine two different masked binary images in order to include the sign in the white region. Although this technique is quite decent, sometimes it may cause overlapped traffic signs. The first image of figure 1 shows a similar situation. If the image in figure 1 is masked using the range of mask Id 1, then the two signs will overlap with each other and will be detected as a single sign. As a result, the classifier will not classify them but if the second range is used, then the signs will not be connected and the classification will be accurate. At the same time, the perfect range for the first image was mask Id 2 but for the second image, it was mask Id 9. So even though the process is computationally expensive, the performance is better.

After that, different morphological processes have been applied to filter out the unnecessary parts, but not all the contours consist traffic signs. Different parameters of contours are utilized to reduce the number of contours further. Finally, the contours those are left after all the processes are bounded by a bounding box. The ROI is then selected by extracting the region bounded by the bounding box. Then the ROI is converted into a 64x64 and 128x128 image patches and fed to the neural networks.

B. Generating the Training Data

In order to make the full use of the neural network, a lot of data is required. Each class requires a large number of images to be trained so that the classification can be as accurate as possible. The signs used for the classification purpose on this paper can be observed in figure 3.

There is a total of 28 signs from different road sign categories such as warning signs, traffic calming signs, speed limit signs etc. The images were taken from the U.K. traffic signs image database [2]. Gathering a sufficient amount of real images for training a neural network can be really challenging and time-consuming. That is why image augmentation has been used in this paper to avoid all the difficulties while getting a robust classifier. Image augmentation is the process of creating new images through different ways of processing or a combination of multiple processing, such as random rotation, shifts, shear and flips, etc. It is a very efficient method if the source material is insufficient. Jia Shijie *et al.* showed that, the more augmented samples added to the original training set, the higher classification accuracy gets [13]. As the signs used in this paper have the circular or triangular shape and the bounding box is rectangular, if the shapes are bounded by the box then certain part of the background will also be included. That is why the traffic signs are superimposed on some random background first before applying any transformation, so that the training images will also have the same attribute. Figure 4 shows some examples that had been used in training. After superimposing the images, different types of transformation have been applied to the images.

Figure 3. Set of signs used in this paper

Figure 4. Example positive images from the training set

Transformations those are commonly used in the augmentation purpose are shear, zoom in, reflection, rotation, histogram equalization, white balance and sharpen [14] but not all the transformations can be applied in this case, such as reflection. If it is used for creating new images, then the true meaning of some of the signs may get altered. For example, if horizontal reflection is applied to a left curved sign, then it will become a right curved sign, which may lead to a corrupt training set. The same thing may happen in the case of vertical reflection. The methods applied for augmentation in this paper are blurring, affine transformation and white balance, so that the classifier can be robust against blurring, different lighting conditions and different perspectives of the camera. The number of images created from the 28 original signs is 28,000, each sign having 1000 samples. Negative images also have been used in the training section to reduce the number of false positives. It has been seen that without using any negative images, the classifier mistakenly labels random images as a sign image, which is highly unacceptable. The distribution can be seen in Table 2. Another dataset also has been used to train another neural network that will classify shape.

TABLE 2. DISTRIBUTION OF THE FIRST DATASET

	Training	Validation
Positive image	25200	2800
Negative image	10800	1200
Total	36000	4000

The dataset used in this case is binary images containing triangular shape, circular shape and random shape. The distribution is expressed in Table 3.

TABLE 3. DISTRIBUTION OF THE SECOND DATASET

	Training	Validation
Triangular	1200	150
Circular	1200	150
Random	1200	150
Total	3600	450

C. Convolutional Neural Networks

TABLE 4. ARCHITECTURE OF THE FIRST CNN

Input image (128,128,1)
Convolution layer with ReLU activation
Maxpooling
Fully connected layer with ReLU activation and dropout
Fully connected layer with Softmax activation

TABLE 5. ARCHITECTURE OF THE SECOND CNN

Input image (64,64,1)
Convolution layer with ReLU activation
Maxpooling
Convolution layer with ReLU activation
Maxpooling
Fully connected layer with ReLU activation and dropout
Fully connected layer with Softmax activation

For the classification purpose, two separate neural networks have been proposed in this paper. One neural network will try to classify if the shape of the contour is triangular or circular or irregular and random. The architecture of this neural network is very simple, which is represented in Table 4. Another neural network will try to classify if the image patch contains any sign or not and if so, then what the sign is. Table 5 shows the architecture of the neural network. Categorical cross-entropy method was used to determine the loss of the network in both cases and Adadelta optimizer was used to achieve the convergence as quickly as possible [16].

IV. RESULTS AND COMPARISON

In order to find the accuracy of the system proposed in this paper, street images of Ukraine and Bangladesh have been used. These images contain different traffic signs, which were used in this paper in different environments. As this paper deals with both traffic sign detection and recognition, the goal is to detect the signs present in the images and identify them. The images of Ukrainian streets were taken from online, while the streets of Bangladesh were captured by mobile phones. The reason for taking two different countries street images is to show the significance of the proposed method despite the slight variation of the signs as well as its ability to label the signs which are used globally throughout the world. The variations of the signs between these two countries can be observed in figure 5.

Figure 5: Differences between signs

Even with this type of variation, the classifier was able to classify the signs correctly. There are a total of 24 Ukrainian street images and 11 Bangladeshi street images. The performance of the proposed system over these images is shown in Table 6. The reason why the number of false positive and false negative is the same is that, every time a sign is classified as another sign, then it will be a false negative for the first class, while at the same time, it will be false positive for the second class. In other words, if there is a false positive, there will always also be a false negative and vice versa. That is why they are the same thing. As a result, the precision and recall are also the same. Images of different streets with different street signs were fed to the system and the system detected and classified the signs (figure 6).

In this paper precision and recall were also taken into account, which are absent in many papers. Some of the papers are only concerned about accuracy [9] [11] [17], which cannot be the only parameter for judging the performance of a classifier.

TABLE 6. PERFORMANCE OF THE SYSTEM

	Total Sign	True Positive	False Positive/ Negative	Precision/ Recall	Accu-racy
Ukraine	29	27	4	0.87	0.93
Bangladesh	11	9	4	0.69	0.81
Total	40	36	8	0.81	0.90

Figure 6: Images showing traffic sign recognition by the system

A classifier can have high accuracy with very low recall or precision, meaning the classifier will not only detect and label signs but it will also incorrectly label random objects as traffic signs, which can be truly misleading. Precision and recall can also give the idea of how well the classifier is detecting possible regions containing signs. Another aspect of this paper is that it is not confined to certain regional traffic signs like [15] [18]. The dataset which was used to train the neural network was augmented from the British traffic sign database, and the images used to measure the performance were from Ukrainian and Bangladeshi streets and are robust to the slight differences between the same sign in separate countries.

V. CONCLUSION

This paper proposed a system that is able to detect and classify a set of 28 traffic signs in different environments. The results are moderate and it can be improved by testing different neural network structures. As a neural network is often called a black box, there is no guarantee that it will perform best with the defined set of parameters stated above. New methods of data augmentation can also be applied to make the classifier more robust. Real-time detection and recognition can also be implemented in the future.

REFERENCES

[1] Zhong LIU, Weihai CHEN, Yuhua ZOU and Cun HU "Regions of Interest Extraction Based on HSV Color Space," IEEE 10th International Conference on Industrial Informatics, July 2012.

[2] Dept. Transp., London, U.K., Traffic Signs Image Database, 2011

[3] Jack Greenhalgh and Majid Mirmehdi "Real-Time Detection and Recognition of Road Traffic Signs," IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 13, NO. 4, DECEMBER 2012.

[4] Jung-Guk Park, Kyung-Joong Kim "A METHOD FOR FEATURE EXTRACTION OF TRAFFIC SIGN DETECTION AND THE SYSTEM FOR REAL WORLD SCENE," IEEE International Conference on Emerging Signal Processing Applications, 12-14 Jan. 2012.

[5] Sungho Kimh and Soon Kwon "Improvement of traffic sign recognition by accurate ROI refinement," 15th International Conference on Control, Automation and Systems (ICCAS 2015) Oct. 13-16, 2015 in BEXCO.

[6] Hurriyatul Fitriyah, Edita Rosana Widasari, Gembong Edhi Setyawan "Traffic Sign Recognition using Edge Detection and Eigen-face," International Conference on Sustainable Information Engineering and Technology (SIET), 2017.

[7] H. Fleyeh, E. Davami "Eigen-based traffic sign recognition," IET Intelligent Transport Systems (Volume: 5 Issue: 3 September 2011).

[8] Ioan Cristian Schuszter "A Comparative Study of Machine Learning Methods for Traffic Sign Recognition," 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2017.

[9] David Soendoro, Iping Supriana "Traffic Sign Recognition with Color-based Method, Shape-arc Estimation and SVM," International Conference on Electrical Engineering and Informatics 17-19 July 2011.

[10] Rongqiang Qian, Bailing Zhang, Yong Yue and Frans Coenen "Robust Chinese Traffic Sign Detection and Recognition with Deep Convolutional Neural Network," 11th International Conference on Natural Computation (ICNC), 2015.

[11] Md. Abdul Alim Sheikh, Alok Kole and Tanmoy Maity "Traffic Sign Detection and Classification using Colour Feature and Neural Network," International Conference on Intelligent Control Power and Instrumentation (ICICPI), 2016.

[12] Tiago Moura, António Valente, António Sousa, Vítor Filipe "Traffic Sign Recognition for Autonomous Driving Robot," IEEE International Conferenceon Autonomous Robot Systemsand Competitions(ICARSC) May14-15, 2014.

[13] Jia Shijie, Wang Ping, Jia Peiyi, Hu Siping " Research on Data Augmentation for Image Classification Based on Convolution Neural Networks," Chinese Automation Congress (CAC), January 2018.

[14] Agnieszka Mikołajczyk, Michał Grochowski "Data augmentation for improving deep learning in image classification problem," International Interdisciplinary PhD Workshop (IIPhDW), 2018.

[15] Prashengit Dhar, Md. Zainal Abedin, Tonoy Biswas1, Anish Datta "Traffic Sign Detection- A New Approach and Recognition Using Convolution Neural Network," IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017, Dhaka, Bangladesh.

[16] An overview of gradient descent optimization algorithms, Accessible on http://ruder.io/optimizing-gradientdescent/?fbclid=IwAR2L2KohtTEVGzkLns2pF4tTswI0WEC-2R8eNBdsHUn31_h6ANHk1NzPAwo.

[17] Aashrith Vennelakanti, Smriti Shreya, Resmi Rajendran, Debasis Sarkar, Deepak Muddegowda, Phanish Hanagal "Traffic Sign Detection and Recognition using a CNN Ensemble," IEEE International Conference on Consumer Electronics (ICCE), 2019.

[18] Mian Mian Lau, King Hann Lim, Alpha Agape Gopalai "Malaysia Traffic Sign Recognition with Convolutional Neural Network," IEEE International Conference on Digital Signal Processing (DSP), 2015.