YOLO-Based Three-Stage Network for Bangla License Plate Recognition in Dhaka Metropolitan City

Sohaib Abdullah
Department of Computer Science and Engineering
Manarat International University Bangladesh
Dhaka, Bangladesh
sohaib@manarat.ac.bd

Md Mahedi Hasan
Institute of Information and Communication Technology
University of Engineering and Technology
Dhaka, Bangladesh
mahedi0803@gmail.com

Sheikh Muhammad Saiful Islam
Department of Pharmacy
Manarat International University
Dhaka, Bangladesh
saifulislam@manarat.ac.bd

***Abstract*—Real-time Automatic License Plate Recognition (ALPR) is a challenging task in computer vision since license plates in the real-world scenarios are difficult to recognize due to the presence of various factors like transparent background, occlusion, existence of multiple plates in an image, variance in illumination and viewing angle, etc. In most of the previous works related to Bangla license plate, images used for ALPR were captured in ideal conditions due to unavailability of dataset representing non-ideal conditions. In this paper, we present an approach for real-time Bangla license plate detection that is robust to enormous variations in complex real-world environment. A novel license plate recognition method is also presented making it a complete end-to-end deep learning-based ALPR system. We introduce a dataset by collecting $1,500$ different Bangladeshi vehicular license plate images that are captured manually from the street resembling various real-world scenarios. In this work, we have employed YOLOv3 algorithm to successfully localize the license plate and recognize the digits. To recognize the character, we have also built a Bangla scene character dataset containing more than $6,400$ characters, with which we have trained a ResNet-20-based deep Convolutional Neural Network (CNN). Our proposed method achieves more than 85% Intersection over Union (IoU) in digit recognition. The ResNet-20-based CNN model achieves 92.7% accuracy in recognizing the Bangla character present in the license plate.**

***Index Terms*—ALPR, Bangla license plate detection, Bangla scene character recognition**

I. Introduction

Automatic License Plate Recognition (ALPR) or Vehicle License Plate Detection (VLPD), plays an important role in many applications such as automatic parking system, toll collection, road traffic monitoring, law enforcement, access and border control and so on. It is the core of an intelligent transportation system around the world. This type of system is also very important for Bangladesh. According to Bangladesh Road Transport Authority (BRTA), Dhaka has around 3.4 million registered vehicles as of March 2018. Traffic law violation, crime and unbearable traffic jam are common in this city. So need for ALPR is even greater than any other cities around the world.

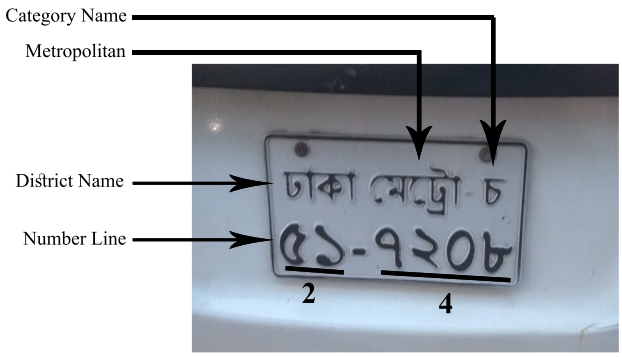

Fig. 1. BRTA license plate format. The first line of the license plate generally contains the name of the district and the word Metro, followed by a hyphen and a character. And the second line contains six Bangla digits.

In Bangladesh, all the license plates are written in Bangla language with a fixed two-line text format and color for different types of vehicles. According to the regulations of BRTA, the first line of the license plate contains the name of the district (vehicle registration area), and the word Metro, if it has been registered in a metropolitan area, followed by a hyphen and a letter indicating the category of the vehicle. Thirty two letters have been assigned to indicate the type of the vehicle. The second line (or Number Line) contains six digits, the first two of which denote vehicle class number and the last four separated from first two by a hyphen constitute unique vehicle registration number.

Earlier researchers used traditional hand-crafted features to detect Bangla license plates. But these techniques are not very useful in real-time license plate detection. In many cases, license plate images have been captured in ideal conditions making it less useful in a busy city like Dhaka. One of the reasons for this is that the dataset containing Bangla license plates representing non-ideal conditions such as presence of different lighting conditions, viewing angles, transparency,

978-1-5386-8207-4/18/$31.00 2018 IEEE

occlusion, etc., were unavailable.

In this paper we present a robust real-time system for Bangla license plate detection and recognition for Dhaka Metropolitan city. The system is robust as we have tested it in a dataset containing images that were collected by us from different camera setup and environmental conditions. Our detection system is real-time as we have used the most latest object detection algorithm namely YOLOv3.

In our proposed novel architecture, there are three stages: first we have used YOLOv3 to detect or localize license plate. In the next stage, YOLOv3 has been employed for two purposes: to recognize six digits at the bottom row of the license plate and to identify the solo Bangla character at the end of the first row. As we are only considering license plates of Dhaka Metropolitan city in this paper, we have only considered detecting and recognizing the single character at the end of the first row, while ignoring the rest. In the last stage, to recognize Bangla character, we have designed a ResNet-20 [21] based CNN network. We have also made a Bangla scene character dataset containing $6,400$ images to train the network.

The prime contributions of our paper are as follows:
1) We have collected a dataset containing $1,500$ images which were taken in many different real-world conditions and from different viewing angles. The dataset will be made publicly available.
2) To the best of our knowledge, we have employed YOLOv3, a real-time state-of-the-art object detection algorithm, for the first time in Bangla license plate detection and recognition problem which achieves comparable performance with other prevalent architectures in literature.
3) We have also made a Bangla scene character dataset containing total $6,400$ images of all 32 characters and trained a ResNet-20-based CNN network with this dataset to recognize the single character present at the end of the first row in Bangla license plate.

The rest of the paper is organized as follows: section II discusses related work and section III describes our Bangladeshi vehicle license plate dataset, while section IV presents our proposed three-stage ALPR system. Performance evaluation is reported in section V. Finally, we summarize our results in section VI along with directions future work.

II. RELATED WORK

In this section we present a brief literature review describing some of the major and/or relevant works.

For Bangla license plate detection and recognition, most of the previous studies used conventional image processing techniques [1]–[4]. In [1], traditional hand-crafted features such as sobel filter, connected component analysis, and morphological operations were used for plate detection. Morphological operation for license plate detection and segment modeling technique for character recognition were also proposed by [2]. Baten et al. [3] and some other researchers [4] used template matching for character and digit recognition. Support Vector Machine (SVM) classifier was used to recognize characters in [5].

Some recent researchers working with Bangla license plate recognition employed deep learning methods [6]–[8]. In [7], comparison was made between sliding window technique and R-CNN, deriving obvious conclusion that R-CNN is better. But R-CNN is in itself slow and not suitable for using in real-time applications. Moreover, the dataset they used was taken from the camera at a single gate leaving the dataset without having any variation. Abedin et al. [8] employed CNN at the last step for recognition purpose and resorted to conventional image processing approaches to detect license plate at the beginning. Their proposed system is not useful for real-time license plate detection purpose.

Many works have been done for detecting license plates in languages other than Bangla. In some of those works, license plate detection in complex real-world conditions has also been focused [9], [12], [15], [16]. Wen et al. [9] combined traditional and machine learning approaches for Japanese license plate detection. They used a method called 'shadow removal'to handle uneven illumination in the license plate and employed SVM for character recognition. In [12], authors used some conventional approaches like color edge detection, connected component analysis and then finally recurrent convolutional neural network was used to recognize character of Chinese license plates. Xie et al. [15] used YOLO modifying it by introducing a new 'angle'parameter to the algorithm. But they only detected license plate and did not discuss anything about recognition. The same is true for [16], in which only detection phase was performed using YOLOv2.

In [10], the authors proposed an approach based on Maximally Stable Extremal Regions (MSER) and Back-Propagation Neural Network (BPNN) to extract and recognize Thai vehicle license plates. Lin et al. [11] proposed a hierarchical license plate recognition system. In the first stage, they detected vehicle using YOLOv2. After detecting the vehicles, they employed SVM to detect license plates. Thereafter, they segmented character and developed a CNN network to recognize this character. YOLO-based approach for detection and recognition of Brazilian license plate was presented in [13]. But, the dataset used by them is not that much challenging, accommodating only frontal view of a car. The authors in [14] employed very similar approach to [13] for Brazilian license plate, but they used YOLOv2 and tested their model in different dataset. Like [11], the authors in [17] also used YOLOv2 only for vehicle detection purpose, while for license plate detection MSER along with BPNN was used and for character recognition CNN was used.

In [13], [14] and [17] the authors used YOLO objected detection algorithm and considered vehicle detection before detecting the license plate. The authors in [11] also did this and rationalized this approach by arguing that it reduced false positives on license plate detection. But our approach is significantly different from the above-mentioned approaches in the sense that we dropped vehicle detection step making the process faster without deteriorating performance. As we have

Fig. 2. Sample images from our collected Bangla license plate dataset.

not used SVM like [11], we have not faced the problem of false positives in license plate detection. As such, our approach is efficient in the sense that we used YOLOv3 directly for license plate detection avoiding the need to detect the vehicle first.

During literature review, we have observed that in some countries, license plate characters are in English such as Brazilian license plates, making it easier to work with. Working with license plates in other languages is difficult but Bangla is even more challenging because this language has many complexities such as stroke over the letters, close similarity among some letters etc.

III. Dataset

We have collected $1,500$ images of different types of vehicles that are commonly found in Dhaka Metropolitan City. Still images were obtained from different places of Dhaka through camera and mobile phones. While capturing images we have considered different conditions such as variation in lighting condition and viewing angle, presence of multiple license plates in the same image. We have also captured different types of vehicles including motorcycle, bus, three wheeler etc. Some characteristics of the dataset are as follows:

- In order to ensure having variation in lighting conditions, we have collected license plate images during daytime, at twilight and at night. We have also collected images from places where high lighting variation is present.
- Different types of vehicles in Dhaka have slight differences in license plate format in terms of color, shape and layout. In order to capture this variability, different types of vehicle images have been taken. Major vehicles that we have covered are car, motorcycle, bus, truck, covered van, human hauler (leguna) and three-wheeler (auto rickshaw).
- Many of the datasets available for different countries obtain only frontal or rear view of a vehicle [13]. We have considered license plates located both at the front as well as the rear of a vehicle.
- One common trend is to have a single license plate per image. But this is not very suitable for real-time license plate detection. Our dataset contains multiple license plates in a single image which is very important for training and testing a real-time detection system.
- ALPR system based on simplistic assumption that a camera remaining static cannot handle the multitudinous possibilities that can appear in the real world scenario. In order to simulate the real-world scenario, we have captured images from different viewing angles.
- Due to multiplicity of conditions, the captured images might be poor and the detection system should be robust enough to detect those also. Our dataset contains blurry and poor quality images captured by mobile cameras. There is occlusion in some of the images also.

Fig. 3. Sample images of our Bangla scene character dataset.

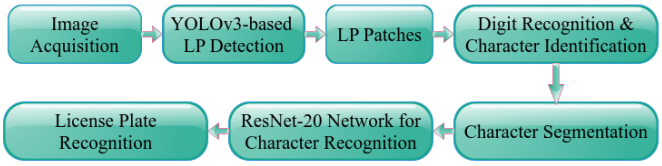

Fig. 4. Pipeline of proposed three-stage network for Bangla license plate recognition. In the first stage, YOLOv3 was used to detect and localize license plates in the collected dataset. In the second stage, all Bangla digits were recognized and existing solo Bangla character was localized using the same algorithm. In the final stage, a ResNet-20 based CNN network was used to recognize the segmented character.

For 32 characters assigned by BRTA signifying the type of a vehicle, we also made a scene character dataset containing more than $6,400$ images, i.e., roughly 200 images per those 32 characters. Characters were basically collected from Google street view, sign boards, banners, etc., and from ISI Bangla Scene Character Database (version 2014) [20] provided by Indian Statistical Institute (ISI), which contains $12,623$ different characters. Some sample characters are shown in Fig. 3.

IV. Proposed Method

Our proposed approach is divided into three stages: license plate detection and localization, digit recognition and character identification, and character recognition. Fig. 4 shows the pipeline of the proposed method.

In the first stage, we used YOLOv3 network to detect and localize license plate. After that we cropped the license plate. In the next stage, we employed YOLOv3 by which we detected and classified 10 Bangla digits and a Bangla character which is located at the end of the first line in the license plate. Finally, we made a Bangla scene character dataset containing $6,400$ images. Then a novel ResNet-20-based CNN network, trained on this dataset, was developed to recognize the segmented character.

A. Number Plate Detection and Localization

Vehicle license plate detection and localization is a crucial task on ALPR system. Some of the earlier researchers had detected vehicle before detecting the license plate. But in our case due to multiplicity of vehicle type and difference in location of license plate in the vehicle, we considered vehicle detection to be of no use. As such we went on detecting the license plate right away.

To detect license plate, we used YOLOv3 algorithm [18]. YOLOv3 is the latest version of YOLO (You Only Learn Once) algorithm [19], which makes several design changes to make it better. It has 53 convolutional layers. Previous YOLO was not that much good in detecting small objects, but because of multi-scale predictions YOLOv3 shows better performance.

We started with $1,500$ images and after data cleaning we dropped some of the subpar images. 80% of those images were used for training. For marking bounding boxes, we used a tool named Yolo_mark. Before training, we have customized YOLOv3 in order to make it suitable for Bangla license plate detection. The batch size was set to 64. Subdivisions were set to 16 for faster training. In the license plate detection phase, we had only one class and as such we took filter size equal to 18. Pre-trained weights for the darknet-53 convolutional layers were used as initial weights for training. The model was iterated 2000 times.

B. Digit Recognition and Character Identification

Using bounding box coordinate provided by YOLOv3 network, we segmented our license plates. At this stage, we performed online data augmentation by random rotation, contrast modification, adding Gaussian noise and blurriness, increasing our images four times.

As we know Bangla license plate has 6 Bangla digits (ranging from 0 to 9) and at the end of the first line there is a character that indicates the type of the vehicle, in this phase our approach was to recognize all six digits at the bottom row and detect the character which will be recognized in the next phase. We did not recognize character in this stage as we did not have enough license plate images to recognize 32 characters individually.

In this stage, we designed another custom YOLOv3 network in order to detect 11 classes (10 digits and 1 character). As there were 11 classes, filter number was set to 48.

80% images were used for training and the rest were used for validation. As our image resolution was good, we did not need to change the input size in the network as done by [13].

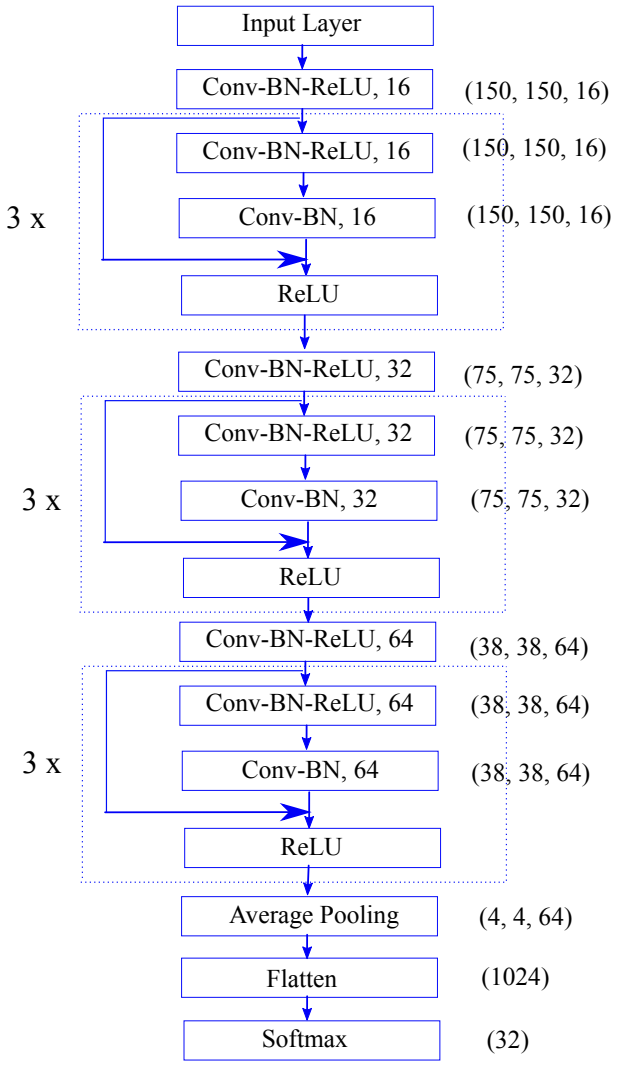

Fig. 5. Proposed ResNet-20 network architecture for Bangla character recognition. At the beginning of each three stage, the feature map is down-sampled by a convolutional layer with strides 2, while the total number of filters is doubled.

Instead, we used gray-scale padding to make it sure that we are feeding fixed size image to the network. Keeping aspect ratio unchanged, we resized every image to 540 x 540 by means of padding. According to the rule of thumb, there should be $2,000$ iterations per class, but after $6,000$ iterations we stopped as we noticed that the model was getting overfitted.

C. Character Recognition

We developed a CNN network based on ResNet-20 architecture for character recognition. Fig. 5 shows the basic building block of our ResNet-20 architecture. We trained the network from scratch to classify 32 Bangla characters available in the license plate using our Bangla scene character dataset. As the performance of deep neural network is strongly correlated with the amount of available training data, we needed to augment our training data to obtain a model which is stable and does not get overfitted. We heavily augmented our data

TABLE I
HYPERPARAMETERS SETTING OF PROPOSED CHARACTER RECOGNITION NETWORK

Hyperparameter	Value
Optimizer	Adam
Epochs	250
Multiple Learning Rates	10^{-3} (for 80 epochs)
	10^{-4} (for next 40 epochs)
	10^{-5} (for next 40 epochs)
	10^{-6} (for next 110 epochs)
Batch size	32

by random rotation, contrast modification, warping, adding Gaussian noise and blurriness. All images were resized to 150 x150, keeping aspect-ratio, by adding gray-scale padding.

Our ResNet-20 based CNN network was trained using Adam optimizer with cross entropy loss function for 250 epochs. We tried several learning rates starting with 10^{-3} and reduced it gradually. Table I summarizes all the hyperparameter settings of our proposed network. It was found that our proposed network was getting overfitted, due to high learning capacity of the model. Hence, we added batch-normalization after every single convolutional layer except after last ReLU layer, to reduce overfitting and boost our model performance.

V. PERFORMANCE EVALUATION

To measure the efficacy of the proposed ALPR system, we performed experiments on our collected license plate dataset. We trained YOLOv3 using 80% of our dataset images for 2000 iterations. The moving average loss was found to be 0.09 in the validation with IoU more than 85%. As per many researchers, license plate identified with IoU $\geq 50\%$ is considered to be a correct detection instance. From that perspective, our proposed method outperforms other state-of-the-art methods achieving total detection accuracy more than 95%.

Fig. 6. Surprisingly successful instances where proposed method correctly detects and localizes images with (a) multiple license plates, (b) varying lighting conditions, (c) high tilt and (d) occlusion

Fig. 6 illustrates that the method has achieved astonishing performance in detecting license plates in highly challenging instances, for example when there were multiple license plates (a) varying lighting conditions, (b) high tilt, (c) and occultation, (d) which justified the effectiveness of our method.

In digit recognition and character identification phase, we used 400 images for validation. The moving average loss for validation was found to be 0.72 and IoU was more than 85% per class. In character recognition, we have achieved 92.7% accuracy on 20% test data from our collected scene character dataset.

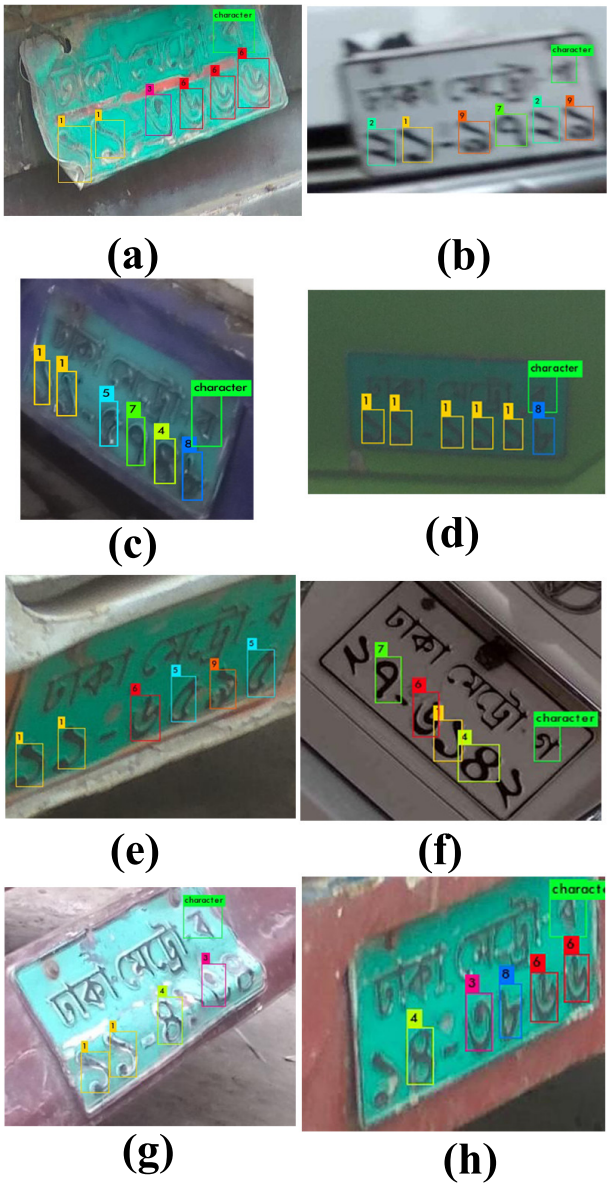

Fig. 7. Some of the predicted samples of digit recognition and character identification of our proposed method. He it is seen that in (a)-(d), tilt, rotation, blurriness, etc. are not restricting the system to recognize digits and identify characters.

Fig. 7 shows the output performance of the proposed digit recognition and character identification phase. Our proposed method correctly recognizes the digits in (a)-(d), while the method failed to recognize some of the digits due to the high slant in viewing angle, and in some cases transparency of the digits and characters.

We evaluated the whole pipeline of our proposed system in

Fig. 8. License plate recognition output of our proposed three-stage network. Proposed method correctly classifies digits and characters in Bangla license plates with higher accuracy in different conditions with a few failure instances marked with red.

video data. Our YOLO-based license plate detection and digit recognition stage yielded real-time output. Fig. 8 illustrates the output of our proposed three-stage network for ALPR. Our method correctly recognizes the characters and digits on images from our collected dataset, with a few failure instances. The failures were seen in the images with partially invisible digits or characters.

VI. Conclusion

In this paper, an accurate, highly efficient, and robust approach is proposed to localize and recognize Bangla license plates in Dhaka Metropolitan city. A challenging dataset of Bangla license plate is also introduced. Our proposed method was evaluated on this dataset and the first two stages of our novel three-stage architecture achieved IoU more than 85% for each class, and our ResNet-20-based model achieved 92.7% accuracy for character recognition. The ALPR architecture we have proposed is highly efficient in terms of its ability to detect and recognize extremely difficult-to-recognize license plate samples. We will, in future, intend to add more images to our present dataset from different districts of Bangladesh to develop a state-of-the-art real-time license plate recognition system for the entire country.

References

[1] A. K. Ghosh, S. K. Sharma, M. N. Islam, S. Biswas, and S. Akter, "Automatic license plate recognition (ALPR) for Bangladeshi vehicles," Global Journal of Computer Science and Technology, December 2011.

[2] N. A. Siddique, A. Iqbal, F. Mahmud, M. S. Rahman, "Development of an automatic vehicle license plate detection and recognition system for Bangladesh," International Conference on Informatics, Electronics & Vision (ICIEV), 2012.

[3] R. A. Baten, Z. Omair, and U. Sikder, "Bangla license plate reader for metropolitan cities of Bangladesh using template matching," 8th International Conference on Electrical and Computer Engineering, 2014.

[4] A. C. Roy, M. K. Hossen, and D. Nag, "License plate detection and character recognition system for commercial vehicles based on morphological approach and template matching," 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2016.

[5] M. A. Uddin, J. B. Joolee, and S. A. Chowdhury, "Bangladeshi vehicle digital license plate recognition for metropolitan cities using support vector machine," International Conference on Advanced Information and Communication Technology, 2016.

[6] M. M. A. Joarder, K. Mahmud, T. Ahmed, M. Kawser, and B. Ahamed, "Bangla automatic number plate recognition system using artificial neural network," Asian Transactions on Science & Technology (ATST), 2(1), 2012, pp. 1-10.

[7] M. Pias, A. K. Mutasim, and M. A. Amin, "Bangladeshi Number Plate Detection: Cascade Learning vs. Deep Learning," 15th International Workshop on Content-Based Multimedia Indexing, June 2017.

[8] M. Z. Abedin, A. C. Nath, P. Dhar, k. Deb,and M. S. Hossain, "License plate recognition system based on contour properties and deep learning model," IEEE Region 10 Humanitarian Technology Conference (R10-HTC), December 2017.

[9] Y. Wen, Y. Lu, and J. Yan, "An algorithm for license plate recognition applied to intelligent transportation system," IEEE Transactions on Intelligent Transportation Systems, vol. 12, pp. 830-845, March 2011.

[10] T. Hong, and A. K. Gopalakrishnam, "License plate extraction and recognition of a Thai vehicle based on MSER and BPNN," 7th International Conference on Knowledge and Smart Technology (KST), pp. 48-53, 2015.

[11] C. H. Lin, Y. S. Lin, and W. C. Liu, "An efficient license plate recognition system using convolution neural networks," 2018 IEEE International Conference on Applied System Invention (ICASI), pp. 224-227, 2018.

[12] Y. Liu, H. Huang, J. Cao, and T. Huang, "Convolutional neural networks-based intelligent recognition of Chinese license plates," Soft Computing, vol. 22, no. 7, pp. 2403-2419, 2018.

[13] S. Montazzolli, and C. Jung, "Real-time brazilian license plate detection and recognition using deep convolutional neural networks," 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 55-62, 2017.

[14] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R.Gonalves, et al., "A Robust Real-Time Automatic License Plate Recognition based on the YOLO Detector," arXiv preprint arXiv:1802.09567.

[15] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A New CNN-Based Method for Multi-Directional Car License Plate Detection," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 2, pp. 507-517, 2018.

[16] G. Hsu, A. Ambikapathi, S. Chung, and C. Su. "Robust license plate detection in the wild," 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6, 2017.

[17] M. Chen, W. Wang, S. Dong, and X. Zhou, "Video Vehicle Detection and Recognition Based on MapReduce and Convolutional Neural Network," International Conference on Sensing and Imaging, pp. 552-562. March 2016.

[18] J. Redmon, and A. Farhadi. "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, April 2018.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.

[20] S. Tian, U. Bhattacharya, S. Lu, B. Su, Q. Wang, X. Wei, Y. Lu and C. L. Tan, "Multilingual scene character recognition with co-occurrence of histogram of oriented gradients," Pattern Recognition, vol. 51, pp.125-134, March 2016.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.