

Triple Detector based on Feature Pyramid Network for License Plate Detection and Recognition System in Unusual Conditions

1stDuy-Linh Nguyen, 2ndMuhamad Dwisnanto Putro, 3rdXuan-Thuy Vo and 4thKang-Hyun Jo
Department of Electrical, Electronic and Computer Engineering, University of Ulsan,
Ulsan, Korea

Email: ndlinh301@mail.ulsan.ac.kr; dputro@mail.ulsan.ac.kr; xthuy@islab.ulsan.ac.kr; acejo@ulsan.ac.kr

Abstract—Vehicle license plate detection and recognition is one application of computer vision which was widely deployed in traffic monitoring, road toll, and parking lot system. Its benefits are reducing service labor and increase processing accuracy. There are many studies in this area such as robotics and machine learning. With the strong development of neural networks and deep learning, application deployment becomes easier and more accurate. This paper focuses on the development of a license plate detector that supports the license plate detection and recognition system in unusual conditions under the influence of light, weather, and camera placement. The license plate detector is built based on the backbone, Feature Pyramid Network (FPN), and triple detector. The network was trained and tested on the CCPD (Chinese City Parking Dataset) dataset and achieved 96.1% of AP (Average Precision) in total, the network can outperform the state-of-the-art detection network in this field. On the other hand, the license plate detection network combined with the LPRNet network (License Plate Recognition via Deep Neural Networks) embedded by the Spatial Transformer to recognize the numbers and characters in the license plate with 98.8% of AP on CCPD-Base subset. As a result, the number plate detector exhibits superior capabilities compared to previous methodologies under unusual conditions, thereby enhancing the overall processing visibility of the system.

Index Terms—Convolutional neural network, License plate detection, License plate recognition, Feature Pyramid Network.

I. INTRODUCTION

As the traffic network grows day by day, the application of automatic tools in traffic monitoring and control becomes essential. Several decades ago, researchers have concerned with testing and deploying intelligent applications, including the system to detect and recognize license plates. However, the biggest obstacle for these systems is the decrease in the accuracy of the system when the resulting license plate is distorted or blurred by various factors. Many studies have employed on processing license plate detection and recognition in different countries in the world. However, these studies have only solved the general problem [1]–[3] or performed on limited datasets [4]–[6], therefore reducing the system processing ability when encountering unusual cases.

In order to build the license plate detection and recognition system, the methodologies are mainly divided into one, two, or three stages. The one-stage approach [7], [8] performs detection and recognition tasks in an entire network. In the two-stage [1]–[3] approach, the system goes through

the license plate detection process, and then recognizing the characters and numbers in the license plate. With a three-stage approach [9], [10], numbers and characters segmentation is performed between detection and recognition tasks. Each approach will have its own advantages and disadvantages on each different dataset. In which, license plate detection task plays an important role in determining the success of the whole system. From that observation, this paper focuses on the design of the license plate detection network applied in a two-stage approach that can well filter the number plate area in unusual conditions under the influence of the environment and camera placement. In addition, this paper also proposes a method to embed the Spatial Transformer after the license plate detection process to obtain the most recognizable area for the next task. The main contributions of this paper are shown as follows:

1 - Proposed network license plate detection which includes the backbone, Feature Pyramid Network (FPN), and triple detector.

2 - Embedded the Spatial Transformer into the license plate recognition network to increase the accuracy of number and character recognition.

The remainder of the paper is organized as follows: Section II presents approaches related to license plate detection. Section III discusses the proposed method in detail. Section IV describes and analyzes the achieved results. Section V is the conclusion and future development direction of the paper.

II. RELATED WORK

In this section, the paper will introduce the methodologies related to license plate detection as well as the pros and cons of each method group. They can be subdivided into traditional methodologies and deep learning-based methodologies

A. The traditional methodologies

Traditional methodologies consider the difference in properties between the background and foreground subjects to extract the license plate area. These features can be obtained by using the gradient [11], Local Binary Pattern and Histogram Matching information [12] or by filtering the background color based on the HSI color system [13]. Besides, some traditional spatial transformation algorithms, classifier and pattern matching are also used such as wavelet transform [14], AdaBoost classifier

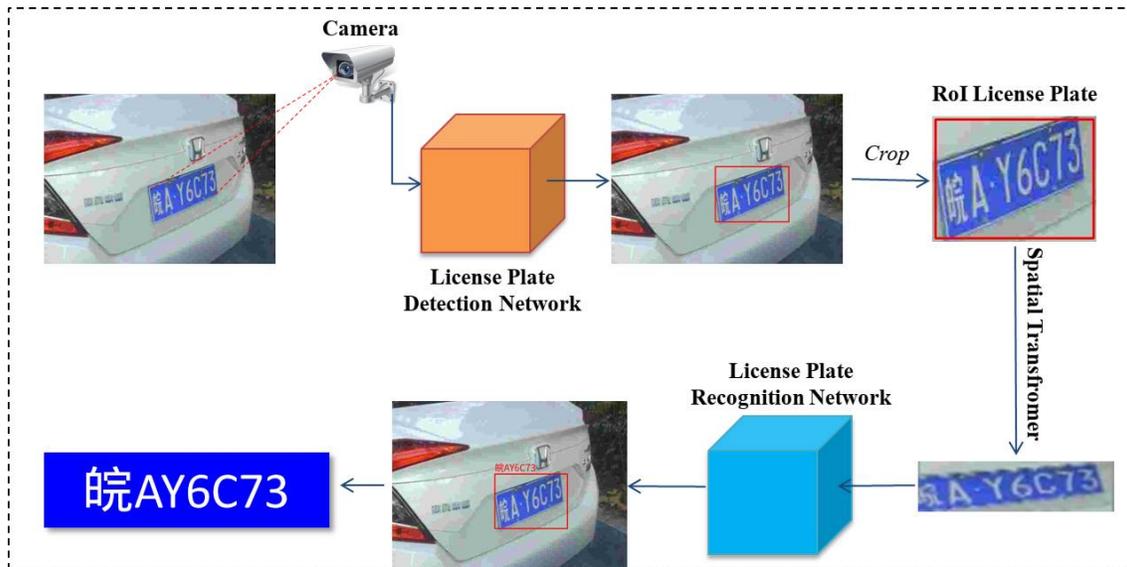


Fig. 1. The overview of two-stage license plate detection and recognition system

[15] and Local Structure Patterns [16]. The strength of these methodologies is that it is possible to extract the license plate characteristics quickly and work with low-profile hardware, but the accuracy is not high.

B. Deep learning-based methodologies

Nowadays, the strong development of deep learning and the application of convolutional neural networks on the image processing dataset has become popular. According to that trend, license plate detection was also extracted based on shallow to deep and further complex networks. Among them is the adoption of several famous network architectures such as YOLO and its variations [3], [4], Faster-RCNN [17], [18], and SSD network [19]. The outstanding advantage of these methodologies is their high speed and accuracy. However, the biggest drawback is that it requires highly configurable hardware and expensive supporting devices. In addition, several studies also combine traditional and deep learning-based methodologies to increase accuracy, processing speed and reduce operating costs [10], [20].

III. METHODOLOGY

The overview of license plate detection and recognition can be described in detail, as shown in Fig. 1. The system divided into two main networks is license plate detection and license plate recognition network. In the license plate detection network, this paper uses benefit of ResNet backbone, FPN and triple detector with depth-wise separable convolutional neural network to extract the Region of Interest (RoI) of the license plate and then crop this area. This cropped area goes to the Spatial Transformer module to reshape and recognize numbers and letters correctly. This section focuses on a detailed description of the proposed license plate detection network architecture.

A. Licence plate detection network

The backbone network used in this case is ResNet50 [21]. This network works in the bottom-up pathway to extract feature maps with multiple levels. As shown in Table I, the backbone uses the first five convolution blocks in the ResNet50 network, each convolution block containing several convolution operations with different numbers of kernels, kernel sizes and skip connection. After each convolution block, the spatial dimensions will be reduced by half. The output of each convolution block is labeled by CoR_i (Convolution of ResNet) with $i = 1$ to 5. In order to reduce computation cost, the bottom-up pathway selects only the last three outputs corresponding to CoR_3 to CoR_5 for the next handler in the top-down pathway.

TABLE I
RESNET50 ARCHITECTURE.

Layer name	Kernel size	Filter	Iter	Feature map
Input layer				$800 \times 800 \times 3$
Conv1(CoR_1)	$7 \times 7 \times 3$	64	$\times 1$	$400 \times 400 \times 64$
Max Pooling	3×3	1	$\times 1$	$200 \times 200 \times 64$
Conv2(CoR_2)	$1 \times 1 \times 64$	64	$\times 3$	$200 \times 200 \times 64$
	$3 \times 3 \times 64$	64		$200 \times 200 \times 64$
	$1 \times 1 \times 64$	256		$200 \times 200 \times 256$
Conv3(CoR_3)	$1 \times 1 \times 256$	128	$\times 4$	$100 \times 100 \times 128$
	$3 \times 3 \times 128$	128		$100 \times 100 \times 128$
	$1 \times 1 \times 128$	512		$100 \times 100 \times 512$
Conv4(CoR_4)	$1 \times 1 \times 512$	256	$\times 6$	$50 \times 50 \times 256$
	$3 \times 3 \times 256$	256		$50 \times 50 \times 256$
	$1 \times 1 \times 256$	1024		$50 \times 50 \times 1024$
Conv5(CoR_5)	$1 \times 1 \times 1024$	512	$\times 3$	$25 \times 25 \times 512$
	$3 \times 3 \times 512$	512		$25 \times 25 \times 512$
	$1 \times 1 \times 512$	2048		$25 \times 25 \times 2048$

Inspired by the advantages of capturing information from different layers, the FPN [22] network was used in the top-down pathway. The FPN network uses a simple approach to

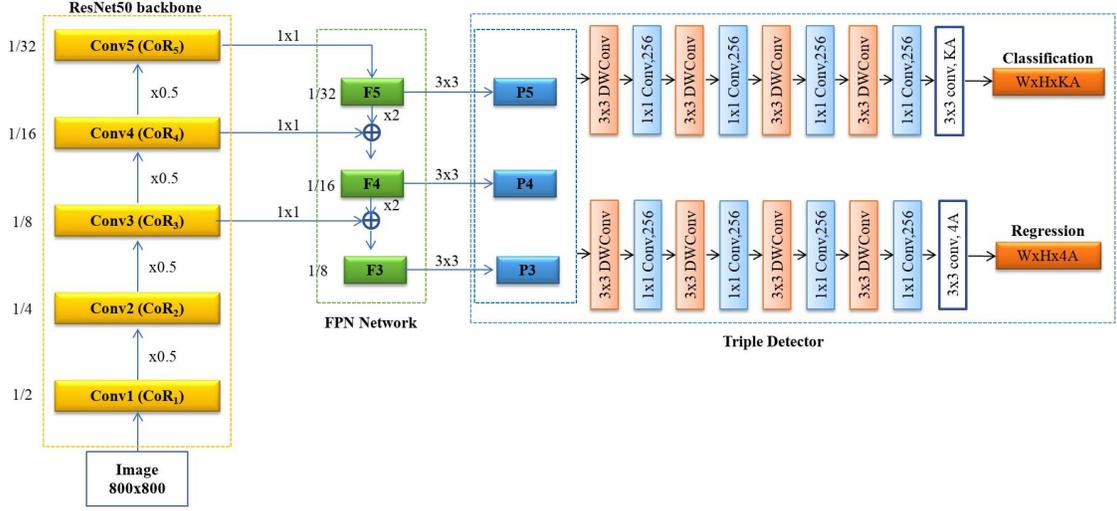


Fig. 2. The proposed license plate detection network. It consist of ResNet50 backbone, Feature Pyramid Network (FPN), and triple detector.

merging feature maps thus enriching the information and the multi-scale features can be obtained. First, the FPN network up-samples the spatial dimension of the input feature map using the nearest neighbor operation and then merged it with the corresponding feature map in the bottom-up pathway which was undergone convolution with a kernel size of 1×1 by the element-wise summation operation. This process is repeated for all feature maps in the bottom-up pathway. Specifically as shown in Fig. 2, a convolution operation with a kernel size of 1×1 was used to reduce the channel depth of the CoR_5 feature map from 2048 to 256 to create the $F5$ feature map. From $F5$ divided into two branches for processing. The first branch performs a 3×3 convolution to obtain $P5$, which is the first feature map for license plate detection. The second branch is up-sampled using the nearest neighbor up-sample operation, then performs the element-wise summation operation with the corresponding feature map from CoR_4 which has been reduced channel depth with 1×1 convolution operation. Then, a 3×3 convolution operation was applied to generate the second feature map for the license plate detection, annotated by $P4$. This same process is repeated to obtain the final feature map $P3$. The obtained feature maps of $P3$, $P4$, and $P5$ are processed by three detectors respectively.

Each detector uses two subnets for classification and bounding box regression. The classification subnet predicts the probability of a license plate appearing in each spatial location. This network takes the feature map with C channels from a pyramid level ($P3$, $P4$, $P5$) and applies four depth-wise separable convolutions [23] with 256 channels followed by the ReLU activation function. Finally, a 3×3 convolution operation is used with the Sigmoid activation function that outputs with KA prediction (A is the number of anchors, K is the number of classes, where $K = 2$, $A = 9$). To perform classification task, Focal loss [24] is used to calculate loss. The architecture of the box regression subnet is similar to the

classification subnet. Its output is the location of the license plate corresponding to the anchor box when that license plate exists. This network generates a parameter $4A$ where 4 is the coordinate of the center point (x, y) and the dimension ($width, height$) of the bounding box, A is the number of anchors, where $A = 9$. The Smooth $L1$ loss is applied to calculate the loss in this subnet.

B. Experimental setup

C. Loss function

The loss function of the license plate detection network is similar to RetinaNet [24], the Focal loss to compute loss for classification and the smooth $L1$ loss for bounding box regression task. The loss function is defined as below:

$$L(c_i, g_i) = \sum_i L_{cls}(c_i, c_i^*) + \sum_i L_{reg}(g_i, g_i^*), \quad (1)$$

where regression loss $L_{reg}(g_i, g_i^*) = R(g_i - g_i^*)$ with R is the Smooth $L1$ loss shown as in Eq. (2), $L_{cls}(c_i, c_i^*)$ is the classification loss using the Focal loss shown as in Eq. (3), g_i is the center coordinates and dimension (width and height) of the prediction and g_i^* is the ground truth coordinates of bounding box, c_i is the predicted classes and c_i^* is the ground truth label.

$$R(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (2)$$

$$FL(c_i) = -\alpha_i(1 - c_i)^\gamma \log(c_i), \quad (3)$$

where α is the balancing parameter and γ is the focusing parameter both are useful for addressing the class imbalance issue.



Fig. 3. The qualitative result of license plate detection network on CCPD dataset.

TABLE II
DESCRIPTIONS OF THE SUBSETS IN CCPD DATASET.

Subset	Description
CCPD-Base	The only common feature of these photos is the inclusion of a license plate.
CCPD-DB	Illuminations on the LP area are dark, uneven or extremely bright.
CCPD-FN	The distance from the LP to the shooting location is relatively far or near.
CCPD-Rotate	Great horizontal tilt degree (20° - 50°) and the vertical tilt degree varies from -10° to 10° .
CCPD-Tilt	Great horizontal tilt degree (15° - 45°) and vertical tilt degree (15° - 45°).
CCPD-Blur	Blurry largely due to hand jitter while taking pictures.
CCPD-Weather	Images taken on a rainy day, snow day or fog day.
CCPD-Challenge	The most challenging images for license plate detection and recognition to date.

IV. EXPERIMENTAL RESULTS

A. Dataset

During training the license plate detection network, the Chinese City Parking Dataset (CCPD) [30] was used. This is the largest dataset in the field of license plate detection and recognition. CCPD contains 250K unique license plate images collected from a parking management company in a capital province in China. The image resolution is 720 (width) \times 1160 (height) with three RGB colour channels. Each image has only one license plate. Each license plate contains a Chinese character, a Latin letter and five Latin letters or numbers. The CCPD is labelled based on license plate bounding box, four vertices locations, horizontal tilt degree and vertical tilt degree, and other information. CCPD provides one training dataset and 9 subsets for the testing phase (CCPD-Base is divided for both training and test) as described in Table II [30].

B. Experimental results

The experiment in this article is trained and tested on a Titan RTX GPU with 24GB graphics memory. The network is implemented using Pytorch and CUDA environment on Ubuntu 18.04 operating system. When training the license

plate detection network, this work uses 100K images from CCPD-Base (This subset is divided for the training and testing phase) and evaluates with eight subsets except for the CCPD-ND subset. The network is trained based on the training settings in RetinaNet [24] with several parameter configurations is used. The Stochastic Gradient Descent optimization method is employed, the batch size of 16, the weight decay is 10^{-4} , the momentum is 0.9, the initial learning rate of 0.01 and then divided by 10 at 60k and again at 80k iterations, and the number of epoch is 30. For the loss calculation phase, α and γ are assigned by 0.25 and 2.0 respectively. The threshold IoU (Intersection over Union) equal to 0.5 to generate the best bounding boxes for testing.

The license plate detection network is trained and evaluated with CCPD dataset as described above. CCPD-Base dataset was randomly selected 100K images for training and the remaining 99,996 images were used for evaluation. In addition, the network was also evaluated with six other data subsets including CCPD-Blur (20611 images), Challenge (50003 images), CCPD-DB (10132 images), CCPD-FN (20967 images), CCPD -Rotate (10053 images), CCPD-Tilt (30216 images), and CCPD-Weather (9999 images). The evaluated results of

TABLE III

THE COMPARISON RESULT OF THE LICENSE PLATE DETECTION NETWORK WITH STATE-OF-THE-ART DETECTION MODELS ON EACH CCPD SUBSET.

Network	AP(%)	Base	DB	FN	Rotate	Tilt	Weather	Challenge
Cascade classifier [25]	47.2	55.4	49.2	52.7	0.4	0.6	51.5	27.5
Faster-RCNN [26]	92.2	98.1	92.1	83.7	91.8	89.4	81.8	83.9
YOLO9000 [27]	93.1	98.8	89.6	77.3	93.3	91.8	84.2	88.6
TE2E [28]	94.2	98.5	91.7	83.8	95.1	94.5	83.6	93.1
SSD300 [29]	94.4	99.1	89.2	84.7	95.6	94.9	83.4	93.1
RPnet [30]	94.5	99.3	89.5	85.3	94.7	93.2	84.1	92.8
Proposed network	96.1	97.5	95.6	89.2	97.2	97.0	97.7	98.6

TABLE IV

THE COMPARISON RESULT OF THE LICENSE PLATE RECOGNITION NETWORK WITH STATE-OF-THE-ART DETECTION MODELS ON EACH CCPD SUBSET. HC DESCRIBE FOR HOLISTIC-CNN REFER FROM [30]

Network	ACC (%)	Base	DB	FN	Rotate	Tilt	Weather	Challenge
Cascade classifier + HC	58.9	69.7	67.2	69.7	0.1	3.1	52.3	27.5
Faster-RCNN + HC	92.8	97.2	94.4	90.9	82.9	87.3	85.5	76.3
YOLO9000 + HC	93.7	98.1	96.0	88.2	84.5	88.5	87.0	80.5
TE2E	94.4	97.8	94.8	94.5	87.9	92.1	86.8	81.2
SSD300 + HC	95.2	98.3	96.6	95.9	88.4	91.5	87.3	83.8
RPnet	95.5	98.5	96.6	94.3	90.8	92.5	87.9	85.1
Proposed system	79.7	98.8	52.8	70.7	79.6	80.1	95.8	80.2



Fig. 4. The qualitative result of license plate detection and recognition system on CCPD dataset.

the license plate detection network are compared with the state-of-the-art detection networks and shown in Table III. This result shows that, when tested on the CCPD-Base subset with the common characteristics of the license plate, the detection accuracy is 97.5% of the AP. Compared with the best network (RPnet), the results were not much different. In contrast, when evaluating subsets with other unusual conditions such as tilt degree (CCPD-Rotate and CCPD-Tilt), illumination (CCPD-DB), blurry (CCPD-Blur), (CCPD-FN), and conditions caused by weather and other factors (CCPD-Weather and CCPD-

Challenge), the proposed network outperforms all other networks. Qualitative results on the image of the different CCPD subsets are shown in Fig. 3. The results prove that the license plate detection network in this paper can work very well under unusual conditions to detect the license plate most accurately.

In addition, this work also implements the embedding of the Spatial Transformer [31] module into the LPRNet [32] network to increase the recognition of characters and numbers on the license plate with near-standard license plate conditions. Finally, the entire license plate detection and recognition

system built based on the proposed detection network is combined with the LPRNet network. The system is also evaluated on the CCDP dataset and the real images were collected from several locations in China with multiple vehicles. The qualitative results are shown in Table IV and Fig. 4. The results showed that the recognition network performed well in the general characteristics of the license plate and under the influence of the weather with the accuracy of 98.8% and 95.8% respectively. This is the basis for improving systems with deeper recognition networks that can capture the unusual properties of letters and numbers on the license plates. On the other hand, it is also the basis for building a dataset to evaluate the unusual conditions of letters and numbers.

V. CONCLUSION AND FUTURE WORK

Because of the importance of extracting number plate area in two-stage detection and recognition system, this paper proposed the license plate detection network based on the advantage of the FPN network with the triple detector. The license plate detection network can generate license plate area with high precision under common and especially conditions. The proposed network can be used in the first stage of the two-stage or three-stage license plate detection and recognition system. The evaluating results serve as the foundation to improve license plate detection and recognition networks, and build the dataset for unusual cases. In the future, this work will continuously develop with the attention mechanisms to increase the ability to focus and extract the license plate area in the image. At the same time, researching and building the dataset for other unusual changing conditions of the license plate area.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the government(MSIT).(No.2020R1A2C200897212)

REFERENCES

- [1] S. Z. Masood, G. Shu, A. Dehghan, and E. G. Ortiz, "License plate detection and recognition using deeply learned convolutional neural networks," *CoRR*, vol. abs/1703.07330, 2017.
- [2] Z. Selmi, M. Ben Halima, and A. M. Alimi, "Deep learning system for automatic license plate detection and recognition," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, pp. 1132–1138, 2017.
- [3] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Gonçalves, W. R. Schwartz, and D. Menotti, "A robust real-time automatic license plate recognition based on the yolo detector," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–10, 2018.
- [4] S. Montazzoli and C. Jung, "License plate detection and recognition in unconstrained scenarios," 09 2018.
- [5] S. Montazzoli and C. Jung, "Real-time brazilian license plate detection and recognition using deep convolutional neural networks," 10 2017.
- [6] J. Zhuang, S. Hou, Z. Wang, and Z.-J. Zha, "Towards human-level license plate recognition," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.
- [7] Q. Huang, Z. Cai, and T. Lan, "A single neural network for mixed style license plate detection and recognition," *IEEE Access*, vol. 9, pp. 21777–21785, 2021.
- [8] Z. Xu, W. Yang, A. Meng, N. Lu, H. Huang, C. Ying, and L. Huang, "Towards end-to-end license plate detection and recognition: A large dataset and baseline," in *ECCV*, 2018.

- [9] C. Henry, S. Y. Ahn, and S. W. Lee, "Multinational license plate recognition using generalized character sequence detection," *IEEE Access*, vol. 8, pp. 35185–35199, 2020.
- [10] I. V. Pustokhina, D. A. Pustokhin, J. J. P. C. Rodrigues, D. Gupta, A. Khanna, K. Shankar, C. Seo, and G. P. Joshi, "Automatic vehicle license plate recognition using optimal k-means with convolutional neural network for intelligent transportation systems," *IEEE Access*, vol. 8, pp. 92907–92917, 2020.
- [11] R. Wang, N. Sang, R. Huang, and Y. Wang, "License plate detection using gradient information and cascade detectors," *Optik*, vol. 125, no. 1, pp. 186–190, 2014.
- [12] A. K. Bachchan, A. Gorai, and P. Gupta, "Automatic license plate recognition using local binary pattern and histogram matching," in *ICIC*, 2017.
- [13] K. Deb and K. Jo, "Hsi color based vehicle license plate detection," in *2008 International Conference on Control, Automation and Systems*, pp. 687–691, 2008.
- [14] S. Yu, B. Li, Q. Zhang, C. Liu, and M. Q.-H. Meng, "A novel license plate location method based on wavelet transform and emd analysis," *Pattern Recognition*, vol. 48, no. 1, pp. 114–125, 2015.
- [15] R. Wang, N. Sang, R. Huang, and Y. Wang, "License plate detection using gradient information and cascade detectors," *Optik*, vol. 125, no. 1, pp. 186–190, 2014.
- [16] Y. Lee, T. Song, B. Ku, S. Jeon, D. K. Han, and H. Ko, "License plate detection using local structure patterns," in *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 574–579, 2010.
- [17] D. Lee, S. Yoon, J. Lee, and D. Park, "Real-time license plate detection based on faster r-cnn," *KIPS Transactions on Software and Data Engineering*, vol. 5, pp. 511–520, 11 2016.
- [18] P. Ravirathinam and A. Patawari, "Automatic license plate recognition for indian roads using faster-rcnn," *2019 11th International Conference on Advanced Computing (ICoAC)*, pp. 275–281, 2019.
- [19] P. Wu and Y. Lin, "Research on license plate detection algorithm based on ssd," in *Proceedings of the 2nd International Conference on Advances in Image Processing, ICAIP '18*, (New York, NY, USA), p. 19–23, Association for Computing Machinery, 2018.
- [20] T. Islam and R. I. Rasel, "Real-time bangla license plate recognition system using faster r-cnn and ssd: A deep learning application," pp. 108–111, 11 2019.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.
- [22] T. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature pyramid networks for object detection," *CoRR*, vol. abs/1612.03144, 2016.
- [23] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *CoRR*, vol. abs/1610.02357, 2016.
- [24] T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *CoRR*, vol. abs/1708.02002, 2017.
- [25] S. Wang and H. Lee, "A cascade framework for a real-time statistical license plate recognition system," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 2, pp. 267–282, 2007.
- [26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [27] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, 2017.
- [28] H. Li, P. Wang, and C. Shen, "Towards end-to-end car license plates detection and recognition with deep neural networks," *CoRR*, vol. abs/1709.08828, 2017.
- [29] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," *CoRR*, vol. abs/1512.02325, 2015.
- [30] Z. Xu, W. Yang, A. Meng, N. Lu, H. Huang, C. Ying, and L. Huang, "Towards end-to-end license plate detection and recognition: A large dataset and baseline," in *ECCV*, 2018.
- [31] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," *CoRR*, vol. abs/1506.02025, 2015.
- [32] S. Zherzdev and A. Gruzdev, "Lprnet: License plate recognition via deep neural networks," *CoRR*, vol. abs/1806.10447, 2018.