Localizing License Plates in Real Time with RetinaNet Object Detector

Ritabrata Sanyal[1], Manan Jethanandani[2], Gummi Deepak Reddy[3], and Abhijit Kurtakoti[4]

[1] Kalyani Government Engineering College, West Bengal, India
sritabrata@gmail.com
[2] LNM Insitute of Information Technology, Rajasthan, India
mananjethanandani01@gmail.com
[3] Jawaharlal Nehru Technological University, Hyderabad, India
gummideepak@gmail.com
[4] Vishweshwarayya Technological University, Karnataka, India
ukabhijit@gmail.com

Abstract. Automatic License Plate Recognition systems have various applications in intelligent automated transportation systems and thus has been a frequent topic of research for the past years. Yet designing a highly accurate license plate recognition pipeline is challenging in an unconstrained environment, where difficulties arise from variations in photographic conditions like illumination, distortion, blurring etc., and license plate structural variations like background, text font, size and color across different countries. In this paper, we tackle the problem of license plate detection and propose a novel approach based on localization of the license plates with prior vehicle detection, using the state-of-the-art RetinaNet object detector. This helps us to achieve real time detection performance, whilst having superior localization accuracy compared to other state-of-the-art object detectors. Our system proved to be robust to all those variations that can occur in an unconstrained environment, and outperformed other state of the art license plate detection systems to the best of our knowledge.

Keywords: Automatic License Plate Detection, Intelligent Transportation, Deep Learning, Convolutional Neural Nets, RetinaNet

1 Introduction

Automatic License Plate Recognition (ALPR) plays a key role in many different aspects of intelligent transportation systems like automated toll collection, stolen vehicle detection, road traffic surveillance, automated parking space allotment and many others. Due to the diversity of applications of an ALPR system, this has been an active topic of research since the past decade. Despite having prolific literature, designing an ALPR system is highly challenging in an unconstrained environment where difficulties may arise from variations in photographic conditions like illumination, orientations, distortion, blurring and also backgrounds,

color, text font disparities across countries. A typical ALPR pipeline consists of a license plate (LP) detection system which entails in localizing the vehicle plates from an input image, followed by an OCR system which reads every plate localized in the previous stage. The localization efficacy of the LP detector is of paramount importance in designing a highly accurate ALPR system—hence in this paper, we propose a LP detection system with high localization accuracy and which also performs in real time. Most previous works entail in engineering handcrafted features for LP detection, and using projections, active contours, CCA etc. to segment the characters of the plate followed by a classical machine learning classifier like SVM to classify the segmented characters. These methods could not perform well in unconstrained settings, where degree of uncertainties are quite large. The recent success of deep learning especially Convolutional Neural Networks (CNN) in image classification, localization, segmentation and many other problems in computer vision, has inspired researchers to employ these techniques in ALPR tasks. Silva et al. [18] and Laroca et al. [6] used the YOLO network [15] for LP detection. YOLO, being an one stage object detector is less accurate, while a lot faster than two stage detectors like Faster RCNN. We use the state of the art RetinaNet [12] object detector for LP detection. It is a one stage detector, thus time taken by it to process a frame is almost comparable to YOLO; yet its localization accuracy is comparable or even better than two stage detectors, thus making RetinaNet the ideal choice for unconstrained LP detection, where both detection accuracy and real time performance are critical. Previously, Safie et al. [17] used RetinaNet for LP detection, but they used Resnet50 network [3] for deep feature extraction, whereas we use VGG19 [19] architecture to that end. This is because VGG19 [19] network has only 19 layers and hence is much lighter than Resnet50 which has 50 layers. Thus it takes much less training and prediction time than Resnet50. Also, their work was based on a custom dataset, having only one vehicle per frame. Our work is aimed at detecting LPs in the wild, where there may be one or more vehicles per frame. Hence we first used RetinaNet [12] to detect vehicles, and then extracted the LPs from the localized vehicles. Lastly, they used only Malaysian LPs in their study, whereas we test our method on LP datasets from multiple countries.

2 Methodologies

The proposed system constitutes of two phases. i) Vehicle detection. ii) License plate (LP) detection. These steps are elaborated as follows:

Vehicle detection is the first stage of the pipeline, which entails in localizing the vehicles from an input image. This is followed by the LP detection phase which extracts the license plate from each of the localized vehicles. Vehicles are detected prior localizing the license plates simply because vehicles, having much larger spatial dimensions than license plates, is easier to locate in a natural scene. Once the vehicles are localized, the search space is reduced drastically to aid accurate detection of license plates. To this end, we use the state of the art one stage object detector namely the RetinaNet [12], for both the detection

purposes. We especially use a one stage detector, rather than a two stage detector because : i) Two-stage detectors like Faster RCNN [16] are generally slower than one stage detectors like YOLO [15], SSD [13], yet the average localization accuracy of the former triumphs that of the latter by a large margin, due to the the extreme foreground-background class imbalance problem. ii) RetinaNet [12] solves this problem by modifying the standard cross entropy loss, such that the loss assigned to well classified samples are down weighted in every iteration. Hence, the impact of easy negative samples to the total loss is reduced, and a sparse set of hard examples contribute the most to the final loss. This novel loss function, namely the focal loss, ameliorates this speed-accuracy tradeoff, thus outperforming all other state-of-the-art one stage and two stage object detectors in terms of real time and localization performance. RetinaNet [12] has been used [2, 10, 14] in various object detection scenarios in different domains with great efficacy. RetinaNet [12] architecture consists of three components, namely a backbone network for feature extraction and two subnets for classification and bounding box regression. Feature Pyramid Network(FPN) is used as the backbone network due to its ability to represent rich multi-scale features. VGG19 [19] network has been to extract high level features maps, from which the FPN backbone generates feature maps at different scales. The classification subnet predicts the probability of presence of a LP at each spatial location of every anchor. In parallel, the regression subnet is trained, in which, a small Fully Connected Network (FCN) is attached to each pyramid level for regressing the offset from each anchor box to a nearby ground-truth object, if it exists. The entire network is trained end to end in a single stage with the focal loss objective. The loss emphasizes more on the hard examples rather than the easy samples, thus solving the class imbalance problem in other one stage detectors. Formally, the focal loss $FL(p_t)$ is expressed [12] as :

$$FL(p_t) = -\alpha(1-p_t)^{\gamma} log(p_t) \tag{1}$$

where α, γ are two hyperparameters.

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{otherwise} \end{cases} \tag{2}$$

where p is the probability predicted by the model, and $y = 1$ represents the ground truth.

3 Experiments

In this section, we conduct experiments to demonstrate the efficacy of our proposed ALPD system. All experiments are performed on NVIDIA 1080 GPU, with 8 GB of memory.

3.1 Datasets

A number of benchmark datasets have been used to evaluate the effectiveness of our proposedt LP detection method. For evaluating the LP detection perfor-

Table 1: Comparison of precision and recall metrics of different license plate detection methods on the Caltech Cars Dataset. The red and the blue highlighted cells indicate the highest and the second highest metric values respectively, achieved by any method.

Method	**Precision(%)**	**Recall(%)**
Le & Li [7]	71.40	61.60
Bai & Liu [4]	74.10	68.70
Lim & Tay [11]	83.73	90.47
Zhou et al. [23]	95.50	84.80
Li & Shen [9]	97.56	95.24
Faster RCNN	97.15	96.30
YOLO	96.67	95.88
RetinaNet	98.50	97.15

Table 2: Comparison of recall rate(%) of various license plate detection methods on the five parts of the PKU Dataset. The red and the blue highlighted cells indicate the highest and the second highest recall value respectively, achieved by any method.

Method	**G1**	**G2**	**G3**	**G4**	**G5**	**Avg**
Zheng et al. [22]	94.93	95.71	91.91	69.58	67.61	83.94
Zhao et al. [21]	95.18	95.71	95.13	69.93	68.10	84.81
Zhou et al. [23]	95.43	97.85	94.21	81.23	82.37	90.22
Li et al. [8]	98.89	98.42	95.83	81.17	83.31	91.52
Yuan et al. [20]	98.76	98.42	97.72	96.23	97.32	97.69
Faster RCNN	98.95	98.50	97.60	95.58	96.90	97.51
YOLO	98.54	98.40	97.35	96.44	97.28	97.60
RetinaNet	99.63	99.15	97.63	98.55	98.81	98.75

mance, we use 3 datasets, namely the Caltech Cars dataset, PKU dataset and the AOLP dataset.

The Caltech Cars 1999 dataset [1] has 126 images, with resolution of 896×592 pixels. The images were captured in Caltech parking lot, consisting of USA license plates in cluttered background like wall, grass, trees etc.

The second dataset is the Application Oriented License Plate (AOLP) database [5]. This dataset consists of 2049 images with Taiwan license plates. It is categorized into 3 subsets with different levels of difficulty and photographic conditions, namely : Access Control (AC) having 681 images, Traffic Law Enforcement (LE) having 757 images and Road Patrol (RP) having 611 images. The images are captured under different illumination and weather conditions and they have a lot of variations in them, like cluttered background, multiple vehicle plates in a single frame and images captured with arbitrary viewpoints and distances.

The third dataset used is the "PKUData" database [20], which consists of 3977 images with Chinese license plates, captured from various scenes and under different settings, like illumination, occlusion, degradation, multiple viewpoints, multiple vehicles etc. This dataset is divided into 5 subsets (G1-G5), corresponding to different conditions, as elaborated in [20].

Table 3: Comparison of detection performance(%) of various license plate detection methods on the three parts of the AOLP Dataset. The red and the blue highlighted cells indicate the highest and the second highest metric values respectively, achieved by any method.

Method	**AC(%)**		**LE(%)**		**RP(%)**		**Avg(%)**	
	Precision	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*
Hsu et al. [5]	91	96	91	95	91	94	91	94
Li & Shen [9]	98.53	98.38	97.75	97.62	95.28	95.58	97.18	97.19
Faster RCNN	99.20	99	98.50	98.20	95.80	96.62	97.83	97.94
YOLO	98.90	99	98	98.33	95.57	95.89	97.49	97.74
RetinaNet	99.65	99.20	98.88	98.58	96.35	96.23	98.29	98

3.2 Evaluation metrics

For evaluating our LP detection model, we use precision and recall rate alike previous methods. Recall rate is defined as the ratio of the number of correctly detected positive regions to the number of labelled positive regions. Precision rate is defined as the ratio of the number of correctly detected positive regions to the number of detected regions. A predicted bounding box is considered to be correct if the license plate is totally encompassed by it and its Intersection over Union (IoU) with the ground truth bounding box is more than or equal to

50% (that is IoU ≥ 0.5), where,

$$IoU = \frac{area(pr \cap gt)}{area(pr \cup gt)} \tag{3}$$

where pr and gt are the predicted and ground truth license plate regions respectively.

3.3 Discussion and analysis

From Table 1, we can see that RetinaNet LP detection model achieved 98.50% and 97.15% precision and recall rates respectively on the Caltech Cars dataset, which are almost $2-3\%$ better than the best previous work by Li & Shen [9]. Faster RCNN and YOLO models performed considerably worse compared to our proposed method. From Table 2, our model achieves 1% more average recall rate than the best previous work by Yuan et al. [20] on PKU dataset. The Faster RCNN and YOLO models achieve about 1% worse recall rate compared to our proposed RetinaNet model. On the AOLP dataset, our RetinaNet model achieves an average precision and recall rate of 98.29% and 98% respectively (Table 3), which is better than the best previous work [9], by a $1-2\%$ margin. From the LP detection results on these varied datasets, we can easily see that RetinaNet has better object detection performance than Faster RCNN and YOLO models by a $1-2\%$ margin. Also from Table 4, it is evident that the RetinaNet model, being a one stage detector, is about twice as fast than a two stage detector like Faster RCNN, and takes about the same time as a YOLO detector. Thus RetinaNet is almost as fast as an one stage object detector, while having superior localization accuracy compared to both one and two stage detectors. It is also empirically seen that our proposed LP detection system can process a full HD video at a rate of 50 FPS, and thus has near real time performance.

Table 4: Comparison of detection time (ms) of our proposed model with different baseline models.

	Dataset	**Method**	**Time (ms)**
Detection	AOLP	Faster RCNN	50
		YOLO	28
		RetinaNet	29
	PKUData	Faster RCNN	45
		YOLO	19
		RetinaNet	19

4 Conclusion

In this paper, we have proposed an automatic license plate detection system, using the state of the art one stage RetinaNet object detector. RetinaNet is trained

using the focal loss objective, which solves the extreme foreground-background class imbalance problem typical in other one stage object detectors. Thus it achieves better localization accuracy than other one or two stage detectors like YOLO or Faster RCNN, while having a real time detection performance. Yet it is to be noted that we had to localize the vehicles prior to LP detection, which takes up extra time. As a future work, we aim to obviate this prior vehicle detection stage without hampering LP localization accuracy.

References

1. Caltech car plates dataset. `http://www.vision.caltech.edu/html-files/archive.html`
2. Cui, Y., Oztan, B.: Automated firearms detection in cargo x-ray images using retinanet. In: Anomaly Detection and Imaging with X-Rays (ADIX) IV. vol. 10999, p. 109990P. International Society for Optics and Photonics (2019)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. computer vision and pattern recognition (cvpr). In: 2016 IEEE Conference on. vol. 5, p. 6 (2015)
4. Hongliang, B., Changping, L.: A hybrid license plate extraction method based on edge statistics and morphology. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. vol. 2, pp. 831–834. IEEE (2004)
5. Hsu, G.S., Chen, J.C., Chung, Y.Z.: Application-oriented license plate recognition. IEEE transactions on vehicular technology **62**(2), 552–561 (2012)
6. Laroca, R., Severo, E., Zanlorensi, L.A., Oliveira, L.S., Gonçalves, G.R., Schwartz, W.R., Menotti, D.: 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. IEEE (2018)
7. Le, W., Li, S.: A hybrid license plate extraction method for complex scenes. In: 18th International Conference on Pattern Recognition (ICPR'06). vol. 2, pp. 324–327. IEEE (2006)
8. Li, B., Tian, B., Li, Y., Wen, D.: Component-based license plate detection using conditional random field model. IEEE Transactions on Intelligent Transportation Systems **14**(4), 1690–1699 (2013)
9. Li, H., Shen, C.: Reading car license plates using deep convolutional neural networks and lstms. arXiv preprint arXiv:1601.05610 (2016)
10. Li, X., Zhao, H., Zhang, L.: Recurrent retinanet: A video object detection model based on focal loss. In: International Conference on Neural Information Processing. pp. 499–508. Springer (2018)
11. Lim, H.W., Tay, Y.H.: Detection of license plate characters in natural scene with mser and sift unigram classifier (11 2010). https://doi.org/10.1109/STUDENT.2010.5686998
12. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
13. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: European conference on computer vision. pp. 21–37. Springer (2016)
14. Milton, M.A.A.: Towards pedestrian detection using retinanet in eccv 2018 wider pedestrian detection challenge. arXiv preprint arXiv:1902.01031 (2019)

15. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)
16. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)
17. Safie, S., Azmi, N.M.A.N., Yusof, R., Yunus, M.R.M., Sayuti, M.F.Z.C., Fai, K.K.: Object localization and detection for real-time automatic license plate detection (alpr) system using retinanet algorithm. In: Proceedings of SAI Intelligent Systems Conference. pp. 760–768. Springer (2019)
18. Silva, S.M., Jung, C.R.: Real-time brazilian license plate detection and recognition using deep convolutional neural networks. In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 55–62. IEEE (2017)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
20. Yuan, Y., Zou, W., Zhao, Y., Wang, X., Hu, X., Komodakis, N.: A robust and efficient approach to license plate detection. IEEE Transactions on Image Processing **26**(3), 1102–1114 (2016)
21. Zhao, Y., Yuan, Y., Bai, S., Liu, K., Fang, W.: Voting-based license plate location. In: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). pp. 314–317. IEEE (2011)
22. Zheng, D., Zhao, Y., Wang, J.: An efficient method of license plate location. Pattern recognition letters **26**(15), 2431–2438 (2005)
23. Zhou, W., Li, H., Lu, Y., Tian, Q.: Principal visual word discovery for automatic license plate detection. IEEE Transactions on Image Processing **21**(9), 4269–4279 (2012)