Automatic License Plate Recognition: An Efficient and Layout-Independent Approach Based on the YOLO Detector

Rayson Laroca[1], **David Menotti**[1]

[1]Department of Informatics, Federal University of Paraná, Curitiba, Brazil

{rblsantos,menotti}@inf.ufpr.br

Abstract. *Automatic License Plate Recognition* (ALPR) *has been a frequent topic of research due to many practical applications, such as border control and traffic law enforcement. This work presents an efficient, robust and layout-independent ALPR system based on the YOLO object detector that contains a unified approach for license plate detection and layout classification and that leverages post-processing rules in the recognition stage to eliminate a major shortcoming of existing ALPR systems (being layout dependent). We also introduce a publicly available dataset for ALPR that has become very popular, having been downloaded more than 600 times by researchers from 77 different countries over the past two years. The proposed system, which performs in real time even when there are* 4 *vehicles in the scene, outperformed both previous works and commercial systems on four public datasets widely used in the literature.*

1. Introduction

A variety of Automatic License Plate Recognition (ALPR) systems and commercial products have been produced over the years due to many practical applications such as automatic toll collection, border control, traffic law enforcement, access control in private spaces, road traffic monitoring, among others. ALPR systems must be capable of recognizing multiple License Plate (LP) layouts since there might be various LP layouts in the same country or region (e.g., Brazilian and Mercosur LPs will coexist for many years). In addition, such systems should operate fast enough to fulfill the needs of Intelligent Transportation Systems (ITS). In the ALPR literature, generally, a system is considered "real-time" if it is capable of processing at least 30 Frames Per Second (FPS).

Despite the importance of having a robust ALPR system, most solutions are still not robust enough to be executed on real-world scenarios, commonly depending on certain constraints such as specific cameras or viewing angles, simple backgrounds, good lighting conditions, search in a fixed region, among others. In this context, there is still a great demand for realistic datasets with vehicle and LP annotations. The SSIG-SegPlate dataset – the best known public dataset of Brazilian LPs for ALPR – contains less than 800 training examples and has several constraints such as the use of a static camera mounted always in the same position, the images have very similar and relatively simple backgrounds, there are no motorcycles and only a few cases where the LPs are not well aligned.

This master's thesis contributes to the field of ALPR in the following ways: **(i)** a new efficient and layout-independent ALPR system using YOLO-based Convolutional

This work is a summary of a master's thesis.

Neural Networks (CNNs), which outperforms previous works and two commercial systems in four public datasets, and achieves competitive results to the baselines in another four[1]; **(ii)** a public dataset for ALPR that includes 4,500 fully annotated images acquired in real-world scenarios, assisting the development and evaluation of new approaches as well as the fair comparison among published works[2]. Compared to the SSIG-SegPlate dataset, the proposed dataset has more than twice the images and contains a larger variety in different aspects; **(iii)** annotations regarding the position of the vehicles, LPs and characters, as well as their classes, in the public datasets used in this work that have no annotations or contain labels only for part of the ALPR pipeline. Precisely, we manually labeled 38,351 bounding boxes on 6,239 images; and **(iv)** a comparative evaluation of the proposed approach with previous works in the literature and two commercial systems in **eight** publicly available datasets. We are not aware of any work in which an end-to-end ALPR system was evaluated on as many publicly available datasets as we have.

2. The UFPR-ALPR Dataset

The dataset contains $4{,}500$ images taken from inside a vehicle driving through regular traffic in an urban environment. These images were obtained from 150 videos with a duration of 1 second and frame rate of 30 FPS. Figure 1 shows the diversity of the dataset.

Figure 1. Sample images of the UFPR-ALPR dataset.

The images were acquired with three different cameras and are available in the PNG format with a size of $1{,}920 \times 1{,}080$ pixels. The cameras used were: *GoPro Hero4 Silver*, *Huawei P9 Lite* and *iPhone 7 Plus*. Images obtained with different cameras do not necessarily have the same quality, although they have the same resolution and frame rate. The dataset is split as follows: 40% of the images for training, 40% for testing and 20% for validation, as other public datasets in the literature (e.g., SSIG-SegPlate).

Every image has the following annotations available in a text file: the camera in which the image was taken, the vehicle's position, type (car or motorcycle), manufacturer,

[1] The entire ALPR system (i.e., the architectures and weights), along with all annotations made by us are ***publicly available*** at `https://web.inf.ufpr.br/vri/publications/layout-independent-alpr/`.

[2] The UFPR-ALPR dataset is ***publicly available*** to the research community at `https://web.inf.ufpr.br/vri/databases/ufpr-alpr/` subject to privacy restrictions.

model and year; the identification and position of the LP, as well as the position of each character. As Brazilian LPs have 7 characters, more than $30{,}000$ characters were labeled.

3. Proposed ALPR System

The nature of traffic images might be very problematic to LP detection approaches that work directly on the frames since (i) there are many textual blocks that can be confused with LPs such as traffic signs and phone numbers on storefronts, and (ii) LPs might occupy very small portions of the image. Therefore, we propose to first locate the vehicles in the input image and then detect their respective LPs in the vehicle patches. Afterward, we detect and recognize all characters simultaneously by feeding the entire LP patch into the network, instead of segmenting and classifying each character individually.

In order to develop an ALPR system that is robust for different LP layouts, we propose a ***layout classification*** stage after LP detection. However, instead of performing both stages separately, we merge the LP detection and layout classification tasks by training an object detection network that outputs a distinct class for each LP layout. In this way, with almost no additional cost, we employ layout-specific approaches for LP recognition in cases where the LP and its layout are predicted with a confidence value above a predefined threshold. For example, all Brazilian LPs have seven characters: three letters and four digits (in that order), and thus a post-processing method is applied to avoid errors in characters that are often misclassified, such as 'B' and '8', 'G' and '6', 'I' and '1', among others. Figure 2 illustrates the pipeline of the proposed ALPR system.

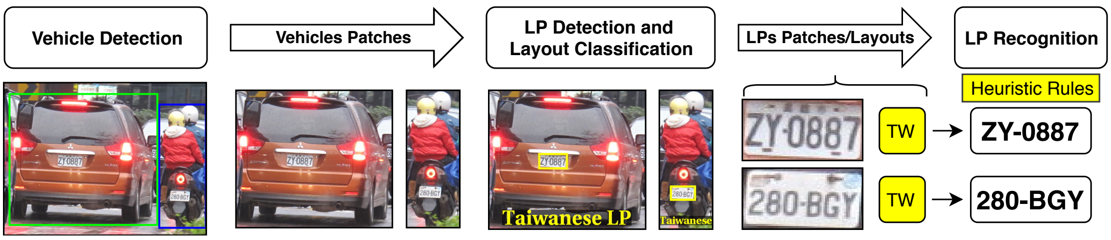

Figure 2. The pipeline of the proposed ALPR system.

In this work, we detect the LP and simultaneously classify its layout into one of the following classes: *American, Brazilian, Chinese, European or Taiwanese* (these classes were defined based on the public datasets found in the literature). We then adapt the results produced by CR-NET – mentioned in the next paragraph – according to the predicted layout class (the *heuristic rules* are described in detail in [Laroca et al. 2019b]). Based on the datasets employed in this work, we defined the minimum and the maximum number of characters to be considered in LPs of each layout. Additionally, inspired by [Silva and Jung 2017], we swap digits and letters on Brazilian and Chinese LPs, as there are fixed positions for digits or letters in those layouts. In this way, as mentioned earlier in this section, we avoid errors in characters that are often misclassified.

As great advances in object detection have been achieved using YOLO-inspired models, we decided to specialize it for ALPR. The models adapted and fine-tuned are *YOLOv2* (for vehicle detection), *Fast-YOLOv2* (for LP detection and layout classification) and *CR-NET* (for LP recognition), which is an architecture inspired by YOLO for character detection and recognition. We performed several data augmentation tricks and modifications to the networks to achieve the best ***speed/accuracy trade-off*** at each stage.

4. Experiments

The experiments were carried out in ***eight*** publicly available datasets: *Caltech Cars*, *EnglishLP*, *UCSD-Stills*, *ChineseLP*, *AOLP*, *OpenALPR-EU*, *SSIG-SegPlate* and *UFPR-ALPR*. For detailed information on these datasets, refer to [Laroca et al. 2019b]. Most of them have no annotations or contain labels for a single stage only (e.g., LP detection), despite the fact that they are often used to train/evaluate algorithms in the ALPR context. Therefore, for training our networks, we manually labeled the position of the vehicles, LPs and characters, as well as their classes in all images of these datasets (i.e., 38,351 bounding boxes on 6,239 images; these annotations are also ***publicly available***).

For each dataset, we compared the proposed ALPR system with state-of-the-art methods that were evaluated using the same protocol as us (we followed protocols used in previous works; see [Laroca et al. 2019b] for details regarding them). In addition, our results are compared with those obtained by *Sighthound* and *OpenALPR*, which are two commercial systems often used as baselines in the ALPR literature.

The results obtained in all datasets by the proposed system, previous works and commercial systems are shown in Table 1. In the average of five runs, across all datasets, our end-to-end system correctly recognized 96.9% of the LPs, outperforming Sighthound and OpenALPR by 9.1% and 6.2%, respectively. More specifically, the proposed system outperformed both previous works and commercial systems in the ChineseLP, OpenALPR-EU, SSIG-SegPlate and UFPR-ALPR datasets, and yielded competitive results to those attained by the baselines in the other four datasets[3].

Table 1. Recognition rates (%) obtained by the proposed system, previous works, and commercial systems in all datasets used in our experiments. Refs. 1-5 refer respectively to [Panahi and Gholampour 2017], [Zhuang et al. 2018], [Silva and Jung 2018], [Gonçalves et al. 2018] and [Laroca et al. 2018].

Dataset	LP Layout	Ref. 1	Ref. 2	Ref. 3	Ref. 4	Ref. 5	Sighthound	OpenALPR	Proposed*	Proposed
Caltech Cars	American	–	–	–	–	–	95.7 ± 2.7	$\mathbf{99.1 \pm 1.2}$	96.1 ± 1.8	98.7 ± 1.2
EnglishLP	European	**97.0**	–	–	–	–	92.5 ± 3.7	78.6 ± 3.6	95.5 ± 2.4	95.7 ± 2.3
UCSD-Stills	American	–	–	–	–	–	**98.3**	**98.3**	97.3 ± 1.9	98.0 ± 1.4
ChineseLP	Chinese	–	–	–	–	–	90.4 ± 2.4	92.6 ± 1.9	95.4 ± 1.1	$\mathbf{97.5 \pm 0.9}$
AOLP	Taiwanese	–	**99.8****	–	–	–	87.1 ± 0.8	–	98.4 ± 0.7	99.2 ± 0.4
OpenALPR-EU	European	–	–	93.5	–	–	93.5	91.7	96.7 ± 1.9	$\mathbf{97.8 \pm 0.5}$
SSIG-SegPlate	Brazilian	–	–	88.6	88.8	85.5	82.8	92.0	96.9 ± 0.5	$\mathbf{98.2 \pm 0.5}$
UFPR-ALPR	Brazilian	–	–	–	–	64.9	62.3	82.2	82.5 ± 1.1	$\mathbf{90.0 \pm 0.7}$
Average	All	–	–	–	–	–	87.8 ± 2.4	90.7 ± 2.3	94.8 ± 1.4	$\mathbf{96.9 \pm 1.0}$

* The proposed ALPR system assuming that all LP layouts were classified as undefined (i.e., without layout classification).
** The LP patches for the LP recognition stage were cropped directly from the ground truth in [Zhuang et al. 2018].

To evaluate the impact of classifying the LP layout prior to LP recognition (i.e., our main proposal), we also report in Table 1 the results obtained when assuming that all LP layouts were classified as undefined and that a generic approach (i.e., without heuristic rules) was employed in the LP recognition stage. The mean recognition rate was improved by 2.1%. We consider this strategy (layout classification + heuristic rules) ***essential*** for accomplishing outstanding results in LPs with fixed positions for letters and digits (e.g., Brazilian and Chinese LPs), as the recognition rates attained in the ChineseLP, SSIG-SegPlate and UFPR-ALPR datasets were improved by 3.6% on average.

In Table 2, we report the time required for each network in our system to process

[3]Qualitative results can be seen in [Laroca et al. 2019b], as well as in the first author's master's thesis.

an input. It is remarkable that although a deep CNN model is used for vehicle detection, our system is still able to process images at 73 FPS on a high-end GPU[4]. As vehicle detection is performed only once, regardless of the number of vehicles in the image, our system is able to process more than 30 FPS even when there are 4 vehicles in the scene. This information is relevant since most ALPR approaches either do not process frames in real time or can only perform in real time if there is at most one vehicle in the scene.

Table 2. The time required for each network in our system to process an input.

ALPR Stage	Time (ms)*	FPS
Vehicle Detection	8.5382	117
LP Detection and Layout Classification	3.0854	324
LP Recognition	1.9935	502
Total	**13.6171**	**73**

* Experiments performed on an NVIDIA Titan Xp GPU.

5. Publications and Impact

This work generated the following publications:

- **[Laroca et al. 2018] (Qualis A1)** - A preliminary version of the proposed method, along with the UFPR-ALPR dataset, was published at the *2018 International Joint Conference on Neural Networks (IJCNN)*, being by far the most cited paper of that edition both according to *IEEE Xplore* and *Google Scholar* (out of 760 papers). This work was covered on the *NVIDIA News Center*, with great importance being given to the results obtained, the dataset introduced by us, and potential applications. In just two years, the UFPR-ALPR dataset was downloaded more than 600 times by researchers from 77 countries around the world, as can be seen **here**.
- **[Laroca et al. 2019b] (Qualis A1)**[5,6] - The proposed approach, which addresses the limitations of the system presented in [Laroca et al. 2018] to considerably improve both the execution time (from 28ms to 14ms) and the recognition results (e.g., from 64.89% to 90% in the UFPR-ALPR dataset), was submitted to *IET Intelligent Transport Systems* (provisionally accepted subject to major revisions).
- **[Laroca et al. 2019a] (Qualis A2)** - We designed a two-stage approach for image-based Automatic Meter Reading (AMR) leveraging many concepts presented in our works on ALPR [Laroca et al. 2018, Laroca et al. 2019b]. In this work, published in the *Journal of Electronic Imaging*, we reported detection and recognition results significantly better than those obtained in previous works, with the networks used by us being able to process impressive 185+ FPS on a high-end GPU.

In addition to the works mentioned above, we would like to highlight three related papers we co-authored during the development of this work: **[Gonçalves et al. 2018, Gonçalves et al. 2019, Oliveira et al. 2019]**, helping to conceptualize the proposed methods and employing our ALPR system as a baseline or part of the novel approaches.

[4]We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

[5]*IET Intelligent Transport Systems* was previously known as *IEE Proceedings - Intelligent Transport Systems* and is not listed in Qualis 2013-2016, but is classified as ***Qualis A1*** in the preliminary version of Qualis 2017-2020.

[6]Our ALPR system is registered, under number BR512020000879-8, with the National Institute of Industrial Property (Brazil).

6. Conclusions and Future Work

In this work, we presented an end-to-end, efficient and layout-independent ALPR system using YOLO-based models that contains a unified approach for LP detection and layout classification and that leverages post-processing rules in the recognition stage to eliminate a major shortcoming of existing ALPR systems (being layout dependent). The robustness of our system is remarkable since it achieved recognition rates higher than 95% in all datasets except UFPR-ALPR (*where it outperformed the best baseline by* 7.8%), which has proved challenging and has become increasingly popular, with more than 600 download requests made by researchers from 77 different countries in the last two years.

As future work, we intend to conduct extensive experiments on cross-dataset scenarios, using for training all available datasets except one – which would be used for testing. Such experiments would simulate real-world situations, in which new cameras are being installed in new locations without existing systems being retrained. In this sense, we plan to design/exploit data augmentation techniques in order to realistically simulate scenarios where the camera's position is known, but there are no labeled images available.

References

Gonçalves, G. R., Diniz, M. A., Laroca, R., Menotti, D., and Schwartz, W. R. (2018). Real-time automatic license plate recognition through deep multi-task networks. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 110–117.

Gonçalves, G. R., Diniz, M. A., Laroca, R., Menotti, D., and Schwartz, W. R. (2019). Multi-task learning for low-resolution license plate recognition. In *Iberoamerican Congress on Pattern Recognition (CIARP)*, pages 251–261.

Laroca, R., Barroso, V., Diniz, M. A., Gonçalves, G. R., Schwartz, W. R., and Menotti, D. (2019a). Convolutional neural networks for automatic meter reading. *Journal of Electronic Imaging*, 28(1):013023.

Laroca, R., Severo, E., Zanlorensi, L. A., Oliveira, L. S., Gonçalves, G. R., Schwartz, W. R., and Menotti, D. (2018). A robust real-time automatic license plate recognition based on the YOLO detector. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–10.

Laroca, R., Zanlorensi, L. A., Gonçalves, G. R., Todt, E., Schwartz, W. R., and Menotti, D. (2019b). An efficient and layout-independent automatic license plate recognition system based on the YOLO detector. *arXiv:1909.01754*. submitted to *IET Intelligent Transport Systems* (provisionally accepted subject to major revisions).

Oliveira, I. O., Laroca, R., Menotti, D., Fonseca, K. V. O., and Minetto, R. (2019). Vehicle re-identification: exploring feature fusion using multi-stream convolutional networks. *arXiv:1911.05541*. submitted to *IEEE Trans. on Intelligent Transportation Systems*.

Panahi, R. and Gholampour, I. (2017). Accurate detection and recognition of dirty vehicle plate numbers for high-speed applications. *IEEE Transactions on Intelligent Transportation Systems (ITS)*, 18(4):767–779.

Silva, S. M. and Jung, C. R. (2017). Real-time brazilian license plate detection and recognition using deep convolutional neural networks. In *SIBGRAPI*, pages 55–62.

Silva, S. M. and Jung, C. R. (2018). License plate detection and recognition in unconstrained scenarios. In *European Conference on Computer Vision*, pages 593–609.

Zhuang, J., Hou, S., Wang, Z., and Zha, Z. (2018). Towards human-level license plate recognition. In *European Conference on Computer Vision (ECCV)*, pages 314–329.