

Character Time-series Matching For Robust License Plate Recognition

Quang Huy Che^{1,2}, Tung Do Thanh^{1,2}, Cuong Truong Van^{1,2}

¹University of Information Technology, Ho Chi Minh City, Vietnam

²Vietnam National University, Ho Chi Minh City, Vietnam

Email: {18520819, 19522491}@gm.uit.edu.vn, cuongtv@uit.edu.vn

Abstract—Automatic License Plate Recognition (ALPR) is becoming a popular study area and is applied in many fields such as transportation or smart city. However, there are still several limitations when applying many current methods to practical problems due to the variation in real-world situations such as light changes, unclear License Plate (LP) characters, and image quality. Almost recent ALPR algorithms process on a single frame, which reduces accuracy in case of worse image quality. This paper presents methods to improve license plate recognition accuracy by tracking the license plate in multiple frames. First, the Adaptive License Plate Rotation algorithm is applied to correctly align the detected license plate. Second, we propose a method called Character Time-series Matching to recognize license plate characters from many consequence frames. The proposed method archives high performance in the UFPR-ALPR dataset which is 96.7% accuracy in real-time on RTX A5000 GPU card. We also deploy the algorithm for the Vietnamese ALPR system. The accuracy for license plate detection and character recognition are 0.881 and 0.979 $mAP^{test}@.5$ respectively. The source code is available at <https://github.com/chequanghuy/Character-Time-series-Matching.git>

Index Terms—Automatic License Plate Recognition, Object detection, Computer vision, Data Association

I. INTRODUCTION

An ALPR system has targeted traffic management in the modern world since its first introduction in [1]. Many approaches have been introduced to make a system possible to predict LP location and identify the information of each LP most accurately. This increases the importance of an ALPR system in various applications such as automatic traffic toll systems, traffic management systems, automatic handling violations systems, etc.

A general ALPR system has two main steps, which are License Plate detection and License Plate recognition. For the License Plate detection step, all LPs were extracted from the image. In this step, most recent works prefer to apply CNN-based object detection methods such as R-CNN [2], Mask R-CNN [3], etc., to locate the region of vehicles that have detected license plates. In the next step, License Plate recognition, characters are recognized by using a Character Segmentation network with a classification head like CRNET [4] or Optical Character Recognition [5]. Although the result of each step achieves state-of-the-art accuracy, however, the overall accuracy of a whole system is still limited.

The authors in [6] proposed an ALPR system that obtained SOTA on many public datasets except for the UFPR-

ALPR dataset. Due to the variety of inclement conditions (lighting, noise, camera distortion) in UFPR-ALPR. We noticed that besides the high performance of their previous stage, there are still some missing characters that devastate the final prediction. Therefore, it influences the character recognition result. The temporal redundancy method, which has been applied in papers [6]–[8] has not yet taken advantage of all the license plate information. To resolve this problem, we propose a Character Time-series Matching method based on the Hungarian method and an Adaptive License Plate Rotation method to improve the ALPR system performance in various conditions.

It is also worth noting that we decide to evaluate the system only on the UFPR-ALPR dataset for many reasons. First, ALPR systems typically analyze videos or image sequences, and it is essential to use appropriate datasets to assess the accuracy of the whole system, such as the UFPR-ALPR dataset, which includes multiple video tracks. Second, this dataset has a lot of complex cases similar to a natural environment. Although, the paper [6] achieved the average end-to-end recognition rate of 96.9% across eight public datasets. However, as regards the UFPR-ALPR dataset, it was the lowest accuracy, at just about 90% [6].

In summary, our contributions can be given as follows:

- A complete ALPR system based on the YOLOv5 model for object detection tasks.
We then propose two methods called Character Time-series Matching and Adaptive License Plate Rotation which increase the accuracy of Character Recognition under challenging conditions.
- Regarding the result assessment, we evaluate our system on the UFPR-ALPR public dataset and our private dataset, which is gathered from multiple IP cameras on the streets of Vietnam. The proposed system archives excellent results on the UFPR-ALPR dataset and can process in real-time, both mid-end and high-end GPUs. In addition, we also implement the system on the Jetson Nano board and Jetson TX2 board for system performance evaluation.

II. RELATED WORK

This section introduces some state-of-the-art approaches to the ALPR system. Most research includes two main tasks:

Vehicle-License Plate Detection and License Plate Recognition.

A. Vehicle and License Plate Detection

License plates can be extracted by using an object detection model. Two popular methods are the one-stage and two-stage object detector model. As for two-stage object detection, Ross Girshick et al proposed R-CNN [2] model that uses the selective search algorithm to extract the region proposals in the first stage. Then a CNN model will be used to classify the objects in the second stage. Similarly, the Fast R-CNN model [9] based on R-CNN has a great soar in both speed and accuracy resulting from applying Region of Interest on feature maps.

Another model named Faster R-CNN [10] uses the Region Proposal Network to reduce the computational time considerably, which is ten times and fifty times compared to Fast R-CNN and R-CNN. However, these models are hard to process in real-time when compared to one-stage object detection; the YOLO [11] and SSD [12] get much better both speed and accuracy.

In the paper [6], [7], the vehicle is first detected, then the license plate is cropped for the next processing step by using another object detection network. It can be seen that the authors need to use two CNNs models for this task which leads to increased resource usage and processing time. In [13], instead of using the vehicle annotation, they expand the LP's annotation and call that extended area the Frontal View. The model learns the frontal views instead of using vehicle annotation and then takes those areas to detect LPs. Paper [14] used a multi-tasking CNN (MTCNN) model, which combines 3 CNNs, P-net, R-net, and O-net, to improve the model performance. As a result, this method is generally slower than single-stage detector models.

B. License Plate Recognition

Regarding license plate recognition, there are two primary trends which are segmentation-based and segmentation-free approaches. In the segmentation-based method, each detected license plate is fed to the segmentation network to locate characters such as CRNet [7] and DeepLabv2 ResNet-101 [15]. After that, to recognize characters, an OCR or a Character Classification network is used. In the segmentation-free method, [14] locates characters by using the Convolutional Recurrent Neural network, and then the Connectionist Temporal Classification is applied to the next step. In [16], they employ the resample and rectified LP and the SRNET instead of Recurrent Neural networks because of growth performance in the same resource usage. While instead of using the segmentation-based method, the authors [6] propose a method for recognizing characters based on the CRNet model. Due to the integration of localization and classification into a single step, the segmentation-free method is faster than segmentation methods. In practical applications, processing on a single frame easily makes erroneous results. To resolve this problem,

in papers [6], [7], the authors apply the Temporal Redundancy technique to the sequential frames to get better results.

III. PROPOSED METHOD

This paper implements an ALPR system for applying our proposed method. Figure. 1 illustrates our proposed system. The system includes two steps which are License Plate Detection and Character Detection. In the first step, all vehicles with match license plates are extracted. Next, all license plate characters will be recognized in the License Plate Recognition. In order to improve the recognition result, the **Adaptive License Plate Rotation algorithm is applied to align the license plate image**. Moreover, we also propose the Character Time-series Matching algorithm based on tracking characters in continuous frames for archiving better results. The detail of our proposed methods is discussed below sections.

A. License Plate Detection

Recent works divide this step to separate stages: vehicle detection and license plate detection [6], [7]. It leads to an increase marginally in both computational cost and processing time. Thanks to YOLOV5 model for detecting small objects well, we design a detection model based on YOLOv5m model with an input size of $1,280 \times 720$. License plates and Vehicles are detected simultaneously in real-time. We train the model based on pretrained YOLOv5m with our custom dataset, which includes all vehicles and license plate annotations.

B. Character Detection

In the Character Detection step, firstly, all detected license plates are stored and sorted by time in a queue. After the tracked object disappears, we propose an additional post-processing algorithm that finds vehicles match for license plate and the Adaptive License Plate Rotation method to align the license plate image. Eventually, we apply the Character Time-series Matching algorithm to get the final result. Regarding the character detection model, we propose a model based on YOLOv5 with the multi-head self-attention block proposed in the paper [17], and the input size is 128×128 . **However, since the characters are relatively large compared to the license plate size, we drop the detection head for the small and medium features and halve the number of filters in every layer, resulting in only slightly reducing the detection accuracy but significantly improving the speed.** Table I shows the architectural design of our character detection model, where a *C3* module is three consecutive Convolution layers with a Bottleneck block at the end, and *C3TR* is the *C3* module that replaced the Bottleneck block with a multi-head self-attention layer.

1) *The Adaptive License Plate Rotation method:* In the practical environment, vehicles can move in many different directions while the camera view is fixed in one spot. Fig. 2 - (a) shows some samples in this case. Therefore, the license plate image may be tilted in a particular direction depending on the vehicle's direction. It makes a considerable reduction in the accuracy of the next step. To handle this issue, we need



Fig. 1: ALPR pipeline

to adjust the angle of the license plate correctly. Assume there is a line d that divides the license plate into two equal parts horizontally, which form by the equation $y = ax + b$, a is the slope, α is the angle formed by the horizontal axis, and the line d , so that $\alpha = \arctan a$. Fig. 2 illustrates the Adaptive License Plate Rotation method, in which (a) are original images and (b) are rotated images.

In the first frame t_0 , we initialize $\alpha_0 = 0$ and estimate the $\beta_0 = \arctan(a_0)$, where a_0 is calculated following equation (1). In the following frame, the LP image is rotated α_1 -degree angle, where α_1 is updated based on equation (2), then the character prediction model predicts the bounding boxes from the rotated LP image at time t_1 and estimates the β_1 value. This process repeats until the final LP frame.

$$a_t = \frac{\sum_{i=1}^{n_t} x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^{n_t} x_i^2 - n(\bar{x})^2} \quad (1)$$

where:

- n_t : number of character bounding boxes from LP image at time t .
- x_i, y_i : x center and y center of the character bounding box i .
- a_t : the slope coefficient of the line d at time t .

$$\alpha_t = \alpha_{t-1} + \beta_{t-1} \quad (2)$$

where:

- α_t : the estimated angle between LP image and the horizontal axis at time $t - 1$.
- α_{t-1} : the estimated angle between LP image and the horizontal axis at time $t - 1$.
- β_{t-1} : the estimated angle between rotated LP image and the horizontal axis at time $t - 1$.

2) *Character Time-series Matching*: In our experiments, processing on a single frame achieves poor results in the case of image quality. Therefore, in this paper, the Character Time-series Matching (CTM) algorithm is proposed to take

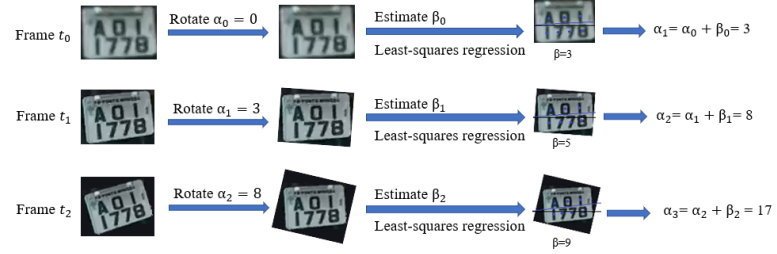


Fig. 2: Visual of The Adaptive License Plate Rotation (a) Original Images (b) Rotated images (c) Alignment images and updated values

TABLE I: Overview of the proposed model.

| Module index | Connected from | Module | Input size | Output size |
|--------------|----------------|-------------|--|---------------------------|
| 0 | | Conv | $3 \times 128 \times 128$ | $16 \times 64 \times 64$ |
| 1 | 0 | Conv | $16 \times 64 \times 64$ | $32 \times 32 \times 32$ |
| 2 | 1 | C3 | $32 \times 32 \times 32$ | $32 \times 32 \times 32$ |
| 3 | 2 | Conv | $32 \times 32 \times 32$ | $64 \times 16 \times 16$ |
| 4 | 3 | C3 | $64 \times 16 \times 16$ | $64 \times 16 \times 16$ |
| 5 | 4 | Conv | $64 \times 16 \times 16$ | $128 \times 8 \times 8$ |
| 6 | 5 | C3 | $128 \times 8 \times 8$ | $128 \times 8 \times 8$ |
| 7 | 6 | Conv | $128 \times 8 \times 8$ | $256 \times 4 \times 4$ |
| 8 | 7 | SPP | $256 \times 4 \times 4$ | $256 \times 4 \times 4$ |
| 9 | 8 | C3TR | $256 \times 4 \times 4$ | $256 \times 4 \times 4$ |
| 10 | 9 | Conv | $256 \times 4 \times 4$ | $128 \times 4 \times 4$ |
| 11 | 10 | Upsample | $128 \times 4 \times 4$ | $128 \times 8 \times 8$ |
| 12 | [6,11] | Concatenate | $[128 \times 8 \times 8]$ $[128 \times 8 \times 8]$ | $256 \times 8 \times 8$ |
| 13 | 12 | C3 | $256 \times 8 \times 8$ | $128 \times 8 \times 8$ |
| 14 | 13 | Conv | $128 \times 8 \times 8$ | $64 \times 8 \times 8$ |
| 15 | 14 | Upsample | $64 \times 8 \times 8$ | $64 \times 16 \times 16$ |
| 16 | [4,15] | Concatenate | $[64 \times 16 \times 16]$ $[64 \times 16 \times 16]$ | $128 \times 16 \times 16$ |
| 17 | 16 | C3 | $128 \times 16 \times 16$ | $64 \times 16 \times 16$ |
| 18 | 17 | Detect | $64 \times 16 \times 16$ | |

advantage of the spatial information obtained. CTM combines two bounding boxes of the same character in the same LP from 2 consecutive frames based on the Euclidean distance

between the center of the two bounding boxes, depicted in Fig. 3. In CTM, if two character bounding boxes are joined together, the Euclidean distance between its center boxes must be less than the threshold ϵ .

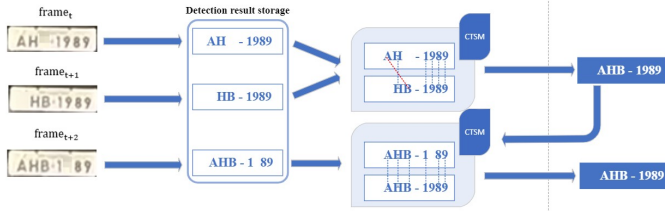


Fig. 3: Character Time-series Matching (CTM)

In detail, we describe our problem as a linear assignment problem, with the objects assigned to each other being the character bounding boxes from I_t^a and I_{t+1}^a , where I_t^a and I_{t+1}^a are the consecutive images of the LP a at time t and $(t+1)$, respectively. The cost for the linear assignment problem is the Euclidean distance between the two bounding boxes. We solve the problem by using the modified **Hungarian algorithm** [18].

At first, a set of track is initialized, $\tau = \{\}$ based on sort algorithm [19], with each track $t \in \tau$ containing the current position (the center of the bounding box), the list cls contains the class of the bounding boxes that have been matched with track t in the previous frame and the list $conf$ stores the confidence corresponding to each class in the cls . We consider three conditions:

- If a bounding box b_j from B_t^a matches with a track t_i from τ , the result of t_i is updated with the position of t_i assigned to the position of b_j ($t_{i,position} = b_{j,position}$), the character class and confidence of b_j is appended to the corresponding list in t_i ; where B_t^a is the bounding boxes from the Character Detection stage of I_t^a .
- If a bounding box b_k from B_t^a does not match any track from τ , then b_k is appended to set τ .
- If a track t_m of τ does not match with any bounding box of B_t^a , t_m is added a ΔC value; where ΔC is the mean distances between the position of all possible track and bounding boxes that can be matched. The idea of updating ΔC is that we want to shift the “unmatchable” tracks in the same direction as the “matchable” tracks.

Fig. 4 shows information about how the Character Time-series Matching update parameters.

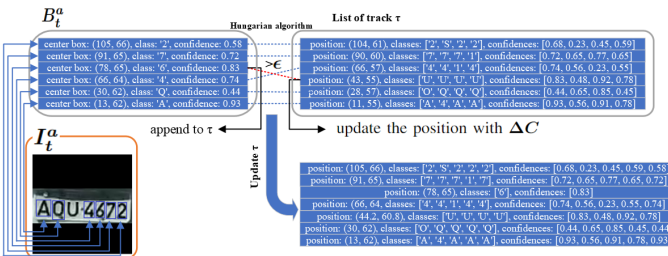


Fig. 4: The update process of Character Time-series Matching

We perform post-processing to get the character detection result of each track in the set τ by calculating the weighted-sum confidence of each class and then select the character class with the most confidence value, as shown in Fig. 5. As shown in Fig. 5, character H is selected, and character M is removed because its weighted-sum value is lower than character H. Equation 3 shows the detail of this step.



Fig. 5: The character result based on the weighted-sum confidence of each class

$$class_i = \arg \max_{c \in class-indices} \{K_i^c | K_i^c = \sum_j class_i^j confidence_i^j\} \quad (3)$$

where:

- $class-indices$: the list of classes' indices
- $class_i$: class of $track_i$
- K_i^c : weighted-sum confidence of class c in $track_i$
- $class_i^j$: is set to 1 if the class of the j^{th} box detection in track i is c
- $confidence_i^j$: confidence of the j^{th} box detection in $track_i$

IV. EXPERIMENTAL RESULTS

This section evaluates the whole pipeline on the UFPR-ALPR and Vietnamese streets to ensure that the suggested ALPR system is effective. All tests were carried out using Pytorch framework on NVIDIA GeForce RTX A5000, RAM 32GB, Intel(R) Core(TM) i9-10900X.

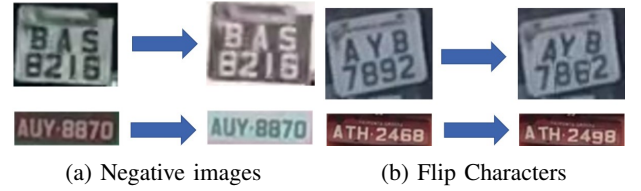


Fig. 6: Illustrated image of data augmentation methods

A. Evaluation on the UFPR-ALPR dataset

1) *Vehicle and License Plate Detection*: Although we used only one model for vehicle and LP detection, we achieved competitive accuracy compared to the [6] using two models, including Vehicle Detection and License Plate Detection. The model accuracy results obtained after applying the filter are shown in Table II.

2) *Character Detection*: Due to data restrictions for characters, we proceed to gather and categorize additional data, including over 900 images of Brazilian LPs¹. Although our

¹The images were downloaded from www.platesmania.com

TABLE II: Accuracy and inference time in Vehicle and License Plate Detection stage.

| Size | Precision | Recall | mAP test@.5 | Inference time |
|----------|-----------|--------|-------------|----------------|
| 1280×720 | 0.968 | 0.959 | 0.968 | 12.4ms |

suggested model only comprises 1213854 parameters and 0.15 GFLOP with an input image size of 128, it has a good level of accuracy on the UFPR-ALPR dataset's test set.

The proposed model has a much lower computational cost than the previous character recognition model [6] [7] so that it can be efficiently run on low-configuration machines and embedded computers.

We consider the characters '1' and 'I', and the characters '0' and 'O' as single classes in the assessments performed in the UFPR-ALPR dataset since they are similar depending on the format of the LP that we adjust accordingly. The Brazilian LP format is AAA-NNNN, where A is an alphabetic character and N is a numeric character.

We use data augmentation methods such as rotation, mosaic, and mix-up and also apply data augmentation by converting negative images and randomly flipping characters directly on number plate images for the training process. The suggested characters to be rotated are listed in Table III. The data augmentation technique automatically transforms the class of character '6' to character '9' and the class of character '9' to character '6' when flip character '6' or character '9' in both directions. Some instances of the proposed data augmentation approach are shown in Fig. 6.

TABLE III: Statistical table of characters that can be flipped in each direction.

| Flip Direction | Characters |
|----------------|--|
| Vertical | 0 (O), 1 (I), 3, 8, B, C, D, E, H, K, X |
| Horizontal | 0 (O), 1 (I), 8, A, H, M, T, U, V, W, X, Y |
| Both | 0 (O), 1 (I), 6, 8, 9, H, N, S, X, Z |

Table IV is the result of the proposed model on the test set of the UFPR-ALPR dataset, with a confidence threshold of 0.1 and an IOU threshold of 0.1.

TABLE IV: Accuracy and inference time in Character Detection stage.

| Size | Precision | Recall | mAP test@.5 | Inference time |
|---------|-----------|--------|-------------|----------------|
| 128×128 | 0.95 | 0.964 | 0.984 | 6.1ms |

3) *ALPR system (end to end)*: Table V shows the result of our proposed system compared to the others on the UFPR-ALPR dataset. In the case of only using the CTM method, the figure for accuracy is 91.67 compared to nearly 90 and 64.9 in paper [6] and [7], respectively. Another interesting point is a considerable increase in case integration of both AR and CTR methods, with 96.7 reaching the highest accuracy. The results are at around 47 fps in our system. Fig. 7 shows some results

when applying the proposed algorithm to the UFPR-ALPR dataset.

TABLE V: The accuracy of ALPR system.

| | [7] | [6] | CTM | AR ^a + CTM |
|-----------------------|------|------------|-------------|-----------------------|
| Number of used models | 4 | 3 | 2 | 2 |
| Accuracy | 64.9 | 90.0 ± 0.7 | 91.67 | 96.7 |
| FPS | - | - | 47.6 | 47.56 |

^a Our baseline with the Adaptive License Plate Rotation method.



Fig. 7: Some results were obtained on [7], [6], and the proposed method.

B. Evaluation on Vietnamese street

In addition to developing a system to provide a powerful LP Recognition method, we collected and labeled 9,887 Vietnamese license plate images, which have two types: square and rectangular license plates and 28180 images for LP recognition on Vietnamese streets.

We also evaluated on the Jetson Nano kit and Jetson TX2 kit, with the results obtained as shown in Table VI. Overall, what stands out from the table is that our proposed method can be implemented well on embedded computers, and the execution time is more noticeable when converted to 16-bit TensorRT floating point inference.

Moreover, we test on a set of 100 videos recorded from IP cameras in real-world conditions. If the system successfully

TABLE VI: Evaluation results on Vietnamese dataset.

| Model | Inference type | Precision | Recall | mAP test@.5 | Inference time | | |
|---------------------|----------------|-----------|--------|-------------|----------------|-------|--------|
| | | | | | Nano | TX2 | Server |
| YoloV5s 640x640 | Pytorch | 0.876 | 0.809 | 0.847 | 198ms | 78ms | 8.3ms |
| | TensorRT | 0.884 | 0.789 | 0.841 | 71ms | 33ms | 0.9ms |
| YoloV5m 640x640 | Pytorch | 0.875 | 0.823 | 0.866 | 515ms | 180ms | 10.9ms |
| | TensorRT | 0.859 | 0.829 | 0.867 | 171ms | 75ms | 1.9ms |
| YoloV5m 1280x1280 | Pytorch | 0.877 | 0.839 | 0.881 | 2162ms | 855ms | 15.6ms |
| | TensorRT | 0.869 | 0.835 | 0.879 | 615ms | 244ms | 5.3ms |
| Yolov5+ Transformer | Pytorch | 0.986 | 0.971 | 0.979 | 30ms | 25ms | 6.1ms |
| | TensorRT | 0.898 | 0.893 | 0.822 | 11ms | 9.5ms | 0.75ms |



Fig. 8: Some qualitative results obtained from our test video

detects the license plate information and vehicle type, the case is said to be true in our test set. Our system achieves 90% accuracy, meaning it could extract correctly 90 LP information and vehicle type from 90 vehicles; Fig. 8 shows some sample images and results.

V. CONCLUSION

In conclusion, we propose the Character Time-series Matching and Adaptive Rotation algorithm, which improves the accuracy of the ALPR system. The algorithms archives high accuracy in case of low quality and small objects. In addition, the processing time of our method can perform in real-time in mid-end and high-end GPU. The experiments on UFPR-ALPR and Vietnamese streets show that our proposed method has excellent performance in various conditions. In the future, we would like to test the effectiveness of our approach on more public datasets and try to optimize CNN architectures for processing on edge computers.

REFERENCES

- [1] R. Lotufo, A. Morgan, and A. Johnson, "Automatic number-plate recognition," in *IEE Colloquium on Image Analysis for Transport Applications*, 1990, pp. 6/1–6/6.
- [2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.
- [3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
- [4] W. Liu, C. Zhang, G. Lin, and F. Liu, "Crnet: Cross-reference networks for few-shot segmentation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 4164–4172.
- [5] M. T. Qadri and M. Asif, "Automatic number plate recognition system for vehicle identification using optical character recognition," in *2009 International Conference on Education Technology and Computer*, 2009, pp. 335–338.
- [6] R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, and D. Menotti, "An efficient and layout-independent automatic license plate recognition system based on the YOLO detector," *CoRR*, vol. abs/1909.01754, 2019. [Online]. Available: <http://arxiv.org/abs/1909.01754>
- [7] 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)*, 2018, pp. 1–10.
- [8] G. R. Gonçalves, D. Menotti, and W. R. Schwartz, "License plate recognition based on temporal redundancy," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 2577–2582.
- [9] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.
- [10] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015. [Online]. Available: <http://arxiv.org/abs/1506.01497>
- [11] G. Jocher, A. Stoken, A. Chaurasia, J. Borovec, NanoCode012, TaoXie, Y. Kwon, K. Michael, L. Changyu, J. Fang, A. V. Laughing, tkianai, yxNONG, P. Skalski, A. Hogan, J. Nadar, imyhxy, L. Mammana, AlexWang1900, C. Fati, D. Montes, J. Hajek, L. Diaconu, M. T. Minh, Marc, albinxavi, fatih, oleg, and wanghaoyang0106, "ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support," Oct. 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.5563715>
- [12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37.
- [13] S. Montazzolli and C. Jung, "Real-time brazilian license plate detection and recognition using deep convolutional neural networks," in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2017, pp. 55–62.
- [14] W. Wang, J. Yang, M. Chen, and P. Wang, "A light cnn for end-to-end car license plates detection and recognition," *IEEE Access*, vol. 7, pp. 173 875–173 883, 2019.
- [15] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *CoRR*, vol. abs/1606.00915, 2016. [Online]. Available: <http://arxiv.org/abs/1606.00915>
- [16] Y. Wang, Z.-P. Bian, Y. Zhou, and L.-P. Chau, "Rethinking and designing a high-performing automatic license plate recognition approach," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2021.
- [17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *CoRR*, vol. abs/1706.03762, 2017. [Online]. Available: <http://arxiv.org/abs/1706.03762>
- [18] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800020109>
- [19] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," *CoRR*, vol. abs/1602.00763, 2016. [Online]. Available: <http://arxiv.org/abs/1602.00763>