An efficient method to improve the accuracy of Vietnamese vehicle license plate recognition in unconstrained environment

***Abstract*—*Background*: Most previous studies in automatic license plate recognition (ALPR) focused on recognizing license plate (LP) in constrained environment where cameras are installed in front of LPs and other conditions such as lighting, weather, and image quality are satisfied. Current image processing techniques including deep learning models have been proved to be effective for ALPR in a constrained environment. Recent studies on ALPR in Vietnam have conducted in small datasets and have not covered various cases of Vietnamese LPs.**

***Aim*: To develop a model for ALPR that is effective in unconstrained environment in Vietnam.**

***Method*: Using multi-stage approach, we propose two improvements: We apply the idea of the key-point detection problem to address the non-rectangular object detection for LP detection part, and use a segmentation free approach based on encoder decoder network for the LP optical character recognition (OCR) part. We train and evaluate models in a dataset that is various in picture types and colours and were collect ed from different times and environment, considering different positions and angles of camera.**

***Results*: Our results show improvements in LP detection accuracy with mean IOU $mIOU = 95.01\%$ and precision $P_{75} = 99,5\%$. The accuracy in LP OCR was up to $Acc_{seq} = 99.28\%$ at sequence level and $Acc_{char} = 99.7\%$ at character level.**

***Conclusion*: This paper provides a large dataset of Vietnamese LP images that can be effectively used to evaluate systems for ALPR in Vietnam, and proposes improvement techniques to tackle problems of ALPR in unconstrained environment in Vietnam.**

***Keywords:* Vehicle plate detection and recognition, keypoints detection, sequence modeling.**

I. Introduction

The increasing number of cars and trucks on the roads and the needs for controls and managing these vehicles has fostered the ALPR. These ALPR systems have been used to monitor vehicles on roads, at the car parks and at toll plazas; to find lost vehicles; and for other specific observation purposes. From images captured by cameras, an ALPR system detects areas containing vehicles and then detects areas containing a license plate. Characters of a licence plate can be identified from these areas. In cases where ALPR systems are used for vehicle monitoring, real-time processing is needed.

In Vietnam, most ALPR systems are deployed at parking stations or toll plazas where cameras are set in a constrained environment. It means the camera placements, lighting conditions, and weather are best designed for recognition; thus, the recognition efficiency is quite high. However, ALPR systems were not effective when cameras were set up outdoor where cameras were located at a distance, in unstable lighting the weather conditions. This leads to capturing of poor quality images that finally affects the recognition efficiency, especially for character segmentation approach. In additions, when a camera is deployed outdoor, camera's viewing angles for vehicles are different, thus rectangular object detection approaches faced difficulties in aligning license plates. In this paper, we address two problems of detecting and recognising licence plates from traffic cameras in unconstrained environment.

Each country has its own standard for that country's license plate, currently most license plate datasets shared in the community are foreign license plate data. Therefore, these datasets can not be applied to train Vietnamese license plate recognition models. Vietnamese license plates include 2-line and 1-line license plates with size, color, font and arrangement are set according to government regulations. Since August 2020, yellow number plates has been applied to vehicles used for business transportation, the plate fonts are also different from those before. Vinh et al. [1] and Duan et al. [2] conducted two studies of ALPR systems on Vietnammese licence plates; However, their datasets were small and they have not addressed problems of unconstrained environment.

This paper proposes a model based on multi-stage approach with improvements focusing on detecting license plate area and recognizing plate characters in unconstrained environment. The main contributions of our paper are:

First, regarding non-rectangular object detection problem, we propose an approach that is based on keypoints detection. We modify the semantic segmentation architecture DDRNet [3] to match the addressed problem.

Second, our system identifies plate based on segmentation-free approach integrated with Attention mechanism [4]. Feature extraction part is designed based on VGG-19 [5] in more fine-grained level.

Third, we build the MTAVLP dataset that includes 15571 photos of vehicles containing license plates and 16012 photos of Vietnamese license plates. This dataset is large and various in image capturing conditions.

This paper is structured as follows: Section II presents research background and related work. Our detail proposed methods are presented in Section III. Section IV shows and discusses obtained results. Conclusions and future work are given in Section V.

II. Research Background and Related Work

There are two main approaches for ALPR: multi-stage and single-stage license plate recognition

A. Multi-stage license plate recognition systems

This approach includes two stages: LP detection and LP optical character recognition (OCR).

1) License plate detection: LP detection uses traditional image processing techniques that are based on the characteristics of license plates such as shape, color, symmetry, and texture. The main limitation of traditional image processing methods is its ability to handle complex backgrounds. With the development of deep learning, Convolutional Neural Network (CNN) have been commonly used to detect license plate objects. Hsu et al. [6] established two networks based on YOLO [7] in order to detect LP in different camera's view angles. Xie et al. [8] uses sliding window and candidate filter integrated with CNN for ALPR problem. However, efficiency of built systems was not high. Jaderberg et al. [9] uses a text processing method to detect LP with different offset angles. However, these methods face difficulties when the background has many other text rather than LP characters.

2) License plate optical character recognition: LP OCR has two main approaches which are: *segmentation-based* and *segmentation-free*.

In segmentation-based approaches, characters are segmented and then classified. Techniques for character segmentation are projection profiles methods, using pixel connectivity, and prior knowledge. These techniques are highly influenced by characteristics and thresholds of input images [10]. A drawback of segmentation-based approaches is that the quality of the segmentation stage has high impact to classification results. When an input image is inclined or uncleared, it will be very difficult to find where to separate the characters.

In segmentation-free approaches, character segmentation stage is omitted in order to reduce the costs of computation and data labelling. System in this approach mostly use a combination of CNN and RNN [11],[12], and were optimised through connectionist temporal classification (CTC) loss [13]. However, CTC loss has drawbacks as it requires conditional independence assumption and it cannot work when the number of characters of an input plate is larger than the maximum number of time steps [13].

In this paper, we use segmentation-free approach and apply encoder-decoder architecture to overcome limitations of CTC loss. We also use an attention mechanism [4] to extract important features which affect a desired output in each part of an image.

B. Single-stage LP recognition systems

Recently, studies on ALPR using single-stage process have recorded good results. Li et al. [14] uses VGG16 as a feature extractor. They modified VGG16, removed three layers, and the output is put into a Region Proposal Network. In these systems, two separate sub-networks were used for LP detection and LP OCR. However, using only one deep neural network will make it more difficult to intervene in every step of the processing. Therefore, it is difficult to apply this approach to Vietnamese license plates that include both 1-line and 2-line plate styles.

C. Keypoints detection

In keypoints detection problem classes, human pose estimation (HPE) problem has attracted more attentions over last decade. HPE includes two main approaches: coordinate regression and heatmap regression [15]. Heatmap regression approaches have shown to be more effective than coordinate regression approaches [16].

Studies based on heatmap regression training includes two main structures: high-resolution structure and feature fusion techniques such as HRNet [17] and DLA network [18]. In which, the first structure recorded the highest accuracy at $AP = 74.4$ in COCO test-dev human pose estimation dataset.

In this paper, we use heatmap regression for LP keypoints detection to handle non-rectangular problem, in which each angle of LP is considered as a keypoint.

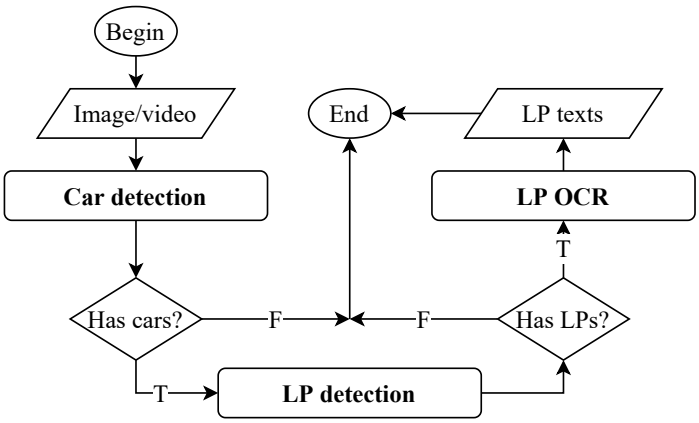

Fig. 1: The illustration of our ALPR systems.

III. Proposed Methods

Our proposed methods includes three main parts: Vehicle detection, LP detection and LP OCR as in Figure 1.

A. Vehicle detection

Vehicles are basic objects presented in many large datasets such as PASCAL-VOC [19] and COCO [20]. We decided to use YOLOv2 network [7] as it can get a balance between performance speed and accuracy [21].

B. License plate detection

Pre-processing: All vehicle images were normalized to the standard normal distribution, and then resized to fixed sizes according to input shapes of the model architectures respectively. The purpose of this pre-processing is to obtain the same range of values for each input image before feeding to the CNN model. It also helps to speed up the convergence of the model.

Keypoints encoding: In our system, the heatmap regression is used for LP keypoints detection. Considering input image $X \in \mathbf{R}^{W \times H \times 3}$ has width of W and height of H. Then, the

output is a heatmap $Y \in [0,1]^{\frac{W}{R} \times \frac{H}{R} \times C}$, in which R is the resolution decrease, C is a number of classes. For each ground truth keypoint $k \in \mathbf{R}^2$ of class C, it is equivalent to $k' = [\frac{k}{R}]$ in Y. The output at (x, y) of a class c in heatmap Y is

$$Y_{xyc} = exp(-\frac{(x - k'_x)^2 + (y - k'_y)^2}{2{r_k}^2}) \quad (1)$$

In which, r_k is the radius of the keypoint k in Y. To diminish $\delta = \frac{k}{R} - k'$, all classes of C shared two offset regression heads $O \in \mathbf{R}^{\frac{W}{R} \times \frac{H}{R} \times 2}$, in which $(O_{xy1}, O_{xy2}) = (\delta_x, \delta_y)$ if $Y_{xyc} > 0, 1 \leq c \leq C$. We choose $R = 4$, $r_k = 2$. Designing the number of keypoints for a LP and the number of classes C is illustrated in the Figure 2. The design $1c$ considers all

Fig. 2: Keypoints design for a licence plate (colours are according to classes). From left to right showed designs $1c$, $2c$, $4c$, $1c3$ accordingly.

keypoints to belong to a class. The design $2c$ considers 2 upper keypoints to belong to one class while 2 below keypoints belong to another class. The design $4c$ considers 4 keypoints to belong to four different classes. The design $1c3$ considers 2 below keypoints and the centre of the LP to belong to the same class, ignoring two upper keypoints as their background varies a lot in different types of vehicles (Figure 2). It has been proved by experiments that the design $1c$ gives the best result. This is contradicted to human pose estimation where all keypoints belong to different classes [15].

Architecture: Beside architectures used for the keypoints detection problem (Section II-C), we found that keypoints detection network architectures are highly correlated with segmentation network architectures. We chose DDRNet [3] and its variants (Figure 3) as DDRNet is one of the few segmentation algorithms that can achieve a balance between accuracy and performance speed.

To be applied into LP keypoints detection problem, we made some changes in the architecture of DDRNet [3]. To inform about lower feature maps, we used residual connection instead of training and collecting information from separated stages as in the original architecture. In addition, the size of an output heatmap that we choose is proportional to the input at $R = 4$. However, DDRNet maintains a high-resolution feature map with size of $R = 8$ compared to the input image. We made two modifications corresponding to two new architectures in order to suit the keypoints detection problem. The first architecture DDRNetsh was obtained via changing the level of down convolution at the first convolution block (Figure 3. The second architecture DDRNetup was obtained by adding one block upconvolution at the end of the original architecture.

Loss function: The heatmap regression loss and offset regression loss is presented in [22].

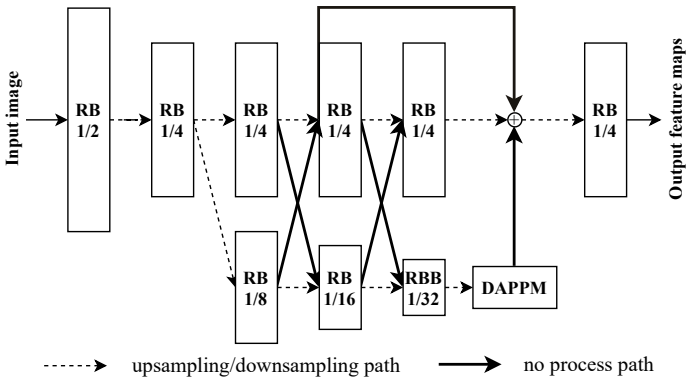

Fig. 3: DDRNetsh Network, "RB" denotes sequential residual basic blocks. "RBB" denotes single residual bottleneck block. "DAPPM" denotes the Deep Aggregation Pyramid Pooling Module [3].

Keypoints decoding: Having obtained feature maps, we used max pooling operator with kernel size of 3×3 according to feature maps of C classes. We removed points that have score less than 0.3 and chose the top 4 points with the highest score. They are the 4 keypoints of a LP needed to find. Then, positions of keypoints in the original image $(x_{ori}, y_{ori}) = ((x + O_x) \times R, y_{ori} = (y + O_y) \times R)$, in which (x, y) is a position of keypoints in the output feature maps.

C. License plate optical character recognition

The characteristics of Vietnamese license plates: Vietnamese vehicle LPs are designed according to government regulations (Figure 4). Vietnamese vehicle LP colors include four main types: blue background or red background with white text, yellow background or white background with black text. The

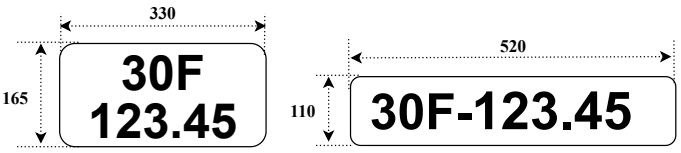

Fig. 4: A 2-line LP (left) and a 1-line LP (right) (unit mm).

characters in the number plate are in a set of 33 characters {ABCDEFGHKLMNPQSTUVXYZ0123456789-.}. These characters are not randomly arranged, but all belong to one of five template types (Figure 5). 1-line LP are obtained from 2-line LP by adding a '-' sign when combining the first and second lines of a 2-line number plate.

Fig. 5: Text formats of Vietnamese LPs.

Pre-processing: All LPs images are normalized to the standard normal distribution and then are resized to fixed sizes according to input shapes of the model architectures respectively.

For 2-line LPs, we use the X-Y Cut algorithm [23] to find out where to split a 2-line LP into two 1-line LPs.

VOCR Architecture: The input of the VOCR is an image of 1-line LP which has size of 32×140. The first part of VOCR is VOCR-CNN which extracts local features based on VGG19 [5] (Figure 6). Due to small input image size 32×140, we chose average pooling instead of max pooling to avoid losing of information. Additionally, the kernel size and stride size were chosen with a small size to compress the information appropriately. The output of VOCR has 256 feature maps, each feature map has size of 1×70. We considered the number of dimensions of each feature map to be the number of time steps, thus time steps $T = 70$. At each time step t, a feature vector of size 1×256 contains the features of all feature maps at that time step.

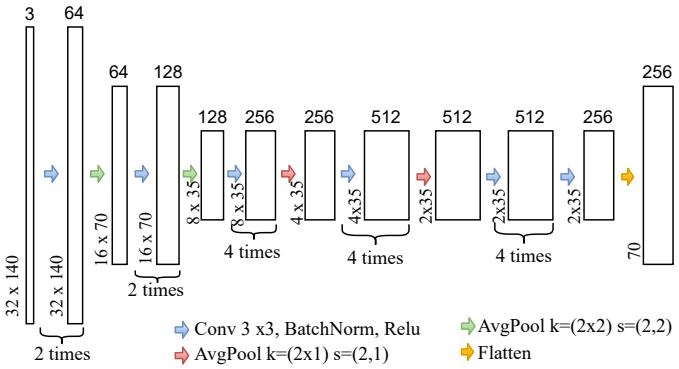

Fig. 6: Architecture of VOCR-CNN.

The second part of the VOCR is the VOCR-RNN in the form of an encoder-decoder using the GRU [24] combined with the attention mechanism [4].

In the encoder part, we chose one GRU-bidirectional layer to extract the inter-dependencies in input sequences. The context vector is obtained by using attention mechanism [4] with additive attention as attention scoring function [25]. The obtained context vector contains important featured information of the input sequence that the current decode step needs.

In the decoder part, we chose one GRU layer for sequential decoding. At each decoding step, the decoder output has probabilities of $N = 36$ (corresponding to the number of characters appearing in Vietnamese LPs plus 3 special characters $< sos >, < eos >, < pad >$). The highest one will the character needed to decode. If we meet $< eos >$ character, the decoding process will complete.

Loss function: The loss function used is a combination of the cross-entropy loss and label smoothing technique [26] to increase the generalisation for our model.

$$L = -\sum_{i=1}^{N}(l_i \times y_i \times (log(\hat{y}_i))) \quad (2)$$

In which, y_i is a groundtruth, $\hat{y}_i$) is a prediction of the model, $l_i = |y_i - \delta|$ is the label smoothing weight at i class. We chose $\delta = 0.1$ based on experimentation of all different thresholds proposed in [26].

IV. Experiment

A. Dataset

As far as we know, studies in ALPR in Vietnam have been conducted using small private datasets. Vinh et al. [1] tested models in a dataset of only 700 images. Duan et al. [2] used 805 images for test dataset. Understanding the important role of dataset size to the generalisation of deep neural networks, we built a large and diverse dataset named MTAVLP.

We installed 2 traffic cameras on two different road tracks. We set up a system as shown in Figure 1–excluding the LP OCR step–onto a server. The collection process was taken within 3 weeks–from 6:00 a.m. to 6:00 p.m. daily–in different weather and light conditions. Vehicle objects captured were all vehicles moving on roads with both old and new LP styles, and in a variety of plate sizes, font styles, and in different degrees of opacity and noise. Some different images of vehicles and plates are shown in Figures 7 and 8.

Fig. 7: Vehicle images captured in MTAVLP dataset.

Fig. 8: License plate images in MTAVLP dataset.

To increase the diversity of distribution for the MTAVLP dataset, we added to our dataset a set of 3000 Vietnamese vehicle images that were collected from different conditions and published in [27]. This brings the number of vehicle images in our dataset to 15571 and the number of LP images is 10773. The number of LP is smaller than the number of vehicle images as we removed images captured in redundant times. LP types were summarised in Table I arranged by their colors. After pre-processing 2-line plate images, we obtained 16012 images of 1-line plate.

TABLE I: Number of plates according to colors

Number of plates	White	Yellow	Blue	Red	Total number
1-line plate	2717	0	13	7	2737
2-line plate	6933	950	143	10	8036
Total number	9650	950	156	17	10773

B. License plate detection results

Conducting Experiment: We randomly divided MTAVLP dataset in to three sets for training, validating and testing according to a rate of 0.8:0.1:0.1.

Validation method: mean IOU ($mIOU$) and precision with $mIOU >= 0.75$ written as P_{75} were chosen as validation criteria [28]. A threshold of $mIOU >= 0.75$ was used instead of $mIOU >= 0.7$ to strengthen the robustness of model.

Training: In order to increase the generalisation of the model, we randomly performed argumentation operations during training. Adam optimization algorithm [29] was used with an initial learning rate of 0.001, batch size of 8, maximum training epoch number of 140, and early stopping patience of 10.

Results and Comparision: We experimented all designs $1c$, $2c$, $4c$, $1c3$ (Section III-B) with the Resnet18 architecture [30].

TABLE II: LP detection results for different keypoints considerations

Techniques	Training size	Testing size	$mIOU$
Resnet18_1c	384×384	384×384	**93.8**
Resnet18_2c	384×384	384×384	93.797
Resnet18_4c	384×384	384×384	91.78
Resnet18_1c3	384×384	384×384	90.5

Table II shows the highest $mIOU = 93.8$ for Resnet18_1c. It strengthens the hypothesis that all keypoints of a LP can be considered as individual objects. Resnet18_2c is as good as Resnet18_1c because the similarities of the upper and lower keypoints pairs are equivalent. Resnet18_4c based on the $4c$ design for the human pose estimation problem gives worst results due to the keypoints have common features that cause the model to be confused during training. Resnet18_3c also produces worse results, indicating that determining a center of a number plate is more difficult in heatmap regression. Thereafter, we used $1c$ design for our experiments.

TABLE III: LP detection results in different sizes of LP images

Techniques	Training size	Testing size	$mIOU$	P_{75}
DDRNet23sup	512×512	512×512	94.3	**99.4**
DDRNet23sup	512×512	384×384	93.1	98.1
DDRNet23sup	384×384	512×512	94.4	99.3
DDRNet23sup	384×384	384×384	**94.5**	99.2

In order to evaluate the effect of input image size on model performance, we studied two image sizes, 384×384 and 512×512. Results are shown in Table III. We chose DDRNet23sup that was originated from DDRNet-23-slim [3] and made modification as proposed in Section III-B. The results suggest that the performance of a model is slightly decreased when applying the model in a dataset that has smaller image size than that of training set.

TABLE IV: LP detection results in different architectures

Techniques	Training size	Testing size	$mIOU$	P_{75}
DDRNet23sup	384×384	384×384	94.5	99.2
DDRNet23sh	384×384	384×384	**95.01**	**99.5**
DDRNet23up	384×384	384×384	94.4	99.0
HRNet18_4s	384×384	384×384	94.2	98.58
HRNet18_3s	384×384	384×384	94.1	98.9
Resnet18	384×384	384×384	93.8	99.1
DLA34	384×384	384×384	90.5	93.6
DLA34	512×512	512×512	93.0	97.4

Using image size of 384×384 for training and test sets, we trained and evaluated different architectures including DDRNet23sh, DDRNet23sup, HRNet18_4s, HRNet18_3s, Resnet18 [30] and DLA34 [18] . DDRNet23sh is originated from DDRNet-23-slim [3] with a modification in downsampling convolution rate. DDRNet23sup is originated from DDRNet-23-slim and DDRNet-23 [3] with a modification in upsampling convolution. HRNet18_4s and HRNet18_3s are 4-stage and 3-stage versions of HRNet18 [17]. Results are showed in Table IV.

DDRNet23sh has highest accuracy with $mIOU = 95.01$, $P_{75} = 99.5$ though it is an architecture for segmentation problem. This proves the efficiency of DDRNet23sh when it is meticulously designed with multi-branch architecture and maintenance of high-resolution. DAPPM module [3] in the architecture enriches extracted information. The results also strengthen the hypothesis that maintaining high-resolution since beginning is better than the use of upsampling convolution at the last layer as in DDRNet23sup và DDRNet23up architectures.

C. Results of LP OCR

Conducting Experiment: We divided the set of 16012 LP images into three sets for training, validating and testing with a rate of 0.75:0.1:0.15. Different architectures with segmentation-free approach were studied in order to compare to VOCR, such as LPRnet-STN [31]-a lightweight architecture used for CNN and CTC loss; CRNN [32]- a combination of CNN và RNN. We added one more architecture CRNN (VGG19) that is based on CRNN with the use VGG19 as the backbone instead of lightweight CNN.

Validation method: We used LP recognition accuracy at sequence level Acc_{seq} to evaluate LP OCR, in which a LP is considered correctly recognised if all OCR characters are correct without any added characters. However, in cases where only one character is wrongly recognised, the above evaluation criteria cannot give meaning. Therefore, we use Acc_{char}, accuracy at character level to evaluation in addition.

Training: In order to increase the generalisation of the model, we randomly performed argumentation operations during training. Adam optimization algorithm was used [29] with an initial learning rate of 0.001, batch size of 32, maximum iteration number of 10000, and early stopping patience of 250.

TABLE V: Results of LP OCR in MTAVLP dataset

Techniques	Training set		Validation set		Testing set	
	Acc_{seq}	Acc_{char}	Acc_{seq}	Acc_{char}	Acc_{seq}	Acc_{char}
VOCR	**99.34%**	**99.8%**	**99.25%**	**99.9%**	**99.28%**	**99.7%**
LPRnet-STN [31]	92.3%	96.9%	94.1%	96.4%	94.8%	96.8%
CRNN [32]	92.9%	96.9%	93.6%	97.2%	93.4%	97.1%
CRNN (VGG19)	92.3%	95.5%	92.5%	95.4%	93.0%	96.2%

Results and Comparison: Table V shows that VOCR has dominant efficiency in this task with $Acc_{seq} = 99.28\%$, $Acc_{char} = 99.7\%$. As limitations of CTC loss (mentioned in Section II-A), LPRNet-STN, CRNN, and CRNN (VGG19) have worsen results compare to VOCR. The differences are more significant in sequence level. This suggests that CTC based approaches having difficulties in correctly recognizing the character string of an input image.

90% of LP images in the MTAVLP dataset have white background. The imbalance is because white plate is a license plate for individuals. To solve the imbalance data problem, we used image negative and histogram matching algorithm

TABLE VI: Affect of unbalanced data to VOCR

Dataset	MTAVLP-color		MTAVLP	
	Acc_{seq}	Acc_{char}	Acc_{seq}	Acc_{char}
Training set	98.8%	99.6%	**99.34%**	**99.8%**
Validation set	98.7%	99.7%	**99.25%**	**99.9%**
Testing set	98.5%	99.5%	**99.28%**	**99.7%**

to generate MTAVLP-color dataset including blue, red, yellow plates from white plates in all 3 training, testing and validation sets. We did not re-train, but used the VOCR model trained in the original MTAVLP dataset to test for it performance in the MTAVLP-color dataset. The results show that the accuracy decreases at only 1% in all datasets as in Table VI. Most color plates that are miss-recognized are usually those that are incorrect in the original set or uncleared in characters due to the data generation process. This suggests that the colors of LP has a slight affect on the results of LP OCR.

Previous studies on ALPR in unconstrained environment[] record an average accuracy at around 89.33% on five different datasets, while our proposed model reaches the accuracy over 99%. It should be noted that these evaluations were performed in different datasets, and our proposed model was performed on the MTAVLP dataset - a dataset of Vietnamese license plates. Applying our proposed models in other datasets is worth to considerer; however, it is out of scope of this paper, so we leave it for future work.

D. Overall performance

Our proposed ALPR system was set up test using a single GPU Nvidia GeForce RTX 2080 and without using any inference optimization techniques such as Tensor-RT. We recorded the processing speeds of stages as follows: The processing speed was 38.28 FPS in the stage of car detection using YOLOv2; In the LP detection stage, DDRNet23sh could be run at 103.5 FPS; and in the LP recognition stage, we got 36 FPS with VOCR. Consequently, our ALPR system can be run at 15.73 FPS in a single Nvidia GeForce RTX 2080 without using any inference optimization techniques. Based on these results, we believe that our ALPR system can run in real-time when being deployed using additional inference optimization techniques.

V. Conclusion and Future Work

This paper proposes a model for ALPR in Vietnam in unconstrained environment that focused on improvements on two stages: Detecting the areas containing a plate number and recognising LP characters. Our improvements have been empirically proved to be effective for Vietnamese ALPR in unconstrained environment. We also built a dataset of Vietnamese car license plates named MTAVLP that is large and covers various conditions of image capturing. To tackle the problem of non-rectangular object detection, an approach based on keypoints detection was used and DDRNet23sh architecture was modified. This improvement get the highest mean IOU $mIOU = 95.01\%$ and precision $P_{75} = 99.5\%$. In this paper, we also propose VOCR based on segmentation-free that uses CNN architecture combined with encoder-decoder RNN and attention mechanism. The performance accuracy of our proposed architecture is up to $Acc_{seq} = 99.28\%$ and $Acc_{char} = 99.7\%$.

In this paper, we focused on ALPR in daytime. In the future, we will expand our system to other environments such as in night time and low-light conditions. Our future work also includes applications of other pre-processing techniques to improve quality of LP detection for images captured in low-light, blurry and noisy conditions, in order to increase the accuracy of license plate recognition in Vietnam in more challenging conditions.

References

[1] Vinh Mai, Duoqian Miao, and Ruizhi Wang. “Building a license plate recognition system for Vietnam tollbooth”. In: *ACM International Conference Proceeding Series* (Aug. 2012).

[2] Tran Duan et al. “Building an Automatic Vehicle License-Plate Recognition System”. In: *Proc. Int. Conf. Comput. Sci. RIVF* (Feb. 2005).

[3] Yuanduo Hong et al. “Deep Dual-resolution Networks for Real-time and Accurate Semantic Segmentation of Road Scenes”. In: *arXiv preprint arXiv:2101.06085* (2021).

[4] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. “Effective Approaches to Attention-based Neural Machine Translation”. In: *CoRR* (2015).

[5] Karen Simonyan and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *arXiv 1409.1556* (Sept. 2014).

[6] Gee-Sern Hsu et al. “Robust license plate detection in the wild”. In: (Aug. 2017).

[7] Joseph Redmon and Ali Farhadi. “YOLO9000: Better, Faster, Stronger”. In: (July 2017).

[8] Lele Xie et al. “A New CNN-Based Method for Multi-Directional Car License Plate Detection”. In: *IEEE Transactions on Intelligent Transportation Systems* (Jan. 2018).

[9] Max Jaderberg et al. “Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition”. In: (June 2014).

[10] C.A. Rahman, Wael Badawy, and Ahmad Radmanesh. “A real time vehicle’s license plate recognition system”. In: (Aug. 2003).

[11] Zied Selmi, Mohamed Ben Halima, and Adel Alimi. “Deep Learning System for Automatic License Plate Detection and Recognition”. In: (Nov. 2017).

[12] Teik Cheang, Yong Shean Chong, and Yong Haur Tay. “Segmentation-free Vehicle License Plate Recognition using ConvNet-RNN”. In: (Jan. 2017).

[13] Alex Graves et al. “Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural ’networks”. In: vol. 2006. Jan. 2006, pp. 369–376. DOI: 10.1145/1143844.1143891.

[14] Hui Li, Peng Wang, and Chunhua Shen. “Toward End-to-End Car License Plate Detection and Recognition With Deep Neural Networks”. In: *IEEE Transactions on Intelligent Transportation Systems* (Sept. 2017).

[15] Feng Zhang et al. “Distribution-Aware Coordinate Representation for Human Pose Estimation”. In: *CoRR* (2019).

[16] Jonathan Tompson et al. “Efficient object localization using Convolutional Networks”. In: June 2015, pp. 648–656. DOI: 10.1109/CVPR.2015.7298664.

[17] Feng Zhang et al. *Distribution-Aware Coordinate Representation for Human Pose Estimation*. 2019.

[18] Fisher Yu et al. *Deep Layer Aggregation*. 2019.

[19] Mark Everingham et al. “The Pascal Visual Object Classes (VOC) Challenge”. In: *International journal of Computer Vision* (June 2010). ISSN: 1573-1405.

[20] Tsung-Yi Lin et al. *Microsoft COCO: Common Objects in Context*. 2015.

[21] Sérgio Montazzolli and Claudio Jung. “License Plate Detection and Recognition in Unconstrained Scenarios”. In: (Sept. 2018).

[22] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. “Objects as Points”. In: *CoRR* (2019).

[23] Faisal Shafait and Thomas Breuel. “The Effect of Border Noise on the Performance of Projection-Based Page Segmentation Methods”. In: *IEEE transactions on pattern analysis and machine intelligence* (Apr. 2011).

[24] Junyoung Chung et al. “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling”. In: *CoRR* (2014).

[25] Dzmitry Bahdanau, Kyunghyun Cho, and Y. Bengio. “Neural Machine Translation by Jointly Learning to Align and Translate”. In: *ArXiv* (Sept. 2014).

[26] Rafael Müller, Simon Kornblith, and Geoffrey E. Hinton. “When Does Label Smoothing Help?” In: *CoRR* (2019).

[27] *https://thigiacmaytinh.com/tai-nguyen-xu-ly-anh/tong-hop-data-xu-ly-anh/*. 2014.

[28] Jing Han et al. “Multi-Oriented and Scale-Invariant License Plate Detection Based on Convolutional Neural Networks”. In: *Sensors* (Mar. 2019).

[29] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017.

[30] Bin Xiao, Haiping Wu, and Yichen Wei. *Simple Baselines for Human Pose Estimation and Tracking*. 2018.

[31] Sergey Zherzdev and Alexey Gruzdev. “LPRNet: License Plate Recognition via Deep Neural Networks”. In: (June 2018).

[32] Baoguang Shi, Xiang Bai, and Cong Yao. *An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition*. 2015.