A Single-Stage Multi-Style License Plate Recognition Method Based on Attention

Longbin Wu LONGBIN.WU@STU.PKU.EDU.CN
Sen Liu 2101212992@STU.PKU.EDU.CN
Yufei Xie MICHAELSHIRE@STU.PKU.EDU.CN
Haowei Lee 2301213363@STU.PKU.EDU.CN
Xiaohui Duan* DUAN@PKU.EDU.CN
Peking University

Editors: Vu Nguyen and Hsuan-Tien Lin

Abstract

Automatic license plate recognition is applied widely in life, but it is still a challenging task in open scenarios. The current ALPR methods require multiple recognitions in multi-license plate scenarios. Furthermore, they have complex recognition structures and insufficient recognition capabilities for multi-style license plates. To solve the above problems, this paper proposes a single-stage multi-style multi-license plate recognition method based on the attention mechanism: CLPRNet. We use a spatial attention module based on UNet to separate the license plate character sequence into each attention heatmap in order. This approach unifies license plates with different character lengths and different character rows into a single processing logic, thereby enabling CLPRNet to recognize multiple styles without additional style judgement branches. At the same time, we abandon the traditional method of cropping RoI from image or feature, and instead combine attention to recognize characters directly, which allows CLPRNet to recognize multiple license plates in a single pass. To address the issue of an inadequate number of multi-style license plate samples, this paper also proposes a multi-style license plate generation method. In the single-stage methods, CLPRNet demonstrates better detection performance on the CCPD dataset and better recognition performance on the FN, Rotate, and Tilt subsets of the CCPD dataset. Compared with the existing license plate recognition methods, CLPRNet can recognize more styles of license plates. Test results and ablation experiments have shown the effectiveness of our proposed method.

Keywords: ALPR, Single-Stage, Attention, Multi-Style License Plate

1. Introduction

Auto License Plate Recognition (ALPR) is an important part of intelligent transportation system and has a wide range of applications in life such as parking lot management, toll station management, and traffic monitoring. In fixed scenarios, ALPR has achieved high accuracy. However, in open scenarios such as autonomous driving, license plate images become more diverse including oblique views, uneven illumination, and blurring. Therefore, ALPR in open scenarios still is a challenging task. Classic ALPR methods use manual extraction of image features for detection and recognition (Hsu et al., 2013; Yuan et al.,

* Corresponding author

2017). With the rapid development of deep learning, deep neural networks have become the main method of ALPR and have shown better performance compared to classic methods.

The goal of ALPR is to detect license plates in an image and recognize the corresponding license plate characters, so we can divide the ALPR methods into two main categories: multi-stage and single-stage. Multi-stage methods divide the task into the detection stage and recognition stage, and cascade the two stages. Some approaches incorporate an image rectification phase in between the two stages (Meng et al., 2018; Wang et al., 2022). The detection stage can use a generic target detection model (Zou et al., 2020; Ke et al., 2023) or redesign the detection network (Wang et al., 2022; Silva and Jung, 2022). Each detected license plate is cropped and rectified from the image according to the coordinates, and then fed to the recognition stage. Because the recognition stage can only recognize one license plate at a time, multi-stage methods need to perform multiple recognitions in multi-license plate scenarios. Multi-stage methods can optimize each stage individually, so multi-stage methods generally have higher accuracy. However, each stage of multi-stage methods requires separate training. Therefore, single-stage methods within a unified architecture are increasingly being focused on, which can be trained end-to-end. Single-stage methods generally obtain license plate coordinates by detection head, and then combine feature cropping methods such as RoIPooling for recognition. In single-license plate scenarios where the number of license plates is fixed, single-stage methods are easier to implement (Xu et al., 2018; Qin and Liu, 2022; Liu et al., 2023). Existing single-stage multi-license plate ALPR methods recognize license plates by adapting the generic target detection model (Duan et al., 2021; Gong et al., 2022; Peng et al., 2023). With feature cropping methods, the features of each license plate are cropped from the feature map according to the detection coordinates and then fed into the recognition sub-network to recognize the characters. By sharing features, feature cropping methods reduce computational complexity, but still require multiple recognition depending on the number of license plates. Therefore, we need a single-stage multi-license plate recognition method that can recognize multiple license plates in a single pass to reduce the inference time in multi-license plate scenarios.

Chinese license plates have multiple styles. The existing license plate recognition methods for different numbers of characters are filling the default value(Fan and Zhao, 2022), connectionist temporal classification (CTC) (Ke et al., 2023), and style judgement branch (Jiang et al., 2023). The recognition methods for different rows of characters currently also use the additional branch(Xu et al., 2022). These methods generally only enhance the recognition capability for particular styles of license plates and cannot be applied to other styles. It is worth noting that realizing the recognition for different styles by adding the extra branch also increases the complexity of network design. Meanwhile, the current large-scale open scenario license plate dataset(Xu et al., 2018) has only a single license plate style, while the existing multi-style license plate datasets (Yuan et al., 2017; Zhang et al., 2021)have a limited number of samples. Therefore, we need a multi-style license plate recognition method that can implement the recognition of multiple styles with a single network, and increase the diversity of license plate styles in the training data.

In this paper, we propose a single-stage multi-style multi-license plate ALPR model: Comprehensive License Plate Recognition Net (CLPRNet). This approach has the following four insights:

1. CLPRNet is a single-stage method that implements detection and recognition under a unified architecture, eliminates intermediate rectification methods, and enables end-to-end training.

2. CLPRNet can detect and recognize multiple license plates simultaneously in an image. Compared to existing methods, CLPRNet abandons image cropping and feature cropping methods, and realizes single-shot recognition for multiple license plates.

3. We introduce an attention-based approach for license plate recognition and enhance the performance of spatial attention module through the UNet structure. With this attention mechanism, the attention heatmap of each character can be generated according to the character position index. It allows the model to recognize different styles of license plates by using a single structure.

4. We provide a multi-style license plate generation method and establish a challenging multi-style license plate test dataset.

The rest of this paper is organized as follows. We first present related works in Section II. Then, the proposed method is described in detail in Section III. Experimental results and ablation studies are discussed in Section IV. Finally, we conclude the paper in Section V.

2. Related Works

We will briefly describe the existing ALPR methods in the following.

2.1. Multi-Stage Methods

The multi-stage ALPR is divided into the detection stage and recognition stage depending on the task implemented in the stage.

The license plate detection task can be considered as a case of the target detection task, so the methods used in the target detection task can be introduced to license plate detection. TE2E(Li et al., 2019) used RPN to obtain the proposed region. Duan et al. (2021) used SSD to detect multiple license plates in an image. The YOLO families have also been widely introduced to the license plate detection task due to their widespread application in the target detection task(Zou et al., 2020; Laroca et al., 2021; Zou et al., 2022; YU and LIU, 2023; Ke et al., 2023). In addition, there are also separately designed structures (Wang et al., 2022; Silva and Jung, 2022) for the license plate detection task. Meanwhile, some of the methods additionally detect other parameters such as rotation angles for the rectification stage (Silva and Jung, 2018; Wang et al., 2022).

The recognition stage is cascaded after the detection stage. Each detected license plate will be cropped from the image based on the coordinates and fed into the license plate recognition stage. Common OCR (Silva and Jung, 2018)is not effective in open scenarios, so the license plate recognition network needs to be redesigned. The backbone of the recognition network can uses CNN(Meng et al., 2018; Zherzdev and Gruzdev, 2018) or Bi-LSTM(Zou et al., 2020) to extract character features. The head of the recognition network applies the FC classifier (Fan and Zhao, 2022), or introduces CTC (Qin and Liu, 2022; Peng et al., 2023) to obtain the sequence of license plate characters. To reduce the difficulty of recognition, pre-processing methods such as rectification(Wang et al., 2022; Jiang et al., 2023), super-resolution(Lee et al., 2019; Pan et al., 2023), and character segmentation (Meng et al., 2018; Chen and Wang, 2022) can be applied before license plates are fed into the recognition network.

2.2. Single-Stage Methods

Single-stage methods can be divided into single-license plate and multi-license plate methods depending on the number of license plates to be detected.

RPNet(Xu et al., 2018) is the first single-stage single-license plate recognition method trained on large-scale datasets. RPNet used CNN for feature extraction, applied RoIPooling to crop the RoI from the feature map according to the coordinates obtained from the detection head, and then fed it to seven classifiers for character recognition. Qin and Liu (2022) added the prediction of the corner points of the license plate and used RoIAlign for rectification, which improves the robustness of recognizing license plates with different rotation angles. The simple classification loss can be replaced by CTCLoss(Wang et al., 2019) for the prediction of license plate character sequences. The above methods can only recognize a single license plate because they are designed to obtain only a single coordinate.

Existing single-stage multi-license plate recognition methods are implemented by adapting existing single-stage generic target detection methods. The post-processing methods detect the license plate and each character simultaneously, and then fuse the coordinate information through post-processing(Hendry and Chen, 2019; Huang et al., 2021b). Unlike the method of detecting license plates and characters simultaneously, Peng et al. (2023) used YOLOX to detect the license plate coordinates, and then sent the features of each license plate into LPRNet to recognize characters by RoIPooling. Gong et al. (2022) used RetinaNet to detect the license plate coordinates and rotation angles, and applied RoIAlign to obtain the RoI of the license plate feature. Although the above methods can detect multiple license plates in a single pass, the recognition head need to run repeatedly in multi-license plate scenarios because the recognition head can only recognize a license plate at a time.

2.3. Multi-Style License Plate

Chinese license plates have a variety of license plate styles including different colors, different numbers of characters, and different rows. The large-scale open scenario license plate dataset: CCPD(Xu et al., 2018) has a variety of images but only one style of license plate (blue, 7-character). The common multi-style license plate datasets are PKUData (Yuan et al., 2017) and CLPD (Zhang et al., 2021). However, the number of samples in these datasets is limited ($<$5k) and the diversity of images is not sufficient. CRPD (Gong et al., 2022) has the most samples in the multi-style license plate datasets, but its styles do not cover all the styles.

Some methods are designed to target multi-style license plates, while most methods include only a small number of multi-style license plates in the training data, which have not been directly tested for their multi-style license plate recognition capability. The methods using CTC(Wang et al., 2019; Qin and Liu, 2022) theoretically support a variable number of characters, but there are no explicit test results in these methods. Fan and Zhao (2022) considered the number of license plate characters to be eight and filled in the default value for license plates with insufficient characters. Jiang et al. (2023) determined the number of characters by using additional branche. Xu et al. (2022) determined whether a license plate has two rows by using additional classifier and used 2D attention to align the features. Huang et al. (2021a) used post-processing to judge whether it is a two-row style.

3. Proposed Method

3.1. Model

The proposed Comprehensive License Plate Recognition Net (CLPRNet) is a single-stage model that can detect multiple license plates and recognize multiple styles of license plates within a unified network. As shown in Fig. 1, CLPRNet contains of five parts: backbone, attention module, detection head, recognition head, and fusion.

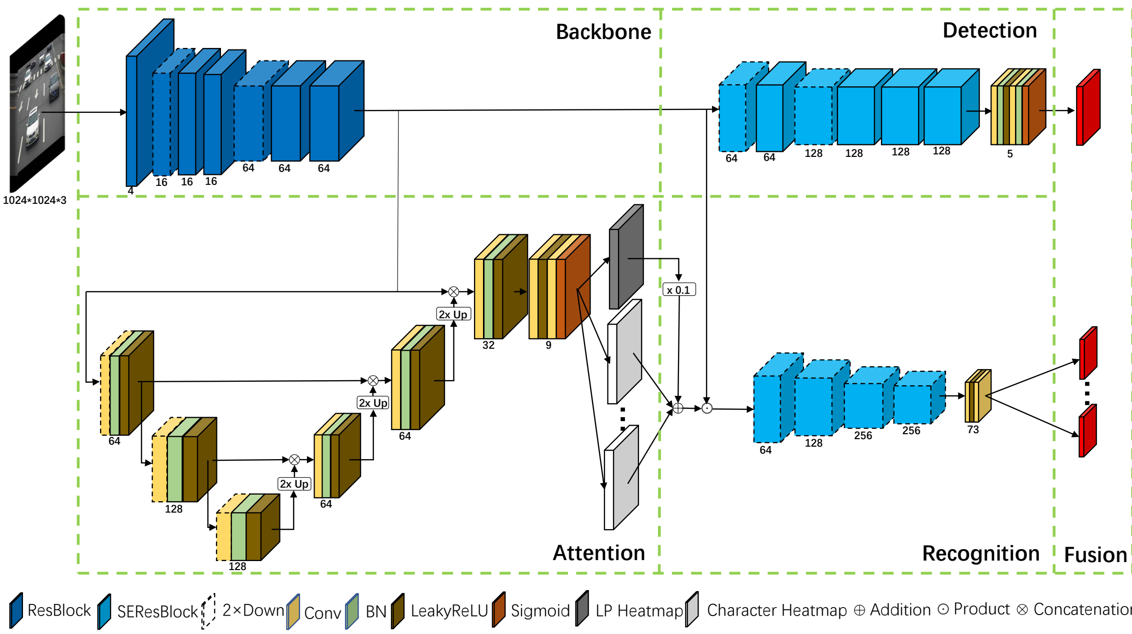

Figure 1: The architecture of CLPRNet.

At the image size of 1920*1080, a single license plate character is usually only a dozen pixels, which is a typical small target in the target detection task. In order to extract and retain the license plate character information, we choose 1024*1024 as the input size of the model. We increase the parameters of the shallow network and reduce the number of downsamples in the backbone. Backbone uses a 7-layer stack of ResBlock (He et al., 2016). We use stride convolution for downsampling instead of using pooling. After two downsamples, the backbone sends the feature F^b to the attention module, detection head, and recognition head, respectively.

3.1.1. Attention Module

The characters in different styles of license plates are standardized and have the same features. Therefore, if we can recognize individual characters sequentially by separating them according to the order of license plate character index, all styles of license plates can be recognized by the same processing method.

We use the UNet(Ronneberger et al., 2015) structure for attention learning. The feature map F^b obtained from the backbone is downsampled three times and upsampled three times to get more contextual information. Then each point in the feature map is classified into license plate or character by the classifly head. Let the output of attention module be $H_a * W_a * \left(M^{lp}, M_i^{ch}\right)$, where $H_a * W_a$ is the size of feature map F^b, M^{lp} represents the heatmap of license plate, and M_i^{ch} represents the heatmap of each character. Let n be the total number of characters and i be the index of the characters, then M_i^{ch} can be formulated as

$$M_i^{ch} = \begin{cases} GT, & 1 \le i \le n \\ 0, & n < i \le 8. \end{cases} \tag{1}$$

The heatmap of the i-th character of each license plate is in M_i^{ch}. For license plates with less than 8 characters, the heatmap of missing characters is empty. As shown in Fig. 2, the license plate character sequence is segmented into individual characters in index order based on this attention mechanism. We only need to recognize a single character and then concatenate the character sequence in order. This makes it compatible to handle different numbers of characters and different numbers of rows without any other additional processing. Meanwhile, we also introduce a loss function on the attention heatmap to guide the learning of attention.

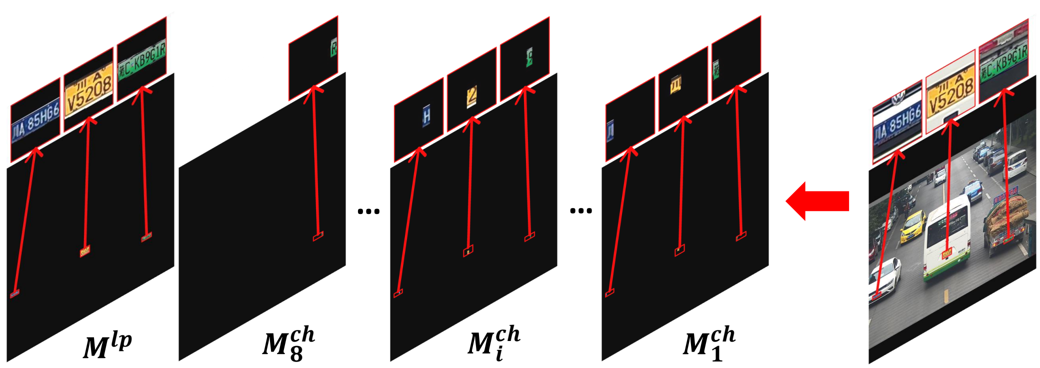

Figure 2: The heatmap of characters and license plates. This image includes a 7-character license plate, an 8-character license plate, and a two-row license plate.

3.1.2. Detection and Recognition

Recognition. The recognition head is responsible for recognizing each license plate character. We divide the input image into $H_r * W_r$ grids. When the license plate image falls within a grid, this grid is responsible for recognizing the characters of this license plate. The recognition head will fuse the feature map F^b and the heatmap of the attention module to recognize each license plate character. The feature of the i-th character of the license plate can be expressed as

$$F_i^{ch} = F^b * M_i^{ch}. \tag{2}$$

As shown in Fig. 3, using the single character feature results in ambiguous samples when the license plate crosses multiple grids. To solve this problem, we propose the edge sight design. When using the heatmap of a single character, the partial heatmap of the license plate is preserved, allowing the network to focus on the features of characters while being able to see the global information of the license plate. Thus, the features sent to the recognition head can be represented as

$$F_i^r = F_i^{ch} + 0.1 * F^{lp} = F^b * M_i^{ch} + 0.1 * F^b * M^{lp}. \tag{3}$$

The recognition head needs to recognize characters at each index in the sequence of license plate characters. Therefore, the recognition head at each index can share parameters. Because the recognition head needs to recognize eight indexes sequentially, it adopts a lightweight design using a 4-layer stack of SE ResBlock (Hu et al., 2018) and downsampling four times. The number of character categories is 73, so the output of the recognition head is $H_r * W_r * (73 * 8)$.

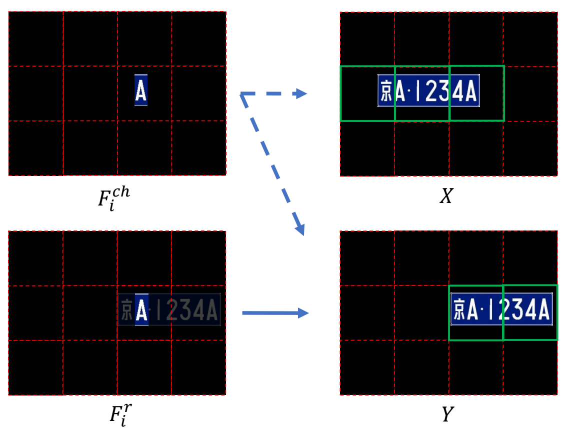

Figure 3: The ambiguousy sample and edge sight design. The feature map F_i^{ch} retains only a single character A. So we cannot confirm whether it is the second character from X or the seventh character from Y.

Detection. The detection head is responsible for predicting the license plate coordinates. The detection head uses the anchor point as a coordinate reference point, which is an anchor-free detector. We adopt the approach of FCOS (Tian et al., 2019) for the anchor point arrangement. The output of detection head is $H_d * W_d * (l, t, r, b, c)$, where l, t, r, b are the distances from the anchor point to the target box. c is the detection confidence, which is labeled as the IoU of the prediction box and the target box. Because most of the license plates are small targets, the number of anchor points is set as 64*64 to ensure that there can be an anchor on every license plate at least. The detection head uses a 6-layer stack of SE ResBlock.

Fusion. The integration of detection and recognition results is done at fusion. If a detection anchor falls on a recognition grid, we distribute the recognition results of this grid to this anchor. The confidence score of each license plate is the product of the detection confidence and the recognition confidence:

$$Score = Pr_{det} * Pr_{rec} = c * \min(p_i^{ch}), \tag{4}$$

where p_i^{ch} is the probability of each character. Finally, we perform Non-Maximum Suppression (NMS) on the output to remove redundant results.

3.1.3. Loss Function

The loss function of CLPRNet consists of three components: attention, detection, and recognition, which are computed as follows:

$$L = \lambda_1 L_{att} + \lambda_2 L_{det} + \lambda_3 L_{reg}, \tag{5}$$

$$L_{att} = BCE(M_{gt}^{lp}, M^{lp}, \beta_1) + \sum_i BCE(M_{gt}^{ch}, M_i^{ch}, \beta_2), \tag{6}$$

$$L_{det} = \frac{\alpha_1}{N_{pos}} \sum -\log\left(1 - IoU(box_{gt}, box)\right) + \frac{\alpha_2}{N_{pos}} \sum (c_{gt} - c)^2 + \frac{\alpha_3}{N_{neg}} \sum (c_{gt} - c)^2, \tag{7}$$

$$L_{reg} = \frac{1}{N_{pos}} \sum CE(C_{gt}, C). \tag{8}$$

N_{pos} denotes the number of positive samples and N_{neg} denotes the number of positive samples. BCE is the binary cross-entropy loss with weights. CE is the cross-entropy loss.

3.2. Style Augmentation

In order to increase the number of multi-style license plate samples, we propose a generative approach to perform style augmentation. First, various styles of license plates are generated based on the national license plate standard and standard character images. Then, the brightness of the generated license plate image is adjusted according to the brightness of the original license plate in the selected image from the existing dataset. Image augmentation, such as noise addition and blurring, is applied to the generated license plate image. Finally, the generated license plate image is placed on the original image based on the four corner points of the original license plate. As shown in Fig. 4, multi-style license plate samples are generated that are as similar as possible to the real image while preserving the background information.

4. Experiment

4.1. Implementation Details and Datasets

We train the CLPRNet by 80 epochs with the batch size of 8 on RTX3090. We use the SGD optimizer with the momentum of 0.9. The initial learning rate is set to 0.01 and reduced

Figure 4: The generated multi-style license plate images.

by half every 5 epochs for the last 20 epochs. Each parameter in the loss function is set as $\lambda_1 = 10, \lambda_2 = 1, \lambda_3 = 0.2, \alpha_1 = 0.2, \alpha_2 = 1, \alpha_3 = 0.1, \beta_1 = 0.1, \beta_2 = 0.05$. The IoU threshold of NMS is 0.3, and the result with a confidence score greater than 0.3 will be output.

CCPD(~250k) as a large-scale open scenario license plate dataset is still the benchmark for the ALPR task due to its diversity and number of images. We train with both CCPD(2018) and CCPD(2020). Following the previous works, we train on the Base subset and test on all subsets of CCPD(2018). CCPD(2020) is also trained according to the publisher's division. Because CCPD contains only one license plate per image, we add CRPD(~30k) as a supplement. Besides style augmentation, we also implement image augmentation including distortion, flipping, scaling, color transformation, blurring, and occlusion.

4.2. Comparison

4.2.1. Test Results of CCPD

Detection. We first compare the license plate detection capability on the CCPD dataset. Following previous works, we set IoU>0.7 as the correct sample. Because CCPD is a single-license plate dataset, we can directly compare the accuracy. As shown in Table. 1, we can see that the current single-stage ALPR methods cannot detect the Green subset which consists of 8 characters. Meanwhile, the detection performance of CLPRNet almost comprehensively exceeds the existing single-stage methods. For the DB subset, CLPRNet can improve the accuracy by 6.5%. There is a 1.1% improvement in the average accuracy of the dataset. Meanwhile, we also compare it with the multi-stage methods. CLPRNet achieves the same performance as the SOTA result of the multi-stage method on the Rotate and Tilt subsets. Although CLPRNet does not completely outperform the SOTA of the multi-stage method, it is only 0.2% behind in the average accuracy indicating that the single-stage CLPRNet has similar detection capability to the SOTA of the multi-stage method. In addition, compared with existing methods, CLPRNet can detect more styles of license plates.

Recognition. Next, we compare the recognition performance on the CCPD dataset. Referring to the previous works, we also set IoU>0.6 and all the license plate characters are correctly recognized as the correct samples. From Table. 2, we can see that the current single-stage methods cannot recognize the Green subset. CLPRNet achieves the best recognition performance of the single-stage methods on the FN, Rotate, and Tilt subsets.

Table 1: Comparison of detection on the test set of CCPD. 'Multi' indicates whether or not it supports the detection of multi-license plates. '-' means the result not provided. 'BGYWDK' denotes blue, green, yellow, white, two-row, and black style respectively. 'Avg' denotes the average accuracy of dateset.

Single-stage	Multi	Style	Avg	Base	DB	FN	Rot	Til	Wea	Cha	Gre
RPNet(2018)	×	B	94.5	99.3	89.5	85.3	94.7	93.2	84.1	92.8	×
TE2E(2019)	√	B	94.2	98.5	91.7	83.8	95.1	94.5	83.6	93.1	×
MTLPR(2019)	×	B	97.7	-	-	-	-	-	-	-	×
Qin and Liu (2022)	√	BYWK	97.8	**99.7**	92.7	94.5	98.5	96.2	99.1	93.5	×
CLPRNet(Our)	√	**BGYWDK**	**98.9**	99.3	**99.2**	**96.8**	**99.9**	**99.8**	**99.6**	**95.7**	**97.4**
Multi-stage											
SSD300(2018)	√	B	94.4	99.1	89.2	84.7	95.6	94.9	83.4	93.1	×
FasterRCNN(2018)	√	B	92.9	98.1	92.1	83.7	91.8	89.4	81.8	83.9	×
YOLOv3(2022)	√	BGYW	96.0	97.1	97.2	93.3	91.6	94.6	97.9	90.5	-
YOLOv4(2022)	√	BGYW	95.1	96.8	93.7	93.1	93.5	94.7	96.6	90.5	-
Silva and Jung (2022)	√	B	89.7	-	86.1	84.3	94.8	93.0	95.7	93.4	×
VertexNet(2022)	×	BGYWK	**99.1**	**99.6**	**99.4**	97.0	**99.9**	**99.8**	**99.9**	95.6	-
MD-YOLO(2023)	√	BGYWK	98.2	98.6	99.1	96.1	99.5	99.5	99.7	93.5	-
LPDNet(2023)	√	BGYWD	98.4	97.3	97.2	**97.7**	98.9	98.9	98.9	**98.6**	**98.8**

Table 2: Comparison of recognition on the test set of CCPD. 'NotRep' indicates no need to repeat recognition in the multi-license plate scenarios. '※' indicates that detecting multiple license plates is not supported. '-' means the result not provided. 'BGYWDK' denotes blue, green, yellow, white, two-row, and black style respectively. 'Avg' denotes the average accuracy of dateset. 'FPS' includes detection and recognition time.

Single-stage	NotRep	Style	Avg	Base	DB	FN	Rot	Til	Wea	Cha	Gre	FPS
RPNet(2018)	※	B	95.5	98.5	96.9	94.3	90.8	92.5	87.9	85.1	×	85
TE2E(2019)	×	B	94.4	97.8	94.8	94.5	87.9	92.1	86.8	81.2	×	3
MTLPR(2019)	※	B	98.8	-	-	-	-	-	-	-	×	64
Qin and Liu (2022)	×	BYWK	97.6	99.5	93.3	93.7	98.2	95.9	98.9	92.9	×	26
Gong et al. (2022)	×	BYWD	97.9	98.3	**98.0**	97.2	92.5	93.7	90.7	87.9	×	35
Xu et al. (2022)	×	BYWD	**98.8**	-	-	-	-	-	-	-	×	-
Peng et al. (2023)	×	B	98.7	**99.9**	91.5	96.2	98.6	96.3	**99.5**	**94.4**	×	-
CLPRNet(Our)	√	**BGYWDK**	98.0	99.4	97.4	**97.3**	**99.2**	**98.9**	97.1	85.6	**91.9**	32
Multi-stage												
SSD300+HC(2018)	×	B	94.4	99.1	89.2	84.7	95.6	94.9	83.4	93.1	×	-
Zou et al. (2020)	×	BGYWK	97.8	99.3	98.5	98.6	92.5	96.4	99.3	86.6	-	-
Zhang et al. (2021)	×	BGYWK	98.5	99.6	98.8	98.8	96.4	97.6	98.5	88.9	-	40
SCR-Net(2022)	×	BYWK	99.5	**99.9**	99.7	99.4	**99.9**	**99.9**	99.4	94.8	×	87
Zou et al. (2022)	×	BGYW	97.5	99.2	98.1	98.5	90.3	95.2	97.8	86.2	-	-
MRNet(2023)	×	BGYWK	**99.8**	**99.9**	**99.9**	**99.8**	**99.9**	99.8	**99.6**	**97.7**	-	**98**
CRNet(2023)	×	BGYWD	98.9	**99.9**	97.7	98.4	99.5	98.6	99.5	97.4	**95.9**	78

Because the recognition stage is optimized individually, multi-stage methods tend to have better recognition performance than single-stage methods. In existing methods, the multi-stage methods crop the license plates from the image based on the detected coordinates, while the single-stage methods crop the features from the feature map. As a result, existing methods have to repeat the recognition according to the number of license plates in the multi-license plate scenarios. CLPRNet does not use cropping methods, making it possible to recognize multiple license plates in a single pass. Meanwhile, CLPRNet can recognize more styles of license plates.

4.2.2. Test Results of CRPD

Finally, we evaluate the multi-license plate and multi-style recognition capability of CLPR-Net on the CRPD dataset. We set IoU>0.6 and all the license plate characters are correctly recognized as the correct samples. As shown in Table. 3, we can see that CLPRNet has higher performance compared to the generic target detection model cascading CRNN. Compared to the method proposed by Gong et al. (2022), CLPRNet has better precision metrics. At the same time, CLPRNet realizes single-shot recognition for multiple license plates and recognizes more styles of license plates.

Table 3: Comparison of recognition on the test set of CRPD. 'NotRep' indicates no need to repeat recognition in the multi-license plate scenarios. 'BGYWDK' denotes blue, green, yellow, white, two-row, and black style respectively. 'R' and 'P' denote recall and precision respectively.

Method	NotRep	Style	Single		Double		Multi		All	
			R	P	R	P	R	P	R	P
YOLOv3+ CRNN(2022)	×	BYWD	73.7	59.4	74.6	64.4	66.2	61.3	73.0	61.0
YOLOv4+ CRNN(2022)	×	BYWD	87.3	68.4	90.4	41.2	88.9	36.8	84.4	60.5
FasterRCNN+ CRNN(2022)	×	BYWD	81.4	71.7	81.1	77.9	69.3	75.2	79.9	73.7
Gong et al. (2022)	×	BYWD	**96.3**	83.6	**95.8**	84.5	**90.8**	85.0	**95.4**	84.1
CLPRNet(Our)	√	**BGYWDK**	92.5	**85.8**	91.8	**87.3**	87.0	**87.6**	90.6	**87.0**

4.3. Ablation Study

To verify the effectiveness of the proposed method in multi-style multi-license plate scenarios, we design ablation experiments on the CRPD dataset, as shown in Table. 4.

Attention. Ablation experiments about the attention mechanism include removing the attention module, removing the edge sight, and abandoning the UNet structure. We can see that the UNet-based spatial attention module is a very efficient structure that can improve the recall by more than 11%. The edge sight design plays the role to reduce the ambiguousy samples.

Backbone. We test increasing the number of layers and the number of downsamples in the backbone. It can be seen that too many downsamples affects the learning of small targets like license plates.

Augmentation. Image and style augmentation enhance the robustness of the model by increasing the sample diversity. In particular, the style enhancement method proposed in this paper increases the multi-style recognition performance by 17%.

Table 4: Ablation study on the test set of CRPD. 'R' and 'P' denote recall and precision respectively.

Method	Single		Double		Multi		All	
	R	P	R	P	R	P	R	P
CLPRNet	**92.5**	**85.8**	**91.8**	**87.3**	**87.0**	**87.6**	**90.6**	**87.0**
w/o attention	78.6	76.0	80.0	81.3	71.6	77.7	77.2↓ 13.4	79.0↓ 8.0
w/o UNet	86.5	83.6	80.9	82.2	72.0	78.5	79.6↓ 11.0	81.5↓ 5.5
w/o edge sight	92.3	86.6	90.4	86.6	84.9	85.7	89.2↓ 1.4	86.4↓ 0.6
w/ 3x downsample	86.2	81.4	88.8	85.4	83.2	84.4	86.6↓ 4.0	84.2↓ 2.8
w/o image augmentation	78.7	73.2	86.8	83.2	84.7	86.0	84.3↓ 6.3	81.5↓ 5.5
w/o style augmentation	43.8	40.4	82.2	77.2	81.0	79.9	72.8↓ 17.8	69.1↓ 17.9

4.4. Results

As shown in Fig. 5, CLPRNet can detect and recognize multiple license plates in an image, and can recognize multiple styles of license plains including blue, green, yellow, black, white and two-row.

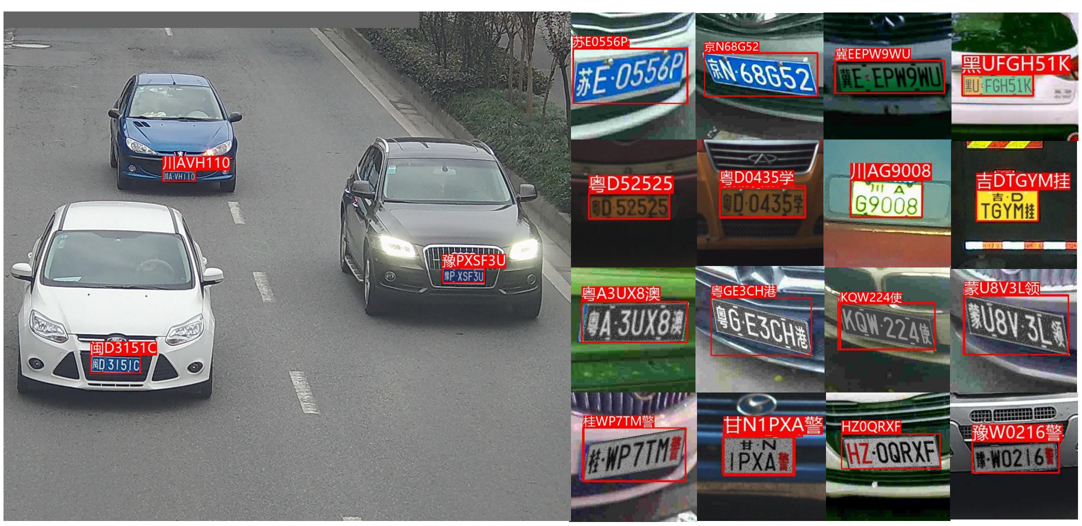

Figure 5: The inference results of CLPRNet.

The license plate styles of the existing license plate dataset do not cover all styles, and most of the samples are the Blue style. In order to better test the performance of

Table 5: The number of samples for each style in MSLPD. Accuracy is the test result of CLPRNet on each style.

Style	Number of samples	Accuracy
Blue	4038	92.4
Yellow	2194	92.3
Green	2311	85.6
White	736	74.5
Black	597	74.0
Two-row	565	71.2

ALPR methods to recognize different styles of license plates, we build a multi-style license plate test dataset: MSLPD. We collect the multi-style license plate samples from the test set of existing datasets and generate more samples of lacking styles through the proposed generative method. This test dataset is challenging with multiple styles of license plates and relatively balanced license plate styles. As shown in Table. 5, we perform tests on MSLPD and show the recognition results of different styles. Both the MSLPD dataset and the CLPRNet model are released at https://github.com/wulb97/CLPRNet.

5. Conclusion

We propose a single-stage multi-style multi-license plate recognition method: CLPRNet. By using an attention-based character recognition method, we achieve unified recognition of multi-style license plates and single-shot recognition in multi-license plate scenarios. We introduce a multi-style license plate generation method and build a multi-style license plate test dataset. Compared to existing methods, CLPRNet can recognize more styles of license plates. Meanwhile, CLPRNet has shown improved performance compared to existing methods on the CCPD and CRPD datasets. This approach can further optimize the model capability and speed according to real needs. It is also noted that the method based on the attention mechanism proposed in this paper can effectively reduce the background information and increase the amount of target information in the features. In the future, the method can be applied to the detection and recognition of small targets, such as remote sensing images and medical images.

References

C. L. Philip Chen and Bingshu Wang. Random-Positioned License Plate Recognition Using Hybrid Broad Learning System and Convolutional Networks. *IEEE Transactions on Intelligent Transportation Systems*, 23(1):444–456, 2022.

Na Duan, Jianjun Cui, Lizheng Liu, and Lirong Zheng. An End to End Recognition for License Plates Using Convolutional Neural Networks. *IEEE Intelligent Transportation Systems Magazine*, 13(2):177–188, 2021.

Xudong Fan and Wei Zhao. Improving Robustness of License Plates Automatic Recognition in Natural Scenes. *IEEE Transactions on Intelligent Transportation Systems*, 23(10): 18845–18854, 2022.

Yanxiang Gong, Linjie Deng, Shuai Tao, Xinchen Lu, Peicheng Wu, Zhiwei Xie, Zheng Ma, and Mei Xie. Unified Chinese License Plate detection and recognition with high efficiency. *Journal of Visual Communication and Image Representation*, 86:103541, 2022.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

Hendry and Rung-Ching Chen. Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning. *Image and Vision Computing*, 87:47–56, 2019.

Gee-Sern Hsu, Jiun-Chang Chen, and Yu-Zu Chung. Application-Oriented License Plate Recognition. *IEEE Transactions on Vehicular Technology*, 62(2):552–561, 2013.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-Excitation Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, Salt Lake City, UT, 2018. IEEE.

Qiuying Huang, Zhanchuan Cai, and Ting Lan. A New Approach for Character Recognition of Multi-Style Vehicle License Plates. *IEEE Transactions on Multimedia*, 23:3768–3777, 2021a.

Qiuying Huang, Zhanchuan Cai, and Ting Lan. A Single Neural Network for Mixed Style License Plate Detection and Recognition. *IEEE Access*, 9:21777–21785, 2021b.

Yu Jiang, Feng Jiang, Huiyin Luo, Hongyu Lin, Jian Yao, Jiaxin Liu, and Jia Ren. An Efficient and Unified Recognition Method for Multiple License Plates in Unconstrained Scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 24(5):5376–5389, 2023.

Xiao Ke, Ganxiong Zeng, and Wenzhong Guo. An Ultra-Fast Automatic License Plate Recognition Approach for Unconstrained Scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 24(5):5172–5185, 2023.

Rayson Laroca, Luiz A. Zanlorensi, Gabriel R. Gonçalves, Eduardo Todt, William Robson Schwartz, and David Menotti. An efficient and layout-independent automatic license plate recognition system based on the YOLO detector. *IET Intelligent Transport Systems*, 15 (4):483–503, 2021.

Younkwan Lee, Juhyun Lee, Hoyeon Ahn, and Moongu Jeon. SNIDER: Single Noisy Image Denoising and Rectification for Improving License Plate Recognition. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 1017–1026, Seoul, Korea (South), 2019. IEEE.

Hui Li, Peng Wang, and Chunhua Shen. Toward End-to-End Car License Plate Detection and Recognition With Deep Neural Networks. *IEEE Transactions on Intelligent Transportation Systems*, 20(3):1126–1136, 2019.

Sen Liu, Yufei Xie, Longbin Wu, Kunpeng Song, Kaiwen Gong, and Xiaohui Duan. A single-stage automatic license plate recognition network with Balanced-IoU loss. *Journal of Physics: Conference Series*, 2504(1):012039, 2023.

Ajin Meng, Wei Yang, Zhenbo Xu, Huan Huang, Liusheng Huang, and Changchun Ying. A robust and efficient method for license plate recognition. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 1713–1718, 2018.

Sen Pan, Si-Bao Chen, and Bin Luo. A super-resolution-based license plate recognition method for remote surveillance. *Journal of Visual Communication and Image Representation*, 94:103844, 2023.

Weiliang Peng, Zhenpeng Hu, and Tianyu Liang. End-to-end efficient cascade license plate recognition system in unconstrained scenarios. In *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*, pages 963–968, 2023.

Shuxin Qin and Sijiang Liu. Towards end-to-end car license plate location and recognition in unconstrained scenarios. *Neural Computing and Applications*, 34(24):21551–21566, 2022.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer, Cham, 2015.

Sergio M. Silva and Cláudio Rosito Jung. A Flexible Approach for Automatic License Plate Recognition in Unconstrained Scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):5693–5703, 2022.

Sérgio Montazzolli Silva and Cláudio Rosito Jung. License Plate Detection and Recognition in Unconstrained Scenarios. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 593–609, Cham, 2018. Springer International Publishing.

Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. FCOS: Fully Convolutional One-Stage Object Detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9626–9635, Seoul, Korea (South), 2019. IEEE.

Wanwei Wang, Jun Yang, Min Chen, and Peng Wang. A Light CNN for End-to-End Car License Plates Detection and Recognition. *IEEE Access*, 7:173875–173883, 2019.

Yi Wang, Zhen-Peng Bian, Yunhao Zhou, and Lap-Pui Chau. Rethinking and Designing a High-Performing Automatic License Plate Recognition Approach. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):8868–8880, 2022.

Hui Xu, Xiang-Dong Zhou, Zhenghao Li, Liangchen Liu, Chaojie Li, and Yu Shi. EILPR: Toward End-to-End Irregular License Plate Recognition Based on Automatic Perspective

Alignment. *IEEE Transactions on Intelligent Transportation Systems*, 23(3):2586–2595, 2022.

Zhenbo Xu, Wei Yang, Ajin Meng, Nanxue Lu, Huan Huang, Changchun Ying, and Liusheng Huang. Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, Lecture Notes in Computer Science, pages 261–277, Cham, 2018. Springer International Publishing.

Lina YU and Shaokun LIU. A Single-Stage Deep Learning-based Approach for Real-Time License Plate Recognition in Smart Parking System. *International Journal of Advanced Computer Science and Applications*, 14(9), 2023.

Yule Yuan, Wenbin Zou, Yong Zhao, Xinan Wang, Xuefeng Hu, and Nikos Komodakis. A Robust and Efficient Approach to License Plate Detection. *IEEE Transactions on Image Processing*, 26(3):1102–1114, 2017.

Linjiang Zhang, Peng Wang, Hui Li, Zhen Li, Chunhua Shen, and Yanning Zhang. A Robust Attentional Framework for License Plate Recognition in the Wild. *IEEE Transactions on Intelligent Transportation Systems*, 22(11):6967–6976, 2021.

Sergey Zherzdev and Alexey Gruzdev. LPRNet: License Plate Recognition via Deep Neural Networks. *arXiv preprint arXiv:1806.10447*, 2018.

Yongjie Zou, Yongjun Zhang, Jun Yan, Xiaoxu Jiang, Tengjie Huang, Haisheng Fan, and Zhongwei Cui. A Robust License Plate Recognition Model Based on Bi-LSTM. *IEEE Access*, 8:211630–211641, 2020.

Yongjie Zou, Yongjun Zhang, Jun Yan, Xiaoxu Jiang, Tengjie Huang, Haisheng Fan, and Zhongwei Cui. License plate detection and recognition based on YOLOv3 and ILPRNET. *Signal, Image and Video Processing*, 16(2):473–480, 2022.