Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline

Zhenbo Xu[1,2], Wei Yang[1](✉), Ajin Meng[1,2], Nanxue Lu[1,2], Huan Huang[2], Changchun Ying[2], and Liusheng Huang[1]

[1] School of Computer Science and Technology,
University of Science and Technology of China, Hefei, China
qubit@ustc.edu.cn
[2] Xingtai Financial Holdings Group Co., Ltd., Hefei, Anhui, China

Abstract. Most current license plate (LP) detection and recognition approaches are evaluated on a small and usually unrepresentative dataset since there are no publicly available large diverse datasets. In this paper, we introduce CCPD, a large and comprehensive LP dataset. All images are taken manually by workers of a roadside parking management company and are annotated carefully. To our best knowledge, CCPD is the largest publicly available LP dataset to date with over 250k unique car images, and the only one provides vertices location annotations. With CCPD, we present a novel network model which can predict the bounding box and recognize the corresponding LP number simultaneously with high speed and accuracy. Through comparative experiments, we demonstrate our model outperforms current object detection and recognition approaches in both accuracy and speed. In real-world applications, our model recognizes LP numbers directly from relatively high-resolution images at over 61 fps and 98.5% accuracy.

Keywords: Object detection · Object recognition
Object segmentation · Convolutional neural network

1 Introduction

License plate detection and recognition (LPDR) is essential in Intelligent Transport System and is applied widely in many real-world surveillance systems, such as traffic monitoring, highway toll station, car park entrance and exit management. Extensive researches have been made for faster or more accurate LPDR.

However, challenges for license plate (LP) detection and recognition still exist in uncontrolled conditions, such as rotation (about 20° onwards), snow or fog

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-01261-8_16) contains supplementary material, which is available to authorized users.

V. Ferrari et al. (Eds.): ECCV 2018, LNCS 11217, pp. 261–277, 2018.
https://doi.org/10.1007/978-3-030-01261-8_16

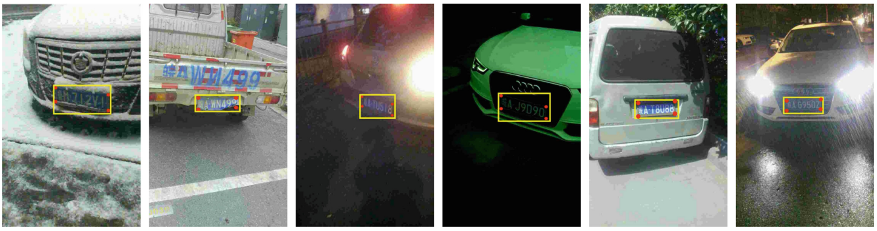


Fig. 1. Sample images from CCPD. Each image above is labelled with its bounding box (the yellow border) and four vertices location (four red dots). Other annotations are omitted here for simplicity. (Color figure online)

weather, distortions, uneven illumination, and vagueness. Most papers concerning LPDR [1–10] often validate their approaches on extremely limited datasets (less than 3,000 unique images), thus might work well only under some controlled conditions. Current datasets for LPDR (see Tables 1 and 2) either lack in quantity (less than 10k images) or diversity (collected from fixed surveillance cameras) because an artificial collection of LP pictures requires a lot of manpower. However, uncontrolled conditions are common in real world. A truly reliable LPDR system should function well in these cases. To aid in better benchmarking LPDR approaches, we present our Chinese City Parking Dataset (CCPD).

CCPD collects data from roadside parking in all the streets of one provincial capital in China where residuals own millions of cars. Each parking fee collector (PFC) works on one street from 07:30 AM to 10:00 PM every day regardless of weather conditions. For each parking bill, the collector is required to take a picture of the car with an Android handheld POS machine and manually annotates the exact LP number. It is worth noting that images from handheld devices exhibit strong variations due to the uncertain position and shooting angle of handheld devices, as well as varying illuminations and different backgrounds at different hours and on different streets (see Fig. 1). Each image in CCPD has detailed annotations in several aspects concerning the LP: (i) LP number. (ii) LP bounding box. (iii) Four vertices locations. (iv) Horizontal tilt degree and vertical tilt degree [11]. (v) Other relevant information like the LP area, the degree of brightness, the degree of vagueness and so on. Details about those annotations are explained in Sect. 3.

Most papers [3–7] separate LPDR into two stages (detection · recognition) or three stages (detection · segmentation · character recognition) and process the LP image step by step. However, separating detection from recognition is detrimental to the accuracy and efficiency of the entire recognition process. An imperfect bounding box prediction given by detection methods might make a part of the LP missing, and thus results in the subsequent recognition failure. Moreover, operations between different stages such as extracting and resizing the LP region for recognition are always accomplished by less efficient CPU, making LP recognition slower. Given these two observations, we come to the intuition that the LP recognition stage can exploit convolutional features extracted in the

LP detection stage for recognizing LP characters. Following that, we design a novel architecture named Roadside Parking net (RPnet) for accomplishing LP detection and recognition in a single forward pass. It's worth noting that we are not the first to design an end-to-end deep neural network which can localize LPs and recognize the LP number simultaneously. However, exploiting Region Proposal Network and Bi-directional Recurrent Neural Networks, the end-to-end model put forward by Li *et al.* [12] is not efficient as it needs 0.3 s to accomplish the recognition process on a Titan X GPU. By contrast, based on a simpler and more elegant architecture, RPnet can run at more than 60 fps on a weaker NVIDIA Quadro P4000.

Both CCPD and the code for training and evaluating RPnet are available under the open-source MIT License at: https://github.com/detectRecog/CCPD.

To summarize, this paper makes the following contributions:

- We introduce CCPD, the largest and the most diverse publicly available dataset for LPDR to date. CCPD provides over 250k unique car images with detailed annotations, nearly two orders of magnitude more images than other diverse LP datasets.
- We propose a novel network architecture for unified LPDR named RPnet which can be trained end-to-end. As feature maps are shared for detection and recognition and losses are optimized jointly, RPnet can detect and recognize LPs more accurately and at a faster speed.
- By evaluating state-of-the-art detection and recognition models on CCPD, we demonstrate our proposed model outperforms other approaches in both accuracy and speed.

2 Related Work

Our work is related to prior art in two aspects: publicly available datasets (as shown in Tables 1 and 2), and existing algorithms on LPDR. Except for [12] which proposed a unified deep neural network to accomplish LPDR in one step, most works separate LP detection from LP recognition.

2.1 Datasets for LPDR

Most datasets for LPDR [13–15] usually collect images from traffic monitoring systems, highway toll station or parking lots. These images are always under even sunlight or supplementary light sources and the tilt angle of LPs does not exceed 20°.

Caltech [13] and Zemris [14] collected less than 700 images from high-resolution cameras on the road or freeways and thus had little variations on distances and tilt degrees. The small volume of images is not sufficient to cover various conditions. Therefore, those datasets are not convincing to evaluate LP detection algorithms. Different from previous datasets, Azam *et al.* [10] and Hsu *et al.* [16] pointed out researches on LP detection under hazardous conditions were scarce and specifically looked for images in various conditions like great tilt angles, blurriness, weak illumination, and bad weather. Compared with CCPD, the shooting distance of these images varies little and the number of images is limited.

Current datasets for LP recognition usually collect extracted LP images and annotate their corresponding LP numbers. As shown in Table 2, SSIG [17] and UFPR [3] captured images by cameras on the road. These images were collected on a sunny day and rarely had tilted LPs. Before we introducing CCPD, ReId [15] is the largest dataset for LP recognition with 76k extracted LPs and annotations. However, gathered from surveillance cameras on highway toll gates, images in ReId are relatively invariant in tilt angles, distances, and illuminations. Lack of either quantity or variance, current datasets are not convincing enough to comprehensively evaluate LP recognition algorithms.

Table 1. A comparison of publicly available datasets for LP detection and CCPD. Var denotes variations.

	Zemris [14]	Azam [10]	AOLPE [16]	CCPD
Year	2002	2015	2017	2018
Number of images	510	850	4200	250k
Var in distance	✗	✗	✗	✓
Var in tilt degrees	✗	✓	✓	✓
Var in blur	✗	✓	✓	✓
Var in illumination	✓	✓	✓	✓
Var in weather	✓	✓	✓	✓
Annotations	✗	✗	✓	✓

Table 2. A comparison of publicly available datasets for LP recognition and CCPD. Var denotes variations.

	SSIG [17]	ReId [15]	UFPR [3]	CCPD
Year	2015	2017	2018	2018
Number of LPs	2000	76k	4500	250k
Var in tilt degrees	✗	✗	✗	✓
Var in blur	✗	✓	✓	✓
Var in illumination	✗	✗	✗	✓
Char dataset	✓	✗	✗	✓
Vertices annotation	✗	✗	✗	✓

2.2 LP Detection Algorithms

LP detection algorithms can be roughly divided into traditional methods and neural network models.

Traditional LP detection methods always exploit the abundant edge information [18–24] or the background color features [25,26]. Hsieh *et al.* [19] utilized morphology method to reduce the number of candidates significantly and thus speeded up the plate detection process. Yu *et al.* [21] proposed a robust method based on wavelet transform and empirical mode decomposition analysis to locate a LP. In [22] the authors analyzed vertical edge gradients to select true plate regions. Wang *et al.* [23] exploited cascade AdaBoost classifier and a voting mechanism to elect plate candidates. In [27] a new pattern named Local Structure Patterns was introduced to detect plate regions. Moreover, based on the observation that the LP background always exhibits a regular color appearance, many works utilize HSI (Hue, Saturation, Intensity) color space to filter out the LP area. Deb *et al.* [25] applied HSI color model to detect candidate regions and achieve 89% accuracy on 90 images. In [26] the authors also exploited a color checking module to help find LP regions.

Recent progress on Region-based Convolutional Neural Network [28] stimulates wide applications [3,4,12,16] of popular object detection models on LP detection problem. Faster-RCNN [29] utilizes a region proposal network which can generate high-quality region proposals for detection and thus detects objects more accurately and quickly. SSD [30] completely eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation in a single network. YOLO [31] and its improved version [32] frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities.

2.3 LP Recognition Algorithms

LP Recognition can be classified into two categories: (i) segmentation-free methods. (ii) segment first and then recognize the segmented pictures. The former [33,34] usually utilizes LP character features to extract plate characters directly to avoid segmentation or delivers the LP to an optical character recognition (OCR) system [35] or a convolutional neural network [15] to perform the recognition task. For the latter, the LP bounding box should be determined and shape correction is applied before segmentation. Various features of LP characters can be utilized for segmentation like Connected components analysis (CCA) [36] and character-specific extremal regions [37]. After segmentation, current high-performance methods always train a deep convolutional neural network [38] or utilize features around LP characters like SIFT [39].

3 CCPD Overview

In this section, we introduce CCPD – a large, diverse and carefully annotated LP dataset.

3.1 Data Creation and Privacy Concerns

CCPD collects images from a city parking management company in one provincial capital in China where car owners own millions of vehicles. The company employs over 800 PFCs each of which charges the parking fee on a specific street. Each parking fee order not only records LP number, cost, parking time and so on, but also requires PFC to take a picture of the car from the front or the rear as a proof. PFCs basically have no holidays and usually work from early morning (07:30 AM) to almost midnight (10:00 PM). Therefore, CCPD has images under diverse illuminations, environments in different weather. Moreover, as the only requirement for taking photos is containing the LP, PFC may shoot from various positions and angles and even makes a slight tremor. As a result, images in CCPD are taken from different positions and angles and are even blurred.

Apart from the LP number, each image in CCPD has many other annotations. The most difficult part is annotating the four vertices locations. To accomplish this task, we first manually labelled the four vertices locations over 10k images. Then we designed a network for locating vertices from a small image of LP regions and exploited the 10k images and some data augmentation strategies to train it. Then, after training this network well, we combined a detection module and this network to automatically annotate the four vertices locations of each image. Finally, we hired seven part-time workers to correct these annotations in two weeks. Details about the annotation process are provided in the supplementary material.

In order to avoid leakage of residents' privacy, CCPD removes records other than the LP number of each image and selects images from discrete days and in different streets. In addition, all image metadata including device information, GPS location, etc., is cleared and privacy regions like human faces are blurred.

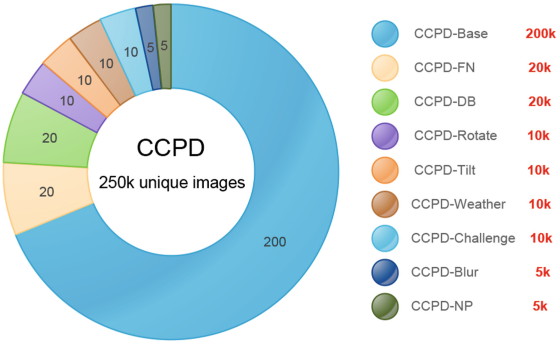


Fig. 2. CCPD layout.

3.2 Dataset Splits and Statistics

CCPD provides over 250k unique LP images with detailed annotations. The resolution of each image is 720 (Width) × 1160 (Height) × 3 (Channels). In practice, this resolution is enough to guarantee that the LP in each image is legible. The average size of each file is about 200 KB (a total of over 48.0 GB for the entire dataset).

Table 3. Descriptions of different sub-datasets in CCPD.

	Description
CCPD-Base	The only common feature of these photos is the inclusion of a license plate
CCPD-DB	Illuminations on the LP area are dark, uneven or extremely bright
CCPD-FN	The distance from the LP to the shooting location is relatively far or near
CCPD-Rotate	Great horizontal tilt degree ($20^\circ \sim 50^\circ$) and the vertical tilt degree varies from -10° to 10°
CCPD-Tilt	Great horizontal tilt degree ($15^\circ \sim 45^\circ$ degrees) and vertical tilt degree ($15^\circ \sim 45^\circ$)
CCPD-Blur	Blurry largely due to hand jitter while taking pictures
CCPD-Weather	Images taken on a rainy day, snow day or fog day
CCPD-Challenge	The most challenging images for LPDR to date
CCPD-NP	Images of new cars without a LP

Each image in CCPD is labelled in the following aspects:

- LP number. Each image in CCPD has only one LP. Each LP number is comprised of a Chinese character, a letter, and five letters or numbers. The LP number is an important metric for recognition accuracy.
- LP bounding box. The bounding box label contains (x, y) coordinates of the top left and bottom right corner of the bounding box. These two points can be utilized to locate the minimum bounding rectangle of LP.
- Four vertices locations. This annotation contains the exact (x, y) coordinates of the four vertices of LP in the whole image. As the shape of the LP is basically a quadrilateral, these vertices location can accurately represent the borders of the LP for object segmentation.
- Horizontal tilt degree and vertical tilt degree. As explained in [11], the horizontal tilt degree is the angle between LP and the horizontal line. After the 2D rotation, the vertical tilt degree is the angle between the left border line of LP and the horizontal line.
- Other information concerning the LP like the area, the degree of brightness and the degree of vagueness.

Current diverse LPDR datasets [10,16,40] usually contains less than 5k images. After dividing these challenging images into different categories [40], some categories contains less than 100 images. Based on this observation, we select images under different conditions to build several sub-datasets for CCPD from millions of LP images. The distribution of sub-datasets in CCPD is shown in the Fig. 2. Descriptions of these sub-datasets are shown in Table 3. Statistics and samples of these sub-datasets are provided in the supplementary material.

We further add **CCPD-Characters** which contains at least 1000 extracted images for each possible LP character. CCPD-Characters is designed for training neural networks to recognize segmented character images. More character images can be automatically extracted by utilizing annotations of images in CCPD.

4 The Roadside Parking Net (RPnet)

In this section, we introduce our proposed LP detection and recognition framework, called RPnet, and discuss the associated training methodology.

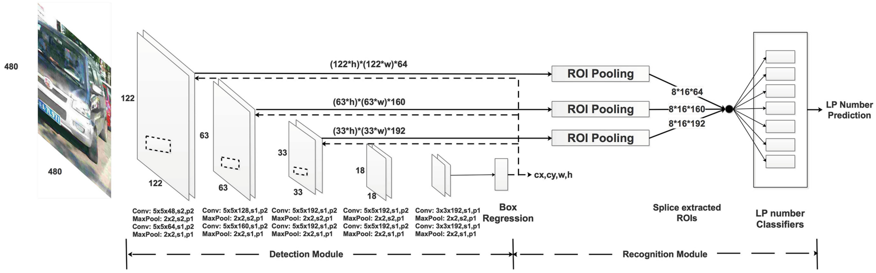


Fig. 3. The overall structure of our RPnet. It consists of ten convolutional layers with ReLU and Batch Normalization, several MaxPooling layers with Dropout and several components composed of fully connected layers. Given an input RGB image, in a single forward computation, RPnet predicts the LP bounding box and the corresponding LP number at the same time. RPnet first exploits the Box Regression layer to predict the bounding box. Then, refer to the relative position of the bounding box in each feature map, RPnet extracts ROIs from several already generated feature maps, combine them after pooling them to the same width and height $(16 * 8)$, and feeds the combined features maps to the subsequent Classifiers.

4.1 Model

RPnet, as shown in Fig. 3, is composed of two modules. The first module is a deep convolutional neural network with ten convolutional layers to extract different level feature maps from the input LP image. We name this module ‘the detection module’. The detection module feeds the feature map output by the last convolutional layer to three sibling fully-connected layers which we name ‘the box predictor’ for bounding box prediction. The second module, named 'the recognition module', exploits region-of-interest (ROI) pooling layers [28] to extract feature maps of interest and several classifiers to predict the LP number of the LP in the input image. The entire module is a single, unified network for LP detection and recognition.

Using a popular terminology 'attention' [41] in neural networks, the detection module serves as the 'attention' of this unified network. It tells the recognition module where to look. Then the recognition module extracts the ROI from shared feature maps and predicts the LP number.

Feature Extraction. RPnet extracts features from the input image by all the convolutional layers in the detection module. As the number of layers increases, the number of channels increases and the size of the feature map decreases progressively. The later feature map has higher level features extracted and thus is more beneficial for recognizing the LP and predicting its bounding box. Suppose the center point x-coordinate, the center point y-coordinate, the width, and the height of the bounding box are b_x, b_y, b_w, b_h respectively. Let W and H be the width and the height of the input image. The bounding box location cx, cy, w, h satisfies:

$$cx = \frac{b_x}{W} \quad cy = \frac{b_y}{H} \quad w = \frac{b_w}{W} \quad h = \frac{b_h}{H}, \quad 0 < cx, cy, w, h < 1$$

Multi-layer Feature Maps for Recognition. Empirically feature maps from different layers within a network are empirically known to have different receptive field sizes [42]. Moreover, previous works such as [43] have shown that using feature maps from the lower layers can improve semantic segmentation quality because the lower layers capture more fine details of the input objects. Similarly, feature maps from relatively lower layers also matter for recognizing LP characters as, just like the object borders in semantic segmentation, the area of the LP is expected to be very small relative to the entire image. After the detection module accomplishes the computation of all convolutional layers, the box predictor outputs the bounding box location (cx, cy, w, h). For a feature layer of size mxn with p channels, as shown in Fig. 3, the recognition module extracts feature maps in the bounding box region of size $(m * h) * (n * w)$ with p channels. By default, RPnet extracts feature maps at the end of three low-level layers: the second, fourth, sixth convolutional layer. The sizes of extracted feature maps are $(122 * h) * (122 * w) * 64$, $(63 * h) * (63 * w) * 160$, $(33 * h) * (33 * w) * 192$. In practice, extracting feature maps from higher convolutional layers makes recognition process slower and offers little help in improving the recognition accuracy. After these feature maps are extracted, RPnet exploits ROI Pooling layers to convert each extracted feature into a feature map with a fixed spatial extent of $P_H * P_W$ (e.g., $8 * 16$ in this paper). Afterwards, these three resized feature maps $8 * 16 * 64$, $8 * 16 * 160$ and $8 * 16 * 192$ are concatenated to one feature map of size $8 * 16 * 416$ for LP number classification.

4.2 Training

RPnet can be trained end-to-end on CCPD and accomplishes LP bounding box detection and LP number recognition in a single forward. The training involves choosing suitable loss functions for detection performance and recognition performance, as well as pre-training the detection module before training RPnet end-to-end.

Training Objective. The RPnet training objective can be divided into two parts: the localization loss (loc) and the classification loss (cls). Let N be the size of a mini-batch in training. The localization loss (see Eq. (1)) is a Smooth L1 loss [28] between the predicted box (pb) and the ground truth box (gb). Let the ground-truth seven LP numbers be $gn_i (1 \leq i \leq 7)$. $pn_i (1 \leq i \leq 7)$ denotes predictions for the seven LP characters and each LP character prediction pn_i contains nc_i float numbers, each representing the possibility of belonging to a specific character class. The classification loss (see Eq. (2)) is a cross-entropy loss. With the joint optimization of both localization and classification losses, the extracted features would have richer information about LP characters. Experiments show that both detection and recognition performance can be enhanced by jointly optimizing these two losses.

$$L_{loc}(pb, gb) = \sum_{N} \sum_{m \in \{cx, cy, w, h\}} smooth_{L1}(pb^m - gb^m) \tag{1}$$

$$L_{cls}(pn, gn) = \sum_{N} \sum_{1 \leq i \leq 7} \{-pn_i[gn_i] + log(\sum_{1 \leq j \leq nc_i} \exp(pn_i[j]))\} \tag{2}$$

$$L(pb, pn, gb, gn) = \frac{1}{N}(L_{loc}(pb, gb) + L_{cls}(pn, gn)) \tag{3}$$

Pre-training Detection Module. Before training PRnet end-to-end, the detection module must provide a reasonable bounding box prediction (cx, cy, w, h). A reasonable prediction (cx, cy, w, h) must meet $0 < cx, cy, w, h < 1$ and might try to meet $\frac{w}{2} \leq cx \leq 1 - \frac{w}{2}, \frac{h}{2} \leq cy \leq 1 - \frac{h}{2}$, thus can represent a valid ROI and guide the recognition module to extract feature maps. Unlike most object detection related papers [29,31] which pre-train their convolutional layers on ImageNet [44] to make these layers more representative, we pre-train the detection module from scratch on CCPD as the data volume of CCPD is large enough and, for locating a single object such as a license plate, parameters pre-trained on ImageNet are not necessarily better than training from scratch. In practice, the detection module always gives a reasonable bounding box prediction after being trained 300 epochs on the training set.

5 Evaluations

In this section, we conduct experiments to compare RPnet with state-of-the-art models on both LP detection performance and LP recognition performance.

Furthermore, we explore the effect of extracting features maps from different layers on the final recognition accuracy.

All data comes from our proposed CCPD, the largest publicly available annotated LP Dataset to date. All our training tasks are accomplished on a GPU server with 8 CPU (Intel(R) Xeon(R) CPU E5-2682 v4 @ 2.50 GHz), 60 GB RAM and one Nvidia GPU (Tesla P100 PCIe 16 GB). All our evaluation tasks are finished on desktop PCs with eight 3.40 GHz Intel Core i7-6700 CPU, 24 GB RAM and one Quadro P4000 GPU.

5.1 Data Preparation

As aforementioned in Sect. 3, CCPD-Base consists of approximately 200k unique images. We divide CCPD-Base into two equal parts. One as the default training set, another as the default evaluation set. In addition, several sub-datasets (CCPD-DB, CCPD-FN, CCPD-Rotate, CCPD-Tilt, CCPD-Weather, CCPD-Challenge) in CCPD are also exploited for detection and recognition performance evaluation. Apart from Cascade classifier [45], all models used in experiments rely on GPU and are fine-tuned on the training set. For models without default data augmentation strategies, we augment the training data by randomly sampling four times on each image to increase the training set by five times. More details are provided in the supplementary material.

We did not reproduce our experiments on other datasets because most current available LP datasets [13–15] are not as diverse as CCPD and their data volume is far fewer than CCPD. Thus, detection accuracy or recognition accuracy on other datasets might not be as convincing as on CCPD. Moreover, we also did not implement approaches not concerning machine learning like [8] because in practice, when evaluated on a large-scale dataset, methods based on machine learning always perform better.

Table 4. LP detection precision (percentage) of state-of-the-art detection models on each test set. AP denotes average precision in the whole test set and FPS denotes frames per second.

	FPS	AP	Base (100k)	DB	FN	Rotate	Tilt	Weather	Challenge
Cascade classifier [45]	32	47.2	55.4	49.2	52.7	0.4	0.6	51.5	27.5
SSD300 [30]	40	94.4	99.1	89.2	84.7	**95.6**	**94.9**	83.4	**93.1**
YOLO9000 [32]	42	93.1	98.8	89.6	77.3	93.3	91.8	84.2	88.6
Faster-RCNN [29]	15	92.9	98.1	92.1	83.7	91.8	89.4	81.8	83.9
TE2E [12]	3	94.2	98.5	91.7	83.8	95.1	94.5	83.6	93.1
RPnet	**61**	**94.5**	**99.3**	**89.5**	**85.3**	94.7	93.2	**84.1**	92.8

5.2 Detection

Detection Accuracy Metric. We follow the standard protocol in object detection Intersection-over-Union (IoU) [12].

The bounding box is considered to be correct if and only if its IoU with the ground-truth bounding box is more than 70% ($IoU > 0.7$). All models are fine-tuned on the same 100k training set.

We set a higher IoU boundary in the detection accuracy metric than TE2E [12] because a higher boundary can filter out imperfect bounding boxes and thus better evaluates the detection performance. The results are shown in Table 4. Cascade classifier has difficulty in precisely locating LPs and thus performs badly under a high IoU threshold and it is not robust when dealing with tilted LPs. Concluded from the low detection accuracy 77.3% on CCPD-FN, YOLO has a relatively bad performance on relatively small/large object detection. Benefited from the joint optimization of detection and recognition, the performance of both RPnet and TE2E surpasses Faster-RCNN and YOLO9000. However, RPnet can recognize twenty times faster than TE2E. Moreover, by analysing the bounding boxes predicted by SSD, we found these boxes wrap around LPs very tightly. Actually, when the IoU threshold is set higher than 0.7, SSD achieves the highest accuracy. The reason might be that the detection loss is not the only training objective of RPnet. For example, a little imperfect bounding box (slightly smaller than the ground-truth one) might be beneficial for more correct LP recognition.

Table 5. LP recognition precision (percentage) on each test set. Apart from TE2E and RPnet, we append a high-performance model to other object detection models for subsequent LP recognition. HC denotes Holistic-CNN [15].

	FPS	AP	Base (100k)	DB	FN	Rotate	Tilt	Weather	Challenge
Cascade classifier + HC	29	58.9	69.7	67.2	69.7	0.1	3.1	52.3	30.9
SSD300 + HC	35	95.2	98.3	96.6	**95.9**	88.4	91.5	87.3	83.8
YOLO9000 + HC	36	93.7	98.1	96.0	88.2	84.5	88.5	87.0	80.5
Faster-RCNN+ HC	13	92.8	97.2	94.4	90.9	82.9	87.3	85.5	76.3
TE2E	3	94.4	97.8	94.8	94.5	87.9	92.1	86.8	81.2
RPnet	**61**	**95.5**	**98.5**	**96.9**	94.3	**90.8**	**92.5**	**87.9**	**85.1**

5.3 Recognition

Recognition Accuracy Metric. A LP recognition is correct if and only if the IoU is greater than 0.6 and all characters in the LP number are correctly recognized.

To our knowledge, before us, TE2E [12] is the only end-to-end network for LPDR. Apart from TE2E and our RPnet, in our evaluations we combine state-of-the-art detection models with a state-of-the-art recognition model named Holistic-CNN [15] as comparisons to TE2E and RPnet. We fine-tuned Holistic-CNN on the training set produced by extracting the LP region from the same

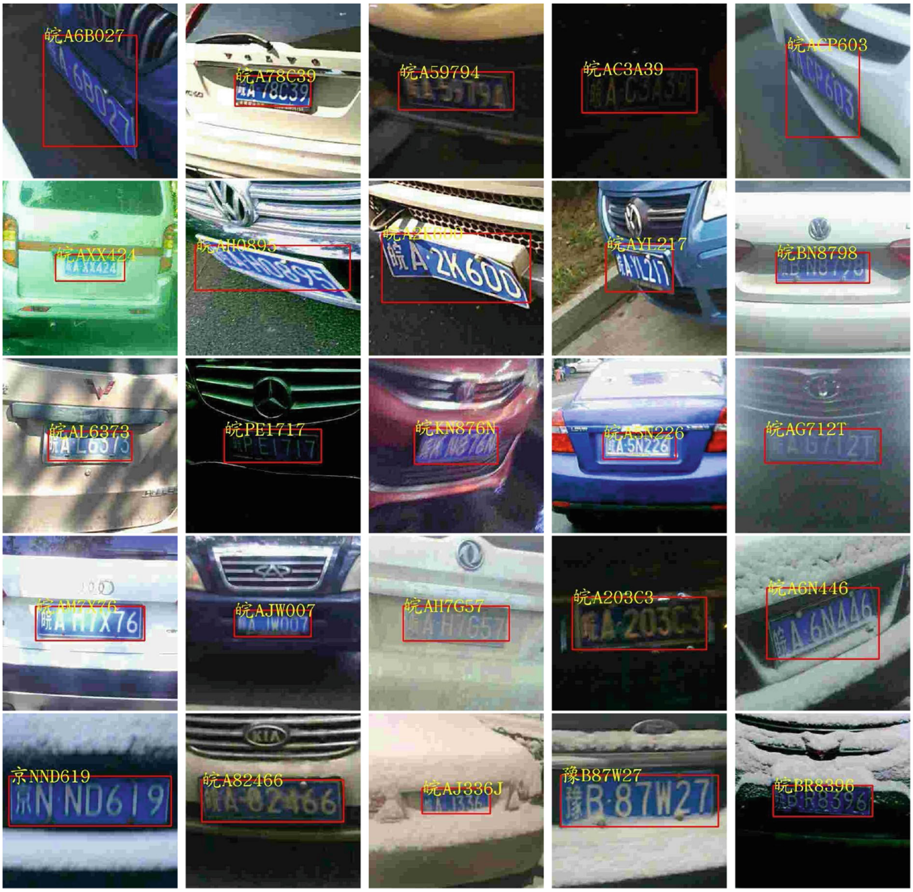
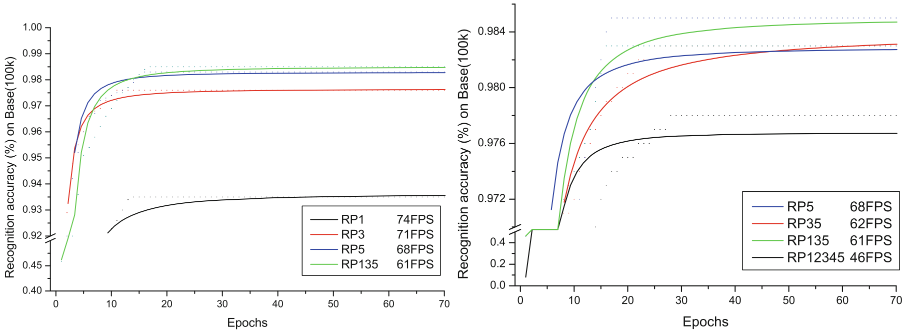


Fig. 4. Detection and recognition results on CCPD with our RPnet model. Each image is the smaller license plate area extracted from the original resolution 720 (Width) × 1160 (Height) × 3 (Channels).

100k images according to their ground-truth bounding box. On the test set produced in a similar manner, Holistic-CNN can recognize over 200 small LP region images per second with a 98.5% accuracy.

As shown in Table 5, these combined models can achieve high recognition speed (36 fps) and high recognition accuracy (95.2%). As a result of precise LP bounding boxes predicted by SSD, the model combining SSD and Holistic-CNN achieves up to 95.2% average precision on CCPD. However, by sharing feature maps between the detection module and recognition module, RPnet achieves a much faster recognition rate 61 FPS and a slightly higher recognition accuracy 95.5%.

In addition, it's worth noting that nearly all evaluated models fail to perform well on CCPD-Rotate, CCPD-Weather, and especially CCPD-Challenge. Difficulties of detection and recognition on these three sub-datasets are partly resulted from the scarce of LPs under these conditions in training data. Their low performances partly demonstrate that by classifying LPs into different subcategories, CCPD can evaluate LPDR algorithms more comprehensively.



(a) Higher layer matters in the recognition module

(b) The more layers, the accuracy is not necessarily higher

Fig. 5. Performance analysis on extracting feature maps from different layers or multiple layers. FPS denotes frames per second.

5.4 Model Analysis

Samples of detection and recognition results produced by evaluating RPnet on CCPD are shown in Fig. 4. To understand RPnet better, we carried out some controlled experiments to examine how each layer affects performance. For all variants of RPnet in experiments, we use the same settings and input size as the original RPnet, except for specified changes.

Higher Layer Matters in the Recognition Module. We number ten convolutional layers in RPnet $C_i, 0 \leq i \leq 9$. Therefore, the original RPnet described in Sect. 4 can be denoted RP135 because it extracts feature maps from C_1, C_3, C_5 for LP character recognition. Similarly, we implement RP1, RP3, and RP5 where feature maps are extracted from only one specific layer. The results are shown in Fig. 5(a). Among these four models, RP1 only reaches the lowest accuracy 93.5%, while all other models achieve an accuracy higher than 97.5%. From RP1 to RP5, with the higher layer is exploited for feature extraction, the recognition accuracy increases and the epochs needed for fitting decreases. With a single layer C_5 for feature extraction, RP5 achieves almost the same accuracy as RP135. Though lower layers can improve semantic segmentation quality, higher order features seem to be more useful for recognition tasks.

Feature Extraction from More Layers Not Necessarily Increases the Accuracy. Based on the knowledge that the sixth convolutional layer C_5 might have a greater impact on the recognition accuracy, we trained two new models RP35 and RP12345 for performance analysis which also exploits C_5 and has different number of layers for feature extraction. As shown in Fig. 5(b), from RP5 to RP35 to RP135, the number of layers for feature extraction increases and the recognition accuracy that can be achieved also increases. However, with five layers for feature extraction, RP12345 not only introduces significantly more recognition time, but its accuracy decreased. Extracting features from too many layers and not having enough neurons for analysing might lead to poor generalization.

6 Conclusions

In this paper, we present a large-scale and diverse license plate dataset named CCPD and a novel network architecture named RPnet for unified license plate detection and recognition. Images in CCPD are annotated carefully and are classified into different categories according to different features of LPs. The great data volume (250k unique images), data diversity (eight different categories), and detailed annotations make CCPD a valuable dataset for object detection, object recognition, and object segmentation. Extensive evaluations on CCPD demonstrate our proposed RPnet outperforms state-of-the-art approaches both in speed and accuracy. Currently, RPnet has been put into practice for road-side parking services. Its accuracy and speed significantly surpass other existing commercial license plate recognition systems.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No. 61572456) and the Anhui Province Guidance Funds for Quantum Communication and Quantum Computers.

References

1. Xie, L., Ahmad, T., Jin, L., Liu, Y., Zhang, S.: A new CNN-based method for multi-directional car license plate detection. IEEE Trans. Intell. Transp. Syst. **19**, 507–517 (2018)
2. Al-Shemarry, M.S., Li, Y., Abdulla, S.: Ensemble of adaboost cascades of 3L-LBPs classifiers for license plates detection with low quality images. Expert Syst. Appl. **92**, 216–235 (2018)
3. Laroca, R., et al.: A robust real-time automatic license plate recognition based on the yolo detector (2018). arXiv preprint: arXiv:1802.09567
4. Montazzolli, S., Jung, C.: Real-time Brazilian license plate detection and recognition using deep convolutional neural networks. In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 55–62 (2017)
5. Selmi, Z., Halima, M.B., Alimi, A.M.: Deep learning system for automatic license plate detection and recognition. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 1132–1138. IEEE (2017)
6. Rizvi, S.T.H., Patti, D., Björklund, T., Cabodi, G., Francini, G.: Deep classifiers-based license plate detection, localization and recognition on GPU-powered mobile platform. Future Internet **9**(4), 66 (2017)
7. Masood, S.Z., Shu, G., Dehghan, A., Ortiz, E.G.: License plate detection and recognition using deeply learned convolutional neural networks (2017). arXiv preprint: arXiv:1703.07330

8. Yuan, Y., Zou, W., Zhao, Y., Wang, X., Hu, X., Komodakis, N.: A robust and efficient approach to license plate detection. IEEE Trans. Image Process. **26**(3), 1102–1114 (2017)
9. Cheang, T.K., Chong, Y.S., Tay, Y.H.: Segmentation-free vehicle license plate recognition using convnet-RNN (2017). arXiv preprint: arXiv:1701.06439
10. Azam, S., Islam, M.M.: Automatic license plate detection in hazardous condition. J. Vis. Commun. Image Represent. **36**, 172–186 (2016)
11. Zhang, Z., Yin, S.: Hough transform and its application in vehicle license plate tilt correction. Comput. Inf. Sci. **1**(3), 116 (2008)
12. Li, H., Wang, P., Shen, C.: Towards end-to-end car license plates detection and recognition with deep neural networks (2017). arXiv preprint: arXiv:1709.08828
13. Caltech: Caltech Licese Plate Dataset. http://www.vision.caltech.edu/html-files/archive.html
14. Zemris: Zemris License Plate Dataset. http://www.zemris.fer.hr/projects/LicensePlates/hrvatski/rezultati.shtml
15. Špaňhel, J., Sochor, J., Juránek, R., Herout, A., Maršík, L., Zemčík, P.: Holistic recognition of low quality license plates by CNN using track annotated data. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6. IEEE (2017)
16. Hsu, G.S., Ambikapathi, A., Chung, S.L., Su, C.P.: Robust license plate detection in the wild. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6. IEEE (2017)
17. Gonçalves, G.R., da Silva, S.P.G., Menotti, D., Schwartz, W.R.: Benchmark for license plate character segmentation. J. Electron. Imaging **25**(5), 1–5 (2016)
18. Haralick, R.M., Sternberg, S.R., Zhuang, X.: Image analysis using mathematical morphology. IEEE Trans. Pattern Anal. Mach. Intell. PAMI **9**(4), 532–550 (1987)
19. Hsieh, J.W., Yu, S.H., Chen, Y.S.: Morphology-based license plate detection from complex scenes. In: Proceedings of the 16th International Conference on Pattern Recognition, vol. 3, pp. 176–179. IEEE (2002)
20. Wu, H.H.P., Chen, H.H., Wu, R.J., Shen, D.F.: License plate extraction in low resolution video. In: 18th International Conference on Pattern Recognition, ICPR 2006, vol. 1, pp. 824–827. IEEE (2006)
21. Yu, S., Li, B., Zhang, Q., Liu, C., Meng, M.Q.H.: A novel license plate location method based on wavelet transform and emd analysis. Pattern Recogn. **48**(1), 114–125 (2015)
22. Saha, S., Basu, S., Nasipuri, M., Basu, D.K.: License plate localization from vehicle images: an edge based multi-stage approach. Int. J. Recent Trends Eng. **1**(1), 284–288 (2009)
23. Wang, R., Sang, N., Huang, R., Wang, Y.: License plate detection using gradient information and cascade detectors. Optik Int. J. Light Electron Opt. **125**(1), 186–190 (2014)
24. Bachchan, A.K., Gorai, A., Gupta, P.: Automatic license plate recognition using local binary pattern and histogram matching. In: Huang, D.-S., Jo, K.-H., Figueroa-García, J.C. (eds.) ICIC 2017, Part II. LNCS, vol. 10362, pp. 22–34. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63312-1_3
25. Deb, K., Jo, K.H.: HSI color based vehicle license plate detection. In: International Conference on Control, Automation and Systems, ICCAS 2008, pp. 687–691. IEEE (2008)
26. Yao, Z., Yi, W.: License plate detection based on multistage information fusion. Inf. Fusion **18**, 78–85 (2014)

27. Lee, Y., Song, T., Ku, B., Jeon, S., Han, D.K., Ko, H.: License plate detection using local structure patterns. In: 2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 574–579. IEEE (2010)
28. Girshick, R.: Fast R-CNN (2015). arXiv preprint: arXiv:1504.08083
29. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
30. Liu, W., et al.: SSD: single shot MultiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016, Part I. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
31. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
32. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger (2016). arXiv preprint: arXiv:1612.08242
33. Ho, W.T., Lim, H.W., Tay, Y.H.: Two-stage license plate detection using gentle adaboost and SIFT-SVM. In: First Asian Conference on Intelligent Information and Database Systems, ACIIDS 2009, pp. 109–114. IEEE (2009)
34. Duan, T.D., Du, T.H., Phuoc, T.V., Hoang, N.V.: Building an automatic vehicle license plate recognition system. In: Proceedings of the International Conference on Computer Science RIVF, pp. 59–63 (2005)
35. Yousef, K.M.A., Al-Tabanjah, M., Hudaib, E., Ikrai, M.: Sift based automatic number plate recognition. In: 2015 6th International Conference on Information and Communication Systems (ICICS), pp. 124–129. IEEE (2015)
36. Maglad, K.W.: A vehicle license plate detection and recognition system. J. Comput. Sci. **8**(3), 310 (2012)
37. Gou, C., Wang, K., Yao, Y., Li, Z.: Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines. IEEE Trans. Intell. Transp. Syst. **17**(4), 1096–1107 (2016)
38. Ciregan, D., Meier, U., Schmidhuber, J.: Multi-column deep neural networks for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3642–3649. IEEE (2012)
39. Abdel-Hakim, A.E., Farag, A.A.: CSIFT: a SIFT descriptor with color invariant characteristics. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), vol. 2, pp. 1978–1983 (2006)
40. Anagnostopoulos, C.N.E., Anagnostopoulos, I.E., Psoroulas, I.D., Loumos, V., Kayafas, E.: License plate recognition from still images and video sequences: a survey. IEEE Trans. Intell. Transp. Syst. **9**(3), 377–391 (2008)
41. Chorowski, J.K., Bahdanau, D., Serdyuk, D., Cho, K., Bengio, Y.: Attention-based models for speech recognition. In: Advances in Neural Information Processing Systems, pp. 577–585 (2015)
42. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Object detectors emerge in deep scene CNNs (2014). arXiv preprint arXiv:1412.6856
43. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
44. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015)
45. Wang, S.Z., Lee, H.J.: A cascade framework for a real-time statistical plate recognition system. IEEE Trans. Inf. Forensics Secur. **2**(2), 267–282 (2007)