



CNN-RNN based method for license plate recognition

ISSN 2468-2322

Received on 20th July 2018

Accepted on 21st July 2018

doi: 10.1049/trit.2018.1015

www.ietdl.org

Palaiahnakote Shivakumara¹ ✉, Dongqi Tang², Maryam Asadzadehkaljahi¹, Tong Lu²,
 Umada Pal³, Mohammad Hossein Anisi⁴

¹Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

²National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, People's Republic of China

³Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India

⁴School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, UK

✉ E-mail: hudempsk@yahoo.com

Abstract: Achieving good recognition results for License plates is challenging due to multiple adverse factors. For instance, in Malaysia, where private vehicle (e.g., cars) have numbers with dark background, while public vehicle (taxis/cabs) have numbers with white background. To reduce the complexity of the problem, we propose to classify the above two types of images such that one can choose an appropriate method to achieve better results. Therefore, in this work, we explore the combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks namely, BLSTM (Bi-Directional Long Short Term Memory), for recognition. The CNN has been used for feature extraction as it has high discriminative ability, at the same time, BLSTM has the ability to extract context information based on the past information. For classification, we propose Dense Cluster based Voting (DCV), which separates foreground and background for successful classification of private and public. Experimental results on live data given by MIMOS, which is funded by Malaysian Government and the standard dataset UCSD show that the proposed classification outperforms the existing methods. In addition, the recognition results show that the recognition performance improves significantly after classification compared to before classification.

1 Introduction

Despite sophisticated license plate recognition systems available online [1], the problem of license plate recognition is considered as a hot research issue. This is due to the lack of uniform format of capturing and displaying number plates across countries. For example, when we look at developed countries, all vehicles follow a particular format or type for displaying number plates. As a result, the developed systems work well for images of respective countries. However, for the countries like Malaysia and India, one can see many types of number plates with different texts. One such example is shown in Fig. 1, where it can be seen in Fig. 1a that private and UCSD have dark background and white foreground, while public has opposite. It is also observed in Fig. 1a that images of public suffer from poor quality due to degradations and distortions. For these images, when we apply the existing method [2] which proposes strokelets for character identification and then recognition of scene texts using a random forest classifier, it does not recognise public and UCSD correctly as shown in Fig. 1b. This is due to the change in background and foreground colour with the distortion effect. The online system, which explores deep learning for Chinese license plate recognition [3] works well for private and public but fails for UCSD images as shown in Fig. 1c. The reason for the poor results is that UCSD suffers from poor quality due to blur. It is noted from Fig. 1 that the method [2] which works well for natural scene texts and the system [3] which uses deep learning fail to perform well.

Therefore, we can conclude that the existing powerful methods and systems may work well for images of particular types but not images affected by multiple adverse factors as shown in Fig. 1a, and hence the latter is challenging. Thus, in this work, we propose to classify images of different types such that we can focus on addressing other challenges, like effect distortion, noise and poor quality. In this work, we explore the combination of convolutional

neural networks (CNNs) and recurrent neural networks to address the above mentioned-challenges.

2 Related work

Since the proposed work involves classification of different types of license plate images and recognition of license plate images, we review the methods on classification and recognition here. As per our knowledge, there are no methods for classification of different license type images. However, one can see several methods for classification of caption and scene text in video. It is true that number plates can be considered as scene text images as they are captured from natural scenes. Therefore, we present the review of caption and scene text classification methods in the following. Raghunandan *et al.* [4] proposed new sharpness features for image type classification based on textual information. The method works based on the fact that caption text images have high quality and clarity because they are edited texts, while the nature of scene texts is unpredictable because it is a part of an image. Xu *et al.* [5] proposed caption and scene text classification in video, which uses gradient and clustering. The method works on the basis of pixels in caption text images share the uniform colour value, while scene text does not. Shivakumara *et al.* [6] proposed the separation of graphics and scene texts in video frames, which explores the contrast and characteristics of text components. Since caption texts have good quality, it is expected that they should exhibit regular patterns of spacing between their characters, while scene texts do not. Roy *et al.* [7] proposed new tampered discrete cosine transform (DCT) features for scene and caption text classification in video frames for classifying two types of texts in video. The method explores these tampered features because editing texts over video is a type of tampering. It is noted from the above review that the methods do not focus on

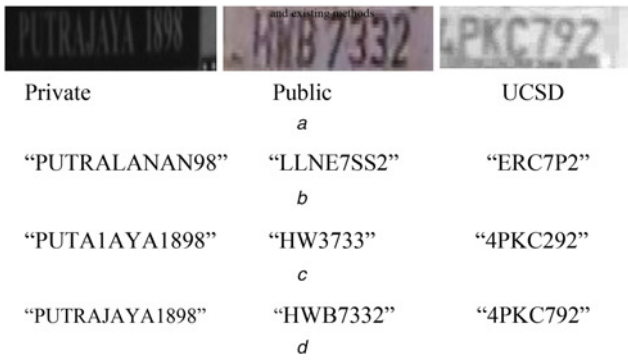


Fig. 1 Illustrating the complexity of the license plate recognition

- a Input license plate images with different complexities
b Recognition result of the existing method [2]
c Recognition result of the existing system [3]
d Recognition result of the proposed method

addressing the issues of adverse multiple factors such as blur, noise and colour change in background and foreground. Therefore, these methods may not perform well for license plate image classification.

In the same way, we also present a review on license plate recognition using deep learning in the following. Yang *et al.* [8] proposed Chinese vehicle license plate recognition using a kernel-based extreme learning machine with deep convolutional features. The method uses CNN for feature extraction and then kernel based extreme learning machines (KELM) for recognition. The main objective of this method is Chinese license plate recognition. Abedin *et al.* [9] proposed a license plate recognition system based on contour properties and a deep learning model. The method segments characters for recognition. However, segmenting characters accurately may be good for images which have high quality but not images which suffer from poor quality. Polishetty *et al.* [10] proposed a next generation secure cloud based deep learning license plate recognition method for smart cities. The method involves edge detection and binarisation for achieving results. It is noted that edge detection and binarisation methods are sensitive to complex background and distortions. Montazzolli and Jung [11] proposed real-time Brazilian license plate detection and recognition using deep CNNs. The method uses CNN for character segmentation and recognition. The method is proposed for high resolution and quality images. Bulan *et al.* [12] proposed segmentation and annotation free license plate recognition with deep localisation and failure identification. The method explores hidden Markov model (HMM) for character segmentation and recognition, which does not require specific character segmentation. However, it is not clear that how the method takes care of distortion effect, poor quality images and noises. Selmi *et al.* [13] proposed a deep learning system for automatic license plate detection and recognition. The method performs many pre-processing steps before applying CNNs. In addition, the method follows the conventional way of character segmentation and then recognition. This idea is good for images that have high quality and particular types. Li *et al.* [14] proposed reading car license plates using deep neural networks. The method explores the combination of CNN and RNN for license plate detection and recognition. However, it is not clear that how the method works well for images affected by multiple adverse factors.

In summary, the review of license plate recognition using deep learning shows that most of the methods focus on particular types of images with the knowledge of country or format of license plates. None of the methods addresses the issues of blur, noise, foreground and background colour changes, and poor quality. Hence, in this work, we tackle the above-challenges through classification of different types of license plate images and then explore the combination of CNN and RNN in a different way for addressing other causes to achieve the results.

3 Proposed methodology

In this work, we consider segmented license plates as the input for recognition. There are methods that detect license plates irrespective of the above-mentioned challenges because segmentation is a pre-processing step for recognition [1]. The proposed method consists of two parts, namely, part 1 presents the classification of private and public license plate images, and part 2 presents the recognition of license plate images. As discussed in Section 1, in Malaysia, license plates of privates have a dark background, while publics have a white background. It is true that values of white pixels in images have high values compared to the values of dark pixels. Based on this, we propose to separate foreground and background information through a Canny edge image of the input image. Edge pixels in the Canny edge image are considered as foreground information, and the rest pixels are considered as background pixels. The proposed method applies K -means clustering with $K=2$ on foreground and background separately, which gives two clusters for foreground and two more clusters for the background. The statistical analysis of the clusters of foreground and background results in two hypotheses for classification of private and public license plate images.

For the classified license plate images, we propose the combination of CNN and RNN, which involves BLSTM (bi-directional long-short term memory), for recognition. The proposed method extracts features using CNN as it has the ability to provide unique features for each character and use BLSTM for finding the sequence of labelling as it has the ability to define context using the past and the present features. Further, the proposed method recognises license plates based on the highest probability scoring with trained samples. More details can be found in the subsequent sections.

3.1 Classification of license plates

It is observed from Fig. 1a that the background and foreground of private plates are represented by black and white colours, respectively. Similarly, background and foreground of public plates are represented by white and black colours, respectively. This shows that the intensity values of the background of a normal plate are lower than those of a taxi plate. In addition, it is also true that the number of pixels which represent background usually is higher than that of pixels which represent the foreground. These observations lead to propose a dense cluster-based voting (DCV) for the classification of private and public license plate images.

It is true that the Canny edge detector gives fine edges regardless of background and foreground colour changes as shown in Fig. 2a, where we can see edges are represented by white pixels for both images. To extract the above observation, we separate edge pixels as foreground and non-edge pixels as background for the input image. Then the proposed method extracts intensity values corresponding to foreground and background pixels from the grey image of the input image. To visualise the difference in intensity distribution, we perform a histogram operation on intensity values of foreground and background of the normal and taxi plate images as shown in Figs. 2b–e. It is observed from Figs. 2b and c that the dense distribution can be seen for the pixels which have intensity values near to 255 in the case of the foreground-normal plate, while the dense distribution can be seen for the pixels which have intensity values near to 0 in case of background-normal plate image. It is vice versa for foreground-taxi and background-taxi as shown in Figs. 2d–e. This is the main basis for the proposing DCV.

To extract such observations, we divide the intensity values of foreground and background into two clusters as Max which gets high values and Min which gets low values using K -means clustering with $K=2$. This results in four clusters for the input image, namely, foreground-Max cluster, Foreground-Min cluster, Background-Max cluster and Background-Min cluster. For each cluster, the proposed method computes the mean, standard deviation and the number of pixels (density) to derive hypotheses to identify license plate images. For example, the product of standard deviation and the number of pixels of the background-min cluster is greater than that of standard deviation and the number of pixels of the background-max

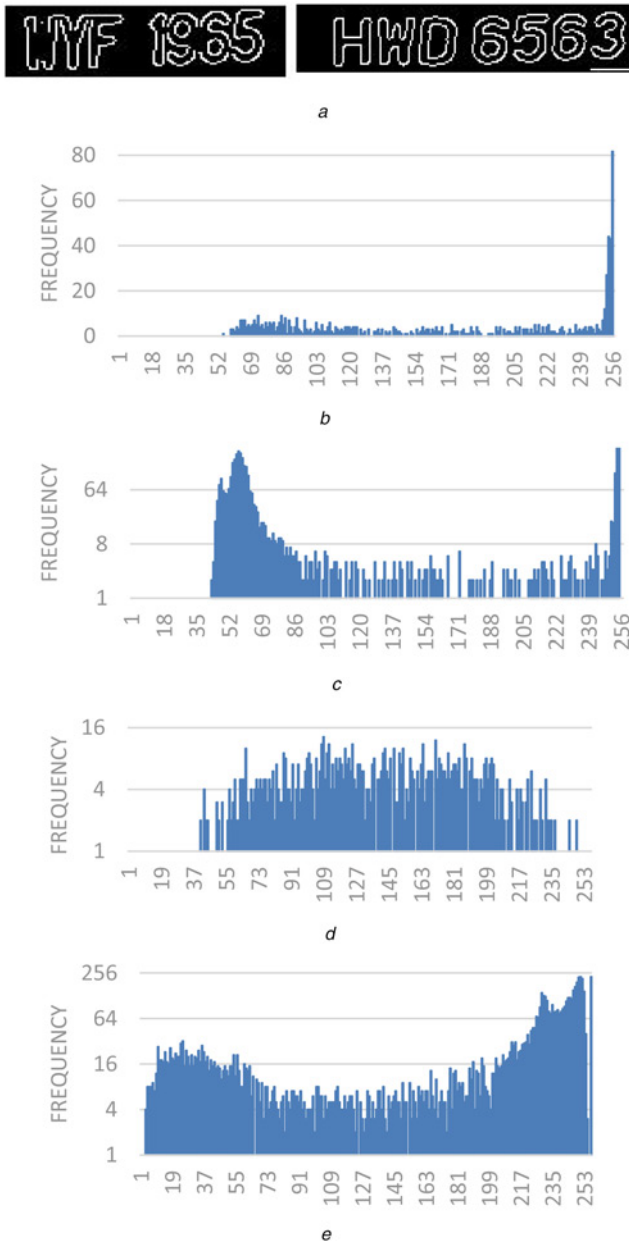


Fig. 2 Intensity distribution of foreground and background of normal and taxi plate images

- a Canny images of normal and taxi plate images
- b Histogram for grey of the foreground of the normal plate
- c Histogram for grey of the background of the normal plate
- d Histogram for grey of the foreground of taxi plate
- e Histogram for grey of the background of taxi plate

cluster for normal plates. This results in response '1'. In this way, the proposed method derives three hypotheses and finds the responses. If the hypothesis gives two responses as '1' out of three, it is identified as a normal plate else a taxi plate.

The foreground and background separation is illustrated in Fig. 3, where Figs. 3a and b of private and public images provide edges without losing shapes. Then the proposed method extracts intensity values in the grey image G corresponding to foreground and background pixels, say GF and GB as defined in (1) and (2) for private and public plate images as, respectively, shown in Figs. 3c and d, where we notice there are colour changes in foreground and background of private and public plate images

$$GF_{x,y} = \begin{cases} G_{x,y}, & \text{if } \text{Canny}(x, y) = 1 \\ 0, & \text{else} \end{cases} \quad (1)$$

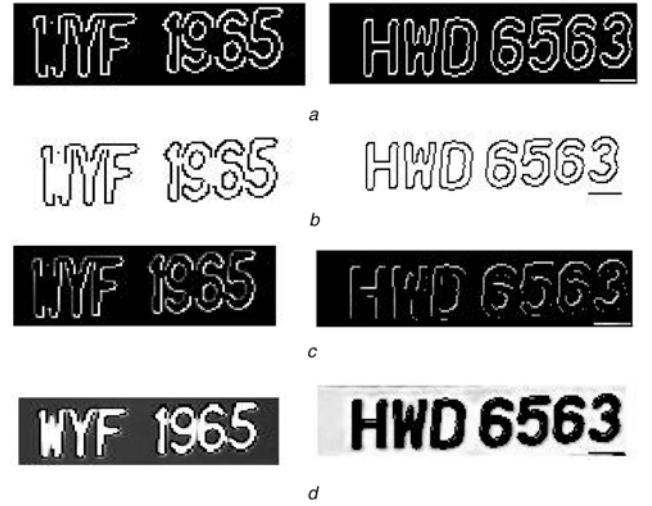


Fig. 3 Foreground and background separation using edge pixels

- a Foreground of the normal and taxi plate images
- b Background of the normal and taxi plate images
- c Grey values for the foreground pixels of the normal and taxi plate images
- d Grey values for the background pixels of the normal and taxi plate images

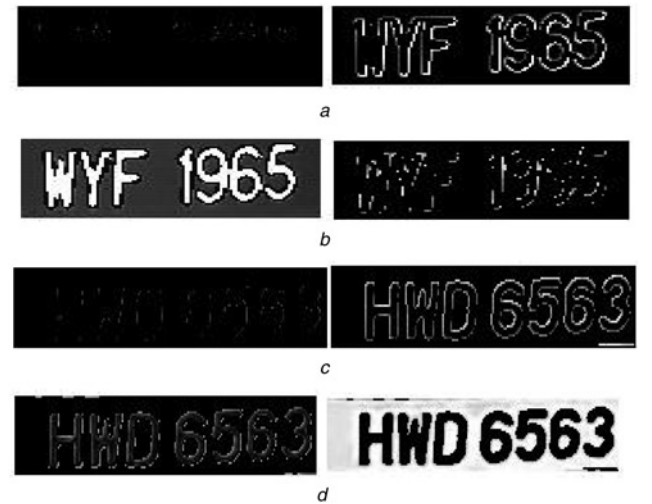


Fig. 4 Min and Max clusters for foreground and background images of normal and taxi license plate images

- a Min and Max clusters for the foreground of normal plate image
- b Min and Max clusters for the background of the normal image
- c Min and Max clusters for the foreground of taxi image
- d Min and Max clusters for the background of taxi image

$$GB_{x,y} = \begin{cases} G_{x,y}, & \text{if } \text{Canny}(x, y) = 0 \\ 0, & \text{else} \end{cases} \quad (2)$$

The proposed method applies K -means clustering with $K=2$ on intensity values of foreground and background of private and public plate images to classify pixels which have high intensity values into Max cluster and pixels which have low intensity values into Min cluster as shown in Figs. 4a–d, respectively. It is noted from Figs. 4a and b that the number of pixels classified into the Min cluster is higher than that of the Max cluster. Though the Max cluster gets high values, the number of pixels in the cluster is lower than that of pixels in the Min cluster. Therefore, the number of pixels in the cluster is considered as the weight, which is multiplied by the standard deviation as defined in (3). On the other hand, it is noted from Figs. 4c and d that the number of pixels which are classified into the Min cluster is lower than that of the

Max cluster. This cue helps us to derive a hypothesis using the number of pixels in clusters and the standard deviations to classify private and public plate images.

The hypotheses H-1, H-2 and H-3 are derived as defined in (4)–(6), where FN_{\min} and FN_{\max} denote the number of pixels in min and max clusters of the foreground, respectively, while BN_{\min} and BN_{\max} denote the number of pixels in background-min and max clusters, respectively. $BStd_{\min}$ and $BStd_{\max}$ denote standard deviations of the background of min-cluster and max cluster, respectively. The proposed method considers each response of the hypothesis as ‘1’ if it satisfies the condition, else it is considered as a response ‘0’. Out of the three responses, if two responses are ‘1’, the input image is identified as a normal one; else it is a taxi image. Therefore, if two responses are ‘0’, the image is identified as a taxi. In this way, the proposed method tests all the eight combinations of three responses for the input image. This process is called as voting as defined in (7), where ϑ is the majority variable, which is set to be greater than or equal to 2 for private and less than 2 for the public

$$Std_j = \sqrt{\frac{\left(\sum_{\gamma=1}^m M_j - X_\gamma\right)^2}{m}} \quad (3)$$

where M_j is the mean of the j cluster, X denotes intensity values, and m is the total number of the pixels in cluster j .

$$H-1 = \begin{cases} 1 & \text{if } FN_{\min} > FN_{\max} \\ 0 & \text{else} \end{cases} \quad (4)$$

$$H-2 = \begin{cases} 1 & \text{if } BN_{\min} > BN_{\max} \\ 0 & \text{else} \end{cases} \quad (5)$$

$$H-3 = \begin{cases} 1 & \text{if } BN_{\min} * BStd_{\min} > BN_{\max} * BStd_{\max} \\ 0 & \text{else} \end{cases} \quad (6)$$

$$\text{Voting} = \begin{cases} 1 & \text{if } (H-1) + (H-2) + (H-3) > \vartheta \\ 0 & \text{else} \end{cases} \quad (7)$$

3.2 LSTM for license plate recognition

The proposed architecture of license plate recognition is shown in Fig. 5, where we can see there are three steps, namely, CNN for feature extraction, LSTM for extracting sequence, and connectionist temporal classification (CTC) for the final recognition.

The architecture for CNN is derived from visual geometry group (VGG) (very deep architecture) as shown in Fig. 5, where the proposed method constructs the components of convolutional layers by considering convolutional and max-pooling layers from the basic VGG [15]. Then the components are used to extract features that represent the input image. Before passing the input image to the network, all the images are scaled to a standard size, namely, $32 \times 100 \times 3$, as shown in Fig. 6. For the input image in Fig. 7a, the proposed method generates a feature map as shown in Fig. 7b, from which the sequence of the feature are extracted given by the convolutional layers. The output of this process is considered as the input for the recurrent layers. In other words, the proposed method scans the feature map from left to right column by column to generate feature vectors of a feature sequence. Therefore, each column of the feature map corresponds to a rectangle region of the input image. The proposed method arranges rectangle regions in the order from left to right with respect columns.

It is noted that traditional gradient vanishing does not have the ability to extract context to predict the next state [14]. To find a solution to this issue, the long-short term memory (LSTM) has been developed, which is a special type of RNN unit. The RNNs are special neural networks, which exploit past contextual

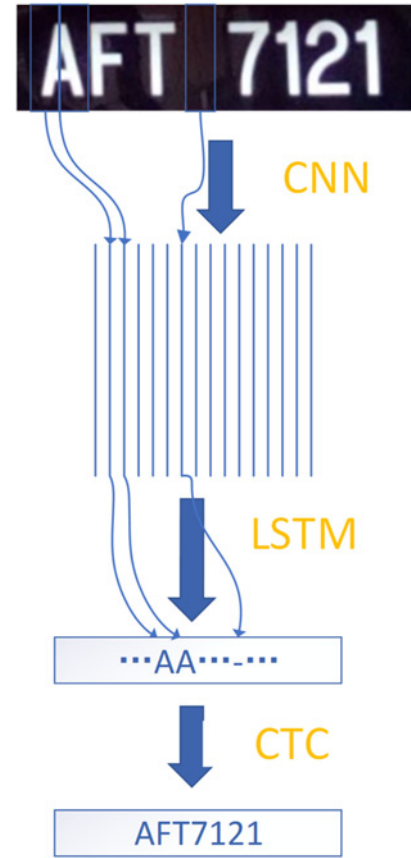


Fig. 5 Proposed architecture of license plate recognition

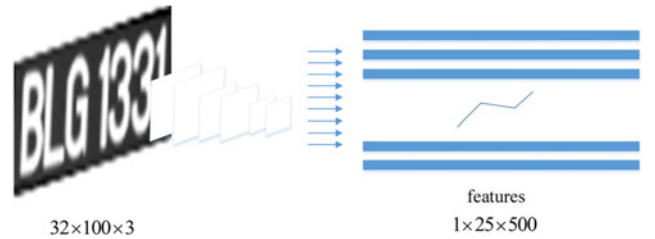


Fig. 6 CNN is used for feature extraction

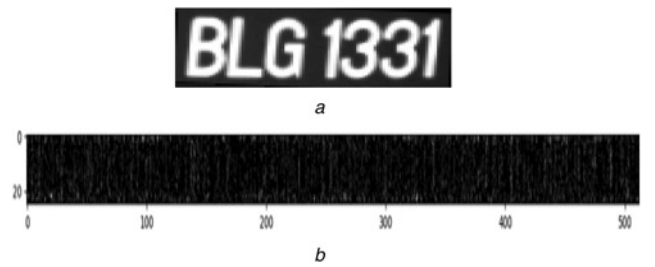


Fig. 7 Feature extraction using CNN

a Input license plate images
b Feature maps of CNN forward propagation

information and in turn help in obtaining stable sequence recognition than treating each feature independently as defined in (8). LSTM is illustrated in Fig. 8a, which comprises memory cells and three multiplicative gates, namely, the input, output and forget gates. In general, memory cells are used to store the past context information, the input and output cells are used to store the context for a long period of time. Similarly, a forget cell is used to clear the context information in the cells. To strengthen the feature extraction and selection, the proposed method combines both

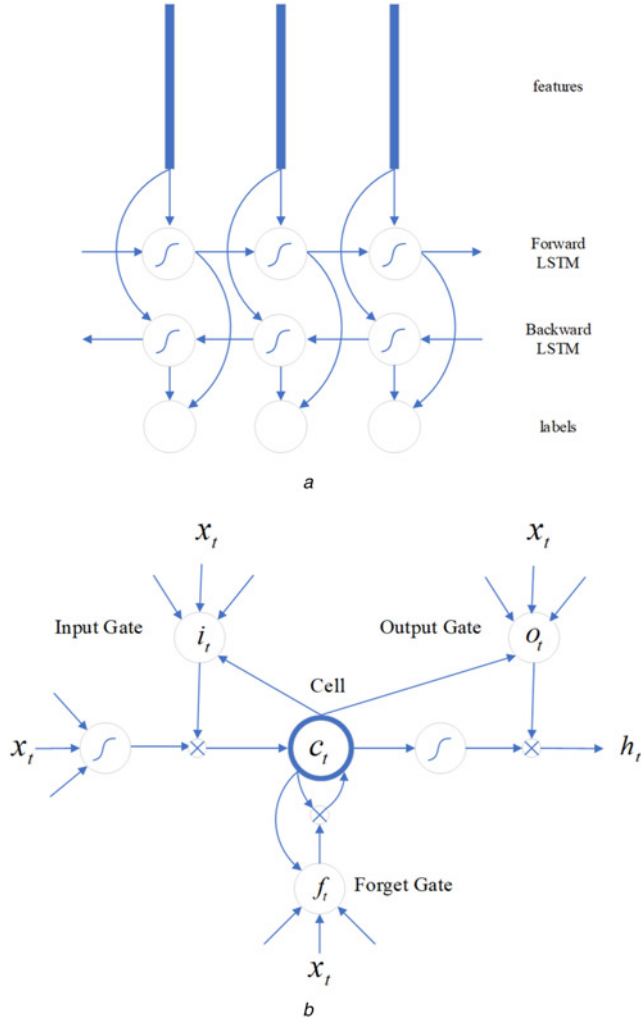


Fig. 8 Bi-directional LSTM structure

a LSTM structure unit
b LSTM forward and backward directions

forward and backward directional LSTM as shown in Fig. 8b as defined in (9), which considers the past and future context. The proposed method uses the soft-max layer for transforming the BLSTM states into probability distribution of each character class.

Further, to convert the sequence of the probability, the proposed method uses a CTC, which does not require character segmentation. This process decodes the predicted probability as output labels [15]. For instance, the proposed method decodes the label sequence, -M-AA-C-33-4-566-7- as MAC34567 by combining two nearest characters which are the same, and deletes the empty reference '-'. Finally, the proposed method chooses the one with the greatest probability for recognition as defined in (10)

$$h_t = \text{RNN}(x_t, h_{t-1}) \quad (8)$$

where h_t denotes the memory state, x_t denotes the current input and h_{t-1} denotes the previous state

$$\begin{aligned} h_t^f &= \text{LSTM}_f(x_t, h_{t-1}^f) \\ h_t^b &= \text{LSTM}_b(x_t, h_{t+1}^b) \end{aligned} \quad (9)$$

where f refers the forward LSTM, b refers the backward LSTM

$$\text{label} = \arg \max_l P(l|s) \quad (10)$$

Table 1 Summary of proposed CNN architecture

Layer type	Parameter
Convolution	Kernel size: 3×3 , Stride: 1, Padding: 'SAME', channel: 64
MaxPooling	Kernel size: 2×2 , Stride: 2, Padding: 'VALID'
Convolution	Kernel size: 3×3 , Stride: 1, Padding: 'SAME', channel: 128
MaxPooling	Kernel size: 2×2 , Stride: 2, Padding: 'VALID'
Convolution	Kernel size: 3×3 , Stride: 1, Padding: 'SAME', channel: 256
Convolution	Kernel size: 3×3 , Stride: 1, Padding: 'SAME', channel: 256
MaxPooling	Kernel size: 2×1 , Stride: 2×1 , Padding: 'VALID'
Convolution	Kernel size: 3×3 , Stride: 1, Padding: 'SAME', channel: 512
BatchNormalization	
Convolution	Kernel size: 3×3 , Stride: 1, Padding: 'SAME', channel: 512
BatchNormalization	
MaxPooling	Kernel size: 2×1 , Stride: 2×1 , Padding: 'VALID'
Convolution	Kernel size: 2×2 , Stride: 2×1 , Padding: 'SAME', channel: 512

The padding type 'Valid' means no padding, 'SAME' means that the output feature map has the same spatial dimensions as the input feature map. Zero padding is introduced to make the shapes match as needed, equally on every side of the input map.

The configuration information of the proposed architecture is summarised in Table 1.

4 Experimental results

Since the proposed method consists of two steps, we conduct experiments on classification and recognition. For evaluating the proposed method, we consider live data given by the MIMOS company funded by Malaysian Government [16], which comprises 874 private images and 961 public images. This dataset includes images affected multiple adverse factors such as low resolution, blur and severe illumination effect, images with different background colour, and images affected by head light, speed of vehicles and so on. We also consider one more benchmark dataset, namely, UCSD [17], which consist of 18,270 images. This dataset includes images captured by different angles, distances and multiple cars in one image. In total, there are 20,105 images for experimentation in this work. Note: since MIMOS dataset contains two different types images, namely, private which has a dark background, and the public which has a white background as shown in Figs. 9a and b, respectively, where one can see a different colour for the background. On the other hand, UCSD images have a white background like public images in MIMOS as shown in Fig. 9c, where we can see images suffer from poor quality and degradations. Therefore, we use the MIMOS dataset for classification and show classification is useful for improving recognition performance. However, we consider UCSD as a benchmark dataset to show that the proposed recognition is effective and generic in this work.

To measure the performance of the classification step, we consider classification rate with a confusion matrix for experiments on classification. Since there is no ground truth for the dataset, we count manually to calculate measures. Similarly, to measure the performance of recognition, we consider the recognition rate. In addition, to validate the classification, we calculate recognition rates before and after classification. Before classification considers images of two classes as the input for calculating recognition rates using different methods. After classification includes images of individual classes for calculating recognition rates using different methods. Besides, the recognition methods are tuned according to the complexity of the classes after classification such that the same method can score better recognition rate compared to before classification.

To show the proposed classification is superior to the existing methods, we implement two latest classification methods. The first one is Xu *et al.*'s method [5], which explores the uniform colour of text components for the classification of captions and scene texts in video. The second one is Roy *et al.*'s method [7], which proposes tampered features for separating captions and scene texts in video. The main reason to choose these two existing methods for the comparative study is that both the methods have the same objective as the proposed method. The methods consider scene

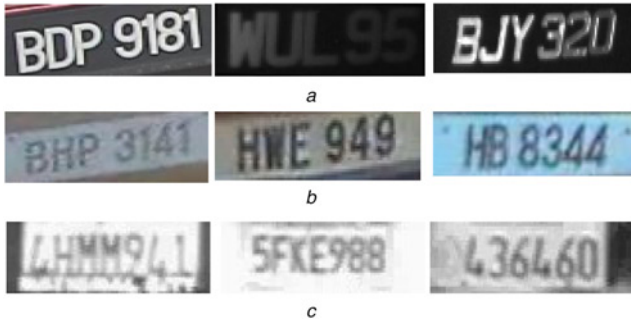


Fig. 9 Sample images of the different database

a Normal plate images from MIMOS
b Taxi plate images from MIMOS
c Sample images from USCD

texts are unpredictable, which suffer from distortions affected by multiple causes as public images in this work. Similarly, the methods consider caption texts have good clarity and contrast, which is the same as private images compared to public images.

For recognition, we implement one state-of-the-art method [2] which proposes strokelets for recognition of texts in natural scene images. Since license plate recognition is almost the same as natural scene text recognition, we use the above existing methods for comparative study in this work. In addition, according to the method, it is robust to the poses caused by distortion and generic. The method [2] uses a random forest classifier for recognition. In order to test the features extracted by the method, we pass the same features to CNN for recognition, which is considered as one more existing method for comparative study. We also use the system which is available online [3] for license plate recognition. This system explores deep learning as the proposed method for recognition. In summary, we choose one state-of-the-art method from natural scene text and one more from license plate recognition to show that existing natural scene text and license plate recognition methods may not work well for license plate images affected by multiple adverse factors.

4.1 Evaluating classification

Sample successful and unsuccessful classification images of the proposed classification are shown in Figs. 10a and b, respectively. Fig. 10 shows that the proposed classification works well for images with blur, low contrast and illumination effects. However, the proposed classification fails to classify images that have too low resolution and poor quality. Therefore, there is a scope for improvement and extension of the proposed classification in the future. Quantitative results of the proposed classification and existing methods are reported in Table 2, where it is noted that the proposed classification gives better results than the existing methods. The reason for poor results of the existing methods is that the existing methods depend on character shapes, while the proposed classification depends on the distribution of foreground and background pixels. For instance, tampered features proposed by Roy *et al.* [7] exist only for caption text but not license plate images.

4.2 Evaluating recognition

For recognition experiments, we split the dataset into two training and testing samples. For all the experiments, we consider 70% for training, 15% for validating and 15% for testing of the datasets. Qualitative results of the proposed recognition and existing methods are shown in Fig. 11, where it can be seen that the proposed method recognises all the input images of three different databases accurately. However, existing methods do not work for all the images. Fig. 11 shows that the strokelet based method [2] misses characters for several input images especially those affected by blur and suffer from poor quality. This is due to that the



Fig. 10 Sample successful and unsuccessful images of the proposed method

a Sample successful images
b Sample unsuccessful images

Table 2 Confusion matrix of the proposed method and existing methods (%)

	Proposed method		Xu <i>et al.</i> [5]		Roy <i>et al.</i> [7]	
	Private	Public	Private	Public	Private	Public
private	77.54	22.46	57.6	42.4	53.5	46.5
public	19.78	80.22	33.69	66.31	31.98	68.02

Bold values indicate, highest results achieved by the respective methods.

Source	Images	Proposed	Strokelet	Strokelet+CNN	HLPR
Private	BDP 9181	BDP9181	BR9AT	BP981	BDP9187
	WUL 95	WUL96	UL9S	UL95	WUL95
	BJY 320	BJY320	BUMA20	BJY320	BJY320
Public	BHP 3141	BHP3141	BPSS	BP31	BHP3P1
	HWE 949	HWE949	AHESA S	HWE949	HWE949
	HB 8344	HB8344	HSSS44	HBS8344	M8344
UCSD	4HMM941	4HMM941	CHE9A1	4HH941	4HMM94L
	5FXE988	5FXE98	SSETS8	5FE988	5FXE98
	436460	436460	S4S6E60	436460	436460

Fig. 11 Recognition results of the proposed and existing methods on different databases

extracted strokelet features are not robust to segment characters from such images. When we use strokelet+CNN, the recognition results are better than the strokelet alone method. However, if the character segmentation step misses characters, recognition by CNN does not make difference in recognition. Similarly, the HLPR system is better than strokelet and strokelet+CNN but lower than the proposed method. When we look at the failure results of the HLPR system [3], the system is not robust to poor quality and degraded images. On the other hand, the proposed recognition is better than existing methods in terms of robustness. The main reason for good results is that the way we use CNN for feature

Table 3 Recognition rate of the proposed and existing methods for a private and after classification

Methods	UCSD	Before classification		
		Private + public	Private	Public
proposed	86.37	71.23	77.03	90.51
Strokelet [2]	23.76	22.72	34.42	24.90
Strokelet + CNN	34.39	34.78	55.31	33.64
HLPR [3]	83.65	69.53	79.34	85.32

Bold values indicate, highest results achieved by the respective methods.



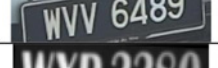
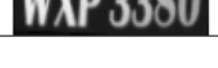
Images	Recognition results	Ground truth
	WBJF1777	BJF1777
	JK14	JKM14
	NCV64	WVV6489
	WXP380	WXP3380

Fig. 12 Limitation of the proposed method

extraction and BLSTM makes sense for tackling the challenges posed by adverse factors. This is the advantage of the proposed recognition compared to existing methods.

Quantitative results of the proposed recognition and existing methods are reported in Table 3. It is observed from Table 3 that the proposed recognition scores the best for UCSD and before-after classification of MIMOS dataset compared to the three existing methods. The HLPR system is the second best compared for all the experiments. Interestingly, when we look at the results of before classification and after classification, all the methods including the proposed method report low recognition rate prior to classification and high recognition rate after classification. The difference between before and after is significant. This is due to the advantage of classification MIMOS data into private and public, which helps us to reduce the complexity. Because of classification, we can tune or modify parameters for tackling the issues with the dataset. In this experiment, we tune the key parameters of CNN and BLSTM based on training samples with respect classes. Therefore, we can conclude that the recognition performance improves significantly when we classify complex data into sub-classes according to complexity. In other words, rather than developing a new method, this is one way of achieving a high recognition rate for complex data. It is also observed from Table 3 that the proposed recognition scores the best recognition rate for the UCSD dataset, which shows the method can be used for any other dataset irrespective of scripts. The main reason for the poor results of the existing methods is the same as discussed earlier.

When the images are oriented in different directions, the proposed recognition method does not work well as shown sample results in Fig. 12. This is not a problem of feature extraction but it is the limitation of CTC, which gets confusing when an image is rotated in different directions. However, it is noted that in the case of license plate images, it is seldom to have different orientations like texts in natural scene images. Therefore, this issue may not be a major drawback of the proposed recognition method.

5 Conclusion and future work

In this work, we have proposed the combination of CNN and RNN for recognising license plate images through classification. For the classification of private and public number plate images, the

proposed method explores foreground and background separation and then clustering in a new way. For recognition, we combine the merits of CNN and RNN to handle the issues of poor quality, complex background, blur and noise. Experimental results on classification show that the proposed classification is better than the existing methods in terms of classification rate. Similarly, recognition experiments on MIMOS dataset and benchmark UCSD dataset show that the proposed recognition is better than existing methods. In addition, the recognition performance of different recognition methods including the proposed method before and after classification shows that classification is useful for achieving high recognition rate for a complex dataset rather than developing a new method. As discussed in Section 4, since there are limitations to the proposed classification and recognition, we have planned to overcome those limitations in the near future.

6 Acknowledgment

This research work was supported by the Faculty of Computer Science and Information Technology, the University of Malaya under a special allocation of Post Graduate Funding for the RP036B-15AET project. The work described in this paper was supported by the Natural Science Foundation of China under grant no. 61672273, and the Science Foundation for Distinguished Young Scholars of Jiangsu under grant no. BK20160021.

7 References

- [1] Du, S., Ibrahim, M., Shehata, M., *et al.*: 'Automatic license plate recognition (ALPR): a state-of-the-art review', *IEEE Trans. Circuits Syst. Video Technol.*, 2013, **1**, pp. 311–325
- [2] Bai, X., Cong, Y., Wenyu, L.: 'Strokelets: a learned multi-scale mid-level representation for scene text recognition', *IEEE Trans. Image Process.*, 2016, **25**, pp. 2789–2802
- [3] 'High Accuracy Chinese Plate Recognition Framework', <https://github.com/zeusees/HyperLPR>, Accessed 18 May 2018
- [4] Raghunandan, K.S., Shivakumara, P., Kumar, G.H., *et al.*: 'New sharpness features for image type classification based on textual information'. Proc. Document Analysis Systems (DAS), Greece, 2016, pp. 204–209
- [5] Xu, J., Shivakumara, P., Lu, T., *et al.*: 'Graphics and scene text classification in video'. Proc. Int. Conf. on Pattern Recognition (ICPR), Sweden, 2014, pp. 4714–4719
- [6] Shivakumara, P., Kumar, N.V., Guru, D.S., *et al.*: 'Separation of graphics (superimposed) and scene text in video'. Proc. Document Analysis Systems (DAS), France, 2014, pp. 344–348
- [7] Roy, S., Shivakumara, P., Pal, U., *et al.*: 'New tampered features for scene and caption text classification in video frame'. Proc. Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), China, 2016, pp. 36–41
- [8] Yang, Y., Li, D., Duan, Z.: 'Chinese vehicle license plate recognition using kernel based extreme learning machine with deep convolutional features', *IET Intell. Transp. Syst.*, 2018, **12**, pp. 213–219
- [9] Abedin, M.Z., Nath, A.C., Dhar, P., *et al.*: 'License plate recognition system based on contour properties and deep learning model'. Proc. Region 10 Humanitarian Technology Conf. (R10-HTC), Bangladesh, 2017, pp. 590–593
- [10] Polshetty, R., Roopaei, M., Rad, P.: 'A next generation secure cloud based deep learning license plate recognition for smart cities'. Proc. Int. Conf. on Machine Learning and Applications (ICMLA), USA, 2016, pp. 286–294
- [11] Montazzolli, S., Jung, C.: 'Real time Brazilian license plate detection and recognition using deep convolutional neural networks'. Proc. SIBGRAPI on Conf. on Graphics, Patterns and Images, Brazil, 2017, pp. 55–62
- [12] Bulan, O., Kozitsky, V., Ramesh, P., *et al.*: 'Segmentation and annotation free license plate recognition with deep localization and failure identification', *IEEE Trans. Intell. Transp. Syst.*, 2017, **18**, pp. 2351–2363
- [13] Selmi, Z., Halima, M.B., Alimi, A.M.: 'Deep learning system for automatic license plate detection and recognition'. Proc. Int. Conf. on Document Analysis and Recognition (ICDAR), Japan, 2017, pp. 1132–1137
- [14] Li, H., Wang, P., You, M., *et al.*: 'Reading car license plates using deep neural networks', *Image Vis. Comput.*, 2018, **72**, pp. 14–23
- [15] Shi, B., Xiang, B., Cong, Y.: 'An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, **39**, (11), pp. 2298–2304
- [16] Shivakumara, P., Konwer, A., Bhowmick, A., *et al.*: 'A new GVF arrow pattern for character segmentation from double line license plate images'. Proc. Asian Conf. on Pattern Recognition (ACPR), China, 2017, pp. 782–787
- [17] Zamberletti, A., Gallo, I., Noce, L.: 'Augmented text character proposal and convolutional neural networks for text spotting from scene images'. Proc. Asian Conf. on Pattern Recognition (ACPR), Malaysia, 2015