

ORIGINAL RESEARCH PAPER

Vehicle license plate recognition for fog-haze environments

Xianli Jin¹ | Rucong Tang¹ | Linfeng Liu^{1,2} | Jiagao Wu^{1,2}
¹ School of Computer, Nanjing University of Posts and Telecommunications, Nanjing, China

² Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing, China

Correspondence

Linfeng Liu, School of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China.

Email: liulinfeng@gmail.com

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 61872191, 61872193

Abstract

The technique of vehicle license plate recognition can recognize and count the vehicles automatically, and thus many applications regarding the vehicles are greatly facilitated. However, the recognitions of vehicle license plates are extremely difficult especially in some fog-haze environments because the fog and haze blur the boundaries and characters of license plates significantly, which makes the license plates hard to be detected or recognised. To this end, this paper proposes a vehicle License Plate Recognition method for Fog-Haze environments (LPRFH). In LPRFH, a dark channel prior algorithm based on the local estimation of atmospheric light value is applied to dehaze the blurred images preliminarily. Then, the images are further dehazed, and the license plate regions are detected through a Joint Further-dehazing and Region-extracting Model on basis of an object detection convolution neural network. Finally, the image super-resolution is accomplished with a convolution-enhanced super-resolution convolutional neural network, and hence the characters of license plates can be recognised successfully. Extensive experiments have been conducted, and the results indicate that LPRFH can recognise the license plates accurately even in some severe fog-haze environments.

1 | INTRODUCTION

With the rapid development of economy and the continuous enlargement of cities, the transportation system has become an indispensable component of modern society. To manage the numerous vehicles conveniently, the Intelligent Traffic System (ITS) [1] is proposed to provide innovative services relating to the transport and the traffic management. Each vehicle around the world has a license number as its principal identifier, as an essential stage in ITS; the technique of vehicle License Plate Recognition (LPR) [2] can recognise and distinguish the vehicles automatically. LPR is widely used by police forces for law enforcement purposes, and it also plays an important role in the applications of vehicle tracking, traffic congestion monitoring, management of city streets and electronic toll collection on pay-per-use roads. As shown in Figure 1, the license plate is blurred by the fogs or hazes, and the characters of this license plate are hard to be identified.

An LPR system typically includes two parts: license plate detection and character recognition. Note that some environmental events (such as weather conditions, light intensities and

complex backgrounds) seriously affect the quality of vehicle images captured by monitoring cameras. Especially under the fog-haze weather which has become progressively more serious [3], the fog and haze blur the boundaries and characters of license plates significantly, which makes the license plates hard to be detected and recognised. Most of the existing license plate detection methods are based on the feature extraction technique, and they cannot detect the license plates accurately in fog-haze environments, due to the fact that the fog-haze interference and image ambiguity make the features very difficult to be extracted. Likewise, the character recognition is also affected by the fog-haze interference significantly.

This paper proposes a vehicle License Plate Recognition method for Fog-Haze environments (LPRFH), and as shown in Figure 2, this method provides a framework which includes the image-dehazing technique based on an Object Detection Convolution Neural Network (ODCNN) [4] and the super-resolution technique based on a convolution-enhanced Super-Resolution Convolutional Neural Network (SRCNN) [5]. Firstly, a dark channel prior algorithm based on the local estimation of atmospheric light value is applied to dehaze the images



FIGURE 1 License plate in fog-haze environment

preliminarily. Then, the images are further dehazed, and the license plates are extracted with an ODCNN to reduce the distortion of image restoration caused by the cumulative errors of multiple parameters, and thus the license plates can be accurately extracted from these improved images. Finally, the image super-resolution technique is realised by a six-layer convolution-enhanced SRCNN to improve the accuracy of the characters recognition of license plates. Especially, note that in our work, the ODCNN is trained on both real images and synthesised images, while the SRCNN is trained on synthesised images because the blurred images of moving vehicles are difficult to

obtain from the real scenes. The main innovations of this work are listed as follows:

- An image-dehazing technique based on an ODCNN which takes into account the fog concentrations is proposed. The images are divided into several regions according to the fog concentrations, and a finer transmission map [6] can be generated by the values of local atmospheric light rather than those of global atmospheric light, and then the fog-haze interference can be filtered from the original images as much as possible.
- A Joint Further-dehazing and Region-extracting Model (JFRM) constructed on basis of an ODCNN is provided. JFRM can dramatically reduce the distortion of image restoration caused by the cumulative errors and accurately detect the license plate regions of the preliminarily-dehazed images.
- In the proposed convolution-enhanced SRCNN, the layer number of convolution is increased from three layers to six layers, which makes much more image features be extracted, and thus the image quality can be significantly improved.

The rest of this paper is organized as follows: Section 2 gives some related works. Section 3 introduces the method of vehicle license plate recognition in fog-haze environments. Section 4 reports the performance evaluation and simulation results. Finally, Section 5 concludes this paper.

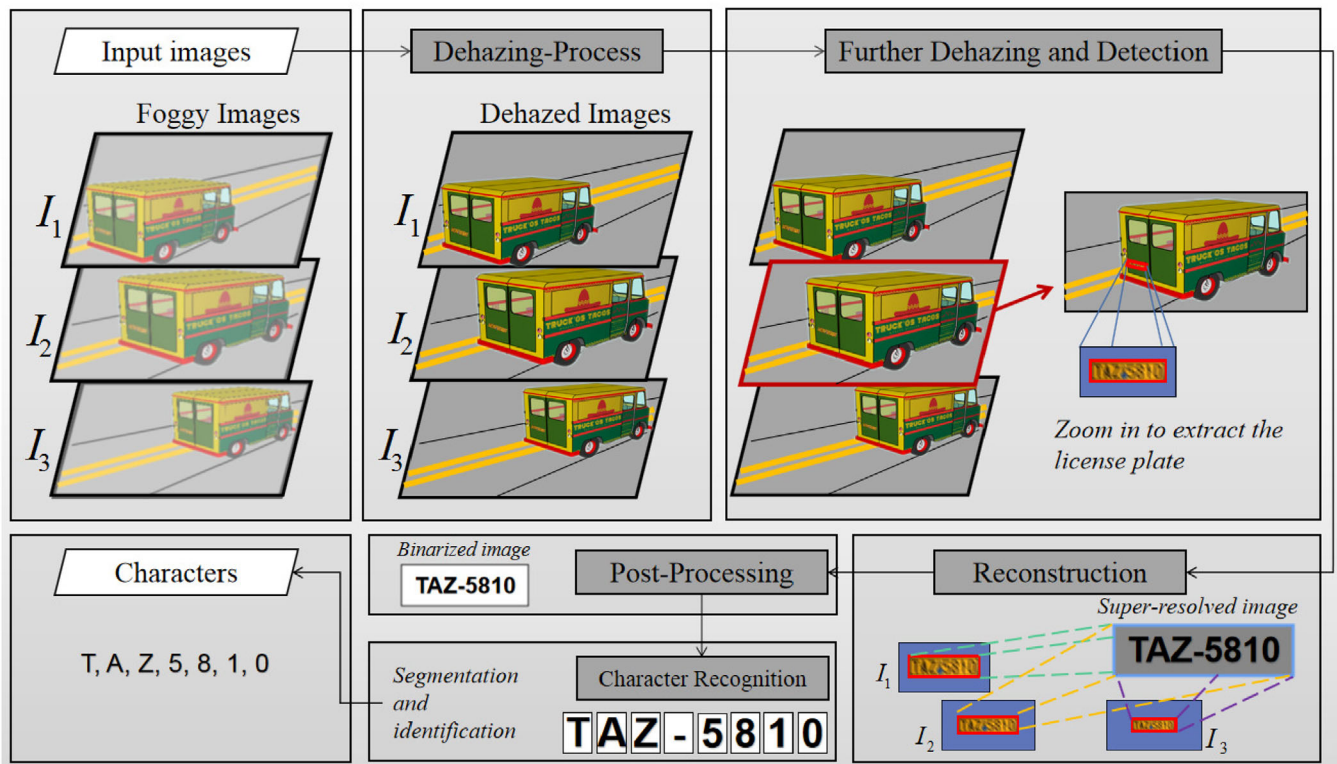


FIGURE 2 The LPRFH framework pipeline

2 | RELATED WORKS

2.1 | Image dehazing and object detection

The problem of image dehazing has been extensively studied [6–9]. In [7], McCartney et al. proposes an atmospheric light scattering model, that is based on the Mie scattering theory, and the incident light attenuation model and the atmospheric light imaging model are used in the imaging mechanism. According to the atmospheric light scattering model and the dark-object subtraction technique, He et al. [6] proposes a dehazing method based on dark channel where at least one color channel has some pixels with low intensities. Reference [8] formulates a haze removal model which integrates with random forest to estimate the transmissions. This model combines the dark channel, local max contrast, hue disparity and local max saturation. In [9], an efficient image haze removal algorithm is proposed, where the color attenuation prior is considered, and a linear model is constructed for the scene depth of the hazy image under this prior. Convolution Neural Networks (CNNs) have witnessed a prevailing success in computer vision tasks and can be introduced for the haze removals [10, 11]. Reference [10] proposes a light-weight CNN, which is designed based on a re-formulated atmospheric scattering model, instead of estimating the transmission matrix and the atmospheric light, respectively. [11] formulates an end-to-end generative method for the image dehazing issue. However, this method does not take into account the atmosphere scattering model.

CNN-based model has exhibited a great progress in terms of the object detection accuracy. In recent years, some light-weight deep learning models for object detection yield some preferable results. Most of these models are constructed based on SSD [12], SqueezeNet [13], AlexNet [14] or GoogleNet [15]. In these models, the pipelines of object detections comprise many components, such as pre-processing, large number of convolution layers and post-processing. The sliding window approach or region proposal method can be taken to evaluate the classifiers. These complex pipelines are computationally intensive and consequently slow. To improve this shortcoming, in [16], the You Only Look Once (YOLO) is proposed, and it is framed as a single regression problem. The outstanding advantage of YOLO is to encode the contextual information, and it makes less mistakes in term of confusing the background patches of images.

2.2 | Image super resolution and character recognition

Single image Super-Resolution (SR) is a process that obtains High-Resolution (HR) images from a single Low-Resolution (LR) image. With regard to the vision problems, the state-of-art super-resolution results can be produced by deep convolutional neural networks (DCNNs). Dong et al. [17] establish a deep learning approach with SRCNN. This method learns an end-to-end mapping between the low-/high-resolution images. SRCNN enhances the spatial resolution of an input LR image

by some hand-crafted upsampling filters, and thus this method optimises all layers instead of handling each component respectively. In [18], a highly accurate single-image super-resolution method is presented. They create a deeper network with 20 layers to learn a residual image based on SRCNN, and this method exhibits a preferable performance. [19] presents SRGAN, a Generative Adversarial Network (GAN) for the image Super-Resolution (SR), to recover the finer texture details. It is a framework capable of inferring photo-realistic natural images for 4× upsampling factors.

Character segmentation is the foundation of the character recognition. Reference [20] presents a character segmentation algorithm, which makes use of priori knowledge and realises the character segmentation based on connected domains. Besides the licence plates, character segmentation is applied in the handwriting text segmentation. Jun et al. [21] propose a new handwriting character segmentation method based on the non-linear clusterings. The whole text line is first divided into some strokes, and the similarity matrix is calculated. Then, the cluster labels of the strokes are obtained by a non-linear clustering method. Based on the cluster labels, the strokes are combined to form the characters. In [22], Roy et al. present an end-to-end real-time text localisation and recognition method to handle the digital documents in real scenes. [23] proposes a segmentation-free license plate recognition algorithm that utilises deep learning techniques in the character detection and recognition process. It improves the recognition results of low-resolution images.

Automatic License Plate detection and Recognition (ALPR) is a quite popular and active research topic in the field of computer vision. Reference [24] presents an automatic framework for License Plate (LP) detection and recognition from complex scenes, which is based on mask region convolutional neural networks used for LP detection, segmentation and recognition. An end-to-end ALPR method based on a hierarchical Convolutional Neural Network (CNN) is proposed in [25]. The core idea of the proposed method is to identify the vehicle and the license plate region using two passes on the same CNN, and then to recognise the characters using a second CNN. Wu et al. [26] proposes a novel PixTextGAN that leverages a controllable architecture that generates specific character structures for different text regions to generate synthetic license plate images with reasonable text details. Specifically, a comprehensive structure-aware loss function is presented to preserve the key characteristic of each character region and thus to achieve appearance adaption for better recognition. In [27], a robust framework for license plate recognition in the wild is designed to solve the problem of significantly reduced license plate recognition performance under feature rotation, distortion, occlusion, blur, shadow or extreme dark or bright conditions. It is composed of a tailored CycleGAN model for license plate image generation and an elaborate designed image-to-sequence network for plate recognition.

2.3 | Motivation of our work

Under the influence of fog-haze weather, the detection and recognition of license plates may become invalid. Considering

the aforementioned fact that the fog and haze blur the boundaries and characters of license plates significantly, it is impossible to achieve an ideal recognition accuracy. An intuitive idea to eliminate the effects of fog and haze is to employ the image dehazing technology and improve the feature extraction of license plate regions. However, much extra cost will be produced in the image dehazing process, while the dehazing effect still cannot be guaranteed.

CNN is widely used in computer vision tasks, and it can be introduced for the image dehazing issue. This paper attempts to improve the dehazing effect especially in the license plate region. To this end, after a preliminary dehazing process, we make the images be further dehazed, and the license plate regions are detected through a convolution neural network, that is, a joint optimisation is implemented to improve the dehazing effect. Finally, a super-resolution network is exploited to improve the image resolution, and hence improve the character recognition accuracy.

3 | LICENSE PLATE RECOGNITION METHOD FOR FOG-HAZE ENVIRONMENTS

3.1 | Improved dark channel prior algorithm

In the existing image dehazing works, the atmospheric scattering model is widely used. According to the atmospheric scattering model, the physical description of the degraded image is approximated as:

$$I(x) = J(x) \cdot t(x) + \mathcal{A} \cdot [1 - t(x)], \quad (1)$$

where $t(x) = e^{-\beta \cdot d(x)}$, and $t(x)$ represents the permeability of the atmospheric medium, β represents the atmospheric scattering coefficient, d is the depth of scene, x is the pixel position, $I(x)$ is the fogged degraded image, $J(x)$ is the restored image, and \mathcal{A} is the atmospheric light value. Obviously, the solution of $J(x)$ requires the prior knowledge of $t(x)$ and \mathcal{A} .

In most local areas of an image, the minimum intensity value of the color channels of pixels is extremely close to 0, and the color channel is referred to as a dark channel. With regard to an input image J , its dark channel can be expressed as:

$$J^{dark}(x) = \min_{j \in \Omega(x)} \left[\min_{c \in \{r, g, b\}} J^c(j) \right], \quad (2)$$

where $J^c(j)$ represents the channel of a color image, and $\Omega(x)$ represents a window centered on pixel x . According to the dark channel prior identification [6]:

$$J^{dark} \rightarrow 0. \quad (3)$$

Besides, by (1) we obtain that:

$$\frac{I^c(x)}{\mathcal{A}^c} = t(x) \cdot \frac{J^c(x)}{\mathcal{A}^c} + 1 - t(x). \quad (4)$$

Assume that the transmittance $t(x)$ in each window is a constant, and thus there is:

$$\min_{j \in \Omega(x)} \left[\min_c \frac{I^c(j)}{\mathcal{A}^c} \right] = t(x) \cdot \min_{j \in \Omega(x)} \left[\min_c \frac{J^c(j)}{\mathcal{A}^c} \right] + 1 - t(x). \quad (5)$$

Combining (1),(2) and (5), we can obtain the following formula:

$$t(x) = 1 - \omega \cdot \min_{j \in \Omega(x)} \left[\min_c \frac{I^c(j)}{\mathcal{A}^c} \right]. \quad (6)$$

ω is a parameter to adjust the dehazing degree. By (1) and (6), we can get that:

$$J(x) = \frac{I(x) - \mathcal{A}}{\max[t(x), t_0]} + \mathcal{A}. \quad (7)$$

The atmospheric scattering model given in (1) indicates that the image dehazing is mainly dependent on the hypothesis or prior knowledge. Thus, the value of atmospheric light value \mathcal{A} will have an impact on the transmittance and the image dehazing performance. In the foggy image, the distributions of the fog and the brightness of each image are typically not uniform. With the estimated global atmospheric light value, the image after a dehazing recovery will produce a halo phenomenon in the area of fog concentration transition, while the local area will produce a high or an overbright phenomenon. To this end, this paper proposes a fog concentration method, which divides each image into several regions with different fog concentrations, and then processes the regions with different fog concentrations, respectively. Note that in this method, the local atmospheric light value is taken to measure the transmittance rather than the global value.

3.1.1 | Regional division

Due to the added atmospheric light in fog-haze weather, the captured foggy image is usually brighter than the fog-free image. Hence, with regard to the different regions of the foggy image, the brightnesses with different fog concentrations are quite different. Besides, the dark primary color in the dark channel of a fog-free image is close to 0, while in the foggy case, the value of the dark channel becomes larger due to the fog interference. Therefore, here, the image brightness and the dark channel are taken as the feature variables to divide the pixels in each foggy image into three regions. Firstly, we gray-process the image, and these grayscale values represent the luminances of each pixel. Then, we calculate the dark channel value for each pixel, with the luminances value and dark channel value being vectorised. Finally, the k -mean clustering algorithm is taken to automatically divide the pixels into different fog concentration regions. With regard to different fog concentrations, the atmospheric light values and transmittance are estimated, respectively.

Algorithm 1 Regional division**Input:** $P = \{p_1, p_2, \dots, p_n\}$, a set of foggy images.**Output:** Different fog concentration regions.

```

1:  $G \leftarrow \{\}$ ;
2:  $D \leftarrow \{\}$ ;
3: for  $p_i$  in  $P$  do
4:    $G \leftarrow \alpha 2\_IMREAD\_GRAYSCALE(P)$ ;
5:    $V1 = np.min(P, 2)$ ;
6:    $D \leftarrow zmMinFilterGray(V1, 7)$ ;
7:   if  $G \neq \text{empty}$  or  $D \neq \text{empty}$  then
8:     return  $p_i$ ;
9:   else
10:     $data = [G, D]$ ;
11:     $[idxs, centers] = kmeans(data, 3)$ ;
12:     $plot(data)$ ;
13:   end if
14: end for

```

3.1.2 | Local estimation

After the division of the fog concentration regions, the local atmospheric value is independently estimated for each region. Then, the dark primary color of the foggy image is estimated, and in each region, we select the pixels with the top-0.1% brightnesses of the dark primary colors. Finally, in the selected pixels, the maximum brightness is taken as the atmospheric light value of the region:

$$A_i = \max_{y \in \Omega(I)} [I_i(y)], i = 1, 2, 3, \quad (8)$$

where $\Omega(I)$ represents the regions corresponding to the top-0.1% pixels. According to formula (8), the local atmospheric light values in different concentration regions are calculated, so the transmittance can be calculated by (6), and then the reconstructed image can be obtained according to (7).

3.2 | Joint optimisation of further dehazing and detection

Under the influence of the fog environment factors, in order to detect the license plate more accurately, against the shortcomings of the traditional dehazing method based on parameter estimation, we introduce the idea of joint optimisation of further dehazing and detection. In the YOLOv3 detection model, the image refinement and dehazing process are embedded, and the dehazing process of the previous step can be improved, which is described as follows:

We provide a JFRM with convolution neural network, which is composed of three modules: (i) multi-scale densely connected dehazing optimisation network; (ii) feature extraction backbone network; (iii) feature interaction detection network. The JFRM network structure is shown in Figure 4.

3.2.1 | JFRM network structure

- (i) In the multi-scale dehazing optimisation network, we add some filters to form the multi-scale features, while the convolution layers and connection layers are combined to compensate for the missing information in the convolution process. Moreover, we endow some residual blocks to deepen the perception of the haze structure in the images, so as to accurately capture the boundary details of objects with different depths, while the perceptual boundary details are expected to be more reliable than the pixel-level details.
- (ii) In the feature extraction backbone network, the Darknet-53 network in YOLOv3 is taken as the main network structure for feature extraction. This network combines the convolution layers and the residual blocks. With regard to a resized image with the size of $416 \times 416 \times 3$, the convolution layer uses 32 different 3×3 convolution kernels, and slides on this image to perform a convolution operation and extract the features. Especially, note that the residual blocks will yields the residual results without changing the sizes of input image and output image. The residual blocks can control the propagation of the gradient, avoid some problems of gradient dispersion and gradient explosion, and strengthen the training speed.
- (iii) In the feature interaction detection, we use the yolo layer to interact with the features extracted by the Darknet-53 multiple downsamplings. It outputs the last three features from five downsamplings to be mapped into three yolo layers with different scales. The feature map size of the minimum scale is 13×13 , the feature map size of the mesoscale is 26×26 , and the feature map size of the maximum scale is 52×52 . Thus, the yolo layers of three different scales can apply a 1×1 convolution kernel to produce an output similar to a fully connected layer.

3.2.2 | JFRM loss function

We give a loss function to measure the performance of JFRM. For the detection module, we improve the prediction according to the coordinates of bounding boxes, the confidences and the conditional class probabilities. In the further dehazing process, we measure the difference between the output image and the original fog-free image.

When the bounding boxes with different sizes are predicted, the prediction error may be completely different. In order to solve this problem, we use the square root of the width and height instead of the original width and height. The prediction error function regarding the coordinates of bounding boxes is written as follows:

$$\lambda_t \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right], \quad (9)$$

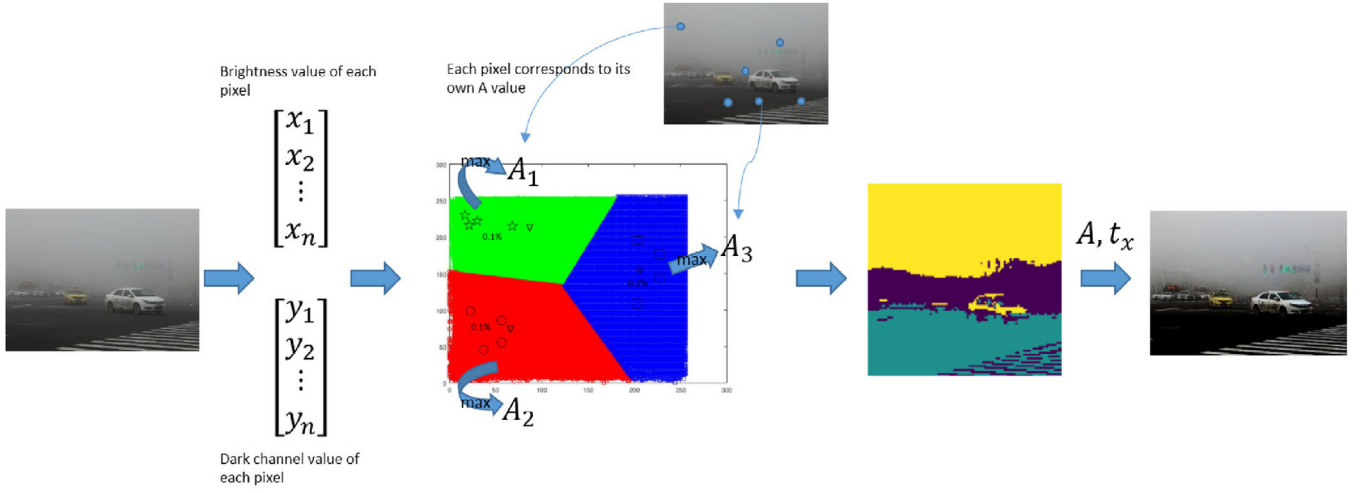


FIGURE 3 Estimation of local atmospheric light value A

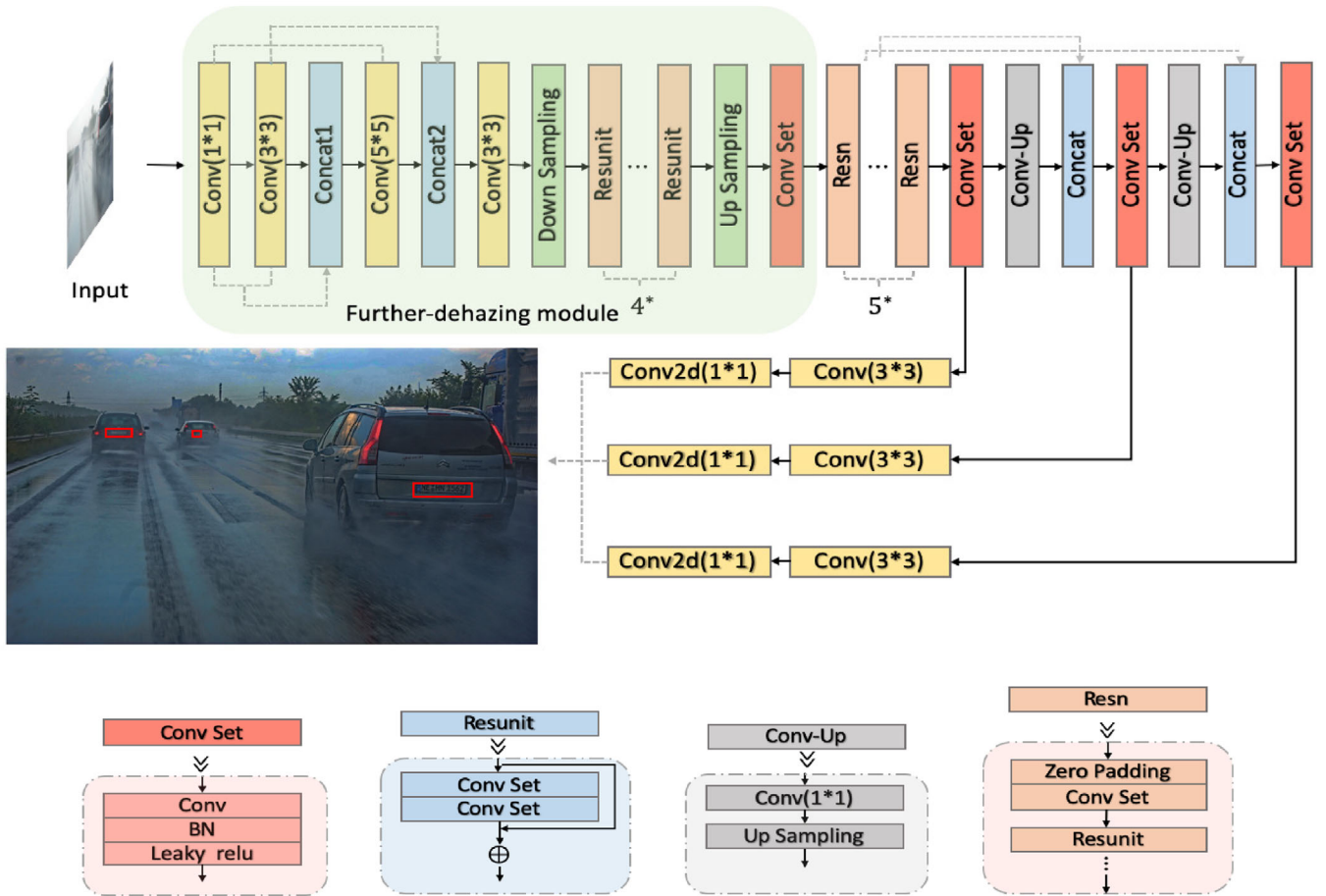


FIGURE 4 JFRM network structure

where $(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i)$ denotes the coordinate of the ground truth, (x_i, y_i, w_i, h_i) denotes the coordinate of the obtained bounding box, S is the scale of feature map, B is the size of the anchor box in a image grid, λ_t and λ_f are the preset weights, and the value of \mathbb{I}_{ij}^{obj} is set 1 or 0, which indicates whether the j -th anchor

box in the i -th grid is responsible for the object prediction or not, respectively.

With regard to the confidence error, the cross entropy is taken to calculate the loss function under two cases: an object and no object. This is because YOLOv3 divides each image by

the grid, and it does not include the object to be detected. If the two cases are identically treated, it will cause the detection network to be unstable or even divergent. Thus, the confidence loss weights of the bounding boxes without the object are reduced, and the loss weights are set to $\lambda_l = 5$ and $\lambda_f = 0.5$, respectively. Under such settings, the bounding boxes in each grid will be more standardised. The prediction error function of bounding box confidence is given as follows:

$$\begin{aligned}
& - \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} \left[\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j) \right] \\
& - \lambda_f \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{noobj}} \left[\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j) \right],
\end{aligned} \quad (10)$$

where \hat{C}_i^j is the confidence, and C_i^j is an IOU (Intersection Over Union) value of the obtained bounding box and the ground truth.

Then, with regard to the classification error, measuring the determination of the class which the bounding box belongs to, we also take the cross entropy as the loss function. When the j -th anchor box of the i -th grid is responsible for an object, the bounding box generated by the anchor box will calculate the classification loss function. The prediction error function of classification is given as follows:

$$- \sum_{i=0}^{S^2} \mathbb{I}_{ij}^{\text{obj}} \sum_{c \in \text{classes}} \left[\hat{P}_i^j \log(P_i^j) + (1 - \hat{P}_i^j) \log(1 - P_i^j) \right], \quad (11)$$

where \hat{P}_i^j represents a true set of class probabilities of the j -th anchor box belonging to the i -th grid, and P_i^j represents a set of class probabilities of the bounding box generated by the anchor box.

Finally, with regard to the error of the further dehazing module, PSNR (Peak Signal to Noise Ratio) is applied to measure the difference between the obtained image and the original fog-free image, which can be achieved by minimizing the MSE (Mean Squared Error) at pixel level, and it is expressed as:

$$L_{\text{MSE}} = \frac{1}{N} \cdot \sum_{x=1}^N \sum_{i=1}^3 \|\hat{J}(x_i) - J(x_i)\|^2, \quad (12)$$

where $\hat{J}(x_i)$ is the original fog-free image, and $J(x_i)$ is the obtained image, i is the channel index, and N is the total number of pixels.

Therefore, the loss function combines the loss of detection module and the loss of further-dehazing module. Besides, in the detection module, the regularization L 2 is added into the loss function for the normalization. The loss function expressed

as:

$$\text{Loss} = (9) + (10) + (11) + (12) + \frac{\lambda}{2S^2B} \|w\|_2^2. \quad (13)$$

3.2.3 | Improved detection algorithm

In the JFRM detection module, the multi-scale features are introduced for generating the bounding boxes with different scales. However, all the bounding boxes share a unified coordinate system, and thus an object may have many overlapping bounding boxes. To solve this problem, a non-maximum suppression (NMS) algorithm is implemented to eliminate the overlapping bounding boxes. Firstly, the NMS algorithm sorts the bounding boxes in a descending order according to their confidences. Then, in the same class, the IOU values falling into the interval of the high-confidence bounding boxes, and the low-confidence bounding boxes will be calculated. If the IOU values are larger than a preset threshold, the bounding boxes with the low confidences are suppressed.

A cross-class suppressed method is given to suppress the low-confidence bounding boxes, no matter whether their classes are consistent with the high-confidence bounding boxes. The IOU threshold is set to 0.55. In the cross-class suppression algorithm, the dehazed images and the bounding boxes extracted from each image are traversed. For the each image, the IOU values between their any two bounding boxes are calculated, while their confidence scores are compared. Then, the label *flag* is initialised by 0. Suppose b_i and b_j are the obtained bounding boxes, when $\text{IOU}(b_i, b_j) \geq \text{thresh}$ and the $\text{Score}(b_i) \geq \text{Score}(b_j)$, the b_j and its confidence score are deleted; otherwise, b_i and its confidence score are deleted, while the label *flag* is set to 1. Then, in the other two bounding boxes, the IOU value is calculated, and the confidence scores are compared again. Finally, the optimal bounding boxes and their confidence scores are output.

For the detection module of JFRM network, the following problems probably occur during the training process: (i) network over-fitting problem; (ii) the accuracy of the test set is very low; (iii) model training is very slow. To this end, the following improvements are made in this paper:

- 1) In order to avoid the network over-fitting problem, the regularisation is added into the loss function for the normalisation;
- 2) The network configuration file of JFRM is modified, and the random parameter is set to 1, so that the image size can be modified randomly at each iteration, which can improve the accuracy of license plate detection;
- 3) In order to accelerate the training process, a momentum update mechanism is adopted;

3.2.4 | Bounding box position correction

With regard to the tilted license plates, the bounding boxes are inaccurate. This is because JFRM network has not considered



FIGURE 5 Position correction of a bounding box

the spatial transformations, and it only generates the rectangular bounding boxes. Hence, the WPOD-NET [28] is introduced to correct the positions of bounding boxes, as shown in Figure 5. The characteristics of Space Transformer Network (STN) are introduced into WPOD-NET, which can be used for detecting the non-rectangular regions.

For a tilted license plate, we first consider a rectangular box with a fixed size around the center of a pixel (m, n) . If the object probability of this pixel is larger than a preset detection threshold, and some of the regressed parameters are utilised to build an affine matrix that transforms a rectangular box into a tilted license plate region.

3.3 | Image super resolution and character recognition

3.3.1 | Super-resolution reconstruction

After the accurate license plate regions having been obtained, many license plate images may produce the phenomena of blur, noise pollution and so on, due to the environment interference, which leads to a poor performance of the character recognition. Then, the convolution-enhanced SRCNN is proposed to obtain the high-resolution images. With regard to a low-resolution image, the SRCNN first resizes the image to the target size by the bicubic interpolation, and then the non-linear mapping is accomplished through the convolution layers.

The structure of the convolution-enhanced SRCNN has six layers, in which the kernel size of the first convolution is increased. In the third to fifth layers, the convolution kernels with different sizes are executed continuously to further extract features, and the sixth layer reconstructs the mapped features and generates the high-resolution images by $5 \times 5 \times 1$ convolution kernels. The parameters of the convolution-enhanced SRCNN are shown in Table 1.

3.3.2 | Character recognition

After the high-resolution license plate images being obtained by the convolution-enhanced SRCNN, the characters of the license plate can be accurately recognised. In our work, the connected domain method and template matching method are exploited to realise the character segmentation, and then the ANN (Artificial Neural Network) is adopted to achieve the character recognition.

TABLE 1 The network parameters of the convolution-enhanced SRCNN

	Conv1	Conv2	Conv3	Conv4
Filter	128	64	32	32
Kernel size	11	9	7	5
Stride	1	1	1	1
Pad	5	4	3	2
	Conv5	Conv6		
Filter	32	1		
Kernel size	1	5		
Stride	1	1		
Pad	0	2		

Algorithm 2 Character segmentation

Input: $P = \{p_1, p_2, \dots, p_n\}$, a set of license plate images.

Output: a set of single character, R .

```

1:   if  $\text{len}(P) == 0$  then
2:     return  $P$ ;
3:   else
4:     for  $p_i$  in  $P$  do
5:        $\text{img} = \text{Preprocess}(p_i)$ ;
6:        $\text{cv.resize}(\text{img})$ ;
7:        $\text{cv.threshold}(\text{img}, 0.0, 255.0)$ ;
8:        $\text{Lc} = \text{findPossibleCharsPlate}(\text{img})$ ;
9:        $\text{Lm} = \text{findListOfMatchingClass}(\text{Lc})$ ;
10:       $\text{Rc} = \text{removeInnerLappingChars}(\text{Lm})$ ;
11:       $\text{result} = \text{recognizeCharsInPlate}(\text{Rc})$ ;
12:       $R \leftarrow \{ \text{result} \}$ ;
13:    end for
14:  end if

```

In the character recognition step, we focus on the region which contains the alphanumerics that we want to recognise discarding the other regions. Thus, we do some post-processes. Firstly, the gray-process and Otsu's binarisation are performed in addition to a Gaussian blur, to remove possible noisy artifacts. Secondly, the connected domain method is utilised to obtain a list of possible characters and the connected domain of alphanumerics. Then, these possible characters are matched with the license plate characters, and the overlapping characters inside the outline are removed. Finally, the characters are split to get a series of single character blocks, and we rely on the OCR network from other people's work to identify the license plate characters.

4 | SIMULATIONS

In this section, the training datasets of JFRM network and the convolution-enhanced SRCNN are first described, and then we



FIGURE 6 License plates in character segmentation process



FIGURE 7 Training samples of object detection

present the performance evaluation of the fog concentration dehazing method. The simulation program running environment is: Ubuntu16.04 system, Inter Core i7-8265U, NVIDIA GeForce GTX1080 based on Opencv-python, Darknet and SRCNN framework.

4.1 | Dataset

The real vehicle images are taken as datasets. These images includes various natural scenes (such as daytime, nighttime) with a total of 3000 images. By adding noise, rotating and other image processes, the number of the images can be increased to 12,000. Then the *xm/* file of the features of the license plates are obtained by manual annotations.

Besides, 300 images with the motion blur are taken as the dataset of the convolution-enhanced SRCNN.

4.2 | A fog concentration dehazing method

In this simulation, a fog concentration method is used to calculate the local atmospheric light value and obtain the trans-



FIGURE 8 Training samples of super-resolution

mittance, which can produce better restored images. In order to verify the performance of the fog concentration method, a comparison is made among the Soft-matting method [6], the Guided filtering method [29] and the fog concentration method, as shown in Figure 9. From Figure 9, the Soft-matting method exhibits a partial distortion phenomenon, and the colors of the restored images which obtained by Guided filtering method are darker than those of the fog concentration method.

Besides, the visible gradient method [30] is also utilised to evaluate the image dehazing results. Table 2 provides the evaluation indexes of the restored images, where e is the ratio of the visible edges, r is the ratio of the visible edges gradient, and σ is the percentage of the saturated black or white pixels. The values of e and r become larger, or the value of the σ becomes smaller, indicating that the restored images are clearer. As shown in Table 2, the quality of the restored images by the fog concentration method are higher than others.

4.3 | Joint optimisation of further-dehazing and detection

As aforementioned, we embed the further-dehazing into the detection network to form a joint optimisation. The JFRM network is trained end-to-end by minimising the loss function which is expressed as the sum of (9), (10), (11) and (12). With regard to the further dehazing module of JFRM, the weights are initialised by using some Gaussian random variables, and the ReLU neuron is utilised instead of the BReLU neuron. To help JFRM network achieve the detection of license plates, the *voc.data*, the number of the classes and the number of the filters are modified. The data of the ImageNet 1000 classes are used to train the first 20 layers of the detection module, which also obtains the initial weights of the YOLOv3. JFRM network takes around 10 training epochs to converge, and usually performs sufficiently well after 10 epochs. The loss, IOU, Class and Recall curves can reflect the preferable accuracy of the JFRM network. Figure 10 indicates that the curves converge to some preferable values as the number of the batches increases.

In order to avoid the over-fitting problem during the training process, the regularisation parameters L_1 , L_2 are added into the loss function. Besides, the Dropout layers are added after the executions of full connection layers. The regularisation parameter L_1 can produce a sparse matrix, which implies that many optimal weights should be set to 0 or close to 0. The regularisation parameter L_2 can produce some small and scattered weights, which indicates that the network can take advantage of more features during the training process, and this mechanism can enhance the generalisation ability of the network.

It can be seen from Table 4 that the accuracy of the test set is 87.11% after adding L_1 into the loss function. In contrast, with regard to another parameter L_2 and the Dropout layers, the accuracy of the test set can reach 91.27%, which indicates that the regularisation method of L_2 + Dropout layers is much better. In Figure 11, the accuracy of the object detection

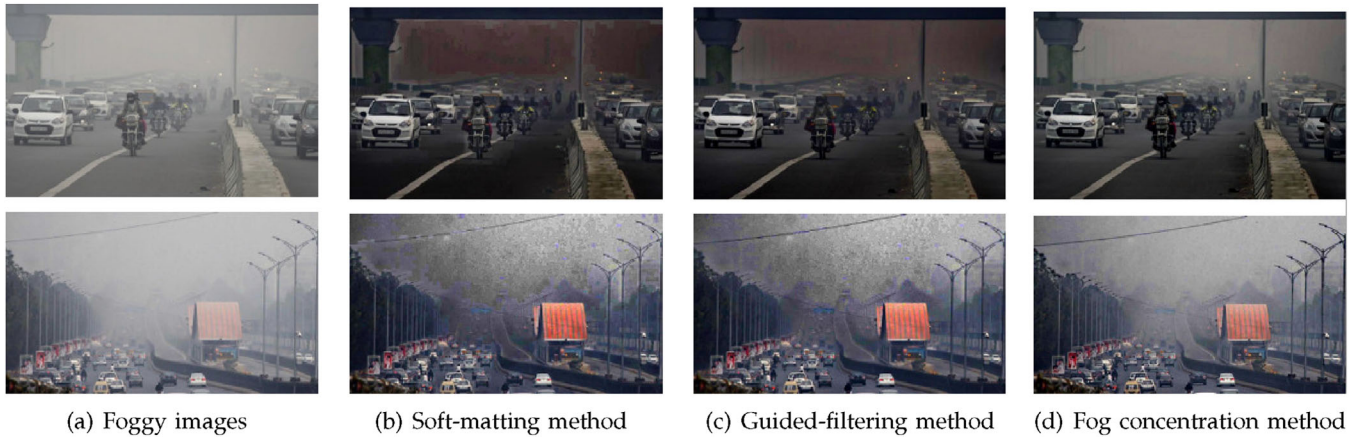


FIGURE 9 Output comparison of different dehazing methods

TABLE 2 Evaluation indexes of the image clarity

Images		Soft-matting	Guided filtering	Fog-concentration
Traffic ₁	e	0.20	0.22	0.24
	r	1.57	1.33	1.85
	σ	0.18	0	0
Traffic ₂	e	0.21	0.19	0.40
	r	1.56	2.35	3.27
	σ	0.09	0	0

TABLE 3 Simulation parameters

Parameters	Description	Values
N	Batch number	64
N_b	Divide the batch	16
r	Learning rate	0.001
p	Learning strategy	step
s_t	Iteration counter	45000
s_c	Scale of learning rate	.1
m	Momentum	0.9
d	Decay	0.0003
f_n	Filter num of last Conv layer	18

TABLE 4 Comparison of the regularisation parameters (number of iterations = 25,000)

Regularisation	Accuracy
L_1	87.11%
$L_2 + \text{Dropout}$	91.27%

becomes higher after our, and this is because the quality of the images are enhanced after dehazing, which makes the object be detected easier.

4.4 | Super-resolution and character recognition

In order to improve the accuracy of the character recognition, the convolution-enhanced SRCNN is utilized to obtain the high-resolution images. Compared with the traditional SRCNN which has three convolution layers, the number of the convolution layers are sequentially increased to four, five and six. The accuracies of the test set with different number of the convolution layers are given in Table 5.

Table 5 indicates that the network structures with more convolution layers can improve the character recognition accuracy. Especially, when the number of convolution layers is four or five, the accuracies of the character recognition are almost the same. The structure with six convolutional layers and seven con-

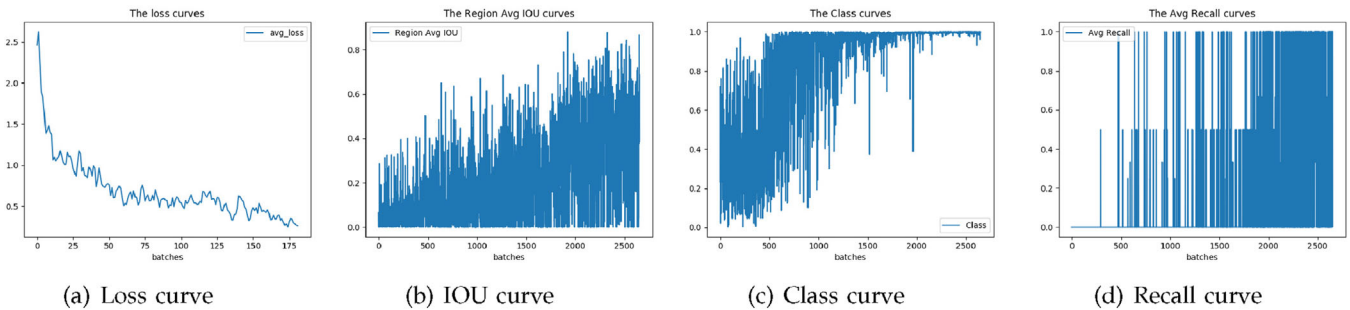


FIGURE 10 Index curves during training process



FIGURE 11 Comparison of the license plates detections

TABLE 5 Accuracy versus the number of the convolution layers (number of iterations=25,000)

Number of network layers	Accuracy
0	78.96%
3	89.21%
4	91.13%
5	91.46%
6	94.10%
7	94.12%

volitional layers achieves almost as high character recognition accuracy, but the complexity of seven convolutional layers is higher. Therefore, the convolution-enhanced SRCNN with six convolution layers is adopted in this paper.

From Table 6, it indicates that the convolution-enhanced SRCNN produces the preferable results, and Figure 12 indicates that all license plates can be correctly recognised in the fog-haze environments. Besides, with regard to a tilted license plate, the non-rectangular bounding box can be output. Meanwhile, the accuracy of the license plates recognition can be guaranteed under the license plates with stains as well.

TABLE 6 Comparison of the image super-resolution reconstruction results

Original images	SRCNN	Convolution-enhanced SRCNN

5 | CONCLUSIONS

This study explores the vehicle license plate recognition problem in the fog-haze environments. To reduce the fog-haze interference, a fog concentration dehazing method is first investigated, and in the proposed JFRM network, the further-dehazing detection is optimised jointly, which dramatically reduces the distortion of image reconstruction caused by the cumulative



FIGURE 12 The output of LPRFH

errors and accurately extracts the license plates from the preliminarily dehazed images. Finally, the image super-resolution is accomplished with a convolution-enhanced SRCNN, and hence the characters of license plates can be recognised successfully. Simulation results demonstrate our method can improve the accuracy of the vehicle license plate recognition in the fog-haze environments.

Future research focuses on the optimisation of transmission maps in the dehazed images. Besides, we will also study an adaptive setting of the IOU threshold in the non-maximum suppression algorithm to obtain the optimal bounding box.

ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China under Grant Nos. 61872191 and 61872193; Six Talents Peak Project of Jiangsu Province under Grant No. 2019-XYDXX-247.

REFERENCES

1. Khekare, G.S., Sakhare, A.V.: A smart city framework for intelligent traffic system using VANET. In: International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing, Kottayam, India (2013)
2. Gou, C., et al.: Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines. *IEEE Trans. Intell. Transp. Syst.* 17(4), 1096–1107 (2016)
3. Liu, W., et al.: Factors affecting relative humidity and its relationship with the long-term variation of fog-haze events in the Yangtze river delta. *Atmos. Environ.* 193, 242–250 (2018)
4. Li, X., et al.: DeepSaliency: Multi-task deep neural network model for salient object detection. *IEEE Trans. Image Process.* 25(8), 3919–3930 (2016)
5. Umehara, K., et al.: Super-resolution Convolutional Neural Network for the Improvement of the Image Quality of Magnified Images in Chest Radiographs. In: Proceeding of SPIE on Image Processing, Bellingham, WA (2017)
6. He, K., et al.: Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(12), 2341–2353 (2011)
7. McCartney, E.J.: Optics of Atmosphere-scattering by Molecules and Particles. *Phys. Today* 30(5), 76–77 (1977)
8. Tang, K., et al.: Investigating Haze-relevant Features in a Learning Framework for Image Database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, 2014, pp. 23–28
9. Zhang, Q., et al.: A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Process.* 24(11), 3522–3533 (2015)
10. Li, B., et al.: AOD-Net: All-in-One Dehazing Network. In: 2017 IEEE International Conference on Computer Vision, Venice, Italy 2017, pp. 4780–4788
11. Liu, Z., et al.: Single image dehazing with a generic model-agnostic convolutional neural network. *IEEE Signal Process. Lett.* 26(6), 833–837 (2019)
12. Liu, W., et al.: SSD: Single Shot Multibox Detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision - ECCV 2016. Lecture Notes in Computer Science, vol. 9905. Springer, Cham. (2015)
13. Iandola, F.N., et al.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and model size. arXiv preprint arXiv:1602.07360, (2017)
14. Krizhevsky, A., et al.: ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 25(3), 1097–1105 (2012)
15. Szegedy, C., et al.: Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 2015, pp. 1–9
16. Redmon, J., et al.: You only Look Once: Unified, Real-time Object Detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016, pp. 779–788
17. Dong, C., et al.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38(2), 295–307 (2016)
18. Kim, J., et al.: Accurate image super-resolution using very deep convolutional networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 2016, pp. 1646–1654
19. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, 2017, pp. 105–114.
20. Liu, Y., et al.: A new license plate character segmentation algorithm based on priori knowledge constraints. In: The 2nd International Conference on Intelligent Systems and Image Processing 2014 (ICISIP2014), Kitakyushu, Japan (2014)
21. Tan, J., et al.: A new handwritten character segmentation method based on nonlinear clustering. *Neurocomputing* 89, 213–219 (2012)
22. McCartney, E.J.: Real-time scene text localization and recognition. 2012 IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, 2012, pp. 3538–3545
23. Elihos, A., et al.: Deep learning based segmentation free license plate recognition using roadway surveillance camera images. arXiv preprint arXiv:1912.02441, (2019)
24. Selmi, Z., et al.: DELP-DAR system for license plate detection and recognition. *Pattern Recognit. Lett.* 129, 213–223 (2020)
25. Silva, S.M., Jung, C.R.: Real-time license plate detection and recognition using deep convolutional neural networks. *J. Visual Commun. Image Represent.* (2020), <https://doi.org/10.1016/j.jvcir.2020.102773>
26. Wu, S., et al.: PixTextGAN: structure aware text image synthesis for license plate recognition. *IET Image Proc.* 13(14), 2744–2752 (2019)
27. Zhang, L., et al.: A robust attentional framework for license plate recognition in the wild. *IEEE Trans. Intell. Transp. Syst.* (2020), <https://doi.org/10.1109/TITS.2020.3000072>
28. Silva, S.M., Jung, C.R.: License plate detection and recognition in unconstrained scenarios. In: European Conference on Computer Vision (ECCV 2018), Munich, Germany (2018)
29. He, K., et al.: Guided image filtering. In: Proceedings of the 11th European Conference on Computer Vision, Berlin, Heidelberg (2010)
30. Hautière, N., et al.: Blind contrast enhancement assessment by gradient rationing at visible edges. *Image Analysis and Stereology Journal* 27, 87–95 (2008)

How to cite this article: Jin X, Tang R, Liu L, Wu J. Vehicle license plate recognition for fog-haze environments. *IET Image Process.* 2021;15:1273–1284. <https://doi.org/10.1049/ipr.2.12103>