ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

Improve the Attention Mechanism to Realize the Clarity of Fuzzy License Plate Images

Xuan Lv[1, a], Biao Wang[2, b] and Jiayun Zou[2, c]

[1]college of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibing, China

[2]Chengdu Shiguantianxia Technology Co., Ltd, Chengdu, China

[a]1546273017@qq.com, [b]274101801@qq.com, [c]zjysnow@outlook.com

Abstract

In order to avoid the impact of low-resolution blurred images on the recognition of license plates, a spatial and channel dual attention network (RSCAN) based on residual network is proposed to recover blurred license plate images. RSCAN adds a spatial attention mechanism to the RCAN (residual channel attention) network, and introduces a new channel attention mechanism to train and test the network through a self-made license plate data set. The RSCAN network is compared with the conventional processed pictures, SRCNN (super-resolution) network, and RCAN network. The test license plate pictures are used for comparison experiments, and the model evaluation is carried out through the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), respectively, 30.312/0.889. The experimental results show that the license plate images processed by RSCAN network achieve the best results.

Keywords

Residual network; Attention mechanism; License plate data set; Evaluation index.

1. INTRODUCTION

The license plate images collected by the public security organs in the process of solving the case cannot be accurately identified due to motion blur, out-of-focus blur, light changes and other reasons, which seriously affects the progress of the case. Therefore, clearing the image has important practical significance.

At present, super-resolution algorithms are mainly divided into three categories: reconstruction-based algorithms, interpolation-based algorithms, and learning-based algorithms [1]. After the advent of deep convolutional neural networks, neural networks began to be applied to image super-resolution processing on a large scale. In 2014, Dong et al. proposed the SRCNN [2] network, which uses a convolutional neural network to super-resolve pictures. Each convolution kernel parameter is updated through backpropagation, rather than artificially set. At the same time, the deep convolutional network will perform multi layer convolution to ensure a large enough receptive field. There have been many network structure algorithms such as VDSR [3], EDSR [4], DRCN [5] and so on.

For the specific field of license plate, super-resolution network is used to realize the clarification of fuzzy license plate. The RCAN [6] network introduces the attention mechanism into the super-resolution field for the first time. The attention mechanism is divided into soft attention and hard attention. Strong attention is not differentiable and is accomplished through reinforcement learning; while soft attention is a continuous distribution problem between [0, 1], it is more suitable to use backpropagation to learn the weight of attention. In soft attention,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

the attention model can be constructed according to the two directions of pixel and channel, namely spatial attention and channel attention. Based on the improvement of the RCAN network, the RSCAN network is proposed to extract and merge the characteristics of all layers in the channel and space, and use the combined module of channel attention and spatial attention to capture the important features of different channels and spatial positions, and obtain the weight parameters of the corresponding positions. The most important information in the license plate is numbers and letters. The application of the attention mechanism [7] makes the network pay more attention to the important information of the license plate. The clarification of the license plate improves the visual effect of the license plate image, strengthens the features required by the image, and suppresses the features that the user does not need. The application of the attention mechanism information can achieve a better recovery effect, and at the same time, the low-frequency information can be more effectively transferred to the image through the jump connection The reconstructed module supervises the output of each level module, speeds up the network convergence, and further improves the quality of image restoration.

Main tasks as follows:

(1) A new module (SCM) combining channel attention mechanism and spatial attention mechanism (SCM) is proposed to improve model efficiency and computational effects.

(2) Based on the CCPD [8] data set, a new license plate data set is made, and ablation experiments are done on the data set. Verify the effectiveness of the proposed algorithm.

(3) Use residual blocks to form a global feature fusion, strengthen the transmission of information at each layer, and obtain better images.

2. RESIDUAL SPACE AND CHANNEL ATTENTION NETWORK (RSCAN)

2.1. Basic Network Structure

Based on the RCAN network, a spatial and channel dual attention network (RSCAN) based on the residual network [9] is proposed. The network is composed of a shallow feature extraction module, a spatial and channel attention fusion module, and an image reconstruction module. Use I_{LR} to represent the input fuzzy license plate image, I_{CI} to represent the output clear image, and use a convolutional layer to extract the shallow features of the input image I_{LR}, as in Equation 1:

$$F_{SF} = H_{SF}(I_{LR}) \tag{1}$$

In the above formula, $H_{SF}(\cdot)$ represents the use of single-layer convolution to achieve shallow feature extraction. Then take the shallowly extracted features as the input of the space and channel attention mechanism modules, and get the mapped high-dimensional features through the space and channel attention mechanism modules, as shown in Equation 2:

$$F_{GF} = H_{RSCAF}(F_{SF}) \tag{2}$$

In the above formula, $H_{RSCAF}(\cdot)$ represents the mapping relationship of the feature fusion module based on the spatial and channel attention mechanism. The proposed feature fusion module of the spatial and channel attention mechanism makes the network more effective use of the extracted useful features, suppresses useless features, enables the network to effectively deepen the network without increasing the calculation, and increase the receptive field of the convolution kernel. The fused feature of the attention mechanism is used as the input of

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

upsampling, and the upsampling operation is performed through the sub-pixel layer [10] to obtain the enlarged feature. As formula 3:

$$F_{US} = H_{US}(F_{GF}) \tag{3}$$

In the above formula, $H_{US}(\cdot)$ represents the up-sampling operation. At present, the most commonly used up-sampling methods in the field of super-resolution reconstruction are interpolation operation [11], deconvolution operation [12] and sub-pixel layer operation. The sub-pixel layer rearranges multiple feature maps into a larger image to obtain a larger receptive field, so that the network can achieve better results after reconstruction. Finally, a simple convolution is used to transform the input features into the corresponding three-channel output image, as shown in Equation 4:

$$I_{CI} = H_{CI}(F_{US}) = H_{RSCAN}(I_{LR}) \tag{4}$$

In the above formula, $H_{US}(\cdot)$ represents the mapping function reconstructed into a color picture, and H_{RSCAN} represents the mapping function from I_{LR} to I_{CI}.

2.2. Global Feature Fusion Module of Spatial and Channel Attention Mechanism

In the network structure design, the feature fusion module based on the space and channel attention mechanism is the main structure to realize the feature mapping. This module includes A spatial and channel attention groups (Spatial and channel attention groups, SCAG) and jump structure

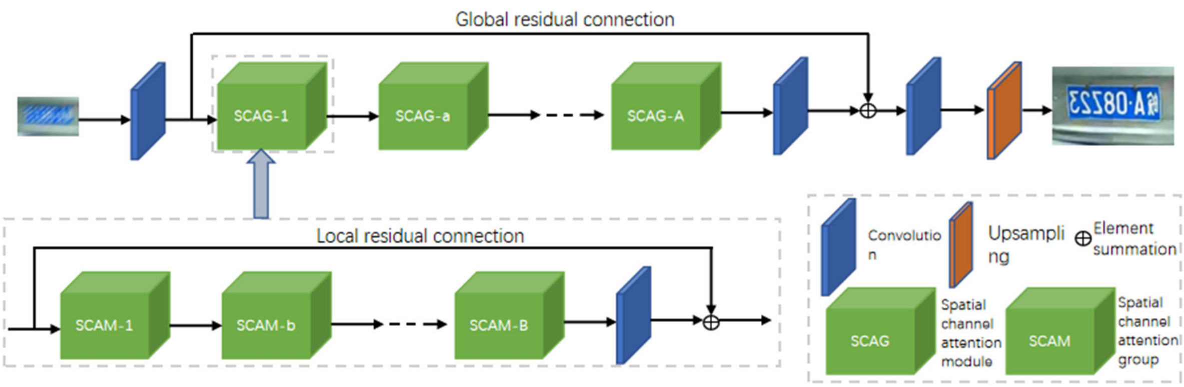


Figure 1. Network structure diagram of residual space and channel attention mechanism

Each SCAG contains B space and channel attention modules (SCAB) with short-hop connections. Such a structure can obtain a deeper algorithm network, as shown in Figure 1.

The SCAG in group A is expressed as formula 5:

$$F_a = H_a(F_{a-1}) = H_a\big(H_{a-1}(\cdots H_1(F_0)\cdots)\big) \tag{5}$$

In the above formula, H_a represents the A-th SCAG,F_{a-1} and F_a are the input and output of the A-th SCAG respectively. Simply stacking multiple SCAGs cannot achieve better performance. To solve this problem, a global residual connection (LRC) is introduced to stabilize the deeper network. The global residual connection is as in formula 6:

$$F_{GF} = F_0 + W_{LR}F_A = F_0 + W_{LR}H_a\big(H_{a-1}(\cdots H_1(F_0)\cdots)\big) \tag{6}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

LR can not only simplify the information flow between SCAGs, but also learn residual information at a rough level. The LR input and features contain a wealth of information, and the goal of the network is to recover more useful information. Abundant low-frequency information can be transmitted to the back through the residual connection.

In addition, B SCABs are stacked in each SCAG, and the b-th SCAB in the a-th SCAG can be expressed as Equation 7:

$$F_{a,b} = H_{a,b}(F_{a,b-1}) = H_{a,b}\left(H_{a,b-1}(\cdots H_{a,1}(F_{a-1})\cdots)\right) \tag{7}$$

In the above formula, $F_{a,b-1}$ and $F_{a,b}$ represent the input and output of the b-th SCAB in the a-th SCAG. Similar to the SCAG module, this B SCAB also has a local residual connection (SRC), as shown in formula 8:

$$F_a = F_{a-1} + W_a + H_{a,B} = F_{a-1} + W_a H_{a,B}\left(H_{a,B-1}(\cdots H_{a,1}(F_{a-1})\cdots)\right) \tag{8}$$

W_a is the weight of a conv at the end of the a-th SCAB module. The existence of LRC and SRC enables richer low-frequency information to be transmitted to a deeper layer during the training process.

2.3. Spatial and Channel Attention Mechanism

The proposal of the attention mechanism enables the network to focus more on useful information features, inhibits useless information, and can greatly improve the utilization of computing resources. Get better quality pictures in the license plate data set.

Spatial attention can be understood as letting the neural network look where, through the attention mechanism, the spatial information in the original picture is exchanged to another space while retaining key information.

Spatial transformer [13] is the realization of the attention mechanism, because training the Spatial transformer can find the area of interest in the information, so that the important information in the picture can be extracted, as shown in Figure 2. Use a two-dimensional convolution to compress the number of channels to 1. In order to ensure the full enhancement of spatial information, the spatial weight information is implemented in this paper through convolution.

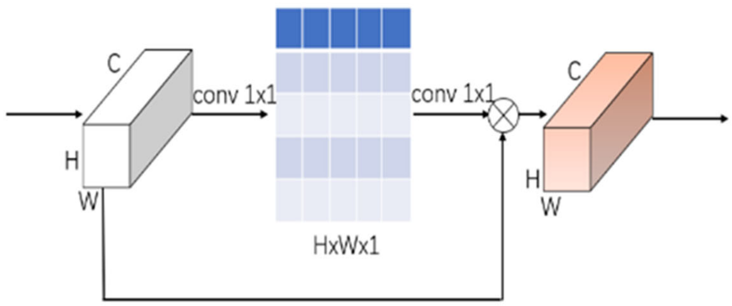


Figure 2. Network structure diagram of spatial attention mechanism

In the initial picture, after different convolution kernels, each channel will generate a new signal, which is equivalent to decomposing each signal into components on the kernel function, and adding a weight to each channel (the weight represents the difference between the channel

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

and the key information). Relevance), the greater the weight, the more attention to the information of the channel. The attention mechanism SE-NET [14] used by RCAN can learn the channel attention of each convolution block, and bring obvious performance improvement to various deep CNN architectures. Specifically, given input characteristics, the SE module first Each channel independently uses global average pooling, and then uses two fully connected (FC) layers and a nonlinear Sigmoid function to generate channel weights. There are flaws. The two FC layers are designed to capture non-linear cross-channel interaction, which involves reducing the dimensionality to control the complexity of the model, but the dimensionality reduction will bring side effects to channel attention prediction and capture the dependencies between all channels. The relationship is not efficient and unnecessary. Therefore, an effective channel attention (ECA-NET [15]) module for deep CNN is used, which avoids dimensionality reduction and effectively captures cross-channel interaction. Channel-by-channel global average pooling is performed without reducing the dimensionality. ECA-NET captures local cross-channel interactions by considering each channel and its k neighbors, as shown in Figure 3.

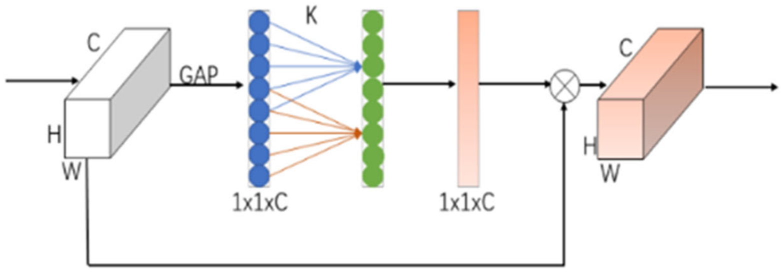


Figure 3. Structure diagram of channel attention mechanism

For the ECA module, the weight of the channel, this module only considers the information interaction between $y^{\{i\}}$ and his k neighbors, the calculation formula is as shown in Equation 9:

$$\omega_i = \sigma\left(\sum_{j=1}^{k} \omega_i^j y_i^j\right), y_i^j \in z_i^k \tag{9}$$

z_i^k represents the set of k adjacent channels of y. In order to further improve the performance, it is also possible to let all the channels share the weight information, that is, Equation 10:

$$\omega_i = \sigma\left(\sum_{j=1}^{k} \omega_i^j y_i^j\right), y_i^j \in z_i^j \tag{10}$$

Based on the above analysis, a new method is proposed, which can realize the information exchange between channels through one-dimensional convolution with a convolution kernel size of K, as shown in Equation 11:

$$\omega = \sigma\big(C1D_K(\mathrm{y})\big) \tag{11}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

Here, C1D stands for one-dimensional convolution, which only involves K parameter information. Therefore, this method of capturing cross-channel information interaction ensures performance results and model efficiency.

Since the ECA module aims to appropriately capture the local cross-track information interaction, it is necessary to determine the approximate range of the channel interaction information (that is, the convolution kernel size k of the 1D convolution). Although it is possible to manually optimize and set the optimal range of information interaction for convolution blocks with different numbers of channels in various CNN architectures, manual cross-validation adjustments will cost a lot of computing resources. Moreover, grouped convolution has been successfully used to improve the CNN architecture. In the case of a fixed number of groups, the high-dimensional (low-dimensional) channel is proportional to the long-distance (short-distance) convolution. Similarly, the coverage of cross-channel information interaction (that is, the kernel size k of one-dimensional convolution) is also proportional to the channel dimension C.

This method guarantees the model efficiency and calculation effect. Given the channel dimension C, there is a mapping relationship between k and c, as shown in Equation 12:

$$C = \emptyset(\mathrm{k}) = 2^{(r*k-b)} \tag{12}$$

In the above formula, the power of 2 is that the design considering the number of channels is generally a power of 2, so that the module K can be calculated more conveniently, and r and b take 2 and 1, respectively.

Then the size of the adaptive convolution kernel (k) can be calculated according to the following formula, as shown in formula 13:

$$k = \varphi(\mathrm{c}) = \left|\frac{log_2(c)}{\gamma} + \frac{b}{\gamma}\right|_{odd} \tag{13}$$

Finally, the proposed space and channel attention are combined together, and a network structure with space and channel attention mechanisms is constructed through the combination. The structure is shown in Figure 4.

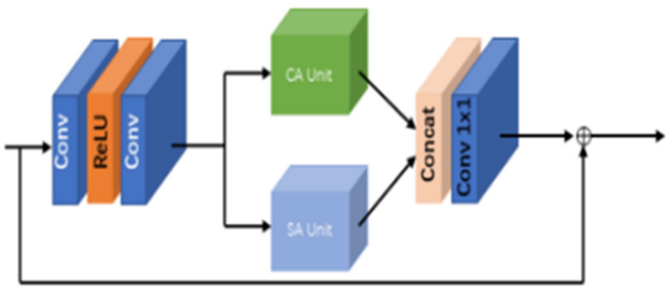


Figure 4. Structure diagram of space and channel attention module (SCAB)

3. TESTS

3.1. Data Set

Based on the proposed RSCAN network, the fuzzy license plate clarification processing experiment is carried out. The pictures used in the experiment are all cropped from the CCPD data set. For the cropped license plate images, they are flipped up and down as high-resolution

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

images. The corresponding low-resolution images are produced by adding Gaussian noise, fogging, motion blur, and media to the image. Blur, etc. to simulate blurry pictures taken in real life.

3.2. Performance Evaluation Index

Loss function: Loss functions include L1, L2[16], MSE [17], GAN [18] loss and texture structure perceptual loss [19]. The main purpose of clarification is to restore it to the naked eye. In order to ensure effectiveness, Choose L1 loss, that is, Equation14:

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left\|H_{RSCAN}\left(I_{LR}^{i}\right) - I_{HR}^{i}\right\| \tag{14}$$

The above formula, θ represents the parameter set of the network in this paper, $H_{RSCAN}\left(I_{LR}^{i}\right)$ is the picture after the fuzzy license plate picture is restored by the network, and I_{HR}^{i} is the high-definition license plate picture.

Model evaluation cannot be based on visual effects alone. It is necessary to use peak signal-to-noise ratio (PSNR) and structural similarity SSIM [16] to evaluate model effect. The formula of peak signal-to-noise ratio (PSNR) is shown in Equation 15:

$$P\text{SNR} = 10 * log_{10}\left(\frac{MAX^{2}}{MSE}\right) \tag{15}$$

In the above formula, MAX represents the maximum possible pixel value of each pixel, which is 255, and MSE represents the mean square error of the corresponding pixel between the image and the image. The larger the PSNR, the better the effect of image generation.

$$M\text{SE} = \frac{1}{M * N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{M}\left(f_{ij} - f_{ij}^{'}\right)^{2}$$

The SSIM formula for structural similarity is shown in formula 16:

$$\begin{gathered} ssim(\mathrm{x},\mathrm{y}) = l(x,y) * c(x,y) * s(x,y) \\ l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_2} \\ c(x,y) = \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \\ s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \end{gathered} \tag{16}$$

In the above formula, $l(x,y)$ is the height comparison, $c(x,y)$ is the contrast comparison, and $s(x,y)$ is the structure comparison. μ_x and μ_y represent the average values of x and y, respectively, and σ_x and σ_y represent the standard deviations of x and y, respectively. σ_{xy} represents the covariance of x and y. And c_1, c_2, c_3 are all constants to avoid system errors caused by the denominator being 0.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

3.3. Hyperparameter Settings of the Network

In order to avoid under-fitting the network during the training process, the training set has a total of 30,000 high-resolution and low-resolution images, and the test set has a total of 4,000 images. The batch-size is set to 16, that is, 16 low-resolution images of 48x48 are drawn during each training. The resolution image set corresponds to the label image for training. The clarification of blurred license plates only needs to restore the blurred license plates to be recognized by the naked eye, so there is no need for too deep neural networks. After experiments, A=10 and B=20 are set in the network structure. The RSCAN network includes 10 SCAG modules, and each SCAG module includes 20 SCAB modules. Except for the 1x1 convolution kernel used in the bottleneck layer of channel compression and amplification and feature fusion in the channel attention mechanism, the other convolution kernels are all 3x3 in size, and the convolution and deconvolution kernels used by the spatial attention mechanism The size is also 3x3, its step size is 3, and epoch=300. In the training process, the network training adopts the strategy of complementing 0 to achieve the same size of the intermediate feature map. For channel attention, the size of the convolution kernel (k)=3.

The network model uses the ADAM optimizer, β1=0.9, β2=0.999, $\varepsilon = 10^{-8}$, the initial learning rate is set to 10^{-4}, after every $2x10^5$ iterations, the network learning rate drops to half of the original, every 1000 After the second iteration, a test is performed on the license plate data set to facilitate the viewing of the network training situation, so that the network hyperparameters can be adjusted according to the training situation. Finally, in order to verify the generalization ability of the network and the effect of license plate recovery, the real fuzzy license plate pictures on the Internet are used to test. The hardware information is shown in Table 1.

Table 1. Software and hardware platform of the experiment

software and hardware	Model/Version
CPU	Intel I9-9700K
GPU	RTX TITAN XP (Video memory 12G)
RAM	16G
Pytorch	0.4.0

3.4. Ablation Experiment

In order to show more clearly whether the innovative measures are effective in improving the quality of image restoration, ablation comparison experiments are carried out on the proposed improvements.

First of all, in order to verify the effectiveness of the improved channel attention mechanism, the following comparative experiments are carried out with only the channel attention mechanism: one is that the network still uses the SE-NET module, the other is that the network uses ECA-NET, see the specific ablation experiment results Table 2:

Table 2. Comparison of the attention of different channels on the license plate data set (x2)

Channel attention	SE-NET	√	×
	ECA-NET	×	√
PSNR value of the license plate data set(dB)		30.132	30.206

In the table, √ means that the network uses this module, and × means that this module is not selected. The experiment compares the effect of the network model in the clarification of fuzzy license plates when different modules are used. Using the evaluation index PSNR as a

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

comparison, it can be found that the use of ECA- Compared with SE-NET's PSNR index, NET has increased by 0.073 dB. It shows that the channel attention mechanism used has a better effect.

Table 3. The influence of spatial and channel attention on the license plate data set (x2)

Channel attention mechanism	√	√
Spatial attention mechanism	×	√
PSNR value of the license plate data set (dB)	30.206	30.312

It can be seen that the algorithm using the channel and space attention mechanism has an improvement of 0.106 dB, indicating that the space and channel attention mechanism used has a better effect on the recovery of blurred license plate images.

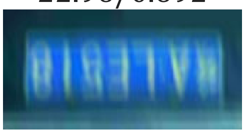
3.5. Experimental Results and Discussion

The test mainly compares the RCAN model and the RSCAN model proposed in this article. First of all, it can be seen by the naked eye that the application of neural network to the clarification of fuzzy license plates has achieved good results. At the same time, the effect of the picture details restored by the RSCAN model is better, as shown in Table 4. Use RSCAN to compare with other classic super-resolution (x2) networks to test the superiority of RSCAN network, as shown in Table 5. Table 6 uses real pictures on the network to test the actual effect of the network.

Table 4. Comparison of RCAN and RSCAN test results

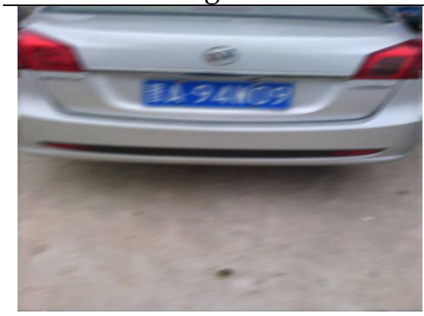
LR	RCAN	RSCAN	HR
			
	22.76/0.823	23.17/0.835	PSNR/SSIM
			
	29.75/0.892	30.06/0.897	PSNR/SSIM

Table 5. Comparison of different super-resolution methods (x2)

LR	BICUBIC	SRCNN	RSCAN	HR
				
	21.78/0.614	22.96/0.692	30.278/0.877	PSNR/SSIM
				
	20.56/0.589	21.65/0.615	24.56/0.845	PSNR/SSIM

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

Table 6. Real pictures of RSCAN test network

Original	License Plate Picture	Restore picture
		
		

4. CONCLUSION

Aiming at some ambiguity problems encountered in license plate recognition, based on the existing super-resolution model using the attention mechanism, it is proposed to use the ECA-NET structure and add the spatial attention mechanism at the same time. The fusion of the two attention mechanisms can effectively obtain the weight values of different features, so that the network can accurately allocate computing resources according to the weights, and effectively improve the quality of super-resolution reconstruction while introducing very few parameters. In order to obtain better results of license plate recovery, using a self-made license plate data set can produce a better recovery effect for the specific situation of license plate. Experiments have proved that the proposed algorithm for license plate reconstruction has improved the evaluation index results, and at the same time, it has also achieved good results in the restoration of some real pictures.

REFERENCES

[1] Wang Z, Chen J, Hoi S C H. Deep learning for image sup-er-resolution: A survey[J]. IEEE Transactions on Pattern Ana-lysis and Machine Intelligence, 2020.

[2] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(2): 295-307

[3] J Kim, J.K Lee, K.M Lee. Accurate image super-resolution using very deep convolutional networks[C]// In CVPR, America: IEEE Xplore, 2016.1:1646-1654.

[4] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 136-144.

[5] J Kim, J.K Lee, K.M Lee. Deeply-recursive convolutional network for image super-resolution[C] //In CVPR, America: IEEE Xplore, 2016.1: 1637-1645.

[6] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 286-301.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0009

[7] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.

[8] Xu Z, Yang W, Meng A, et al. Towards end-to-end license plate detection and recognition: A large dataset and baseline[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 255-271.

[9] He K, Zhang X, Ren S, et al. Deep residual learning for im-age recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[10] W Shi, J Caballero, F Huszar, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//In CVPR, America: IEEE Xplore, 2016.1:1874-1883

[11] S Fadnavis. Image Interpolation Techniques in Digital Image Processing: An Overview[J]. Shreyas Fadnavis Int. Journal of Engineering Research and Applications.2014.10,4(10):70-73.

[12] M Zeiler, D Krishnan, G Taylor, et al . "Deconvolutional net-works"[C]//In CVPR, America: IEEE Xplore, June 2010: 2528-2535. oncept for ubiquitous transpor

[13] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[C]//Advances in neural information processing systems. 2015: 2017-2025.

[14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision an-d pattern recognition. 2018: 7132-7141.

[15] Wang Q, Wu B, Zhu P, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11534-11542.

[16] Girshick R. Fast rcnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[17] Willmott C J, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance[J]. Climate research, 2005, 30(1): 79-82.

[18] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681-4690.

[19] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for realtime style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.

[20] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.