2018 3rd IEEE International Conference on Image, Vision and Computing

License Plate Image Super-Resolution Based on Convolutional Neural Network

Yang Yang, Ping Bi, Ying Liu
Center for Image and Information Processing, Xi'an University of Posts and Telecommunications, Xi'an, 710121, China
Key Laboratory of Electronic Information Application Technology for Scene Investigation, Ministry of Public Security, Xi'an, 710121, China
International Joint Research Center for Wireless Communication and Information Processing Technology of Shaanxi Province, Xi'an, 710121, China
e-mail: biping@xupt.edu.cn

***Abstract*—To improve the visual quality of the low-resolution license plate image in the video surveillance system, this paper proposes a new method of super-resolution, namely multi-scale super-resolution convolutional neural network (MSRCNN), it was inspired by Inception architecture of GoogLeNet. The proposed method uses different sizes of filters for parallel convolution to obtain different features of the low-resolution license plate image, then the features can be fused by the layer of concatenation. Finally, the high-resolution images can be reconstructed through non-linear mapping. Experimental results prove that the proposed method can improve the quality of low-resolution license plate image obviously.**

Keywords-super-resolution; convolutional neural network; license plate image

I. INTRODUCTION

With the rapid increase in the number of motor vehicles, the license plate plays an important role in the field of public security. However, in the real-world video surveillance, due to the camera's location, the light condition and other reasons, the license plate images obtained are of poor visual quality with low resolution mostly. It affects the applications of the license plate image for public security. Hence, improving the license plate image's resolution is important in practical application.

In order to obtain high-resolution license plate images, the most direct way is to improve the hardware device, but the cost is high. Therefore, super-resolution (SR) technique has become the main way. SR uses signal-processing techniques to reconstruct a high-resolution (HR) image from a low-resolution (LR) image [1]. Usually, SR could be divided into three categories: interpolation-based method [2] [3], reconstruction-based method [4] [5] and learning-based method [6] [7]. Interpolation-based method is easy and fast, but the result is too smooth and ringing artifact seriously. Reconstruction-based method uses mathematical models to reconstruct high resolution images, but the computation is complicated. Learning-based method makes full use of the inherent priori knowledge of images. It can well preserve details of the image and it is suitable for dealing with special images, such as spectral images [8], infrared images [9] and medical images [10]. But learning-based method requires a large amount of training samples, and its time cost on training is high.

In this paper, we propose a new super-resolution method for license plate image, namely multi-scale super-resolution convolutional neural network (MSRCNN). The method consists of three convolutional layers, the first layer extracts and fuses features of LR image, the second layer maps features from LR image to HR image, and the third layer reconstructs HR image. Experiments show that the proposed model is effective to improve the quality of super-resolution image, especially for reconstruction of edge information. The rest of the paper is organized as following. In Section II, the proposed SR method is explained in details. Section III provides our experimental results and Section IV concludes this paper.

II. PROPOSED METHOD

A. GoogLeNet Convolutional Network

Now the convolution neural network is getting bigger and bigger, and its performance is getting better and better. In generally, the performance of neural networks can be improved by adding layers and the number of filters per layer.

GoogLeNet [11] is the champion of ImageNet 2014. Its architecture is considered as network in network. The main module is called Inception architecture, which is a tiny network. In this network, the image is simultaneously convolved with different size filters to get different feature maps. Then these feature maps connect together to form a new fusion feature map. This architecture reduces the influence of filter size on neural network performance and can extract more features. Therefore GoogLeNet have an extraordinary good performance. Inspired by Inception architecture, the proposed method uses different filter size on first layer to extract more features and improve the image reconstruction performance.

B. Multi-scale Super-resolution Convolutional Neural Network

Super resolution convolutional neural network (SRCNN) [12] [13] is end-to-end mapping between LR images and corresponding HR images. SRCNN has experimentally verified that adjusting the number of network's layer, the size of filter and the number of filter in each layer can effect on reconstruction results. But which size of filter should be used for some special kind of image (such as the license plate image) has no fixed rule, so in this paper we will use three different filter sizes to deal with the license plate image simultaneously, then connecting different feature maps together to form a new fusion feature map, which could improve the quality of image reconstruction, as shown in Fig. 1.

In the next sections, X represents high-resolution image, Y represents low-resolution image interpolation by Bi-cubic, X and Y are the same size. $F(Y)$ is the reconstruction of image Y, $F(Y)$ is most similar to X.

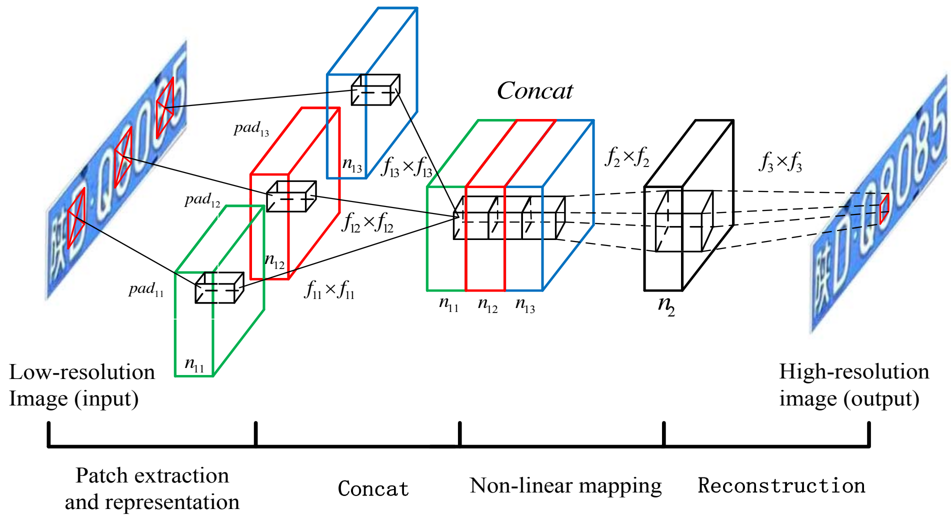

Figure 1. Multi-scale input network structure

1) Feature extraction and representation

In first layer, image is simultaneously convolved by three different filters f_{11}, f_{12}, and f_{13} (corresponding pad_{11}, pad_{12}, pad_{13}, respectively). These three feature maps connect together to form a new fusion feature map, which is the input of the next layer. The feature extraction and representation are expressed as:

$$F_1(Y) = \max(0, W_{11} * Y + B_{11}) \oplus \max(0, W_{12} * Y + B_{12}) \oplus \max(0, W_{13} * Y + B_{13}) \quad (1)$$

where, $f_{11} \times f_{11}$, $f_{12} \times f_{12}$, $f_{13} \times f_{13}$ are the filter size, W_{11}, W_{12}, W_{13}, B_{11}, B_{12}, B_{13}, n_{11}, n_{12}, n_{13} are filter weights, filter biases, and the number of filter, respectively. Here, B_{11}, B_{12}, B_{13} are n_1-dimensional vector, suppose $n_{11} = n_{12} = n_{13} = n_1$, '$*$' denotes convolution operation, '$\oplus$'denotes concatenation.

2) Non-linear mapping

The non-linear mapping means that $3n_1$ -dimensional image feature are mapped to n_2-dimensional feature. Non-linear mapping is expressed as:

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2) \quad (2)$$

where, $f_2 \times f_2$ is the filter size, n_2 is the number of filter. W_2 and B_2 are the filter weight and the filter bias, respectively. n_2 -dimensional vector represents high-resolution features of image and it is used for the final reconstruction.

3) Image reconstruction

The final HR image is obtained from the reconstruction filter convolution n_2 -dimensional high-resolution features. Image reconstruction is expressed as:

$$F(Y) = W_3 * F_2(Y) + B_3 \quad (3)$$

where, W_3 is the filter weight corresponding to $f_3 \times f_3$ filter, B_3 is a scalar.

C. *Loss function*

SRCNN [12] [13] uses Mean Squared Error (MSE) as a loss function for testing the accuracy of the training model and updating the parameters of the model. In this paper we still use MSE as the loss function of the neural network , the expression is:

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \left\| F(Y_i; \Theta) - X_i \right\|^2 \quad (4)$$

where n is the number of training images patches. Y_i is the *ith* LR image patch which is interpolated using bicubic interpolation. X_i is the HR image patch corresponding to LR image patch. $F(Y_i; \Theta)$ is the reconstructed HR image patch by the network model Θ.

III. Experiments

A. *Experimental Data and SRCNN Model Results*

In this paper, datasets are collected by video surveillance systems. Dataset1 includes 96 license plate images, training set is 91, testing set is 5 (Set5), as shown in Fig. 2. Dataset2 includes 986 license plate images, training set is 976, testing set is 10 (Set10), five of them are the same as Set5. The upscale factor is 3 in this paper.

Figure 2. Testing set of Dataset1 (Set5)

This experiment uses Dataset1 and Dataset2 to train SRCNN model, respectively. Set5 is tested by SRCNN model [12], SRCNN Dataset1 model, SRCNN Dataset2 model, as shown in Fig. 3. Notice that SRCNN model [12] was trained by author using ImageNet, so only the final reconstruction results can be obtained, but in order to compare with our experiment results, we draw a line in Fig. 3. Experiment shows that the more training data, the better property. Besides, for license plate image reconstruction, license plate image training sets is better than natural image sets.

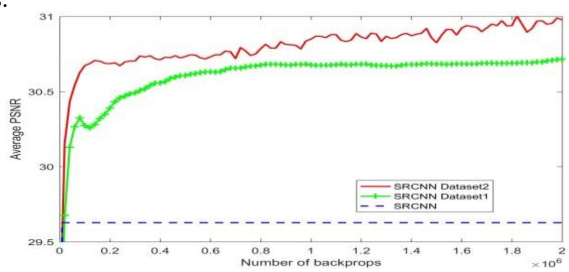

Figure 3. SRCNN Dataset2 vs SRCNN Dataset1 vs SRCNN [12]

B. MSRCNN

In order to improve the reconstruction quality of license plate image, three different filter sizes are designed in first layer, MSRCNN1 ($f_{11}=5$, $f_{12}=7$, $f_{13}=9$, $f_2=3$, $f_3=5$, $n_{11}=n_{12}=n_{13}=32$, $n_2=32$, $n_3=1$), MSRCNN2($f_{11}=7$, $f_{12}=9$, $f_{13}=11$, $f_2=3$, $f_3=5$, $n_{11}=n_{12}=n_{13}=32$, $n_2=32$, $n_3=1$), and MSRCNN3($f_{11}=9$, $f_{12}=11$, $f_{13}=13$, $f_2=3$, $f_3=5$, $n_{11}=n_{12}=n_{13}=32$, $n_2=32$, $n_3=1$). The testing results on Dataset2 are shown in Fig. 4.

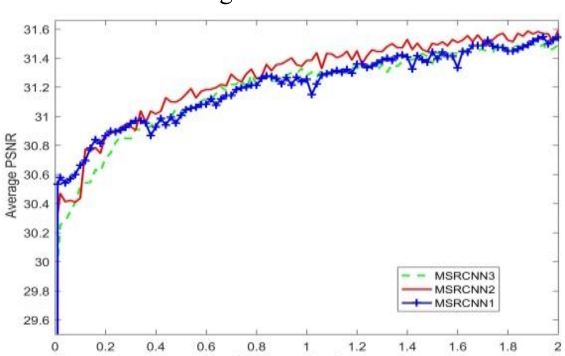

Figure 4. MSRCNN1 vs MSRCNN2 vs MSRCNN3

Experiment shows that the best average PSNR is obtained by MSRCNN2 model ($f_{11}=7$, $f_{12}=9$, $f_{13}=11$). Therefore, parameters of MSRCNN2 are adjusted to further improve the quality of reconstruction in the next section.

C. Model Parameters

1) The number of filters

In MSRCNN2, the filter number of each layer is $n_{11}=n_{12}=n_{13}=32$, $n_2=32$, $n_3=1$. This experiment changes filter number in first layer and second layer, that is MSRCNN4 ($n_{11}=n_{12}=n_{13}=64$, $n_2=64$, $n_3=1$) and MSRCNN5 ($n_{11}=n_{12}=n_{13}=16$, $n_2=16$, $n_3=1$). The testing results on Dataset2 are shown in Fig. 5.

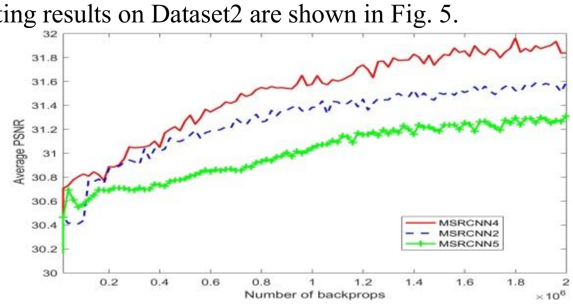

Figure 5. MSRCNN2 vs MSRCNN4 vs MSRCNN5

Experiment shows that MSRCNN4 is better than MSRCNN2, PSNR value is improving 0.38dB, but the training time is greatly increased. In order to balance performance and time consumption, MSRCNN2 model is the better choice.

2) Filter size

In MSRCNN2, the filter size of the non-linear mapping layer is $f_2=3$. The filter size of the non-linear mapping layer is adjusted to MSRCNN6 ($f_2=1$) and MSRCNN7 ($f_2=5$). The testing results on Dataset2 are shown in Fig.6. Experiment shows that the average PSNR value of MSRCNN2 is higher than the other two models, so MSRCNN2 model is still the best choice.

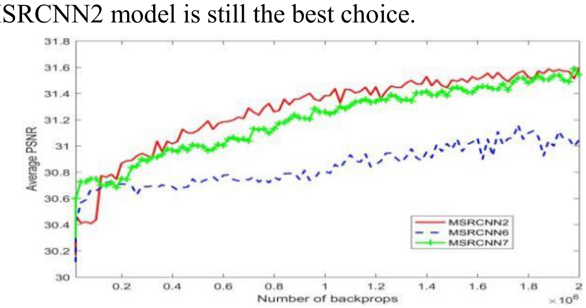

Figure 6. MSRCNN2 vs MSRCNN6 vs MSRCNN7

3) The number of layers

A non-linear mapping layer is added to MSRCNN2 model, that is MSRCNN8 model ($f_{11}=7$, $f_{12}=9$, $f_{13}=11$, $f_{21}=3$, f_{22}=3 , $f_3=5$). The testing results on Dataset2 are shown in Fig.7. Experiment shows that the average PSNR value of MSRCNN8 and MSRCNN2 are almost the same. However, MSRCNN8 model is more complex and its training time is longer than MSRCNN2 model. So, MSRCNN2 model is still the best choice. PSNR value and SSIM value of models are shown in Table I.

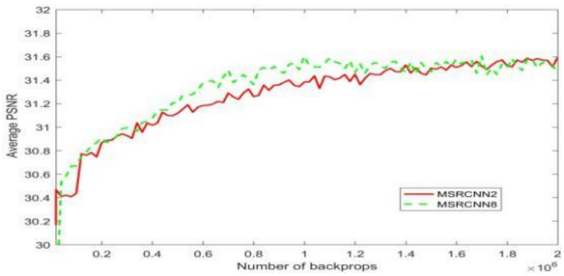

Figure 7. MSRCNN2 vs MSRCNN8

Table I shows that compared with MSRCNN2 model, PSNR value of MSRCNN4 model, MSRCNN7 model and MSRCNN8 model are respectively improving 0.38dB, 0.30dB and 0.37dB. But the training time of MSRCNN4 model, MSRCNN7 model and MSRCNN8 model are longer than MSRCNN2 model. In order to balance time consumption and reconstruction quality, MSRCNN2 model is the best choice. The license plate image reconstruction results are shown in Fig. 8. The proposed method is effective on the license plate image reconstruction.

Figure 8. License plate image reconstruction results

TABLE I. TEST RESULTS

PSNR /SSIM	**A466W5**	**A087DN**	**A062AN**	**A12L37**	**A66BL7**	**Average of Set5**	**Average of Set10**
Bi-cubic	26.4929	24.9867	24.6982	25.6827	28.1865	26.0094	26.9214
	/0.8023	/0.7421	/0.7643	/0.8127	/0.8518	/0.7946	/0.7592
SRCNN [12]	30.1528	26.9906	29.5549	28.6380	31.6472	29.3967	29.9824
	/0.8811	/0.8679	/0.8647	/0.8807	/0.919	/0.8826	/0.8589
SRCNN (Dataset1)	30.5470	28.3892	30.7347	29.5043	32.6336	30.3617	30.6559
	/0.8870	/0.9037	/0.8997	/0.9196	/0.9346	/0.9088	/0.8826
SRCNN (Dataset2)	31.3017	28.7263	31.4079	30.5656	33.2725	31.0548	31.1410
	/0.9118	/0.9106	/0.9055	/0.9267	/0.9425	/0.9194	/0.8888
MSRCNN1 (Dataset2)	32.1100	28.8110	32.5817	31.7758	34.0367	31.8631	31.8866
	/ 0.9243	/ 0.9180	/ 0.9267	/ 0.9394	/ 0.9515	/0.9319	/0.9055
MSRCNN2 (Dataset2)	32.1152	29.5620	32.5834	32.1238	34.3242	**32.1417**	**32.1107**
	/0.9287	/ 0.9317	/ 0.9286	/ 0.9384	/ 0.9539	/**0.9363**	**/0.9103**
MSRCNN3 (Dataset2)	32.2647	29.4115	32.6338	32.0824	34.2397	32.1064	32.1014
	/0.9340	/0.9285	/0.9296	/0.9414	/0.9537	/0.9276	/0.9100
MSRCNN4 (Dataset2)	32.7546	29.3482	32.9775	32.6556	34.6398	**32.4751**	**32.4934**
	/ 0.9353	/ 0.9276	/ 0.9305	/ 0.9445	/ 0.9558	/**0.9387**	**/0.9177**
MSRCNN5 (Dataset2)	31.5263	29.1004	32.0730	31.2869	33.7416	31.5456	31.5916
	/ 0.9107	/ 0.9207	/ 0.9216	/ 0.9336	/ 0.9470	/0.9267	/0.9874

MSRCNN6 (Dataset2)	31.4126 / 0.9115	28.7354 / 0.9107	31.8657 / 0.9108	30.9261 / 0.9298	33.6527 / 0.9467	31.3185 /0.9219	31.2830 /0.8916
MSRCNN7 (Dataset2)	32.7722 / 0.9375	29.2593 / 0.9217	33.1222 / 0.9380	32.4821 / 0.9423	34.5336 / 0.9560	**32.4388** /**0.9391**	**32.4116** **/0.9152**
MSRCNN8 (Dataset2)	32.8434 / 0.9362	29.4160 / 0.9240	33.4530 / 0.9354	32.7366 / 0.9451	34.4259 / 0.9542	**32.5750** /**0.9389**	**32.4876** **/0.9170**

IV. CONCLUSION

In this paper, a super-resolution reconstruction method for license plate images is proposed. Based on SRCNN [12], [13] method, the proposed method uses multi-scale parallel filters in the first layer to obtain more features of image. Non-linear mapping is implemented in the second layer, and then super-resolution reconstruction is done in the third layer. Experiment shows that the average PSNR value of the proposed method is 2.1dB higher than SRCNN [12]. The proposed super-resolution reconstruction method is effective. But for practical forensic images, the size of license plate image is usually very small even just only a dozen pixels. Therefore high-magnification reconstruction method will be studied in the follow-up work.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (No. 61601362), Shaanxi International Joint-Research Project (No. 2015KW-005, No. 2017KW-016).

REFERENCES

[1] Hayat K. Super-Resolution via Deep Learning[J].2017. Computer Vision and Pattern Recognition.arXiv: 1706.09077[cs.CV]

[2] Chen S L, Huang H Y, Luo C. A Low-Cost High-Quality Adaptive Scalar for Real-Time Multimedia Applications[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2011, 21(11):1600-1611.

[3] Chavez-Roman H, Ponomaryov V. Super Resolution Image Generation Using Wavelet Domain Interpolation With Edge Extraction via a Sparse Representation[J]. IEEE Geoscience & Remote Sensing Letters, 2014, 11(10):1777-1781.

[4] Li X, Hu Y, Gao X, et al. A multi-frame image super-resolution method[J]. Signal Processing, 2010, 90(2):405-414.

[5] Gao X, Wang Q, Li X, et al. Zernike-Moment-Based Image Super Resolution[J]. IEEE Transactions on Image Processing, 2011, 20(10):2738-2747.

[6] Yu J, Gao X, Tao D, et al. A unified learning framework for single image super-resolution[J]. IEEE Transactions on Neural Networks & Learning Systems, 2014, 25(4):780-792.

[7] Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches[C]// IEEE Conference on Computer Vision & Pattern Recognition, 2008:1-8.

[8] Wang C, Liu Y, Bai X, et al. Deep Residual Convolutional Neural Network for Hyperspectral Image Super-Resolution[C]// International Conference on Image and Graphics. Springer, Cham, 2017:370-380.

[9] Shao M, Wang Y, Wang Y, A Super-Resolution Based Method to Synthesize Visual Images from Near Infrared[C]// International Conference on Image Processing, 2009:2453-2456.

[10] Ker J, Wang L, Rao J, et al. Deep Learning Applications in Medical Image Analysis[J]. IEEE Access, 2017,PP(99):1-1.

[11] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// IEEE Conference on Computer Vision & Pattern Recognition, 2015:1-9.

[12] Dong C, Chen C L, He K, et al. Image Super-Resolution Using Deep Convolutional Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(2):295-307.

[13] Dong C, Chen C L, He K, et al. Learning a Deep Convolutional Network for Image Super-Resolution[C]// European Conference on Computer Vision. Springer International Publishing, 2014:184-199.