Super-Resolution Restoration of License Plate Images Based on Deep Learning

Written by Tian-Zheng Li, Yu-Bo Zhou, Hao Lan, Zong-Tao Zhou, Meng-Hao Wang

[1] Class of School of Informatics & Class of AI, Xiamen University
23020221154096, 23020221154153, 31520221154210, 23020221154154, 30920221154284

Abstract

With the rapid development of the economy and society and the improvement of people's living standards, the number of cars is also increasing. Therefore, super resolution (SR) reconstruction of low resolution license plate images is of great significance for improving the level and efficiency of vehicle safety management. With the development of depth convolution neural network (DCNN), the research of super-resolution has made progress in recent years. We try to use the deep convolutional neural network to reconstruct low resolution license plate images. We delete unnecessary modules in the traditional residual network, and our network can achieve better performance under the same computing resources. The method we have implemented can successfully carry out super-resolution reconstruction of fuzzy license plate images, obtain high-definition license plate images, facilitate subsequent license plate recognition and other work, and show good performance on the benchmark dataset.

Introduction

An important part of smart cities is smart traffic that can automatically manage the traffic affairs such as traffic violations. Hence, we need to process many pictures of accident scenes and extract useful information. However, not all pictures are perfect to do so because of their low resolution and few available areas. High-resolution licence plate images can improve the quality of research on licence plate detection, licence plate recognition and licence plate segmentation. However, the blurring of licence plate images due to the hardware and the environment in which they are captured poses a significant challenge. But lucky enough, super-resolution can solve this problem. To address the problems of the high economic cost of improving image resolution through hardware enhancement and the ineffectiveness of traditional resolution enhancement methods, a method of super-resolution reconstruction of blurred licence plate images using convolutional neural networks is proposed. A neural network is used to enhance the resolution of the blurred license plate image, and low-cost software is used to improve the resolution.

In the last couple of years, super-resolution is burgeoning due to the rapid development of deep learning. Some algorithms, like VDSR, SRResNet, performed excellently solving this problem. And, consequently, those algorithms are also sensitive to hyper-parameters and net architecture which would trigger heavy computation and memory use if they are not designed carefully.

Based on SRResNet, we make a lot of improvements including modifications of residual block and rebuilding the architecture to solve these problems. We removed some unnecessary layers in the residual block to decrease the number of parameters on the premise that the effectiveness of the residual block still remains. Besides, we adjust the architecture with the improved residual block mentioned previously and find some suitabe hyper parameters to achieve a higher performance.

While on training period, we collected a batch of license plate images of vehicles in a low resolution and scale them to a high resolution. We also choose the DIV2K dataset as a part of the training set to gain more training samples to improve the performance of our model. When evaluating our model, PSNR and SSIM criteria are used to test images of DIV2K datasets and collected license plate images of vehicles.

Related Work

Image Super-Resolution (SR), especially single-image super-resolution (SISR),is an important class of image processing techniques to enhance the resolution of images and videos in computer vision.It refers to the process of recovering a high-resolution (HR) image from its degraded lowersolution (LR) one. Recent years have witnessed remarkable progress of image super-resolution using deep learning techniques. In general, we can roughly group the existing studies of SR techniques into three major categories: supervised SR, unsupervised SR, and domain-specific SR. In this report, we focus on supervised SR, i.e., trained with both LR images and corresponding HR images.

The existing super-resolution restoration models based on depth learning can be divided into four categories in chronological order: convolutional neural networks (CNNs) based models; generative adversarial networks (GANs) based models; channel attention (CA) based models and Transformer based models. SRCNN is the first CNN-based SISR model, it extracts feature vectors from the image patches by convolution, which are then non-linearly mapped to find the most representative patches to reconstruct the HR image. Motivated by SRCNN, building a deep neural network to directly learn the mapping between LR and HR images has become the mainstream method in SISR. Inspired by ResNet, EDSR models the residual image between LR and HR images to relieve vanishing/exploding gradients and degradation problems caused by directly stacking the layers.

In contrast to traditional machine learning methods, GANs are known for their ability to preserve texture details in images, create solutions that are close to the real image, and appear perceptually convincing. Thus, GANs are also suitable for SISR. Christian et al. proposed the most well-known and first successful GAN-based SISR model, the Super-Resolution Generative Adversarial Network (SRGAN). In SRGAN, the generator G is essentially as SR model that trained to fool the discriminator D, and D is trained to distinguish SR images from HR images. Therefore, the generator can learn to produce outputs that are highly similar to HR images, and then reconstruct more perceptual and natural SR images.

The attention mechanism can be viewed as a tool that can allocate available resources to the most informative part of the input. In order to improve the efficiency during the learning procedure, some works are proposed to guide the network to pay more attention to the regions of interest. Zhang et al. incorporate the channel attention mechanism with SR and propose RCAN, which markedly improves the representation ability of the model and SR performance by focusing on more useful channels.

The key idea of Transformer is the self-attention mechanism, which can capture long-term information between sequence elements. Nowadays, some researchers try to introduce Transformer to image restoration tasks. Liang et al. proposed the SwinIR for image restoration based on the Swin Transformer. Specifically, the Swin Transformer blocks (RSTB) is proposed for feature extraction and DIV2K+Flickr2K are used for training. It's also the SOTA SISR model at present with parameters of more than 1M (M=million).

Method

In this section, we mainly introduce our overall idea of super-resolution. First, we define the problem we want to solve so that our work content can be clearly defined. Then we will introduce the basic structure and improved structure of the residual block. In the following part, we introduce the super-resolution prediction network structure composed of residual blocks, convolution layer, upper sampling layer, etc. Finally, we introduce the loss function and evaluation methods.

Problem Definition

We consider the problem of super resolution the pictures which contain license plates of vehicles so people or other model networks can distinguish the license plate number clearly and easily. To solve this problem, we aim at improving the performance of an existing super-resolution network not only in the field of training speed and usage of resource, but also in model effect.

Figure 1: A sample of training data: high resolution, 2x, 3x, 4x downsample

Dataset Preprocessing

DIV2K dataset is a classically proposed high-quality (2K resolution) image dataset for image restoration tasks. The DIV2K dataset consists of 800 training images, 100 validation images, and 100 test images. As the test dataset ground truth is not released, we choose some license plate images processed through bicubic degeneration as the testing dataset to verify the performance. The actual effect of image restoration can be tested by comparing the downsampled image with the original image. The SR results are evaluated with PSNR and SSIM on Y channel (i.e., luminance) of transformed YCbCr space.

The detail of data augmentation are as follows:

- The original resolution of the dataset is 2K which is high time-consuming for the training process. For all the images we need to crop them to the size 192*192 in order to increase the amount of training data and reduce the computational complexity.
- We rotate all the images by 90 degrees increments to increase the amount of training data further and improve the robustness of the model.
- To generate the counterparts of the high resolution images, we obtain low resolution ones by downsampling with bicubic interpolation (2x, 3x and 4x).

Residual block

We plan to improve the SRResNet and remove the BN layer in the residual block as Fig.2 shows. The BN layer is used to regularize batches and prevent over-fitting. However, the BN layer regularizes batches by normalization operations, which will destroy the original color distribution of the image and lose the contrast information. Moreover, in the field of super-resolution, the BN layer will slow down the process

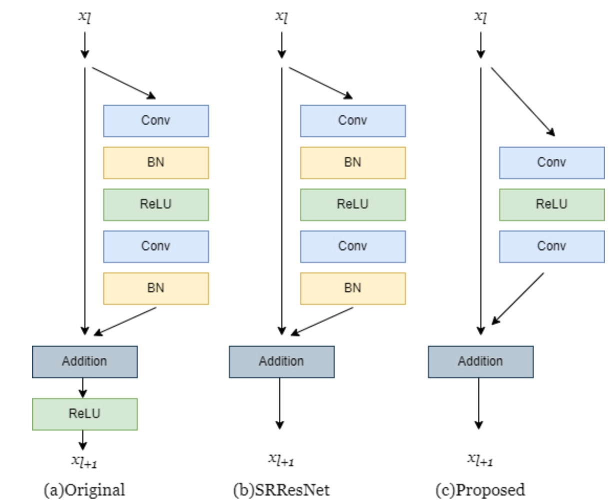

Figure 2: Comparison of residual blocks in original ResNet, SRResNet, and ours.

of training and even make it divergent. At Fig.2, we compare the architectures of original ResNet, SRResNet and our improved SRResNet. By removing useless layers, our model can be trained faster and could be larger to gain a better performance.

Channel Attention

The CA (Channel Attention) module is used to improve the feature representation of the network by modeling the dependencies of the channels of the feature map, firstly by global pooling in the channel dimension, then by adaptively modeling the interdependencies of the channels using full connectivity and non-linear activation operations, and finally by weighting the channels of the original feature with the new interdependencies. Finally, the new interdependencies are used to weigh each channel of the original feature for the purpose of feature response and recalibration. In this way, the network learns to selectively enhance features containing useful information and suppress features that are not useful or have little effect by using global information.

Single-scale model

We use residual blocks mentioned in the previous section to build the baseline model. Its structure is similar to SRResNet, but the external Relu layer is removed. Since we use only 64 feature maps for each convolution layer, we also remove the residual scaling layers. We set parameter B(the number of layers) = 32, parameter F(the number of feature channels) = 256 with a scaling factor 0.1 to expand the baseline model. The model architecture is displayed in Fig.3.

Loss Function

It is crucial to choose an appropriate loss function in feature learning for license plate recognition. In this work, we choose MAE (i.e. L1Loss) as the loss function, which is the error obtained by taking the absolute value of the target

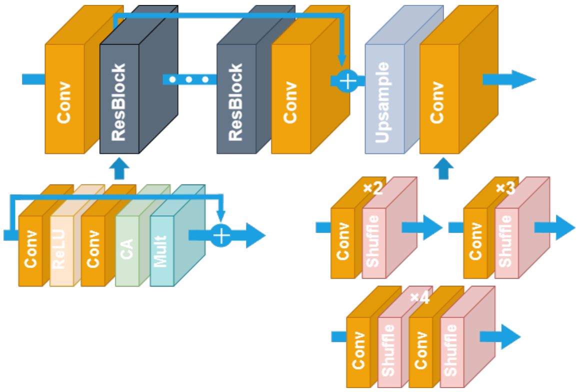

Figure 3: The architecture of the proposed single-scale SR network.

value y_i and the model output (i.e. estimated value) $f(x_i)$, as shown in equation 1.

$$Loss(x, y) = \frac{1}{n} \sum_{i=1}^{n} (\mid y_i + f(x_i) \mid) \qquad (1)$$

The main reasons for using MAE are: 1. Compared with MSE, it will be no exploding gradient. 2. Its degree of convergence is more stable in this work.

Evaluation

We train all this model with 3×105 updates in this experiment. Evaluation is conducted on the 10 images of DIV2K validation set, with PSNR and SSIM criteria. PSNR, Peak Signal to Noise Ratio, which compares every pixel of the original image with the corresponding one of the generated image, is a generally used quality criterion of images. When it is higher than 40dB, the generated image is very close to the original one. And when it is lower than 20dB, the generated image is unacceptable. But as it is much objectively and globally that the criterion compares the images, there are some cases that human think the image is acceptable because they just focus on some part while PSNR holds the opposite opinion. Therefore, we also use the SSIM, a criterion which uses structural similarity to compare the original images and the generated ones. It uses luminance, contrast and structure to judge the quality of generated one. Due to its behavior being close to human's, it acts very well in the computer vision field. For the evaluation, we plan to use full RGB channels and ignore the (6+scale) pixels from the border.

Experiment

Training Details

For training, we use the RGB input patches from the LR image with the corresponding HR patches. Before training, we specifically augment the training images with random horizontal flips and 90 rotations. We pre-process all the images by subtracting the mean RGB value of the DIV2K dataset. We use the Adam optimizer, an efficient stochastic optimization method that requires only first-order gradients and a

Figure 4: The car plate images of origin, ×2 scale, ×3 scale and ×4 scale.

small amount of internal memory. For the learning rate, we set the initial learning rate to be 10^{-4}, and set the minibatch size at 16. We train our networks using L1 loss instead of L2. Minimizing L2 is generally preferred since it maximizes the PSNR. However, based on a series of experiments, we empirically found that L1 loss provides better convergence than L2, and thus makes the training more stable. We use PyTorch to implement our models with an NVIDIA 3050 GPU and the number of epochs is 100.

Result

We trained three models in total with the scale of ×2, ×3, and ×4. We tested all the models with specific car plate photos downsampled by ×2, ×3, and ×4. We chose SSIM and PSNR to evaluate our models. The evaluation result is shown in Table 1. It shows that the bigger the scale is, the lower the quality of the enhanced image is. This result is acceptable and predictable, because bigger scale means more information to restore from the input that contains limited information due to its low resolution.

We input downsampled car plate images and the outputs are shown in Figure 3. As we can see, the super resolution result is good that we can clearly recognize the plate number in the image. The Chinese character, however, is not enhanced well due to its complicated structure that is hard to predict.

Conclusion

In this paper, a neural network structure based on depth convolution neural network is used to reconstruct low resolution license plate images. We have made a lot of improvements to the traditional SRResNet. On the premise of retaining the validity of the residual block, we have removed some unnecessary layers in the residual block, and reduced the number of parameters, so that we can obtain improved results while making the model more compact. Our model can successfully carry out super-resolution reconstruction of fuzzy license plate images to obtain high-definition license plate images and has achieved good results on the DIV2K dataset.

Scale	SSIM ↑	PSNR ↑
x2	0.94	29.66
x3	0.92	28.33
x4	0.9	28.16

Table 1: The evaluation of different super resolution strategies

Reference Examples

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2014. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, 184–199. Springer.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Kim, J.; Lee, J. K.; and Lee, K. M. 2016. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1637–1645.

Lim, B.; Son, S.; Kim, H.; Nah, S.; and Mu Lee, K. 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 136–144.

Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681–4690.

Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; and Fu, Y. 2018. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, 286–301.

Dai, T.; Cai, J.; Zhang, Y.; Xia, S.-T.; and Zhang, L. 2019. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11065–11074.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1833–1844.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022.

Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.-H.; and Zhang, L. 2017. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 114–125.

Sung; Cheol; Park; Min; Kyu; Park; Moon; Gi; and Kang. 2003. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine IEEE*.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.