Super-resolution Implementation of Anime Pictures

Written by Qing-Song Wu, Rui Nangong, Li-Jian Zhuang, Wen-Liang Chen, Chi Shang

[1]Xiamen University
350581199508072715, 31520211154072, 31520211154120, 31520211154027, 31520211154075

Abstract

With the continuous development of electronic display devices, super-resolution (SR) reconstruction of previous low-resolution anime is of great importance to those who like anime. With the booming development of deep learning, single image super-resolution (SISR) has been progressed in recent years, so we try to use a neural network structure based on deep convolutional neural network to reconstruct one frame of anime image with super-resolution, and then synthesize a new video to realize the high definition of old anime. We use a deep convolutional neural network structure to reconstruct the frames of anime in super-resolution, and then synthesize the new video to realize the high definition of old anime, so that more people can access these excellent old anime. After our processing, the super-resolution reconstructed anime images are more detailed and more comfortable than the original anime images.

Introduction

In the current digital era, a high resolution for storing and publishing art images is essential. Anime images are a kind of artworks, which are popular among young people. However, many images or animations published in the online media, may have lower resolution. They may look pixelated or blur. Low-resolution images are used in part to save storage, reduce upload or download time, deter illegitimate use of images, and use intentionally them as thumbnails(Zou and Yuen 2011). As for those classic animation of the past, its low quality mainly due to hand-drawn factor and the limitations of era. Hence, reproducing existing anime with larger sizes and higher quality is challenge. Now image super-resolution can help us to improve the quality of pictures and even videos. Image super-resolution (SR), which refers to the process of transforming a low-resolution (LR) image into a high-resolution (HR) image, is an important class of image processing techniques in computer vision and image processing. It enjoys a wide range of real-world applications, such as medical imaging (Greenspan 2009), (Isaac and Kulkarni 2015), (Huang, Shao, and Frangi 2017), surveillance and security (Zhang et al. 2010),(Rasti et al. 2016). With the rapid development of deep learning techniques in recent years, deep learning based SR models have been actively explored and often achieve the state-of-the-art performance on various benchmarks of SR(Dong et al. 2014),(Dong et al. 2016),(Ledig et al. 2017). So we use a modified deep convolutional neural network model with attention mechanism to improve the resolution of anime images. We collected high resolution anime pictures from the web and blurred them to get low resolution image sets. By using HR and LR datasets for supervised learning, we obtained the final model. Finally, we also applied this model to an old cartoon and found that the resolution of the video was indeed improved.

Figure 1: Similar images in $1024 * 1024$ resolution.

Related Work

Interpolation-based SR reconstruction method. When it comes to image SR reconstruction, we have to start from interpolation-based SR reconstruction, which is a traditional interpolation method between pixels on the existing image to supplement the missing pixels in the HR image:

- Upsampling, obtaining some points of the HR image directly from the LR image.
- Interpolation, inserting the missing pixel points of the HR image.
- Deblurring, smoothing the image.

Figure 2: Similar images in $256 * 256$ resolution.

Traditional interpolation methods include nearest neighbor interpolation, bilinear interpolation, bicubic interpolation, etc.

Although the method is simple and easy to calculate, with the increase of image magnification, the reconstruction results will have defects such as edge smoothing, blurring and ringing and jagged effects, especially for complex images, the reconstruction quality will not be good, so we abandon this method.

SR reconstruction method based on traditional learning. The learning-based SR method is a single-image SR method with samples, which learns the statistical relationship between high- and low-resolution images and applies this relationship to the reconstruction process to achieve SR reconstruction of images. The SR reconstruction method based on sample learning can be subdivided into three methods based on sample learning, sparse representation and regression. Because it is difficult to express the complex feature data between image blocks and to consider the complexity and diversity of image scenes, the accuracy of the reconstructed HR images is still not high and there are more edge blurring and texture detail problems.

Deep learning-based SR reconstruction method. With the continuous updating of deep learning techniques, SR reconstruction methods based on deep learning have been flourishing in recent years. With deep learning, there is no need for separate pre-processing operations such as feature extraction and later HR image block aggregation for image blocks. By automatically learning multi-level features using linear transforms, it is possible to uncover deeper intrinsic connections between high and low resolution images. By using deep convolutional neural networks [1], it is possible to directly learn the direct mapping relationship between LR images and HR images, and reconstruct the missing details of LR images to achieve the SR effect. The network used in this course is the attention in attention network (A2N) based on deep convolutional neural network [5], which the authors introduce an attention mechanism to distinguish texture regions from smoothed regions and perform high frequency compensation after locating the location of high frequency details, making the SR effect better.

Method

Data Processing

Original Data Set. We collected 420 high-definition anime images on the Internet for training. The resolutions of these images are all above 1024*1024. In addition, we searched for another 24 low-resolution images on the web for testing. The resolutions of these images are all below 256*256.

For solving practical application problems, we selected the first episode of the animated film "Journey to the West" from the ancient period of China for super-resolution recovery. The animation video has a frame rate of 25 and a resolution of 576*432.

Preprocessing.

- For 420 images, we first need to crop them. The benefits of cropping are twofold: first, it can increase the amount of data for training, and second, it can reduce the computational overhead during training. For each image, we intercept 480*480 size subimages in steps of 200. The number of subgraphs that can be cropped from each original image is uncertain, depending on the resolution of the original image. After this step, we obtain a set of 6139 images with a fixed size of 480*480, which are used as supervised information HR (i.e., high resolution images) for training.
- Regarding the generation of LR (low resolution images), we chose to obtain them directly by interpolating downsampling (2x and 4x) on the basis of the high resolution images.
- In addition, for the construction of the dataset, we recommend using the h5 format, which has great advantages for storing large amounts of data and is very efficient for processing. Therefore, we jointly transformed HR and LR data pairs into h5 files during the data pre-processing phase.

Create Dataset. We used the Dataset class of pytorch as the parent class and constructed our own MyDataSet class using the h5 file generated earlier.

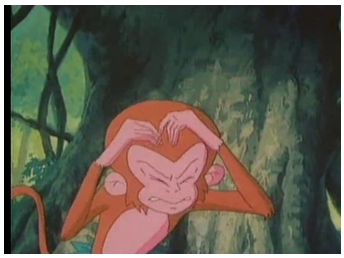

Figure 3: Video of "Journey to the West" with frame rate of 25 and resolution of $576 * 432$.

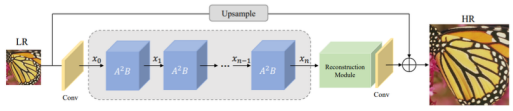

Figure 4: Network structure.

In this Dataset, we also need to perform a crop of the HR image into a 256*256 image, and accordingly, the LR image into a 64*64 image (hyperscale of 4). In this crop, the selection of the crop position is a random process. After that, we use some data enhancement methods, including horizontal flip, vertical flip, clockwise rotation by 90 degrees, etc., and the probability of these operations is set to 0.5. Thus, our own DataSet returns the cropped and data enhanced HR and LR data pairs.

Network Structure

Attention in Attention Network Overall. For a low-resolution image (LR), the first layer of 3*3 convolution is used to extract the initial feature information feature map, and then the feature map is fed into n cascading Attention in Attention Blocks, and then the output goes through a Reconstruction Module and a layer of 3*3 convolution to obtain the final network learning content. Finally, the SR (super-resolution image) is obtained by adding the content learned by the network with the result of upsampling the original LR image by traditional interpolation.

Among them, Attention in Attention Block is an improvement of the attention mechanism by the authors in the reference paper, while the Reconstruction Module mainly uses the Pixel Attention mechanism with the traditional interpolation upsampling to improve the image resolution. In other words, the entire network does not use subpixel convolution or deconvolution, but only traditional interpolation upsampling. In particular, both subpixel convolution and deconvolution are classical upsampling methods that have been frequently used in many previous works.

Attention in Attention Block. This module is one of the main contributions of this reference paper. There are three types of attentional mechanisms: channel attentional mechanisms, spatial attentional mechanisms, and channel spatial attentional mechanisms. The authors propose improvements to the spatial channel attention mechanism. This module performs a global average pooling of the input, followed by two fully-connected layers, and finally activated by a softmax function, which corresponds to the generation of two probabilities, one corresponding to the Non-attention Branch and the other to the Attention Branch. In general, this module adds an attention mechanism for whether to use the attention mechanism or not, so it is named Attention in Attention Block. The input is then combined with the probabilities of the Dynamic Attention Block output to perform a weighted summation, and the summation result is added to the original input after a layer of 1*1 convolution to obtain the output of the module.

Training configuration

We use the Adam optimizer, an efficient stochastic optimization method that requires only first-order gradients and a small amount of internal memory. The method calculates adaptive learning rates for different parameters by estimating the first and second gradients, and is well suited for solving problems with large-scale data or parameters, as well as for solving non-stationary problems with large noise and sparse gradients. For the learning rate, we set the initial learning rate to be 5e-4, and set the batch size to be 32 (usually the size of the learning rate and the batch size show a positive correlation). For the super-resolution recovery task, there are various designs of loss functions, among which Pixel loss is the most common kind of loss. The Charbonnier loss is used in the Pixel loss, which can make the loss more stable compared to the L1 loss and L2 loss of the same Pixel loss, and thus make the training more stable:

$$\mathcal{L}_{pixel-Cha}(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} \sqrt{\left(\hat{I}_{i,j,k} - I_{i,j,k}\right)^2 + \epsilon^2} \tag{1}$$

The last parameter is a very small constant, which we take as 1e-6. It can be observed that the Charbonnier loss is actually a variant of the L1 loss.

Pixel loss is simple and easy to implement, and also enables a concise and efficient training process, but it does not take into account the image quality (e.g., perceptual quality of texture), often lacks high-frequency details, and produces textures that are too smooth and difficult to satisfy. Nonetheless, pixel loss in super-resolution rate recovery task cannot be ignored.

Experiments

In order to prove the effectiveness of the SR method, we used the 24 LR anime images we downloaded for testing, and the results are shown below (original images on the left, SR reconstructed images on the right). We can see that the clarity of the processed images has been improved, and the effect of improving the quality of anime images has been well achieved.

We also searched for the first episode of the Journey to the West cartoon from more than 20 years ago to test its effect on continuous video frames. The test procedure is to reconstruct the SR frame by frame, and then combine all the processed images one by one into one video.

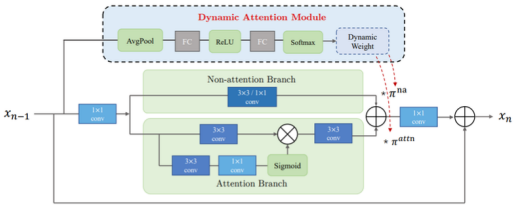

Figure 5: Attention Block.

Figure 6: Experimental comparison results.

The result is surprisingly good, and we show a few frames of the comparison below. (The left is the original video image, and the right is the image after SR reconstruction)

The experimental results show that our trained model can achieve good results in SR reconstruction of LR anime images, and the old anime processed by this SR method can appear much clearer to a certain extent, which can improve the viewing experience.

Conclusions

In this paper, we use a deep convolutional neural network with an integrated attention mechanism to train a model that can directly learn the direct mapping relationship between LR and HR images and reconstruct the missing details of LR images, and to some extent distinguish between textured and smoothed areas. The model can achieve good results in SR reconstruction on most of the test sets of LR animation images and improve the image quality to some extent. In addition, it is also very good in the SR reconstruction of real LR animation videos, and successfully improves the video quality significantly.

Reference Examples

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2): 295–307.

Fang, Z.; Dongxu, Z.; Zhitao, X.; Lei, G.; Jun, W.; and Yanbei, L. 2021. Research progress of single-image super-resolution reconstruction technology. *Journal of Automation of China*, 47: 10001–10021.

Zhang, Y.; Chen, H.; Chen, X.; Deng, Y.; Xu, C.; and Wang, Y. 2021. Data-Free Knowledge Distillation for Image Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7852–7861.

Wang, L.; Dong, X.; Wang, Y.; Ying, X.; Lin, Z.; An, W.; and Guo, Y. 2021. Exploring Sparsity in Image Super-Resolution for Efficient Inference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4917–4926.

Chen, H.; Gu, J.; and Zhang, Z. 2021. Attention in Attention Network for Image Super-Resolution. *arXiv preprint arXiv:2104.09497*.

Zou, W. W.; and Yuen, P. C. 2011. Very low resolution face recognition problem. *IEEE Transactions on image processing*, 21(1): 327–340.

Greenspan, H. 2009. Super-resolution in medical imaging. *The computer journal*, 52(1): 43–63.

Isaac, J. S.; and Kulkarni, R. 2015. Super resolution techniques for medical image processing. In *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, 1–6. IEEE.

Huang, Y.; Shao, L.; and Frangi, A. F. 2017. Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6070–6079.

Zhang, L.; Zhang, H.; Shen, H.; and Li, P. 2010. A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, 90(3): 848–859.

Rasti, P.; Uiboupin, T.; Escalera, S.; and Anbarjafari, G. 2016. Convolutional neural network super resolution for face recognition in surveillance monitoring. In *International conference on articulated motion and deformable objects*, 175–184. Springer.

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.

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.