The Research of Single Image Super-Resolution Based on the Generative Adversary Networks

Chunjiang Duanmu1*, Zhe Zheng2

1Departments of Electrical and Information Engineering, School of Physics and Electrical Information Engineering, Zhejiang Normal University, 321004, P. R. China

Email of all the authors: duanmu@zjnu.cn, 652323083@qq.com

* Corresponding Author: Chunjiang Duanmu; email: duanmu@zjnu.cn; phone: 13967485426; fax: 0579-82291384.

Abstract: In order to enhance the visual effects of the reconstructed images in the single image super-resolution, and solve the problem of instability of the training phase of super-resolution with generative adversary network (SRGAN), a new super-resolution model is built with more depth and width, and the corresponding super-resolution algorithm is proposed. The network structure of the original SRGAN is modified, so that deeper and wider convolutional networks can be used with high efficiency. Furthermore, the loss function is employed to balance the training of the generative network and the discriminator network, and solve the instability problem in the training phase. The training images for the network are obtained from the database of ImageNet on the web as other researchers. Experimental results show that images reconstructed by the proposed algorithm has better visual effects than those by the original SRGAN. In addition, the objective measures of PSNR and SSIM of the reconstructed images have also been improved.

1. Introduction

1.1. Review of the existing super-resolution algorithms

The objective of single image super-resolution is to reconstruct a high resolution image from a single low resolution image. This technique has wide applications in medical imaging, public security area, remote sensing, industry imaging, internet usages, etc.

Currently, the single image super-resolution methods can be divided into the following three classes [1], i) the methods based on interpolation [2], ii) the methods based on rebuilding [3-4], and iii) the methods based on learning [5]. The bi-cubic interpolation method [6] is a typical interpolation method, it assumes that the image pixel values vary smoothly, and estimates the unknown pixel values from its surrounding known pixel values. However, this assumption is unsuitable for a lot of cases. The back projection method [7-8], the maximum posterior probability method [9], the convex set projection method [10] are typical methods based on rebuilding. These methods have a close relationship with prior high-resolution image knowledge. If the input image is too small or the magnification factor is too large. The rebuilding results will be poor. The methods based on learning is to learn the relationship between high-resolution and low-resolution images to rebuild the high-resolution image. They have a lot of advantages over other methods and currently is the trend for the research of single image super-resolution.
In the recent few years, since the deep learning have obtained good results in many areas, more and more researchers begin to use deep learning for super-resolution. In 2014, Dong etc.[11] first proposed the single image super-resolution method based on convolution neural network, and named their method as SRCNN. The performance of SRCNN is better than many traditional methods. The convolution neural network in this method has three layers. Later, Dong etc. [12] proposed a revised method to accelerate the online super-resolution time. In this method, the layers of SRCNN is revised from 3 layers to 8 layers, and the de-convolution layer is used in the last layer. The online super-resolution speed of this method is much faster than that of the SRCNN. In 2017, Ledig etc. [13] first proposed to use the generative adversary network (GAN) for single image super-resolution, and named their method as SRGAN. A GAN is proposed to learn the relationship between low-resolution and high-resolution images and generates many high-resolution images. A discriminator is used to judge if an image is obtained from the generative network or from real high-resolution image database. When the discriminator cannot work to judge the high-resolution images, the corresponding generative GAN is employed for the super-resolution. From the experiments of SRGAN, the visual effects of the generated high-resolution images are more real.

1.2 Review of the generative adversary network
The GAN is proposed by Goodfellow etc. [14], it is illuminated from the game in the game theory. The high abilities of GAN to generate images make it have wide applications in image fusion [15], image in-painting [16], super-resolution [13], scratch image restoration [17], etc.

The basic framework of GAN has two components. One is the generative model and the other is the discriminator model.

The steps of the GAN is described in the following. x is sampled from the real training data as the input for the generative model D(x), and the D(x) will learn by itself and tries to make the output close to 1. At the same time, z is sampled from the prior distribution, and a forge sample G(z) is made by the generative model as the input to the discriminator model. The aim of the discriminator model is to let D(G(z)) as close to zero as possible, i.e. to judge the image as false. At the same time, the aim of the generative model is to let D(G(z)) as close to one as possible, i.e. to judge the image as true. Finally, a balance is reached in the game of these two models.

1.3 Review of the residual network
The residual network (ResNet) and the skip connection were proposed by He etc. [18]. This network structure makes the training of more deep network easier. The basic structure of it is shown in the figure 2.

The skip connection is added in the original convolution layer in the ResNet to make the output function from F(x) to H(x)=F(x)+x. This structure makes the aim of the learning network to learn F(x) instead of H(x). The learning of F(x) is much easier than the learning of H(x), since the output of F(x) has close relationship with the input of x. A residual network is built based on the layers of the residual ones. It solves the problem of degeneration of the network when the layers of the network is large and boosts the performance of the network.

This paper is organized as follows. In the Section 1, relative background information is reviewed. The proposed algorithm is described in details in Section 2, and the experiment results are shown and analyzed in Section 3. The conclusions are made in Section 4.

2. Material and Methods
The proposed algorithm is described in this section. First, we introduce the revision ideas to improve on the exiting algorithms. Then, the steps of the proposed algorithm named as WD-GAN are described. Its loss function is described to balance the training of GAN.
2.1 Revision ideas
The influence on the network structure’s performance includes the network structure, its depth, and its width. Thus, the proposed network tries to optimize the network structure, and decides optimal depth and width.

In this paper, it employed the idea in the reference papers [19-20] to remove the batch normalization (BN) in the SRResNet to boost the performance. The BN performs not well in the super-resolution. The addition of BN will make the training slower and more unstable. By removing the BN, under the same computing resource, more layers can be added and each layer can retrieve more features to boost the performance, since it is pointed by papers [22-24] that deeper network can improve the network accuracy greatly.

This paper also employs the idea in [21]. In this paper, it is pointed that the ReLU activation function in the super-resolution network will block the information exchange. Thus, the ReLU functions in the network are removed and the number of channels in the network features are increased.

In the traditional GAN, there is a very big problem [25], i.e. when to stop the training of the generator, and when to stop the training of the discriminator. If the discriminator is over trained, since the generator cannot continue to learn, it will produce bad network model. On the contrary, if the generator is over trained, since the discriminator cannot continue to learn, it will also produce bad network model. If a good loss function can be employed in the training, the difficulty in the training can be greatly reduced.

In summary, the proposed WD-GAN has the following revision ideas based on the network structure of SRGAN. 1. Improve the residue block in the generator of SRGAN and remove the BN layers in it. 2. Make the network have deeper structure, i.e. increase the layers of residue blocks. 3. Make the channels before the activation functions have more numbers to extract more features. 4. Replace the loss function to make the GAN more stable.

2.2 The explanation and the description of the proposed network structure of WD-GAN
In details, the proposed WD-GAN is different from the original SRGAN in the following points.
1. The first convolution layer in the generator has four 3x3 convolution kernels instead of the original one 9x9 convolution kernel. In this way, the computational complexity can be reduced.
2. The residue blocks in the original SRGAN is 16. In this paper, the number of residue blocks is increased to 24 to make the network deeper.
3. The BN layers in the residue blocks of generator and the discriminator are all removed.
4. The activation function is changed from the original PReLU to the function of ReLU.
5. The number of channels before the activation functions in the residue blocks of the generator is changed to be 96 instead of the 64.
6. The last sigmoid layer of the discriminator in the SRGAN is removed, since the change of the loss function requires to delete the last layer.

2.3 The design of the loss function
In the WGAN[25], it proposed to use the Wasserstein distance to replace the JS divergence as the network optimization objective. It also pointed that the problems of model collapse and instability etc. in the training phase are caused by the unreasonable measure of the distance between two distributions using the JS divergence.

The Wasserstein distance is defined as:

$$W(p_{data}, p_{g}) = \inf_{\gamma \in \Pi \{p_{data}, p_{g}\}} E_{\gamma(x,y)}[|x-y|]$$  \hspace{1cm} (1)

where, $$\Pi \{p_{data}, p_{g}\}$$ is a set, it has the elements of all the possible joint distribution of $$p_{data}$$ and $$p_{g}$$, $$\gamma(x,y)$$ is a specific joint distribution of $$p_{data}$$ and $$p_{g}$$, and its marginal probability distribution is $$p_{data}$$ and $$p_{g}$$, respectively, $$E_{\gamma(x,y)}[|x-y|]$$ is the mathematical expectation of the distance between
x and y, inf is the operator to compute the lower bound. Since the \( \inf_{p \in \mathcal{P}_{\text{data}}} \) in (1) cannot be computed directly, the duality is used to change (1) as
\[
W(p_{\text{data}}, p_g) = \sup_{\psi \in \mathcal{L}} E_{x \sim p_{\text{data}}} [f(x)] - E_{x \sim p_g} [f(x)]
\]
(2)

where Sup is the operator to compute the upper bound, \(|f|_L \leq 1\) is a Lipschitz continuous condition that \( f \) must be satisfied with a constant of 1. This condition limits the largest local variation of the function \( f \).

A group parameters of \( w \) can be used to define a serial possible function of \( f_w \). In this way, the computation of the equation (2) can be the approximate computation of the following equation (3).
\[
W(p_{\text{data}}, p_g) \approx \max_{\psi \in \mathcal{L}} E_{x \sim p_{\text{data}}} [f_w(x)] - E_{x \sim p_g} [f_w(x)]
\]
(3)

where \(-E_{x \sim p_g} [f_w(x)]\) is the loss function of the generator, \( E_{x \sim p_g} [f_w(x)] - E_{x \sim p_{\text{data}}} [f_w(x)]\) is the loss function of the discriminator. The training phase in the GAN is indicated by the computation of (3), where the less its value the high its training accuracy.

2.4 The description of the training phase
In the training of the proposed model, the ImageNet database is used as the training data set. In the training phase, each real high-resolution image is randomly cropped to make 96x96 sub-images with high-resolution. Each cropped 96x96 sub-image is down-sampled by the factor of 4 using bi-cubic method to obtain the 24x24 sub-image with low-resolution. Then, the sub-images of 96x96 pixels and its corresponding low-resolution sub-images of 24x24 pixels are the input for the generative network and the discriminator network. The initial learning rate is set as 1x10^-4. After 50000 iterations, the learning rate is set as one tenth of its original value.

After 75000 iterations, the learning rate is set as one hundredth of its original value. The training is carried out for 100000 iterations, and the batch size is set as 16.

3. Results
The operating system for the experiments is Windows 10 with 64 bits. The type of the GPU is GTX1050 Ti, the GPU’s memory is 4GB. The version of the Tensorflow is Tensorflow 1.13.1, the version of the CUDA is CUDA 10.1.0, and the version of the python is python 3.7.

The experiments use the ImageNet as the data set, where 50000 pictures with high resolution and different sizes are selected for training.
3.1 Objective measure
The PSNR (Peak Signal-to-Noise Ratio) [26] and the SSIM (Structural Similarity Index Measure) [27] are widely used to assess the quality of images. Thus, the objective measure for PSNR and SSIM are also used in the paper to judge the quality of the reconstructed high-resolution images, and the performance of the proposed super-resolution method.

3.2 Experiment results
In Figure 1, the reconstructed images of different algorithms for super-resolution are shown, where picture (a) is the result from bi-cubic algorithm, picture, (b) is from SRCNN algorithm, (c) is from the SRGAN algorithm, (d) is original high-resolution image where the down-sampling is made for the input of different images, (e) is from the proposed WD-GAN algorithm. It can be seen that except the original image, the result of the proposed algorithm looks more clear with detail information. The PSNR values of pictures (a), (b), (c), and (e) are 21.59dB, 23.53dB, 21.15dB, and 23.89dB, respectively, where the proposed algorithm performs best in term of the PSNR, since it reaches the highest value of PSNR. The SSIM values of pictures (a), (b), (c), and (e) are 0.6423, 0.7532, 0.6868, and 0.7695, respectively, where the proposed algorithm performs best in term of the SSIM, since it reaches the highest value of SSIM.

4. Conclusion
In this paper, it proposes a new network model of WD-GAN with deeper network layers and wider channels in each layer than the SRGAN. It adds more residual blocks and revises the structure and parameters for the residual blocks. The discriminator’s last sigmoid layer has been removed to use the proposed loss function. From the visual effects of the reconstructed high-resolution images, it can be seen that the images produced by the WD-GAN are more clear than those produced by the SRGAN. From objective measure of PSNR and SSIM, it can also be seen that the PSNR and SSIM have been improved by using the WD-GAN instead of the SRGAN.

In the future research, the pursuit of more optimal network will be conducted to make the reconstructed high-resolution images more clear and have less detail loss.

Acknowledgement
This research was financially supported by Natural Science Foundation of the Zhejiang Province (Grant No. Y15F010032).

References
[1] SU H, ZHOU J, ZHANG Z H. Survey of super-resolution image reconstruction methods. Acta Automatica Sinica, 2013, 39(8): 1202-1213.
[2] BATZM, EICHENSEER A, SEILER J, et al. Hybrid super-resolution combining example-based single-image and interpolation-based multi-image reconstruction approaches. ICIP 2015: Proceedings of the 2015 IEEE International Conference on Image Processing. Piscataway, NJ: IEEE, 2015: 58-62.
[3] LIN Z C, SHUM H Y. Fundamental limits of reconstruction-based super-resolution algorithms under local translation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(1): 83-97.
[4] ZHOU J H, ZHOU C, ZHU J J, et al. A method of super-resolution reconstruction for remote sensing image based on non-subsampled contourlet transform. Acta Optica Sinica, 2015, 35(1): 106-114.
[5] LIAN Q S, ZHANG W. Image super-resolution algorithms based on sparse representation of classified image patches. Acta Electronica Sinica, 2012, 40(5): 920-925.
[6] KEYS R. Cubic convolution interpolation for digital image processing. IEEE Transactions on Acoustics Speech and Signal Processing, 1981, 29(6): 1153-1160.
[7] IRANI M, PELEG S. Super resolution from image sequences. Proceedings of the 10th International Conference on Pat-tern Recognition. Piscataway, NJ: IEEE, 1990: 115 -120.

[8] IRANI M, PELEG S. Improving resolution by image registration. CVGIP: Graphical Models and Image Processing, 1991, 53(3): 231-239.

[9] SCHULTZ R R, STEVENSON R L. Extraction of high-resolution frames from video sequences. IEEE Transactions on Image Processing, 1996, 5(6): 996-1011.

[10] STARK H, OSKOUI P. High-resolution image recovery from im-age-plane arrays, using convex projections. Journal of the Op-tical Society of America, 1989, 6(11): 1715-1726.

[11] DONG C, CHEN C L, HE K, et al. Learning a deep convolutional network for image super-resolution, Computer Vision—ECCV 2014, LNCS 8692. Berlin: Springer, 2014: 184 – 199.

[12] DONG C, CHEN C L, TANG X. Accelerating the super-resolution convolutional neural network. Proceedings of the 2016 14th European Conference on Computer Vision. Berlin: Springer, 2016: 391-407.

[13] LEDIG C, THEIS L, HUSZAR F, et al. Photo-realistic single im-age super-resolution using a generative adversarial network / CVPR 2017: Proceedings of the 2017 IEEE Conference on Com-puter Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2017: 105-114.

[14] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets. NIPS’14: Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2014: 2672-2680.

[15] RADFORD A, METZ L, CHINTALA S. Unsupervised representa-tion learning with deep convo-lutional generative adversarial net-works. 2017-07-2. http://www.xuehu.baidu.com/s?wd=paperuri%3A%289cf80c74b1e5d54b1872f902dabf8124%29&filter=se_long_sign &tn=SE_xueshu_source &sc_vurl=http%3A%2F%2F2Faxiv.org%2Fpdf%2F1511.06434&ie=utf-8&sc_us=7608204741076306321.

[16] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting [C. CVPR 2016: Pro-ceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2016: 2536 -2544.

[17] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial net-works [EB / OL]. [2017-12-09]. http://sсе.tоngji.edu.cn/ yingshen / course / PR2017Fall / read-ings / Image-To-Image_Translation. Pdf.

[18] HE K, ZHANG X, REN S, et al. Deep residual learning for im-age recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Computer, 2016: 770-778.

[19] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution. CVPRW 2017: Proceed-ings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington, DC: IEEE Computer Socie-ty, 2017: 1132-1140.

[20] NAH S, KIM T H, LEE K M. Deep multi-scale convolutional neural network for dynamic scene deblurring [EB / OL].[2017-06-09]. http://www.Openaccess. Thecvf. com / content_cvpr_2017 / papers / Nah_Deep_Multi-Scale _Convolutional_CVPR_2017_paper.pdf.

[21] Jiahui Yu et al. “Wide Activation for Efficifient and Accurate Image Super - Resolution”. In: arXiv preprint arXiv:1808.08718 (2018).

[22] SIMONYAN K, ZISSE R MAN A. Very deep convolutional net-works for large-scale image recognition, 2016-12-09. http://www.x-algo.cn / wp-content / uploads /2017 /01 / VERY-DEEP-CONVOLUTIONAL-NETWORK- SFOR-LARGE- SCALE- IMAGE- RECOGNI- TION. Pdf.

[23] SUN J, SUN J, XU Z. Image super-resolution using gradient pro-file prior. CVPR 2008: Proceed-ings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Computer, 2008: 1-8.
[24] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolu-tions. CVPR 2015: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 1-9.

[25] Martin Arjovsky. “Wasserstein GAN”. In: arXiv preprint arXiv:1701.07875 (2017).

[26] TONG Y B, ZHANG Q S, QI Y P. Image quality assessing by combining PSNR with SSIM. Journal of Image and Graphics, 2006, 11(12): 1758-1763.

[27] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.