ATLANTIS PRESS

Advances in Intelligent Systems Research (AISR), volume 155

2018 International Conference on Advanced Control, Automation and Artificial Intelligence (ACAAI 2018)

Single Image Super-Resolution Based on Improved WGAN

Lei Yu[1,*], Xiang Long[2] and Chao Tong[3]

[1]School of Computer Science and Engineering, Beihang University, Beijing, China

[2]State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China

[3]School of Computer Science and Engineering, Beihang University, Beijing, China

[*]Corresponding author

***Abstract*—SRGAN has successfully applied the Generative Adversarial Network to the single image super-resolution reconstruction, which has achieved good results. But the loss function based on feature space in SRGAN objectively sacrifices the pursuit of high peak signal-to-noise-ratio (PSNR), which is the result of a tradeoff. At the same time, Improved Training of Wasserstein GANs makes the training process more stable. We redesign the SRGAN, using VGG16 network for feature extraction, setting discriminator network's working space as feature space, and adding the loss function based on the mean square error of pixel space, then gain more details and high PSNR in the reconstruction at the same time. We use the design of WGAN-GP for reference to make the training more stable.**

Keywords-Super-resolution; WGAN-GP; VGG

I. INTRODUCTION

Super-resolution is the reconstruction of a high-resolution image using information from one or more low-resolution images [1]. It well solved the problem of low resolution caused by sensor, make up for the deficiency of the hardware, and also overcome the influence from fuzziness, noise and other factors in the process of imaging. Therefore, super-resolution has wide application prospects of remote sensing, medical imaging and security monitoring.

SRCNN [2] creatively map the steps of traditional learning method in the super-resolution to the convolutional neural network and use CNN to complete the task. It's the first time that deep learning model is applied to SISR problem. Generative Adversarial Nets (GAN) is one of the most popular deep learning model in recent years. SRGAN [3] successfully applies GAN in SISR problem, and puts forward a kind of new loss function, which can achieve more details in the super-resolution with a 4× down-sampling factor compared with the loss function based on mean-square error (MSE). But SRGAN sacrifices peak signal-to-noise-ratio (PSNR) for details. At the same time, we can know that GAN has problems of training and the gradient disappearing always appear according to WGAN and WGAN-GP [4,5] as well as our own practice. SRGAN, based on GAN, also has the same problem.

Compared with SRGAN, we redesign the SR network model based on WGAN-GP, adjust the working space of the Discriminator Network as feature space, and append the MSE loss function based on pixel space for Generator Network. Compared with SRGAN, we improved the PSNR of generated image while getting the details.

II. RELATED WORK

In general, the method of super-resolution can be divided into three main categories:

A. Super-Resolution Based on Interpolation.

This method [6,7] which is based on the technique of frame images interpolation obtains the pixel values of high resolution image on non-uniform spacing sampling points by estimating the relative motion between frames. Then high resolution image is obtained by non-uniform interpolation.

B. Super-Resolution Based on Reconstruction.

The method is mainly divided into frequency domain method and spatial domain method. The frequency domain method [8] is to improve the spatial resolution of images by eliminating spectrum aliasing in the frequency domain. The spatial domain method is to model space element affecting the imaging of low resolution image (including optical blur, motion blur, etc.), including non-uniform sampling interpolation, the iterative back projection method, the method based on probability and convex set hybrid projection method and the MAP/POCS algorithm, etc.

C. Super-Resolution Based on Learning.

This is a method [9] using machine learning techniques to learn a priori map between low resolution and high resolution image blocks by the given example as the basis of super-resolution reconstruction. The preceding two kinds of method just consider the image as a signal, but the algorithm based on the learning pays more attention to the understanding of image content and structure, which makes use of the prior knowledge of problems and related data to provide stronger constraints. It can often get better results.

Compared with the other two methods, the third learning method can recover more abundant information. SRCNN uses the convolution neural network model to realize the super-resolution of single image, and opens a new path for the SISR problem. SRGAN firstly applies the Generative Adversarial Network to SISR problem, which provides a new idea for the design of loss function.

GAN is one of the most popular deep learning model in recent years [10]. It solves the prominent problem in unsupervised learning: given a batch of samples, training a system to produce similar new samples. Compared with other generative models, the images GAN generated are more realistic and natural. But GAN itself also has some disadvantages. GAN is normally hard to train and WGAN and WGAN-GP also prove that GAN has the problem of vanishing gradient in great degree through strict mathematical derivation. Therefore, WGAN put forward the corresponding modification, such as remove sigmoid from the last layer of discriminator, don't take logarithmic function in generator and discriminator, truncate the absolute value of every discriminator's parameter to no more than a fixed constant c, etc. The same year, WGAN-GP was proposed as the improvement scheme based on WGAN. Although WGAN reduces the difficulty of GAN's training, but it's still difficult to converge, and the effect of image generated is poorer than DCGAN. WGAN-GP penalize the norm of the gradient of the critic with the new one to its input instead of weight clipping to solve the above problems.

In previous SR studies, most researchers use MSE between generated image and original image as loss function which is based on pixel space [11]. It is obvious that such a loss function will result in a higher PSNR value for the generated image. SRGAN uses the loss function based on the feature space. It adopts the first 5 convolution layers of VGG19 network [12] as feature extractor and calculate the MSE loss between features of generated image and original image as content loss. It gains more details in generated image in the 4× SISR task.

III. OUR WORK

A. Network Architecture

In SISR task, we need to make low resolution image I^{LR} as input to obtain the high resolution image I^{SR}. We use I^{HR} for the original high resolution image, it is only used in the stage of training. In the training phase, we get I^{LR} from I^{HR} with down-sampling factor r. For one image whose channel number is C, we described the size of I^{LR} as $W \times H \times C$, so the size of corresponding I^{SR} and I^{HR} is $rW \times rH \times C$.

By reference to the design of GAN's network, we also set up generator and discriminator networks. In the design of generator, we adopt the design of the residual convolution neural network for faster convergence. Each residual block contains two convolution layers. The enlargement and reconstruction part of the generator adopts periodic shuffle for sub-pixel which refers to ESPCN, and generates I^{SR} based on the preceding feature map. In the design of the discriminator, we use for reference SRGAN and WGAN-GP to change the activation function of the last layer of the network from Sigmoid to Leaky-ReLU. The network structure is shown in the Figure 1 and Figure 2.

In SRGAN, the author adopts a new loss function:

$$l^{SR} = l^{SR}_{VGG} + 10^{-3} l^{SR}_{Gen} \tag{1}$$

$$l^{SR}_{VGG} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}\left(G_{\theta_G}(I^{LR}) \right)_{x,y} \right)^2 \tag{2}$$

The loss function is a result of weighted sum of content loss l^{SR}_{VGG} and adversarial losses l^{SR}_{Gen}. l^{SR}_{VGG} represents the MSE between features extracted from I^{SR} and I^{HR} by the first 5 convolution layers of VGG16 (hereinafter, VGG16). Because of l^{SR}_{VGG}, the image generated will be more realistic and get more abundant details. We improved the network structure of SRGAN. The network structure as shown in Figure 2, VGG16 is used to extract features of image I^{SR} generated by generative model and I^{HR}. And make the extracted features as discriminator's input. In conclusion, we put VGG16 as feature extractor into our model.

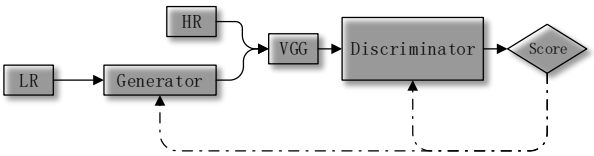

FIGURE I. THE OVERVIEW OF OUR MODEL. THE DISCRIMINATOR SCORE THE FEATURES OF I^{SR} AND I^{HR} RESPECTIVELY

B. Loss Function

The loss function is divided into generator's loss function l_G and discriminator's loss function l_D:

$$l_G = l_{MSE} + 10^{-6} l_{Gen} \tag{3}$$

$$l_D = \frac{1}{B} \sum_{i=1}^{B} \left(D_{\theta_D}\left(\phi(I^{SR})\right) - D_{\theta_D}\left(\phi(I^{HR})\right) \right) + \frac{\lambda}{B} \sum_{i=1}^{B} \left(\left\| \frac{\partial D_{\theta_D}(y)}{\partial y} \right\|_{y = \varepsilon_i \phi(I^{HR}) + (1-\varepsilon_i)\phi(I^{SR})} - 1 \right)^2 \tag{4}$$

We refer to the design of WGAN-GP's loss function, we define l_G by l_{MSE} between I^{SR} and I^{HR}, as while as l_{Gen} produced by discriminator:

$$l_{MSE} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} \left(I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_{x,y} \right)^2 \tag{5}$$

$$l_{Gen} = -D_{\theta_D}(\phi(I^{SR})) \tag{6}$$

At the same time, we define l_D as shown in Equation 4. The precede item refers to the difference of score from discriminator between $\phi(I^{SR})$ and $\phi(I^{HR})$. The aim is to minimize the $D_{\theta_D}\left(\phi(I^{SR})\right)$ and maximize $D_{\theta_D}\left(\phi(I^{HR})\right)$ when minimize l_D. $\phi(X)$ represents the feature extracted by VGG16. The subsequent term of l_D is the penalty item in WGAN-GP, which is used to constrain discriminator.

IV. EXPERIMENTS

A. Data sets and Evaluation Index

We experimented on three widely used data sets (Set5, Set14, BSD100), and all of the down-sampling factors were set to 4. The data sets we used can be obtained from SelfExSR [13]. We use PSNR and structural similarity (SSIM) as the evaluation index for super-resolution. For the sake of fairness, all PSNR [dB] and SSIM are calculated based on the y-channel in YCbCr space. All images are cropped based on center and remove a 4-

pixel wide strip from each border. To compare the effects, we do experiments of our model and the results of bicubic, SRCNN and SRGAN are gained from SelfExSR.

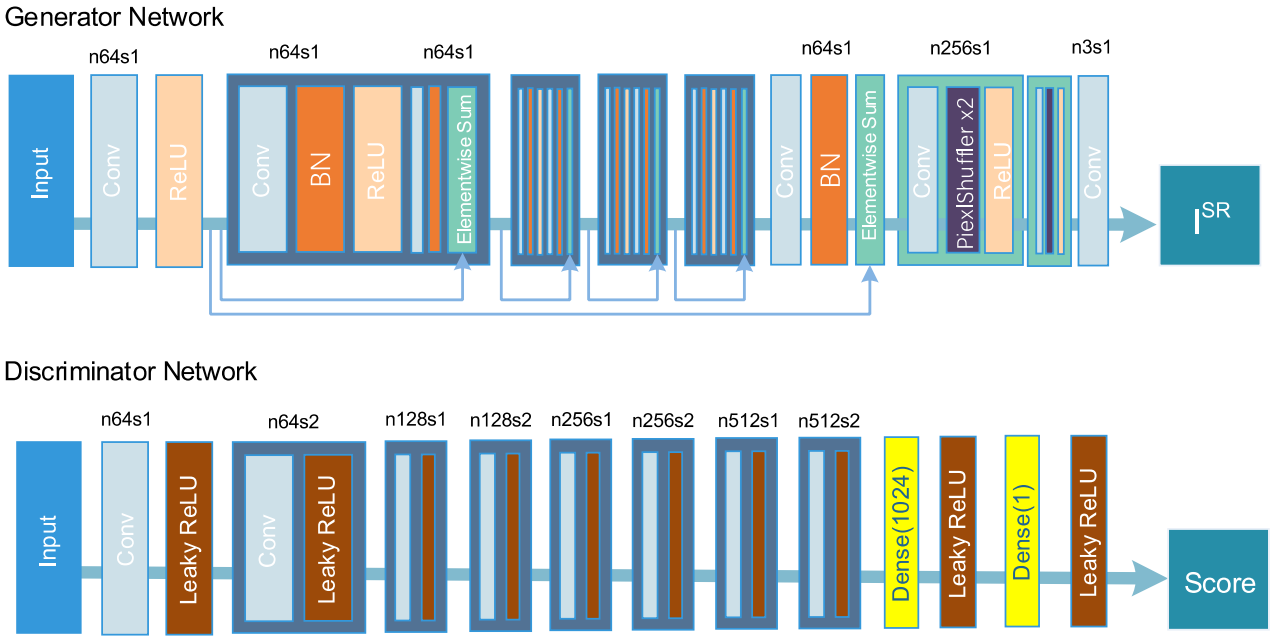

FIGURE II. ARCHITECTURE OF GENERATOR AND DISCRIMINATOR NETWORK WITH CORRESPONDING NUMBER OF FEATURE MAPS (N) AND STRIDE(S) INDICATED FOR EACH CONVOLUTIONAL LAYER.

SRWGAN-MSE SRWGAN-GP Original HR image

FIGURE III. THE RESULTS OF OUR MODEL JUST BASED ON MSE(SRWGAN-MSE), OUR MODEL(SRWGAN-GP), AND THE GROUND TRUTH

B. Training Details and Parameters

We trained all networks on a NVIDIA GeForce GTX 980 GPU. We create training data set based on PACSCAL VOC2012. In the practical training process, we take the sub image (96 × 96) from PACSCAL VOC2015 as the training data I^{HR} (BGR, C = 3). And get I^{LR} from I^{HR} by the way of bicubic down-sampling, the down-sampling factor r = 4. We take the sample by sliding window to get the sub image and the stride is 10.

The pixel values' scope of input image model is [0, 255], for convenience, we map it to [0, 1]. Therefore, when calculating l_{MSE}, the values' scope is [0, 1] also. In order for l_{Gen} to be comparable with l_{MSE}, we also qualify the output of VGG16 with the limiting factor of 1/12.75. As for the optimizer, we use Adam optimizer for training. The initial learning rate is set to 0.0001, the beta1 is set to 0.9, and the beta2 is set to 0.99.

We train the generator with loss function just base on l_{MSE} firstly, and initialize the entire network with it in order to avoid the local optima.

C. Performance of Our Model

In our experiment, we tested the trained model with Set5, Set14 and BSD100 respectively. For each test data set, we get the performance of bicubic, SRCNN, SRGAN and our model. The results show that our model achieves a higher PSNR and recover more details at the same time. The experimental results are shown in Table 1.

In order to verify the advantage of our model in restoring more details in images, we adjusted the loss function and do experiments with the model just based on MSE. The loss function can be seen in Equation 7. The experimental results show that the image generated by model just based on MSE is

overly-smooth. And our model can perform better in details, making the image more realistic. Figure 3 shows the results of our model just based MSE (SRWGAN-MSE), our model (SRWGAN-GP), and the ground truth.

$$l_G = l_{MSE} \tag{7}$$

D. Additional Experiment

We also test the super-resolution model based on WGAN (SRWGAN). The experimental parameters and network's design are same with the above model. Following WGAN, we just modified the constraint for the parameters of discriminator and limit the number to [-0.01, 0.01]. The number out of range will be clipped. We also pre-train the generator based on l_{MSE} to prevent local optima. During the experiment, we found that the exploding gradient phenomenon is extremely easy to occur on SRWGAN. The model cannot be trained effectively. And it's consistent with the judgment to WGAN from WGAN-GP. After a lot of experiments, the best experimental data obtained finally as shown in the Table 2. We can see that it is not better than the result based on WGAN-GP. The model based on WGAN is hard to be trained well.

TABLE I. THE EXPERIMENTAL RESULT ON DIFFERENT DATA SET

Set5	bicubic	SRCNN	SRGAN	ours
PSNR	28.43	30.07	29.40	**30.76**
SSIM	0.8211	0.8627	0.8472	**0.8791**
Set14	bicubic	SRCNN	SRGAN	ours
PSNR	25.99	27.18	26.02	**27.48**
SSIM	0.7486	0.7861	0.7397	**0.7982**
BSD100	bicubic	SRCNN	SRGAN	ours
PSNR	25.94	26.68	25.16	**26.81**
SSIM	0.6935	0.7291	0.6688	**0.7403**

TABLE II. THE RESULT ON SRWGAN AND SRWGAN-GP

	Set5	**Set14**	**BSD100**
bicubic	28.43	25.99	25.94
SRWGAN	28.28	26.13	25.84
SRWGAN-GP	**30.76**	**27.48**	**26.81**

V. CONCLUSION

We try to improve the structure of SRGAN, take VGG16 as a feature extractor on I^{LR} and I^{HR} for feature extraction to be the input of discriminator. We set discriminator's workspace as feature space and achieve more details in the process of reconstruction. In the design of loss function, we refer to WGAN-GP, and make the training more stable and easy to converge. The results of experiment show that the design of network structure has achieved good results in the 4-fold super-resolution task.

Due to time pressures, we neither adopt a deeper network structure, nor add training skills such as Dropout. In addition, according to GAN's theory, we should set the size of batch to be larger to achieve better results. However, limited by device performance, we set up a smaller size. In subsequent experiments, the network structure can be improved, and better results can be obtained.

ACKNOWLEDGMENT

The work is supported in part by the research fund of the State Key Laboratory of Virtual Reality Technology and Systems.

REFERENCES

[1] Glasner D, Bagon S, Irani M. Super-resolution from a single image[C]// IEEE, International Conference on Computer Vision. IEEE Xplore, 2009:349-356.

[2] Dong C, Chen C L, He K, et al. Learning a Deep Convolutional Network for Image Super-Resolution[J]. 2014, 8692:184-199.

[3] Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[J]. 2016.

[4] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved Training of Wasserstein GANs[J]. 2017.

[5] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. 2017.

[6] Rajan D, Chaudhuri S. Generalized interpolation and its application in super-resolution imaging[J]. Image \& Vision Computing, 2001, 19(13):957-969.

[7] Anbarjafari G, Demirel H. Image Super Resolution Based on Interpolation of Wavelet Domain High Frequency Subbands and the Spatial Domain Input Image[J]. ETRI Journal, 2010, 32(3):390–394.

[8] Vandewalle P, Süsstrunk S, Vetterli M. A frequency domain approach to registration of aliased images with application to super-resolution[J]. Eurasip Journal on Advances in Signal Processing, 2006, 2006(1):1-14.

[9] Tang Y, Yuan Y, Yan P, et al. Single-image super-resolution based on semi-supervised learning[C]// Pattern Recognition. IEEE, 2012:52-56.

[10] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// International Conference on Neural Information Processing Systems. MIT Press, 2014:2672-2680.

[11] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2010, 19(11):2861-2873.

[12] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[13] SelfExSR[EB/OL]. https://github.com/jbhuang0604/SelfExSR.