Deep Against Net Image Super-Resolution Reconstruction Algorithm Based on W Distance

Yang Liu

School of Computer Science, Guangdong University of Technology, China.

906328986@qq.com

Abstract. This paper proposes an image super-resolution algorithm based on the condition of Wasserstein distance generation against network, aiming at improving the quality of reconstructed images and improving the condition against the stability of neural network in image super-resolution research. Wasserstein distance algorithm is used to solve the instability problem of the traditional GAN network generator and make the model more stable. In the test sets such as Set5 and Set14, the three evaluation indexes PSNR, SSIM, and IFC of the SRWCGAN algorithm are superior to the VDSR algorithm, and the convergence speed and stability are better than those without W-distance.

Key words: super-resolution; GAN; Wasserstein distance.

INTRODUCTION

In 2014, Ian Good fellow proposed Generative Adversarial Networks (GAN) [1], which has now become a hot research topic in deep learning. In [2], the Deep Convolutional Generative Adversarial Networks (DCGAN) method is proposed, which introduces the convolutional network into the generative model to perform unsupervised training and uses the powerful feature extraction of the convolutional network. Ability to improve the learning effect of the generation network. The network generator uses a deconvolution; the literature [3] proposes a Conditional Generative Adversarial Networks (CGAN) method. A condition-constrained GAN is proposed. Conditional variable y is introduced in the modeling of the generating model (D) and the discriminative model (G). Using the additional information y to add conditions to the model can guide the data generation process. It can be seen that CGAN is an improvement of a purely unsupervised GAN into a supervised model, which effectively solves the problem that if the data picture is large, simple GAN is not controllable. However, CGAN still has the problem that the model training does not converge and the gradient disappears.

In this paper, based on CGAN research object, an image super-resolution reconstruction algorithm based on WCGAN deep residual network is proposed. The algorithm mainly includes the following features: (1) The generator input is modified to a low-resolution image instead of a random noise input, without pre-processing the image The formation of an end-to-end supervisable network model (2) The generator uses a deep residual network ResNet [4] to effectively improve the network's learning ability, reduce training time, deepen feature learning, and improve image reconstruction quality; (2) Loss function The optimization uses Wasserstein distance instead of KL divergence and JS divergence to improve the stability of the network. (4) It can handle complex structures such as three-channel images. The objective quantification results and visual effects of images in test sets such as Set5 and Set14 are better than VDSR.

RELATED WORK

The traditional GAN is an unsupervised model whose generator is to make completely random noise z training to generate specific images that are not related to it, but this results in the network being too free and difficult to train.

The CGAN algorithm can effectively solve the problem of generating difficult network training. The change of CGAN is mainly to add the instructive label y to the GAN generator and discriminator (the algorithm added by the generator is a low-resolution image). Supervised end-to-end model. The output data of the generator and the real input data of the discriminator will have a large overlap, so that the network training will become directional, training will be easier and the effect will be better.

The Wasserstein distance, also called the Earth-Mover distance (EM distance), is defined as follows:

$$W\,(P1,\,P2) = \inf\gamma\sim\Pi\,(P1,\,P2)\,E(x,y)\sim\gamma[\|x\text{-}y\|] \tag{1}$$

In (1), Π (P1, P2) is a set of all possible joint distributions of P1, P2 combined. Conversely, the distribution of edges for each distribution of Π (P1, P2) is P1, P2. $E(x,y)\sim\gamma[\|x\text{-}y\|]$ can be understood as the consumption required to move the P1 pile of "sand soil" to P2 "location" under the path planning of γ, and W(P1, P2) is the minimum consumption under the optimal plan, so it is also called the bulldozing distance.

According to the discriminator loss defined by the original GAN, under the optimal discriminator, we can transform the generator loss defined by the original GAN equivalent to minimize the JS dispersion between the real distribution P(x) and the generated distribution P(z). Degree, JS divergence is a constant when there is no overlap or overlap between P(x) and P(z). Eventually it will cause the generator gradient to disappear. This is why the generator network is unstable.

To solve the above problems, Wasserstein distance is superior to KL divergence and JS divergence. It has superior smoothness characteristics. Wasserstein distance can still reflect their distance even if there is no overlap between real samples and generated samples. The JS divergence is constant in this case, and the KL divergence may be meaningless. Therefore, the EM distance is used to design the loss function to replace the KL divergence and JS divergence used by the original GAN, so that the problem of disappearing the gradient can be solved.

ALGORITHM IN THIS PAPER

The Wasserstein CGAN (WCGAN) algorithm's activation function, loss function, and optimization method are as follows:

1. Activation function: The activation function of each layer of the WCGAN generator and discriminator uses the Parametric Rectified Linear Unit (PReLU). Compared to the ReLU activation function, PReLU only adds a very small number of parameters and can be more accurate. effect.

2. Loss function: The last layer of the discriminator does not need sigmoid [24]: The original CGAN discriminator performs the true and false two-category task, and the last layer uses the sigmoid activation function. The WCGAN discriminator does approximate EM distance and belongs to the regression task. Therefore, the last sigmoid of D needs to be removed. Therefore, the discriminator and generator loss functions are as follows:

$$\text{D\_loss} = -\,E_{x\sim pdata(x,y)}D(x,y) + E_{z\sim pz(z)}[D\,(G\,(x)\,)\,] \tag{2}$$

$$\text{G\_loss} = -E_{z\sim pz(z)}D\,(G\,(x)\,) \tag{3}$$

In the above two formulas, x and y represent the input low-resolution image and the generated super-resolution image, respectively. From this, it can be seen that the calculation of the loss function is more convenient than the calculation of the cross-entropy of the loss function of the traditional GAN, and the operating speed is improved. During training, the smaller the value of the loss function is, the smaller the EM distance between the true distribution and the generated distribution is, and the better the WCGAN training is.

4. Use RMSProp optimization algorithm: Compared to Adam optimization algorithm, RMSProp is more stable for WCGAN. If Adam is used, the loser of the discriminator sometimes collapses. When it collapses, the cos value of the angle between the update direction and the gradient direction given by Adam becomes negative, and the direction of the update and the direction of the gradient change. This means that the loss gradient of the device is unstable, so it is not suitable for momentum-based optimization algorithms such as Adam.

EXPERIMENT

The training of neural networks requires a lot of matrix operations, and the graphics processing unit (GPU) is more suitable for such operations than the central processing unit (CPU), so the training of SRWCGAN is implemented on the GPU. The experimental environment is Ubuntu 16.04 operation system, GPU is configured as NVDIA GeForce GTX1080., and training is done on deep learning platform TensorFlow.

The experimental training data set is 91-image dataset (a total of 91 pieces) +General-100 dataset (100 pieces) + BSD100 (total 100 pieces), and a total of 291 different color images. The training set data test data set is Set5+Set14 + BSD200. In order to enhance the generalization ability of the model, we carry out data enhancement through random clipping and random horizontal reverse input images (for example, the input image is 64 x 64, the cutting step is 16, and then 4156 pictures can be obtained by reversing). In order to compare with the existing image super-resolution algorithm, this paper only trains and tests the Y channel (YCrCb color space, Y for luminance channel).

The current image super resolution objective quantization evaluation criteria include the peak signal to noise ratio (Peak Signal to Noise Ratio, PSNR). The higher the PSNR value between the two images, the better the image super resolution recovery is, the better the structure similarity (Structural Similarity Index Method, SSIM), the higher the similarity of the two images, the higher the above, the higher the above, the higher the similarity of the two images, through the above the comparison of two evaluation indicators can quantitatively evaluate the superiority of the algorithm. At the same time, we evaluate the network advantage by evaluating the loss value during training.

This algorithm also increases the subjective evaluation of the human eye observation reconstruction image output as the evaluation standard, training set and test set, such as the above, the different is to the RGB three channel color map training and testing.

CONCLUSION

Considering the different hardware configuration of the system, this paper does not compare the training time. In this chapter, we use 152-layer depth residual SRWDGAN to compare with existing four-super-resolution algorithms including Bicubic, A+[7], SRCNN[5], and VDSR[8]. Each algorithm includes this algorithm to magnify the image by two times. 3 times, 4 times to test, the test set is Set5, Set14, BSD200. The experimental results are shown in Table 3. The PSNR and SSIM in this algorithm are higher than the other four algorithms. And the experimental results of the 152th layer in this chapter are slightly better than the effects of the same layer in the third chapter.

TABLE 1. Comparison of Average PSNR (dB)/SSIM for Different Image Super Resolution Algorithms

Datasets	Scale	Bicubic	A+ [7]	SRCNN [5]	VDSR [8]	OURS
Set5	×2	33.66 / 0.9299	36.54 / 0.9544	36.66 / 0.9542	37.53 / 0.9587	37.85 / 0.9611
	×3	30.39 / 0.8682	32.58 / 0.9088	32.75 / 0.9090	33.66 / 0.9213	34.32 / 09353
	×4	28.42 / 0.8104	30.28 / 0.8603	30.48 / 0.8628	31.35 / 0.8838	31.92 / 0.9037
Set14	×2	37.53 / 0.9587	32.28 / 0.9056	32.42 / 0.9063	33.03 / 0.9124	34.13 / 0.9226
	×3	27.55 / 0.7742	29.13 / 0.818	29.28 / 0.8209	29.77 / 0.8314	30.76 / 0.8521
	×4	26.00 / 0.7027	27.32 / 0.7491	27.49 / 0.7503	28.01 / 0.7674	29.12 / 0.7991
BSD200	×2	29.56 / 0.8431	31.21 / 0.8863	31.36 / 0.8879	31.90 / 0.8960	33.27 / 0.9108
	×3	27.21 / 0.7385	28.29 / 0.7835	28.41 / 0.7863	28.82 / 0.7976	30.03 / 0.8127
	×4	25.96 / 0.6675	26.82 / 0.7087	26.90 / 0.7101	27.29 / 0.7251	28.49 / 0.7717

The main purpose of this chapter's algorithm is to solve the problem of instability of GAN generator. The generator has better effect with 152-layer residual network. It trains W-distance 152-deep residual network and W-distance-free network. Training 150,000 observations the evolution of the generator's loss value during the training process is compared with the case where no W-distance is added. As shown in the figure, the red line represents the change in generator loss during the training of this chapter. The green line is the experimental result in the third chapter. In the training process, the value of the loos value of the generator can be used to determine whether the network training is good or bad, and the model can be trained as the value of the generator's loss gradually decreases.
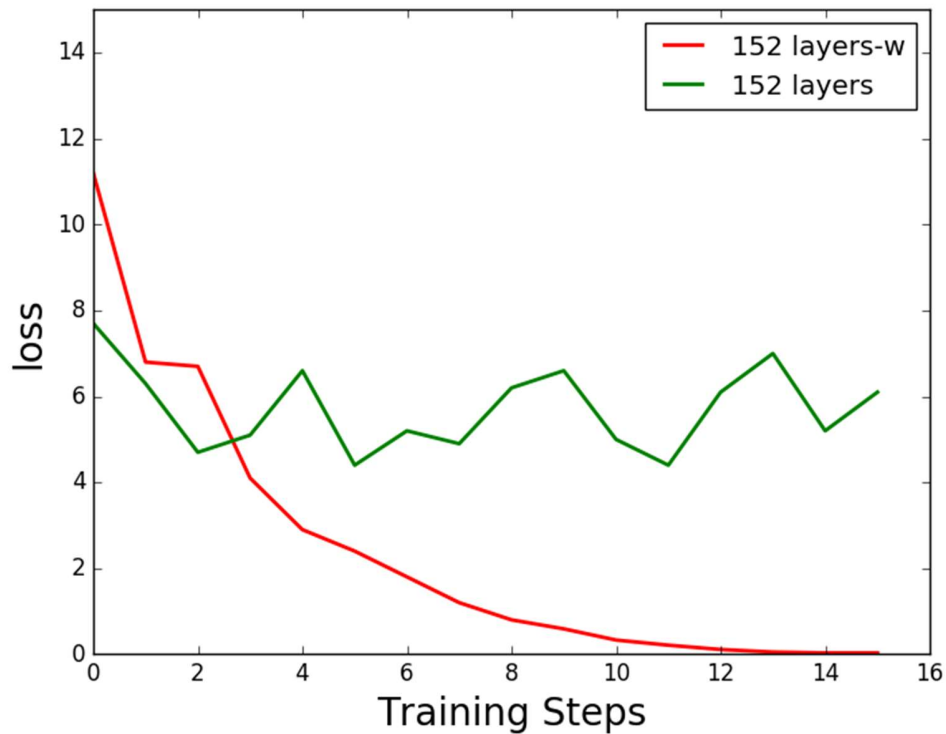
FIGURE.1. Loss value

ACKNOWLEDGMENTS

REFERENCES

1. Goodfellow I, Pougetabadie J, Mirza M, et al. Generative Adversarial Nets[J]. Advances in Neural Information Processing Systems, 2014:2672-2680.
2. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. Computer Science, 2015.
3. Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.
4. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
5. Dong C, Loy C, He K, et al. Image Super-Resolution Using Deep Convolutional Networks. [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 38(2):295-307.
6. Martin D, Fowlkes C, Tal D, Jitendra Malik J. Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(5): 530-549.
7. Timofte R, De Smet V, Van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution. Asian Conference on Computer Vision, Springer, 2015, 111-126.
8. Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// Computer Vision and Pattern Recognition. IEEE, 2016:1646-1654.