ORIGINAL PAPER

Inter-frame video image generation based on spatial continuity generative adversarial networks

Tao Zhang[1] · Peipei Jiang[1] · Meng Zhang[1,2]

Abstract

This paper proposes a method for generating inter-frame video images based on spatial continuity generative adversarial networks (SC-GANs) to smooth the playing of low-frame rate videos and to clarify blurry image edges caused by the use of traditional methods to improve the video frame rate. Firstly, the auto-encoder is used as a discriminator and Wasserstein distance is applied to represent the difference between the loss distribution of the real sample and the generated sample, instead of the typical method of generative adversarial networks to directly match data distribution. Secondly, the hyperparameter between generator and discriminator is used to stabilize the training process, which effectively prevents the model from collapsing. Finally, taking advantage of the spatial continuity of the image features of continuous video frames, an optimal value between two consecutive frames is found by Adam and then mapped to the image space to generate inter-frame images. In order to illustrate the authenticity of the generated inter-frame images, PSNR and SSIM are adopted to evaluate the inter-frame images, and the results show that the generated inter-frame images have a high degree of authenticity. The feasibility and validity of the proposed method based on SC-GAN are also verified.

Keywords GAN · Adversarial training · Spatial continuity · Adam · Inter-frame image generation

1 Introduction

The technology to improve the frame rate of video is generally called motion estimate [1]/motion compensation [2] (ME/MC), and its basic idea is inserting a generated image frame between two consecutive frames in the video sequence to increase frame rate. There are two methods to improve frame rate, namely non-motion compensation algorithm and motion compensation algorithm [3], and non-motion compensation algorithm generally is realized as frame repetition and frame averaging [4, 5]. Featured with lower complexity, non-motion compensation algorithm is easy to integrate into the product, but that does not work well in non-stationary video sources. Block matching algorithm is utilized for ME/MC [6] technology in the motion estimation to obtain motion vector (MV) [7] which is calculated as 'motion compensation frame.' The motion compensation frame is inserted into two consecutive frames to improve the video frame rate. ME/MC technology eliminates jitter and smear and enhances clarity in video playing, but will cause the vagueness in the image edge in movement. Therefore, on the premise of ensuring image clarity, improving the frame rate of video has become an urgent problem.

Great progresses have been made for inter-frame image generation in the field of deep learning. In [8], Guo proposed the image generation model, which can predict the future frames according to the past frames and the current frames in the video. [9] proposed a method to generate an optimal low dynamic range image based on neural network. Hou et al. [10] based on convolution neural network, proposed a self-learning method to enhance frame rate. Gucan et al. [11] proposed a deep convolution neural network, which learns the motion of objects in video by convolution network to generate the middle frame of the video. Although these studies have contributed to the generation of inter-frame images, the quality and authenticity of the generated images are not ideal.

As a new image generation algorithm, generative adversarial networks (GANs) are able to generate face images [12], digital images, and other objects, translate text into image,

✉ Tao Zhang
  zhtao@ysu.edu.cn

1   School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei, China

2   Neusoft Software Co., Ltd., Qinhuangdao, Hebei, China

complete semantic annotation, and generate high-resolution images from low resolution images [13]. However, GAN has been less researched on inter-frame image generation at present. The most typical one among those researches, proposed by Mathieu et al., is a deep multi-scale video prediction method which is beyond the mean square error and can predict the next frame of the video [14]. It is the first application of the antagonistic idea in GAN into the field of video research, which makes GAN play an important role in the video research and promotes the development of GAN in the field of video research. However, the poor quality of predicted frames generated in the experiment often leads to distortion or blurring of moving objects.

In order to solve these problems, this paper proposes a new method to improve the frame rate based on GAN, that is, to generate inter-frame images based on SC-GAN. Inter-frame images, which are generated by trained SC-GAN model, are inserted into the corresponding two frames to improve the frame rate. It provides a new method to improve the frame rate of video.

2 Basic principle of GAN

GAN, whose basic idea originates from 'zero-sum game' in game theory, consists of a generator and a discriminator [15]. Both the generator and the discriminator use the neural network [16]. A random noise signal z is input into the generator to produce $G(z)$. The discriminator takes the real data x and the generated sample data $G(z)$ as inputs, and the output is the probability of the discriminator treating $G(z)$ as real data. This probability is used to judge the quality of the generated model.

$$\min_{G} \max_{D} V(D, G) = E_{x \sim P_{\text{data}}(x)}[\log D(x)] \\ + E_{z \sim P_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

GAN model optimization is a 'minimax game' process, and Eq. (1) indicates its objective function [17]. The discriminator aims at maximizing the D-value between $D(x)$ and $D(G(x))$, that is, maximizing the $\log D(x)$, to distinguish between the sample data and real data during the training. The generator tries to deceive the discriminator as far as possible and minimize the $\log(1 - D(G(x)))$, which is to maximize the loss of the discriminator. The generator implicitly defines a data distribution P_g, and the real-data distribution is defined as P_{rd}. After the adversarial training of the discriminator and the generator, the optimal result is that the discriminator cannot distinguish whether the sample is a real sample or a generated, that is $P_g = P_{\text{rd}} = 0.5$. At this point, it can be considered that the generator has learned the data distribution of the real sample.
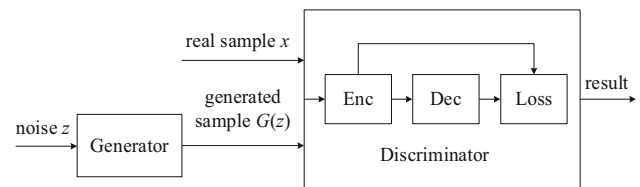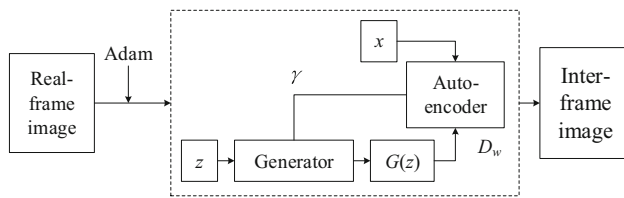

Fig. 1 Image generation model of SC-GAN

3 Inter-frame image generation

Based on the improved GAN model, an auto-encoder, first proposed in EBGAN [18], is used as a discriminator in this paper. While typical GANs try to match data distributions directly, this paper aims to match auto-encoder loss distributions using a loss derived from the Wasserstein distance [19]. The equilibrium term γ is introduced to balance the discriminator and the generator. The dataset used in this paper consists of continuous video frame images, and the images are spatially continuous. Therefore, the generated model can generate inter-frame images with the input of two continuous frames images.

3.1 Model frame

Figure 1 shows our image generation model based on the typical GAN model. We use the auto-encoder which consists of encoder and decoder as a discriminator. Encoder converts the input signal into the encoded signal through the encoding function, and decoder converts the encoded signal into the output signal through the decoding function. An auto-encoder is a neural network that reproduces the input signal as much as possible. To accomplish this reproduction, the auto-encoder must capture the most important features that represent the input signal. The generator input is n-dimensional random noise z, and then, the generator maps z to the image space to get the generated sample $G(z)$. The real samples x are input into the discriminator, and the image features are acquired through encoder subsampling. Then, the image features are mapped into image space through decoder's up-sampling process, and the distribution of the generated sample data is reconstructed, and the reconstructed loss function is obtained. The reconstructed loss distribution belongs to the same distribution as the real sample data distribution. By optimizing the distance between the two loss distributions, the generation model is optimized. In the experiment of inter-frame image generation, due to the spatial continuity between consecutive video frames, we use Adam to find an optimal data distribution between two consecutive frames and map it to the image space through the generated model to get the final inter-frame image. The frame diagram of the inter-frame image generation is shown in Fig. 2.

Fig. 2 SC-GAN inter-frame image generation framework

As is shown in Fig. 2, the SC-GAN model uses an auto-encoder as a discriminator and introduces a hyperparameter γ to keep the balance between the generator and discriminator. The real samples x are input to the discriminator to obtain the reconstructed loss distribution $\Gamma(x)$, and the random noise z is input to the generator to generate the data $G(z)$. Then, the $G(z)$ is input to the discriminator to obtain the reconstructed loss distribution $\Gamma(G(z))$. The parameters of the model are adjusted to minimize the value of D_w through the Wasserstein distance $W(\Gamma(x), \Gamma(G(z)))$ between the two distributions $\Gamma(x)$ and $\Gamma(G(z))$.

$$D_w = W(\Gamma(x), \Gamma(G(z))) \tag{2}$$

At the same time, in order to ensure the stability of the training process and the diversity of the generated samples, the hyperparameter γ is introduced to balance the generator and discriminator.

3.2 Wasserstein distance

Wasserstein distance, also known as earth-mover (EM) distance, between the auto-encoder loss two distributions of real and generated samples is computed. Wasserstein distance is defined as:

$$W(P_1, P_2) = \inf_{\lambda \sim \prod(P_1, P_2)} E_{(x,y) \sim \lambda}[||x - y||] \tag{3}$$

where $\prod(P_1, P_2)$ denotes the set of all joint distributions of the two distributions $\lambda(x, y)$ whose marginals are P_1 and P_2, and for each joint distribution λ, a pair of samples x and y is obtained from the sample of $\lambda(x, y)$, and the distance between these two samples is $||x - y||$. In the joint distribution of λ, the expectation for the distance is $E_{(x,y)}[||x - y||]$. In all joint distributions, the lower bound of the sample's expectation on distance is the Wasserstein distance.

In the training process of the typical GAN model, the generated data distribution matches the distribution of real samples directly, so it is difficult to learn all the key features of the sample data, and convergence speed is relatively slow. In the SC-GAN model, the real sample data are reconstructed by encoder and decoder, and the reconstructed loss distribution $\Gamma(x)$ is obtained. If two data distributions are identical, their loss distributions should also be same. The Wasserstein

distance between $\Gamma(x)$ and $\Gamma(G(z))$ can reflect the difference between the two distributions so that the generated model can be further optimized through the difference.

3.3 Hyperparameters γ

In this paper, we define two normal distributions $\eta_1 = N(m_1, c_1)$ and $\eta_2 = N(m_2, c_2)$, with the means $m_1, m_2 \in \mathbf{R}^p$, and the covariances $c_1, c_2 \in \mathbf{R}^{p \times p}$. According to the Wasserstein formula, the squared Wasserstein distance between the two normal distributions is defined as:

$$W(\eta_1, \eta_2)^2 = ||m_1 - m_2||_2^2 + \text{trace}(c_1 + c_2 - 2(c_2^{1/2} c_1 c_2^{1/2})^{1/2}) \tag{4}$$

In Eq. (4), trace($\bullet$) indicates the trace. In the case of $p=1$, Eq. (4) can be simplified to:

$$W(\eta_1, \eta_2)^2 = ||m_1 - m_2||_2^2 + c_1 + c_2 - 2(c_1 c_2)^{1/2} \tag{5}$$

Wasserstein distance of the discriminator reconstruction loss distribution is optimized in order to optimize the model. Therefore, as long as Eq. (5) satisfies the monotonicity, that is, $\frac{c_1 + c_2 - 2(c_1 c_2)^{1/2}}{||m_1 - m_2||_2^2}$ is constant or monotonically increasing, the squared Wasserstein distance between two distributions can be optimized. This can simplify the problem to:

$$W(\eta_1, \eta_2)^2 \propto ||m_1 - m_2||_2^2 \tag{6}$$

The loss distribution of the generated sample data through the discriminator is $\Gamma(G(z))$. The loss distribution of the real sample data through the discriminator is $\Gamma(x)$. And when the two distributions satisfy $m_1 = m_2$ during the training, the generator and the discriminator are considered to be in the balanced and stable training status, which is shown in Eq. (7).

$$E[\Gamma(x)] = E[\Gamma(G(z))] \tag{7}$$

However, considering the optimization of Wasserstein distance between the real sample data and the generated sample data, when $m_1 = m_2$, $\frac{c_1 + c_2 - 2(c_1 c_2)^{1/2}}{||m_1 - m_2||_2^2}$ is close to infinity, and the model cannot be optimized and even collapses. Therefore, this paper introduces a hyperparameter $\gamma \in [0, 1]$ to balance the generator and the discriminator so that one will not win against the other, making the training process of the model more stable.

$$\gamma E[\Gamma(x)] = E[\Gamma(G(z))] \tag{8}$$

Equation (8) shows how parameter γ balances the generator and the discriminator during training. In our model,
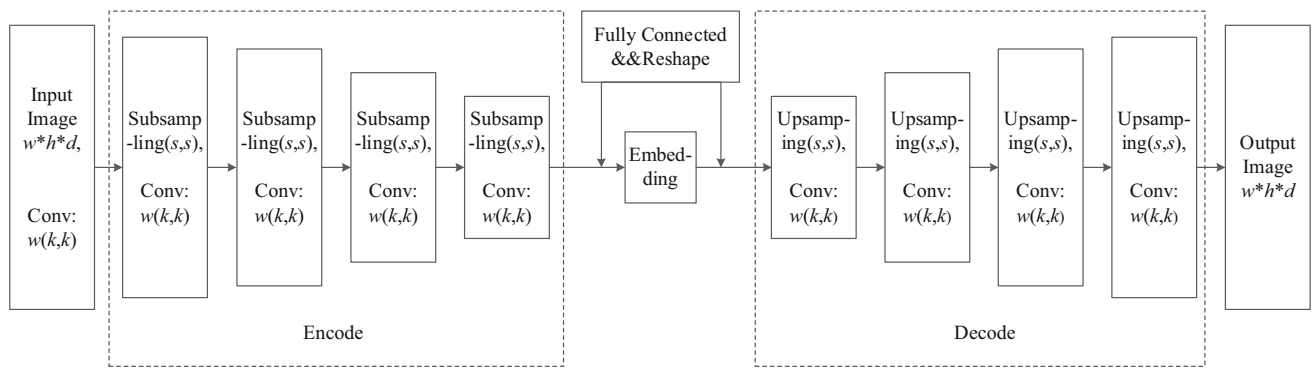
Fig. 3 Discriminator model framework of SC-GAN

the discriminator has two functions: It can encode the real image automatically and distinguish the real image from the generated sample image. The parameter γ can guarantee the stability of the training process of the discriminator and the generator. When the value of γ is low, the discriminator focuses on the auto-encoding of the real image, and therefore, diversity of the generated images will be reduced. Since the model focuses more on the quality of the generated image, within a certain range, the higher the value of γ is, the better the generated model is.

3.4 Spatial continuity

The dataset in this paper consists of every frame of video, which can be considered spatially contiguous. We use the Adam to find an optimal value z_r between the two consecutive images to minimize the e_r value.

$$e_r = ||x1_r - G(z_r)| - |x2_r - G(z_r)|| \tag{9}$$

In Eq. (9), $x1_r$ and $x2_r$ correspond to the previous frame image and the next frame image, respectively, and z_r is mapped to the image space to obtain the interpolation result of two consecutive frames. This method realizes the image generation between the video frames. At the same time, it proves that the trained generation model does not simply memorize the image, but actually learns the characteristics and content of the image during the training process. Image generation between real images provides a novel way to increase the frame rate of the video.

In this model, instead of matching the data distribution of the samples directly in the typical GAN, an auto-encoder is used as a discriminator, and the samples input is reconstructed by encoder and decoder to obtain a reconstructed loss distribution $\Gamma(x)$. Using the Wasserstein distance to measure the difference between the two distributions $\Gamma(x)$ and $\Gamma(G(z))$, the convergence speed is faster and the quality of the generated image is better than that of the traditional
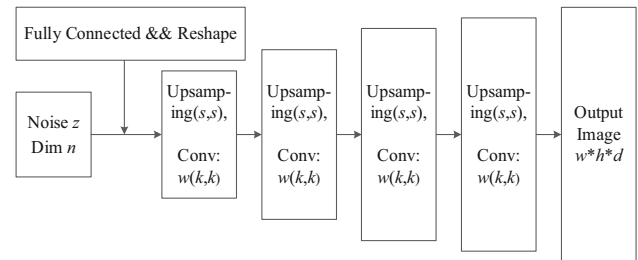
Fig. 4 Generator model framework of SC-GAN

GAN which directly matches the sample data distribution. The introduction of the hyperparameter γ is to balance the generator and the discriminator, which effectively solves the problem that the model in the typical GAN is too free and uncontrollable, making the training model more stable and greatly avoiding the model collapse. The features of consecutive video frame images are spatially continuous. Adam is used to find an optimal value between two consecutive frames, and this value is mapped to the image space to get the inter-frame images.

3.5 Network model

The frame of discriminator and generator in SC-GAN model is shown in Figs. 3 and 4. The process of encoding and decoding in the discriminator is shown in Fig. 3. The input of encoder is the d-dimensional image of $w * h$. The size of the convolution core used in this paper is $k * k$. Full connection layer is used in both the output layer of encoder and the input layer of decoder, where w is the width of the input image, h is the height of the input images, and s is the number of sampling steps. In this model, $w = 64$, $h = 64$, $s = 2$, $k = 3$. In Fig. 4, the input z of the generator is n-dimensional uniformly distributed noise, where $z \in [-1, 1]$.

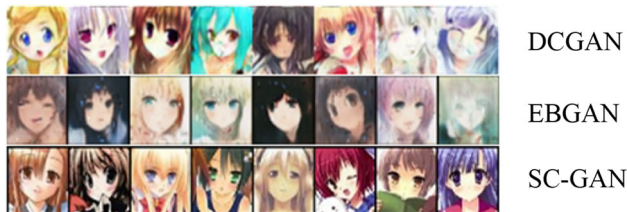**Fig. 5** Comparison of generated results from CelebA

Fig. 6 Comparison of generated results from CartoonFaces

4 Inter-frame image generation

4.1 Model assessment

In this paper, to illustrate the ability of the SC-GAN model, the quality of the generated model is tested by using the 200k celebrity faces images dataset of CelebA and the 50k CartoonFaces dataset. These two datasets are images with different angles, expressions and brightness.

In order to objectively evaluate the generated image, this paper evaluates the generated model by using the commonly used evaluation index peak signal-to-noise ratio (PSNR) [20] and structural similarity index (SSIM) [21]. PSNR is an objective criterion for evaluating image quality. The PSNR value of the two images reaches 30 or even 31, which shows that they are very close and have less distortion. SSIM measures similarity between two images. The range of SSIM value is in [0, 1]. When the two images are exactly same, the SSIM value is 1.

Adam with an initial learning rate in $[5 \times 10^{-5}, 10^{-4}]$ is used in this paper. The resolution of the input images is 64×64, while batch_size $= 16$ and epoch $= 300$ are set in the experiment. In these conditions, this paper compares the quality of different generated models.

Figures 5 and 6 show the random samples generated by both the CelebA and CartoonFaces datasets based on different generated models. In the same case, our model has some advantages compared with DCGAN [22] and EBGAN [18] in terms of image sharpness and diversity, and the visual effect is smoother and more natural.

Table 1 shows the assessment of the quality of the models by using PSNR and SSIM. The evaluation results show that the images generated by our model using different datasets

Table 1 PSNR and SSIM assessment results

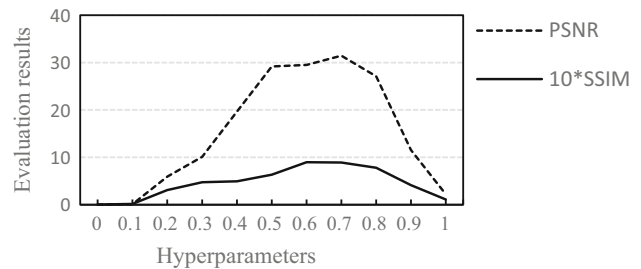| Dataset	Method	PSNR	SSIM
CelebA	DCGAN [22]	30.01	0.89
	EBGAN [18]	31.11	0.90
	SC-GAN	32.51	0.92
CartoonFaces	DCGAN [22]	31.08	0.91
	EBGAN [18]	30.00	0.88
	SC-GAN	33.02	0.94

Fig. 7 Assessment results of different γ values

are better than those by DCGAN and EBGAN, and the quality of our model is also proved to be superior to other models.
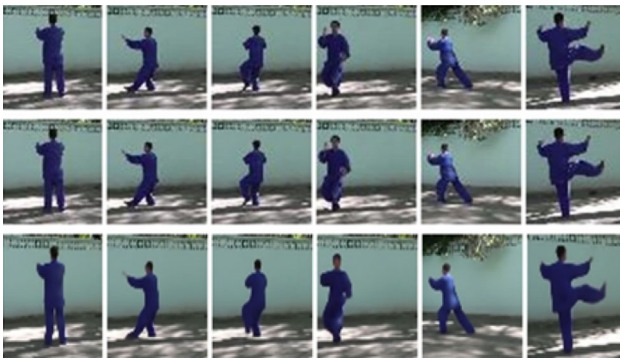
4.2 Inter-frame image generation

In this paper, the optimizers of both generator and discriminator adopt gradient-based optimization algorithm Adam and the learning rate lr is set in $[5 \times 10^{-5}, 10^{-4}]$ to make the gradient descent method perform better. In order to illustrate the influence of the hyperparameters on the training process of the model, and make the selection of parameters more convincing, using Taiji dataset based on Taiji instructional video, 11 groups of experiments under the same conditions were carried out with 0.1 as the step size. The evaluation results according to PSNR and SSIM were used to get the broken line diagrams of different γ values, as shown in Fig. 7.

To display the trend of evaluation results more directly, the graph uses value of 10 times SSIM to compare. According to the results of PSNR and SSIM, the γ value of 0.7 is set in this paper.

In the inter-frame image generation experiment, Taiji dataset based on Taiji instructional video and Ball dataset based on American animated video elastic ball are used. Each dataset contains 50K images with resolutions of 64×64.

To illustrate the generative ability of the model, in the inter-frame image generation experiment, six sets of images with two consecutive frames of different scenes are selected arbitrarily from the datasets of Ball and Taiji, and then, these images are input into the model to get six sets of different inter-frame images which are numbered InFNo. 1 to InFNo.

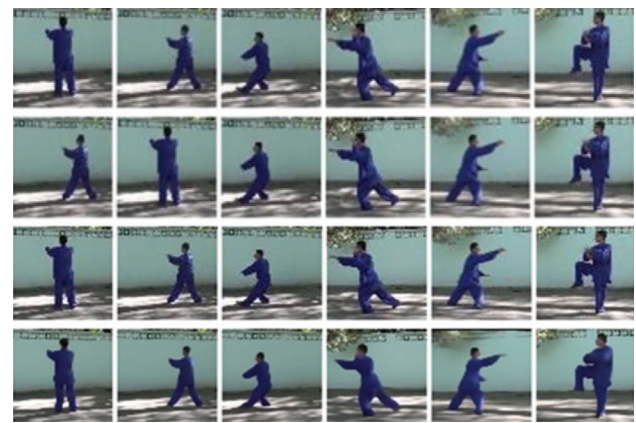**Fig. 8** Inter-frame image generation results of Taiji dataset
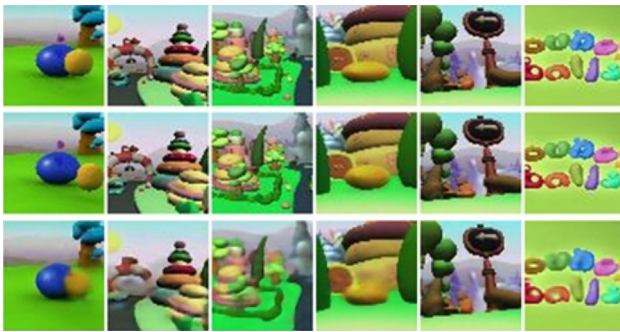
Fig. 9 Inter-frame image generation results of Ball dataset

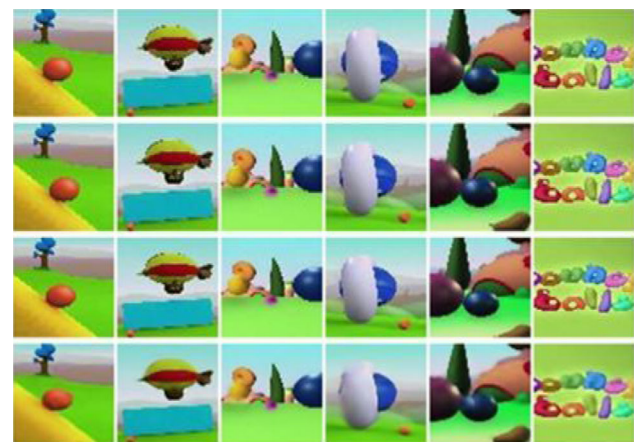**Table 2** Comparison of inter-frame image evaluation results

Frame no.	Taiji		Ball	
	PSNR	SSIM	PSNR	SSIM
InFNo. 1	31.64	0.96	29.74	0.88
InFNo. 2	32.04	0.96	30.00	0.88
InFNo. 3	40.72	0.97	30.23	0.88
InFNo. 4	41.38	0.98	30.10	0.90
InFNo. 5	28.83	0.97	31.55	0.88
InFNo. 6	41.04	0.98	29.66	0.85

6. The experimental results of the Taiji dataset and the Ball dataset experiment are shown in Figs. 8 and 9. The first row shows the first-frame image; the second row shows the second-frame image; the third row shows the generated inter-frame image. To show the quality of the generated image, PSNR and SSIM are used to evaluate the generated inter-frame. The evaluation results are shown in Table 2.

In order to illustrate the similarity between the generated inter-frame image and the real image, the quality verification experiment of inter-frame generation is carried out. Firstly, six sets of images with three consecutive frames of different angles, scenes and hues are selected arbitrarily from the dataset. Then, the first frame and the third frame are input into the model, and the corresponding six sets of inter-frame images are obtained, which are numbered from InCNo. 1 to

Fig. 10 Input image of Taiji dataset model

**Fig. 11** Input image of Ball dataset model

InCNo. 6. This experiment includes a comparison with real video frames, that is, the generated inter-frame images are compared with the real second-frame images in six groups, and evaluated by PSNR and SSIM.

Experimental results of inter-frame image generation on Taiji and Ball are shown in Figs. 10 and 11. The first and second rows are the first-frame image and the third-frame image; the third and fourth rows are the real second-frame image and the generated inter-frame image, respectively. The evaluation results of the inter-frame images generated by Taiji and Ball datasets are shown in Table 3. In the process of experiment, the convergence of the model is shown in Fig. 12. (a) is the convergent trend of the discriminator, and (b) is the convergent trend of the generator.

Compared with other models, the quality of images generated by GAN model is higher in the verification experiment. Therefore, this paper uses the same input image for inter-frame image generation based on GAN model in the contrast experiment, and the contrast experiment results are shown in Figs. 13 and 14.

Table 3 Comparison of inter-frame image evaluation results

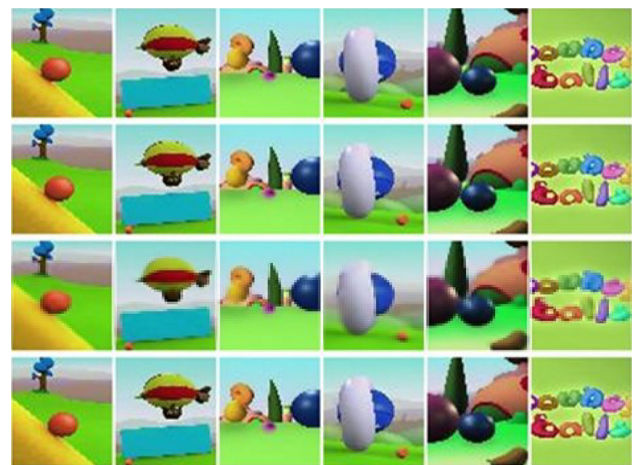| Frame no.	Taiji		Ball	
	PSNR	SSIM	PSNR	SSIM
InCNo. 1	33.05	0.97	29.81	0.88
InCNo. 2	33.40	0.96	30.54	0.92
InCNo. 3	34.53	0.96	30.24	0.91
InCNo. 4	36.81	0.97	29.87	0.86
InCNo. 5	31.40	0.97	27.22	0.67
InCNo. 6	35.36	0.97	29.94	0.89
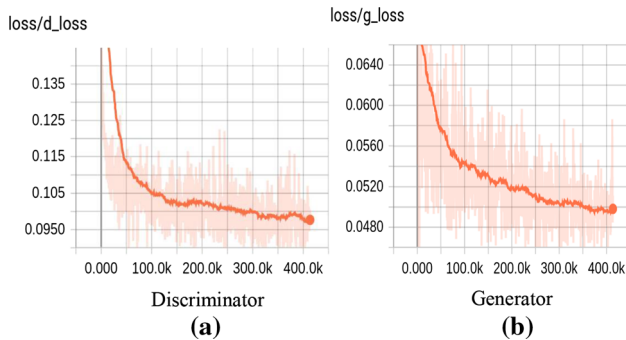
Fig. 12 Convergence trend in training process

Fig. 13 Taiji dataset based on GAN generation results

In Figs. 13 and 14, the first row shows the first-frame image; the second row shows the second-frame image; the third row shows the generated inter-frame image; the fourth row shows the third-frame image. In GAN model, the evaluation results of the inter-frame images generated by Taiji and Ball datasets are shown in Table 4.

From the results of model quality evaluation experiment, inter-frame image generation experiment, and the contrast experiment of GAN model, we can see that the SC-GAN model can produce higher quality images than the traditional methods in video-based inter-frame image generation, without the problems of edge blurring caused by the traditional

Fig. 14 Ball dataset based on GAN generation results

Table 4 Comparison of GAN model evaluation results

Frame no.	Taiji		Ball	
	PSNR	SSIM	PSNR	SSIM
InCNo. 1	24.44	0.70	22.77	0.63
InCNo. 2	26.64	0.79	24.74	0.68
InCNo. 3	25.84	0.83	25.29	0.74
InCNo. 4	25.46	0.73	24.75	0.65
InCNo. 5	23.75	0.58	23.85	0.63
InCNo. 6	26.37	0.75	25.97	0.77

methods. And the convergence speed is fast, and the image generation is more efficient. Compared with GAN model, SC-GAN model proves its good ability of generating inter-frame images in terms of visual effects and evaluation results.

In each stage of the experiment, several groups of non-repeated sampling experiments were carried out, and the quality of the generated results was evaluated by PSNR and SSIM. Visually, the SC-GAN model can produce higher quality images than other models. In the video with simple scenes, it is impossible to distinguish the generated image from the real image, and it will not produce image contour distortion and image blurring. The quantization results show that the inter-frame images generated by SC-GAN model have high authenticity and high structural similarity with real video frames.

5 Conclusion

In order to solve the problem of low-frame rate video playback and edge blurring caused by using traditional methods when improving the video frame rate, this paper proposes a video inter-frame image generation method based on space continuous generation antagonistic network. This method

trains SC-GAN model by video-based image dataset and uses the spatial continuity of the image features to complete the generation of inter-frame images for low-frame rate video, and inserts these generated images into two corresponding frames as a new video frame, so as to improve the frame rate. This way of improving the frame rate is effective in both static and dynamic video sources. Under the premise of guaranteeing the sharpness of the generated image, the problem of blurring image edges caused by traditional methods is avoided, and a new method is provided for improving the video frame rate. Experiments show that, in the selection of hyperparameter, when the hyperparameter is low, the generated samples are single and the diversity is obviously insufficient; as the hyperparameter is gradually increasing, the generated images are more diverse and clearer. When the hyperparameter is close to the critical value, the quality of the generated samples becomes worse and the model becomes unstable. In terms of results of image generation, PSNR and SSIM evaluation prove that the inter-frame image generated by the model has high authenticity, and also verify the feasibility and validity of the proposed method for video inter-frame image generation based on SC-GAN.

The model can generate good quality inter-frame images in relatively simple video and has good applicability in animation and simple comic video. But for complex scene datasets, the training needs more time, and the network structure is more complex. This is where the method needs to be improved. It is also the content of our next research.

References

1. Paramkusam, A.V., Reddy, V.S.K.: Multilayer reference frame motion estimation for video coding. Signal Image Video Process. **9**(8), 1851–1860 (2015)
2. Purwar, R.K., Prakash, N., Rajpal, N.: A matching criterion for motion compensation in the temporal coding of video signal. Signal Image Video Process. **5**(2), 133–139 (2011)
3. Ha, T., Lee, S., Kim, J.: Motion compensated frame interpolation by new block-based motion estimation algorithm. IEEE Trans. Consum. Electron. **50**(2), 752–759 (2004)
4. Dikbas, S., Altunbasak, Y.: Novel true-motion estimation algorithm and its application to motion-compensated temporal frame interpolation. IEEE Trans. Image Process. **22**(8), 2931–2945 (2013)
5. Jeong, S.G., Lee, C., Kim, C.S.: Motion-compensated frame interpolation based on multi-hypothesis motion estimation and texture optimization. IEEE Trans. Image Process. **22**(11), 4497–4509 (2013)
6. Wei, C.: Frame rate up conversion algorithm research based on motion estimation and motion compensation. Doctoral dissertation (2016)
7. Taşdemir, K., Çetin, A.E.: Content-based video copy detection based on motion vectors estimated using a lower frame rate. Signal Image Video Process. **8**(6), 1049–1057 (2014)
8. Qi, G.: Research and application of image generation based on deep learning. Doctoral dissertation (2017)
9. Park, K., Yu, S., Park, S., Lee, S., Paik, J.: An optimal low dynamic range image generation method using a neural network. IEEE Trans. Consum. Electron. **64**(1), 69–76 (2018)
10. Jingxuan, H., Yao, Z., Chunyu, L., Meiqin, L., Huihui, B.: CNN-based frame rate up-conversion algorithm. Appl. Res. Comput. **35**(2), 611–614 (2018)
11. Gucan, L., Xiaohu, Z., Qifeng, Y.: Deep convolutional neural network for motion compensated frame interpolation. J. Natl. Univ. Def. Technol. **38**(5), 143–148 (2016)
12. Goodfellow, I.: Nips 2016 tutorial: generative adversarial networks. arXiv preprint arXiv:1701.00160 (2016)
13. Kunfeng, W., Chao, G., Yanjie, D.: Generative adversarial networks: the state of the art and beyond. Acta Autom. Sin. **43**(3), 321–332 (2017)
14. Mathieu, M., Couprie, C., Lecun, Y.: Deep multi-scale video prediction beyond mean square error. Electr. Eng. Syst. Sci. **32**(24), 1091–1105 (2017)
15. Berthelot, D., Schumm, T., Metz, L.: BEGAN: boundary equilibrium generative adversarial networks. arXiv preprint arXiv:1703.10717 (2017)
16. Xueqi, C., Xiaolong, J., Yuanzhuo, W.: Survey on big data system and analytic technology. J. Softw. **25**(9), 1889–1908 (2014)
17. Mirza, M., Osindero, S.: Conditional generative adversarial nets. computer. Science **10**(21), 2672–2680 (2014)
18. Zhao, J., Mathieu, M., LeCun, Y.: Energy-based generative adversarial network. arXiv preprint arXiv:1609.03126 (2016)
19. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan. arXiv preprint arXiv:1701.07875 (2017)
20. Tanchenko, A.: Visual-PSNR measure of image quality. J. Vis. Commun. Image Represent. **25**(5), 874–878 (2014)
21. Zhu, R., Zhou, F., Xue, J.H.: MvSSIM: a quality assessment index for hyperspectral images. Neurocomputing **272**, 250–257 (2018)
22. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)