

# Super-resolution Reconstruction Based on Capsule Generative Adversarial Network

Ziyi Wu

School of Computer Science and Engineering  
Xi'an Technological University  
No.2 Xuefu Middle Road, Weiyang district,  
xi'an, Shaanxi, China  
E-mail: wuziyi\_5817@163.com

Hong Jiang

School of Computer Science and Engineering  
Xi'an Technological University  
No.2 Xuefu Middle Road, Weiyang district,  
xi'an, Shaanxi, China  
E-mail: 249479898@qq.com

Hongge Yao

School of Computer Science and Engineering  
Xi'an Technological University  
No.2 Xuefu Middle Road, Weiyang district,  
xi'an, Shaanxi, China  
E-mail: yaohongge@xatu.edu.cn

Wei Zhang

School of Computer Science and Engineering  
Xi'an Technological University  
No.2 Xuefu Middle Road, Weiyang district,  
Xi'an, Shaanxi, China  
E-mail: 913917873@qq.com

Hualong Yang

School of Computer Science and Engineering  
Xi'an Technological University  
No.2 Xuefu Middle Road, Weiyang district,  
xi'an, Shaanxi, China  
E-mail: 1067726236@qq.com

Jun Yu

School of Computer Science and Engineering  
Xi'an Technological University  
No.2 Xuefu Middle Road, Weiyang district,  
xi'an, Shaanxi, China  
E-mail: 763757335@qq.com

**Abstract**—Using each part of the image's spatial information to generate better local details of the image is a key problem that super-resolution reconstruction has been facing. At present, mainstream super-resolution reconstruction networks are all built based on convolutional neural networks (CNN). Some of these methods based on Generative Adversarial Networks (GAN) have good performance in high-frequency details and visual effects. However, because CNN lacks the necessary attention to local spatial information, the reconstruction method is prone to problems such as excessive image brightness and unnatural pixel regions in the image. Therefore, using the capsule network's excellent perception of hierarchical spatial information and local feature relationships, the author proposes a super-resolution reconstruction based on capsule network CSRGAN. The experiment's final result shows that compared with the pure convolution method RDN,

the PSNR value of CSRGAN is increased by 0.14, which is closer to the original image.

*Keywords*-Generative Adversarial Network; Capsule Network; Capsule Generative Adversarial Network; Capsule Discriminator; Super-resolution Reconstruction

## I. INTRODUCTION

In computer image processing, obtaining more necessary information from low-resolution input images to achieve super-resolution reconstruction (SR) is always a critical issue. To solve this problem, it needs to use the low-resolution image's existing information to infer its own mapping of the high-resolution image. The essence of this mapping is an ill-conditioned reasoning problem,

so super-resolution reconstruction of low-resolution images is an arduous task. Artificial intelligence methods have been intensely studied in such fields as intelligent cognition technology [1], efficient image retrieval technology [2], blockchain technology [3], communication virtual mobile network resource management [4], machine motion detection analysis [5], fault identification [6], and so on, and have achieved exciting research results. Moreover, CNN [7] and reinforcement learning are the most widely used in these fields among the artificial intelligence methods and have achieved exciting research results. To obtain better reconstruction results, many scholars have proposed several super-resolution reconstruction methods based on deep learning. Among them, the deep network based on CNN has shown great promise in reconstruction accuracy [8].

He Kaiming and others first proposed a super-resolution reconstruction network SRCNN that uses a three-layer convolutional network to fit nonlinear mapping in 2014 and achieved excellent experimental results at that time [9]. This is also the first successful attempt of CNN in the field of super-resolution reconstruction. In 2016, Wenzhe Shi and others proposed the ESPCN framework [10]. Compared with those methods that use interpolation to enlarge the image to the target size before doing the convolution operation, the sub-pixel convolution mentioned in the paper directly performs the upsampling process on the low-resolution image, which makes the ESPCN significantly reduced the calculation complexity and parameter amount [11]. The running speed is also improved, and it can even support real-time super-resolution reconstruction in the video. In 2017, SRGAN [12] proposed by Christian Ledig et al. was the first successful attempt to use the GAN [13] to achieve super-resolution reconstruction. It performed well in improving the realism of the picture, generated better high-frequency texture details, and had good generalization, and it performs well in a scene with 4x resolution magnification. The emergence of SRGAN proves that GAN has excellent potential in the field of super-resolution reconstruction [14]. Based on the research of super-resolution reconstruction of deep

convolutional networks [15] and deep spatial feature transform [16,17]. In 2018, Yulun Zhang et al. proposed the RDN [18]. The RDB structure used in this framework can make full use of all convolutional layers' hierarchical information to achieve more efficient feature learning and finally obtain outstanding visual effects. In 2020, the dual regression scheme used by Dual Regression Networks proposed by Yong Guo et al. added additional constraints to the LR data [19]. This framework allows the network to learn not only the mapping of low-resolution images to high-resolution images, but also the inverse mapping of super-resolution images to low-resolution images. These better mapping combinations can better guide the network to reconstruct images.

However, these super-resolution networks simply based on CNN usually pay more attention to the target's overall features but ignore the relationship between spatial level information and local features. In super-resolution reconstruction, enhancing the perception ability of local feature spatial information is of great significance for the reconstruction of local details. This characteristic of CNN hinders a super-resolution reconstruction network's process to create closer to the real details through this information and further leads to local over-smoothness of the generated image and sometimes even unnatural texture.

In order to solve the above problems, we propose a new generation of confrontation super-resolution reconstruction model CSRGAN. CSRGAN takes advantage of the Capsule Network's ability [20,21] to infer the relationship between the various parts of the image and makes up for the lack of CNN in dealing with the feature space connection. The CSRGAN model uses the RDN architecture as a generator to enhance the acquisition of each convolutional layer's hierarchical information and uses the Capsule network to replace the CNN network as a discriminator to enhance the model's ability to judge spatial level details. And by adding a vector inner product loss function to improve the picture quality of the generated picture. CSRGAN mainly has the following two contributions:

- 1) A dual-route capsule network discriminator is constructed, and the parameter matrix of the

capsule network in the discriminator is restricted to small parameters through a clip function. It enables the discriminator network to extract features from coarse-grained to fine-grained through a dual-routing process, so that the capsule network can improve the efficiency and stability and also enhance the perception of the local spatial posture information of the generated image.

2) The vector inner product loss is proposed and added to the generator's loss function to train the generator network. The feature set extracted by the capsule network is also added to the loss calculation so that the generation network has a better representation of the local feature content and ultimately improves the performance of the local detail texture.

This article is composed of six chapters. The first chapter INTRODUCTION puts forward the motivation and contribution of this article; The second chapter RELATED WORK introduces previous work related to this article and their contribution to this article; The third chapter METHOD introduces the method proposed in this article and the network operation mechanism in detail; Chapter 4 LOSS FUNCTION introduces the theoretical basis of the network; Chapter 5 EXPERIMENTS is the specific experimental process and experimental data analysis of this article; Chapter 6 CONCLUSION is a summary of the full text.

## II. REALATED WORK

**SRGAN** The SRGAN model inherits the GAN model's generative confrontation idea, and on its basis, improves the cost function for super-resolution reconstruction. One part of the improvement is the content-based cost function, and the other part is the cost function based on confrontation learning. The content-based cost function contains a minimum mean square error based on the feature space, calculated using the high-level features of the image extracted by the VGG [22]. These improvements make SRGAN better at processing high-frequency texture details. In the generation network, SRGAN uses the sub-pixel convolution in ESPCN for up-sampling, which improves the quality of the generated pictures and reduces the complexity of reconstruction calculations.

to

**RDN** The RDN model combines the ideas of Res-Net [23] and Dense-Net [24] and improves on it. RDN realizes the local feature fusion and global feature fusion of all feature maps' levels through the combined use of the Residual dense blocks (RDB) module and Dense feature fusion (DFF), achieving the maximum reuse of convolutional network-level information. This enables RDN to perform a more detailed feature understanding of the input low-resolution images when performing super-resolution tasks. And it can maintain a stable training effect even when the network level is deepened.

**Capsule Network** Capsule Network is a new type of network proposed by the Hinton team in 2017. Hinton believes that although the "translation invariance" of CNN brings the robustness of classification, it inevitably loses the spatial hierarchical information of the feature, and this information is just helpful to the understanding of the feature. The capsule network uses vectors to express feature information. Its "identity" can represent the spatial hierarchical information in the feature, and it has better recognition in image classification from different angles. The modulus of the capsule vector represents the probability of the feature's existence, and the direction of the vector represents the spatial hierarchical relationship between the features. These characteristics of the capsule vector make the capsule network more accurate in feature recognition. Moreover, the capsule network can complete network training with less image training data.

Based on the advantages of the capsule network and the above two networks, we use RDN to build the basic structure of the generation network and use the capsule network to build the discriminator's basic structure. Besides, we have implemented dual routing improvements based on the capsule network. The two routing processes of the capsule cooperate so that the network has a better perception of the local correlation of the image feature space, and finally promotes the generator to generate a closer high-resolution image of the original image.

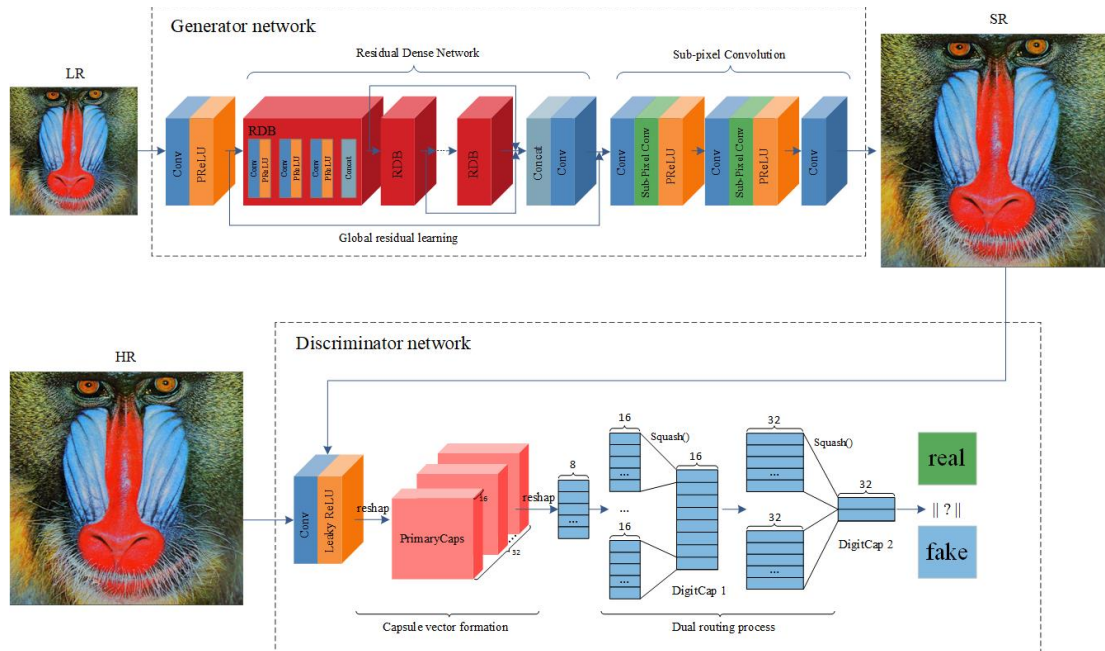


Figure 1. Structure diagram of CSRGAN.

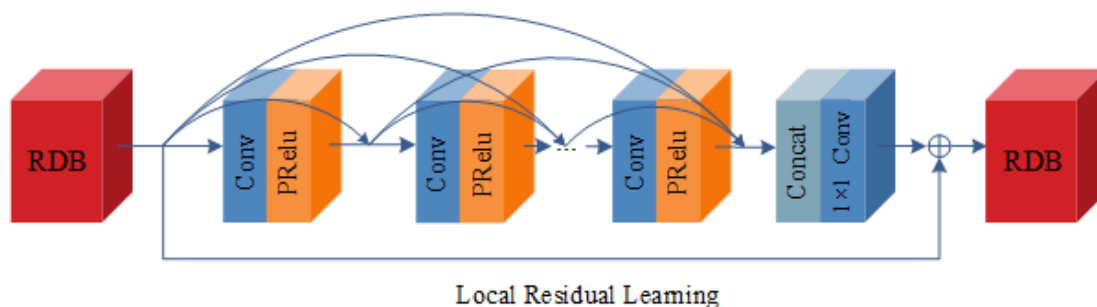


Figure 2. Structure diagram of RDB.

### III. METHOD

This paper constructs a capsule generation adversarial super-resolution reconstruction network CSRGAN based on the capsule network. The network framework includes two parts: the generator network and the discriminator network. The structure of the network is shown in Fig. 1.

#### A. Generator network

The generator network contains two parts: a dense residual network layer and a sub-pixel convolutional layer. The dense residual network contains 16 residual blocks composed of a

convolutional layer and an activation function—each residual dense block structure, As shown in Fig. 2. using this structure can avoid the disappearance of the gradient during backpropagation while using Global Residual learning to obtain global dense features from the original LR image, and further combine the shallow features and deep features to fully invoke the layering of low-resolution images features to improve feature extraction capabilities so that the network learns more effective features.

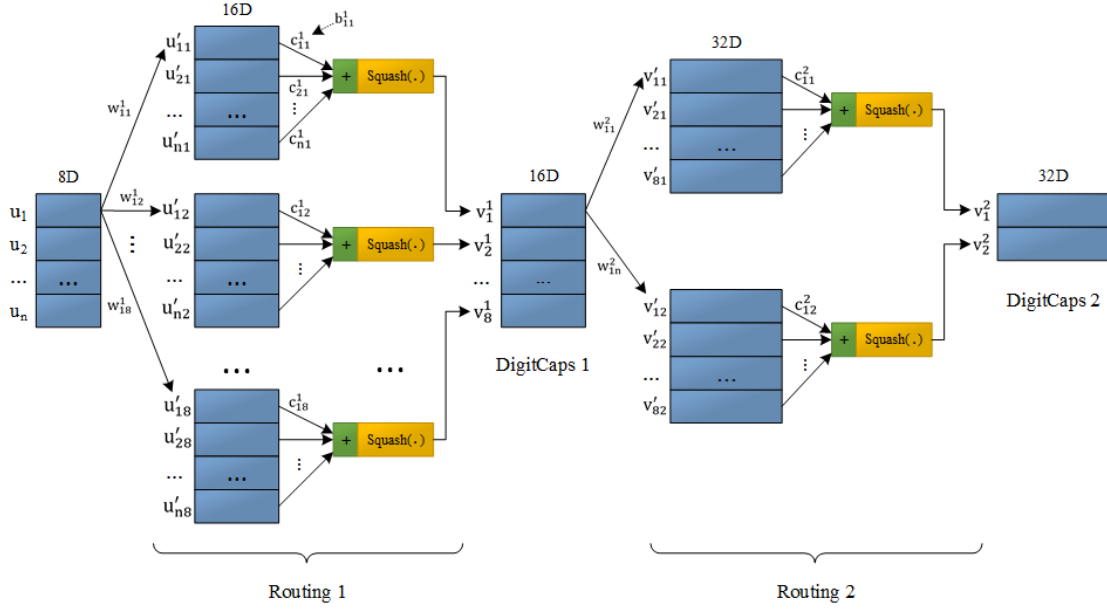


Figure 3.  $u$  is  $n$  8-dimensional vectors obtained by encapsulating and flattening the data through the PrimaryCaps layer,  $v^1$  is a 16-dimensional digital capsule vector after a routing process and  $v^2$  is the 32-dimensional vector output at the end of the second routing process.  $w_{ij}$  is the fully connected weight matrix of the connected vector. The vectors  $u'$  and  $v'$  are the prediction vectors of  $u$  and  $v^1$  obtained through a linear hierarchical relationship, the subscripts  $i, j$  correspond to the number of vector  $u$  and the next layer vector  $v^1$ , respectively. The parameter  $b_{ij}$  is the log prior probability from the lower capsule to the upper capsule. The parameter  $b_{ij}$  is the logarithmic prior probability from the low-level capsule to the high-level capsule and is used to update each  $c_{ij}$  correspondingly.  $c_{ij}$  is the coupling coefficient connecting the two layers' vectors before and after, representing the degree of correlation between the  $i$ -th vector of the previous layer and the  $j$ -th vector of the next layer. Squash is a linear rectification function, which normalizes the results obtained.

### B. Discriminator network with double routing

The discriminator network uses a dual-route capsule network to achieve true or false classifications, including a capsule vector forming part and a dual-routing part. The capsule vector forming part includes a Conv layer, a PrimaryCaps layer, and two DigitCaps layers.

The capsule vector formation part is that after the Conv layer performs preliminary feature extraction on the input image, the original scalar neuron of the convolutional network is repackaged into a vector neuron during the parameter transfer process PrimaryCaps layer. The length of each vector represents the estimated probability of whether the object exists, and its direction records the posture parameters of the object.

After the PrimaryCaps layer, the obtained multi-dimensional tensor is flattened into a one-

dimensional array and then input into the two Routing processes to obtain two DigitCaps layers, respectively. These two DigitCaps layers implement the feature's coarse extraction process and the feature's fine extraction process to classify the features. There is a more gradual transition, which is beneficial to improve the classification accuracy. The dual routing is shown in Fig. 3.

It should be emphasized that the importance of low-level feature capsules to high-level feature capsules can be measured by  $c_{ij}$  and the corresponding prediction vector.  $c_{ij}$  is the normalized result of  $b_{ij}$ . The formula is as follows:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ij})} \quad (1)$$

The update of  $c_{ij}$  is affected by  $b_{ij}$ . The initial value of  $b_{ij}$  is 0. The  $c_{ij}$  normalized by the formula (1.1) is assigned the average probability value, such as:  $c_{11}^1 + c_{12}^1 + \dots + c_{18}^1 = 1$ .

The update of  $b_{ij}$  uses the following formula:

$$b_{ij} = b_{ij} + \hat{u}_{ij} \cdot v_j \quad (2)$$

Relying on (1) and (2) the capsule network can iteratively improve the coupling coefficient  $c_{ij}$ .

During training, the input of the capsule discriminator is the generated high-resolution image and the original high-resolution image of the training set. Its output is the binary classification result calculated by the second normal form of the vector output from the second DigitCaps layer, corresponding to the generated image (fake) and real image (real).

The dimension expansion of the feature vector through  $v^1$  and  $v^2$  corresponds to the rough extraction and fine extraction of the dual routing process. That allows the dual-route capsule network to deepen the network's understanding of image features in a higher dimension and enhance the discriminant network's discriminative ability to assist the generation of the network better.

#### IV. LOSS FUNCTION

This article is based on GAN network construction. The overall network loss function depends on the GAN loss function and the vector inner product loss function we proposed. To achieve a complete network function, we need to discriminant network, using the training's discriminator loss function and generator loss function [25].

##### A. D Loss

Since the discriminator network is built by the capsule network as a whole, and the final actual output result is a multi-dimensional capsule vector, the impact of the capsule network's loss function needs to be considered.

$L_M$  is the edge loss function used by the capsule network to implement parameter training, which is defined as follows:

$$L_M = \sum_{k=1}^K T_k \max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k) \max(0, \|v_k\| - m^-)^2 \quad (3)$$

Among them,  $k$  is the number of classifications, and  $T_k$  is a function of the classification.  $T_k = 1$  when and only when the classification of  $k$  appears, and  $T_k = 0$  when it does not exist.

The loss function of the GAN discriminator is as follows:

$$\max V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (4)$$

After combining the super-resolution task with the above loss function and making improvements based on WGAN [26,27], we can define the final discriminator loss function, as in (5):

$$L_D = \max E_{I^{HR} \sim p_{train}(I^{HR})} [-L_M(D_{\theta_D}(I^{HR}), T=1)] + E_{I^{LR} \sim p_G(I^{LR})} [-L_M(D_{\theta_D}(G_{\theta_G}(I^{LR})), T=0)] \quad (5)$$

The two parts of the formula represent the discriminative loss of the two source data, respectively. Among them,  $D$  is the discriminator network. Since the capsule network is used as the discriminator in this article, the actual output result is a vector.  $G$  is the generator network, and the output is a high-resolution image.  $I^{HR}$  represents the incoming discriminant network  $D$  is the training set high-resolution (HR) image,  $I^{LR}$  represents the incoming generation network  $G$  is the pre-processed low-resolution image (LR), The LR is processed by the  $G$  network into a high-resolution image SR and then transferred to the  $D$  network. The source is the real training set whose  $T$  value is set to 1, and the source is the generated image whose  $T$  value is set to 0.

##### B. G Loss

When the generator is trained, the minimum value of the generated adversarial is calculated. Since the first half of the formula has nothing to do with the  $G$  network, the second half of the (1) is used during actual training, and the  $T$  value is set to 1, as follows:

$$L_{Adversarial} = \min_{E_{I^{LR} \sim p_G(I^{LR})}} [L_M(D_{\theta_G}(G_{\theta_G}(I^{LR})), T=1)] \quad (6)$$

In addition to the generator counter loss of (4), in order to improve the perception of the texture details of the generated image, this paper proposes a vector inner product loss function based on the capsule network as follows:

$$L_{V_{ecoter}}^{SR} = \sum_i^m (V_i(I^{HR}) \cdot V_i(G_{\theta_G}(I^{LR})) - \|V_i(I^{HR})\|^2)^2 \quad (7)$$

$V$  represents the vector before the compressed rectification function of the capsule inner product loss function, that is, a 16-dimensional vector from the DigitCaps layer of the capsule network. The subscript  $i$  of  $V$  represents the number of sequences to be classified, and  $m$  is the total number of classifications. There are two classifications of true and false in the experiment, and the value of  $i$  is 0 or 1.

Therefore, combining (3) and (4), the final cost function  $L_G$  of CSRGAN generator is as follows:

$$L_G = L_{V_{ecoter}}^{SR} + 10^{-3} \cdot L_{Adversarial} \quad (8)$$

To make the final reconstructed pixels of the generated image closer to the HR, we need to train the main constraints to focus more on the pixels' accuracy while relatively weakening the dependence on the "creativity" of the adversarial

network. The adjustment coefficient in (8) is set to 10<sup>-3</sup>.

## V. EXPERIMENTS

### A. Experimental results

The DIV2K data set is used in the experiment, which contains 800 images of the training set, 100 images of the test set, and 100 images of the verification set. Random horizontal flipping is used to enhance the data set. The batch size of the network is set to 16. Before the experiment, it is necessary to obtain the LR image by bicubic interpolation of the HR and then use the super-resolution reconstruction algorithm to reconstruct the LR image into an SR image. The processes from LR to HR and HR to SR are all performed with a 4 times scale factor. The network's ultimate goal is to train a generator network  $G$ , which can generate a corresponding HR image from a given LR image. The experiment finally realizes the optimization of the above-mentioned mini-max problems through alternate training of D-network and G-network. The RMSProp algorithm, which is more suitable for gradient instability, is used for parameter optimization during training. In order to make the discriminator have a certain discriminative performance in the experiment, the experiment is set to start training the  $G$  network when the  $D$  network is trained 25 times. After many experiments, it is found that the training ratio of the  $D$  to  $G$  network is adjusted to 3:1. The two networks can have a better convergence effect.

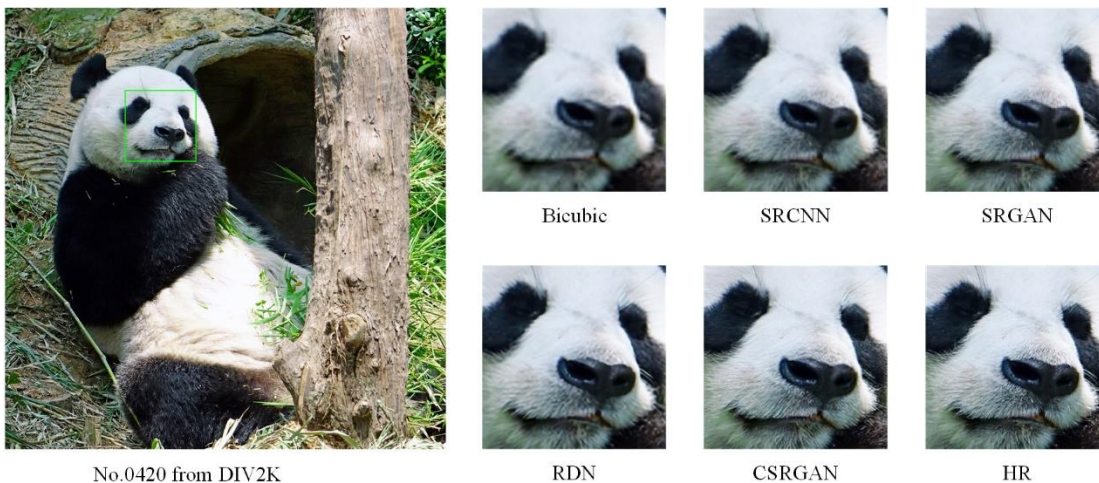
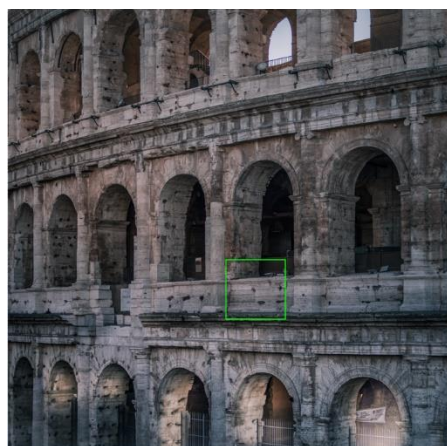


Figure 4. Panda.



No.0864 from DIV2K



Figure 5. Rome.



No.0872 from DIV2K



Figure 6. Nut.

The experimental content mainly focuses on the three dimensions of image texture effect, image brightness, and local features. The comparison with Bicubic, SRCNN, SRGAN, and RDN is shown in figures below:

### B. Analysis of results

Fig. 4, Fig. 5 and Fig. 6 respectively capture some of the  $250 \times 250$  pixels,  $185 \times 185$  pixels, and  $100 \times 100$  pixels corresponding to the experimental results.

In pixel-level reconstruction tasks, the texture is an important part of describing the image's details, usually in the reconstruction effect of the partial lines of the image. It can be seen from the experimental results that the reconstructed image obtained by the Bicubic method is blurred,

especially in the image where the pixel difference is dense, and a lot of texture details are missing. For example, in Fig. 6, the lines on the surface of the reconstructed fruit in the image are even disconnected and discontinuous. The image reconstructed by SRCNN cannot well restore the image's texture, and there is a lack of smooth transition between local edge pixels in the high-multiple reconstruction task. For example, in Fig. 5, the image reconstructed by SRCNN does not accurately restore the straight lines of smaller wall tile targets, and some curves are reconstructed into straight lines.

In the experimental evaluation of super-resolution reconstruction, brightness is an important parameter of the similarity index of image structure. It is also the image element that



human visual senses are most likely to perceive. As shown in Fig. 5, SRGAN, which uses GAN as the network structure, has a significant deviation from the label data in the restoration of the brightness space. The reconstructed building image produces a generally high brightness result. On the other hand, although the sharpness of SRGAN is greatly improved compared with SRCNN, the general grid-like texture shown in Fig. 6 is generated in the image after reconstruction, which also greatly impacts the look and feel of the image. RDN has high clarity, but it will produce excessively sharp edges in the image's local details. As shown in Fig. 5, the reconstructed image's wall lines are excessively sharp, which is not consistent with the style of ancient Roman buildings, which makes the image not realistic in appearance. CSRGAN significantly weakens the raster effect produced by the reconstruction of the SRGAN architecture and has richer local texture details, ensuring that the image is clear enough during high-magnification reconstruction.

Unlike other reconstruction methods that lack attention to the correlation between local features, CSRGAN using the capsule network is more in line with the natural distribution of the image in terms of local features. As shown in Fig. 4, we can see that in the reconstruction results of the panda's nose and surrounding hair, CSRGAN has the most natural and smooth transition and can achieve more accurate reconstruction in different parts of the hair with similar color and texture. As shown in Fig. 6, in the original image of HR, the lines and grooves on the nut's surface also have a large number of different parts of the same color but different color depths. CSRGAN also accurately achieves the reconstruction and has a look and feel closer to the real image.

In the experiment, we use PSNR as the key image evaluation index. The PSNR training results of CSRGAN based on the Set5 dataset are shown in Fig. 7.

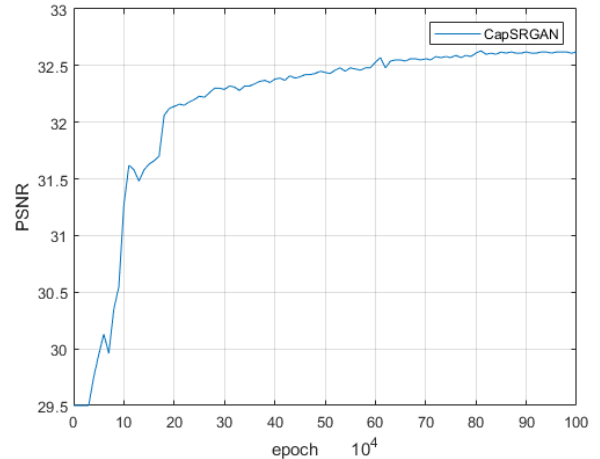


Figure 7. Train PSNR is evaluated on Set5.

As shown in Fig. 7, CSRGAN has a relatively stable convergence process and can stably generate high-quality super-resolution images after reaching a certain number of training times.

In this experiment, the two standards of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) comprehensively evaluate the results of each experimental method as follows:

TABLE I. PSNR AND SSIM

Method	Bicubic	SRCNN	SRGAN	RDN	CSRGAN
PSNR	28.39	30.45	29.43	32.44	32.58
SSIM	0.8102	0.8616	0.8477	0.8988	0.9003

It can be seen from Table I that Bicubic's PSNR and SSIM achieved 28.39 and 0.8102, respectively, which is the lowest score for this experiment. The PSNR and SSIM parameters of SRGAN are 29.43 and 0.8477, respectively, and both indicators are weaker than SRCNN and RDN. The two parameters of CSRGAN's PSNR and SSIM's evaluation value in the experiment are compared with the leading RDN reconstruction algorithm in the field. The two parameters have increased by 0.14 and 0.0015, respectively, which shows that CSRGAN has the ability to generate more realistic super-resolution images than RDN.

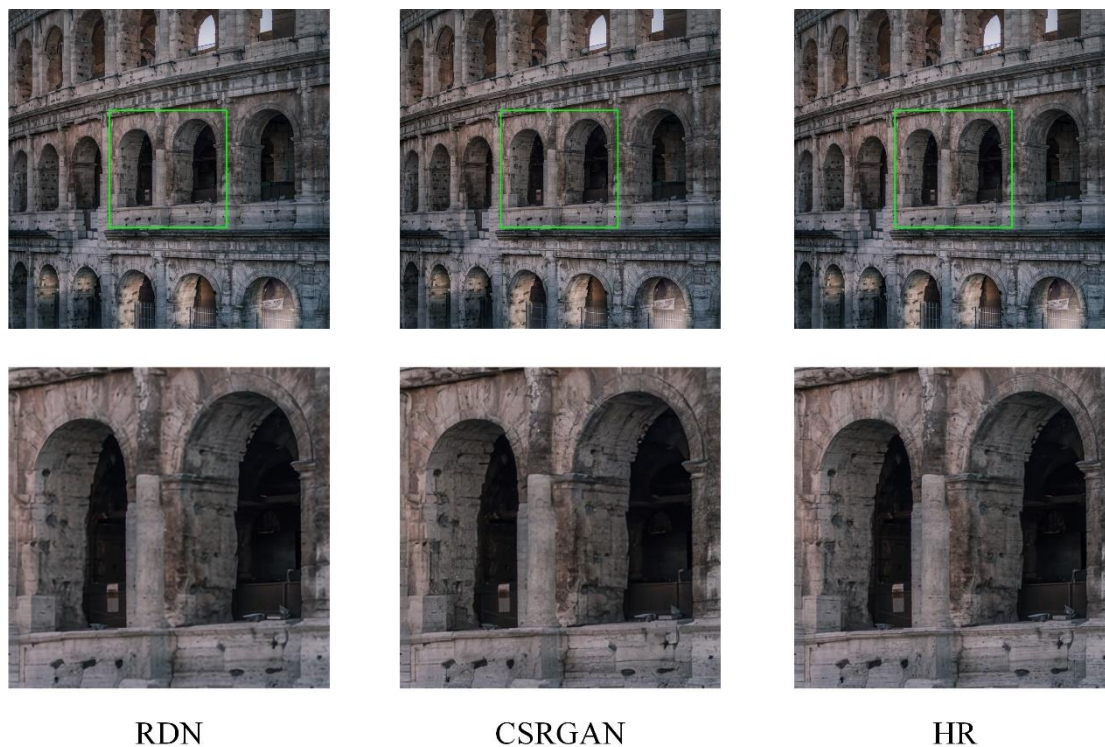


Figure 8. The Roman Colosseum(a).

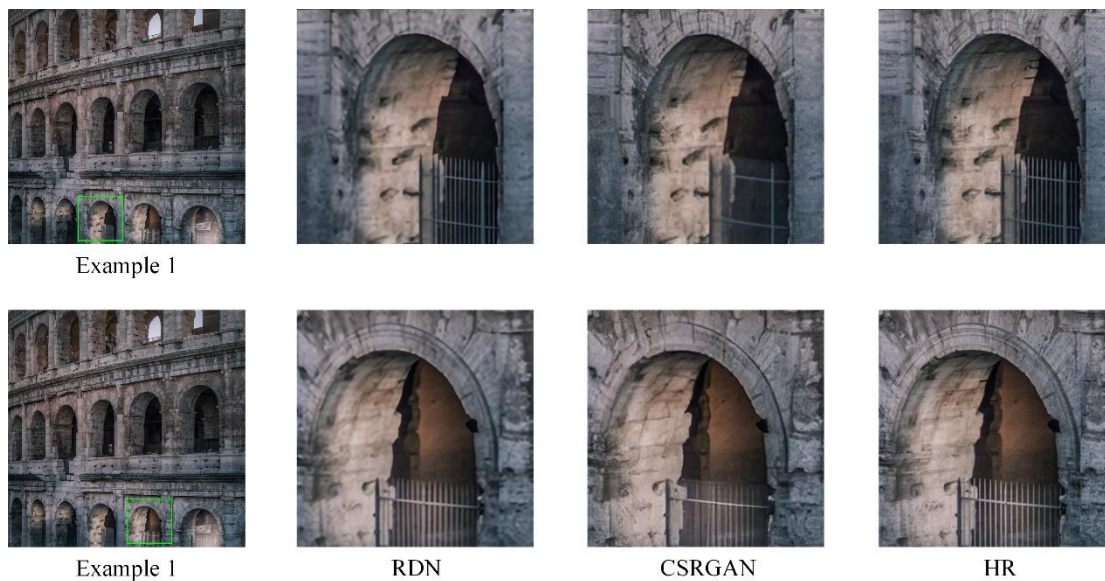


Figure 9. The Roman Colosseum(b).

*C. Experimental negative sample analysis*

Nevertheless, CSRGAN still has a lot of work worthy of in-depth research in the future, and there are also unresolved factors that negatively affect the imaging effect. In the few images in the

experiment, we also found some samples with poor performance. As shown in Fig. 8 and Fig. 9 below.

Although the images captured in Fig. 8 and Fig. 9 are from the same set of experimental data, the

image area CSRGAN shown in Example 1 exceeds the image generated by RDN in terms of the doorway's details and the overall texture.

As shown in Fig. 10 and Fig. 11, the image data is still taken from the same experimental test sample. In Fig. 10, the clarity of the reconstructed image of the trees generated by CSRGAN is far better than that of the image generated by RDN, and even the restoration of the stairs and walls in Fig. 11 gives the audience a clearer subjective feeling. However, we can observe the details of the bricks of the steps and the wall's decoration, which produces a larger deviation compared with the HR image.

We believe that the situation in Fig. 8 and Fig. 9 is highly likely to be caused by the fact that the capsule network has few relevant features in the training set and weak local correlation (such as the strong correlation between the overall features

of the human face and the features of local facial features), and that the level of the capsule network in CSRGAN is relatively shallow and no deeper features are extracted. Improving the capsule network's performance by deepening the capsule network is a direction of our future related research. In addition to the above factors, some of the influencing factors in Fig. 10 and Fig. 11 may come from the GAN model. The selection of effective features between capsule layers is achieved through clustering algorithms, which will slow down the overall training speed during training. Therefore, how to optimize the capsule network clustering algorithm and improve its training speed in the future is also an important research direction in the research of super-resolution reconstruction of capsule network.



Figure 10. castle(a).

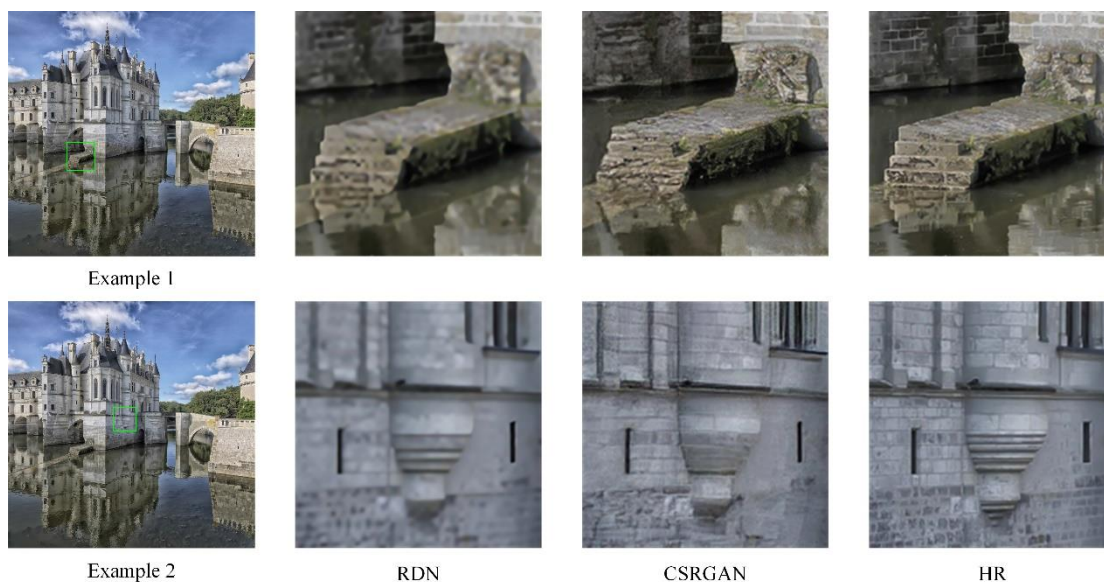


Figure 11. castle(b).

## VI. CONCLUSION

To improve the image quality of super-resolution reconstruction of images and obtain better local detail texture performance, this paper proposes a GAN model CSRGAN that uses a capsule network as a discriminator. This model uses a capsule network to replace the traditional CNN discriminator network, which enables the network model to have a more detailed grasp of local spatial information, and enables the generation network to achieve the more accurate reconstruction of regions with complex local pixel distribution; Simultaneously, the vector inner product loss function is added, and the images generated by CSRGAN achieve better detail performance through the network's training.

In summary, although there are still many areas for improvement in the network, CSRGAN is still a more effective attempt. The final PSNR and SSIM results of the experiment show that the CSRGAN model has a better reconstruction effect than the CNN-based super-resolution model on the DIV2K data set, which proves that the attention of capsule network to local spatial information in images is effective, which will contribute to the generation of images with

higher reduction accuracy in super-resolution reconstruction tasks.

## REFERENCES

- [1] Lu H, Li Y, Chen M, et al. Brain intelligence: go beyond artificial intelligence [J]. *Mobile Networks and Applications*, 2018, 23(2): 368-375.
- [2] Lu H, Zhang M, Xu X, et al. Deep fuzzy hashing network for efficient image retrieval [J]. *IEEE Transactions on Fuzzy Systems*, 2020.
- [3] Lu, H., Member, S., Tang, Y., & Sun, Y. (2020). DRRS-BC: Decentralized Routing registration system based on blockchain. *IEEE/CAA Journal of Automatica Sinica*, 1–9.
- [4] Lu H, Zhang Y, Li Y, et al. User-oriented virtual mobile network resource management for vehicle communications[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [5] Lu H, Li Y, Mu S, et al. Motor anomaly detection for unmanned aerial vehicles using reinforcement learning [J]. *IEEE internet of things journal*, 2017, 5(4): 2315-2322.
- [6] Chen Z, Lu H, Tian S, et al. Construction of a Hierarchical Feature Enhancement Network and Its Application in Fault Recognition [J]. *IEEE Transactions on Industrial Informatics*, 2020.
- [7] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [8] H. Lu, R. Yang, Z. Deng, Y. Zhang, G. Gao, R. Lan, "Chinese image captioning via fuzzy attention-based DenseNet-BiLSTM", *ACM Transactions on Multimedia Computing Communications and Applications*, 2020.
- [9] Dong Chao, Loy Chen Change, He Kaiming, Tang Xiaoou. Image Super-Resolution Using Deep Convolutional Networks. [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 38(2).

- [10] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1874-1883.
- [11] Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637-1645.
- [12] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681-4690.
- [13] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.
- [14] Wang X, Yu K, Wu S, et al. Esrgan: Enhanced super-resolution generative adversarial networks[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 0-0.
- [15] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.
- [16] Wang X, Yu K, Dong C, et al. Recovering realistic texture in image super-resolution by deep spatial feature transform[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 606-615.
- [17] Timofte R, Rothe R, Van Gool L. Seven ways to improve example-based single image super resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1865-1873.
- [18] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2472-2481.
- [19] Guo Y, Chen J, Wang J, et al. Closed-loop matters: Dual regression networks for single image super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5407-5416.
- [20] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//Advances in neural information processing systems. 2017: 3856-3866.
- [21] Jaiswal A, AbdAlmageed W, Wu Y, et al. Capsulegan: Generative adversarial capsule network[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 0-0.
- [22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [23] 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.
- [24] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
- [25] Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks [J]. arXiv preprint arXiv:1701.04862, 2017.
- [26] Arjovsky M, Chintala S, Bottou L. Wasserstein gan[J]. arXiv preprint arXiv:1701.07875, 2017.
- [27] Cao J, Mo L, Zhang Y, et al. Multi-marginal wasserstein gan [J]. arXiv preprint arXiv:1911.00888, 2019.