**ERDBNet: Enhanced Residual Dense Block Net**  
--A New Net to Rich ESRGAN Image Details

Lizhuo Gao*

College of Physics and Electronic Engineering, ShanXi University, Taiyuan, China.

*Corresponding author e-mail: 201801102010@email.sxu.edu.cn

**Abstract.** Super resolution is applied in many digital image fields. In many cases, only a set of low-resolution images can be obtained, but the image needs a higher resolution, and then SR needs to be applied. SR technology has undergone years of development. Among them, SRGAN is the key work to introduce GAN into the SR field, which can truly restore a large number of details on the basis of low-pixel pictures. ESRGAN is a further improvement on SRGAN. By removing the BN layer in SRGAN, the effect of artifacts in SRGAN is eliminated. However, there is still a problem that the restoration of information on small and medium scales is not accurate enough. The proposed ERDBNet improve the model on the basis of ESRGAN, and use the ERDB block to replace the original RRDB block. The new structure uses a three-layer dense block to replace the original dense block, and a residual structure of the starting point is added to each dense block. The pre-trained network can reach a PSNR of 30.425 after 200k iterations, and the minimum floating PSNR is only 30.213. Compared with the original structure, it is more stable and performs better in the detail recovery of many low-pixel images.

**1. Introduction**

The single image super-resolution, which has attracted the attention of many research institutions in the world, and has been rapidly developed in recent years. SR aims to convert low-resolution images (LR) into high-resolution images (HR). Because of many applications in daily life, so it has attracted wide attention. This paper is to study this field and update the network structure on the basis of ESRGAN.

The earliest research in this field was the SRCNN network proposed by Dong [2]. It was the first time that deep learning was introduced into super-resolution. Since then, more and more people have studied and improved on his basis. A variety of new network structures and judgment standards have been introduced. After years of development of SR technology, SRGAN introduced GAN into the SR field. The most outstanding one is Xintao Wang’s project ESRGAN, which can restore low-resolution images well. ESRGAN has made further improvements in the structure of SRGAN, including improving the network structure, removing all BN layers, using five-layer convolution dense blocks and residual blocks to replace the original basic blocks to deepen the depth of the network; and The Relativistic average GAN was used instead of GAN, which uses relative authenticity to judge image quality [3].

The images processed by the ESRGAN model not only perform well on metrics such as PSNR or PI, but also eliminate artifacts on the generated images, greatly improving the quality of image restoration. The model has a good effect on detail restoration, but the performance of detail restoration at different
distances is not uniform, and the restoration of some details will be blurred. Therefore, based on ESRGAN, the proposed method improves its basic network and changes the structure. The RRDBNet is replaced by a new structure called ERDBNet, which uses a three-layer convolution dense block to replace the original five-layer dense block, and after each group of dense blocks are connected to the starting point to form a residual. The block structure allows the structure to be rich in details at multiple scales and recover better and higher-quality images.

2. Related Work
In the field of SR research, many research methods were born. Dong et al. first proposed the use of neural networks to study SRCNN in the SR field, and mainly used a three-layer convolution network to transform LR to HR [2], firstly introduced deep learning methods into the SR field. After this, many deep learning methods were born. Including FSRCNN, which mainly uses the deconvolution layer to obtain different up-sampling models, with faster training speed and wider application range [4]; and VDSR, which introduces ResNet into the model, and solves the previous deep The problem that the model cannot be trained [5]; there is also the RED model, through a symmetrical convolution and deconvolution structure, each convolution layer corresponds to a deconvolution layer, and then the convolution layer is connected to the deconvolution through a jumper Layers can solve the disappearance of gradients and restore a cleaner image [6]; and SRDenseNet, through the connection of dense blocks, connects the features of each layer, so as to finally restore the best results [7].

With the development of GAN, Ledig et al. introduced GAN into the SR field, and proposed SRGAN [1], which uses perceptual loss and confrontation loss to restore the realism of HR images. It can be seen in the thesis that GAN is used to determine the quality of photos. The authenticity is carried out through the continuous generation of judgments between the generator and the discriminator. With the birth of SRGAN, there have been important advances in the SR field. Although the images generated by SRGAN have better restoration effects than previous images, there are still many problems, such as artifacts.

In order to solve these problems, based on SRGAN, the ESRGAN model established by Xintao Wang et al. improves the original SRGAN model, solves the problem of artifacts in the original paper, and improves the quality of the final image. The ESRGAN model removes the BN layer in the network structure, and replaces the original convolution-BN-Activation function with a five-layer convolution dense block structure, the BasicBlock is changed to an RRDB structure, and the high-level architecture of the original SRGAN is retained. Because the BN layer is removed, the normalization of the mean variance of the data set features is missing, which reduces the complexity of the calculation; improves the generalization ability of the model; at the same time, it reduces the influence of artifacts caused by the BN layer, and improves the image quality [3].

In this paper, a new structure is proposed to replace the original RRDB network. The restoration effect of ESRGAN is very good, but the restoration effect on the image is not uniform enough. On the pictures with richer distance levels, the restoration at close distances is more blurred. Therefore, the new structure of the ERDB structure is used, and three three-layer convolution dense blocks and continuous residual blocks are used to train the model on multiple levels to obtain better results.

3. Proposed Methods
Like ESRGAN, the model proposed in the paper still uses the high-level architecture of SRGAN, as shown in Fig1.

![Fig. 1 Get better structural performance through the improvement of BasicBlock](image)
On the basis of this structure, ESRGAN replaces the Basic Block with an RRDB block. The structure of the RRDB block is shown in Fig. 2, and a residual structure composed of three Dense Blocks is inserted into a residual structure. Compared with SRGAN's Basic Block, it uses the Conv-BN-ReLU-Conv-BN structure. RRDB removes all BN layers, increases the convolution numbers to five, and uses dense block links, compared to the SRGAN, ESRGAN has larger network capacity, better fitting and generalization effects, and eliminates the influence of artifacts.

Fig. 2 The RRDB block is composed of a basic DenseBlock

The new Net proposed in this paper also proposes a new block called ERDB Block. Make the following changes based on the structure of SRGAN
1) Reduce the number of convolution layers in Dense Block, and use a new dense block with 3 convolution layers instead of the original one.
2) Add a residual structure to the dense blocks of each layer, connecting from the starting point to the back of each dense block.
3) The two structures mentioned in 2) are connected to form a whole. Then a new residual structure is formed.

Fig. 3 ERDB structure

The network structure of ERDBNet, still retains the high-level architecture design of SRGAN, but uses the ERDB block in Fig 3 to replace the Basic Block in SRGAN. According to the results of a number of studies [8] [9], more uses of convolution layers and links can better improve performance and increase network capacity. The use of RRDB blocks in ESRGAN also proves that dense blocks can replace the basic blocks of the BN layer to obtain better experimental results [3], especially in the elimination of artifacts. The use of residual blocks can better deepen the structure of the network, and better combine the input information of the image for comprehensive recovery; the dense block structure can effectively remove the influence of artifacts and improve the quality of the picture, and continuous dense blocks have conducive to increase network capacity and improve the effect of the model.

In summary, the basic structure of the ERDB block can be designed. The overall structure is shown in Fig. 3, and two ERDBBase structures are connected by a residual block. The 23 ERDB blocks are connected as the main structure and connected to the residual block in the high-level architecture. Each ERDBBase structure contains a continuous residual structure composed of three convolutional layer dense blocks. Each convolutional dense block contains three convolutional layers, forming an ERDB containing a total of 18 convolution layers block structure.

According to Fig3, it can be observed that the new structure uses the residual block more frequently. In the original RRDB structure, a multi-level residual network is also used, and the residual network is
used at different levels, but its main path still uses dense blocks as the main path, which also greatly increases the capacity of the network. In the new structure proposed in the paper, the residual block is used more frequently in the network, and the size of the dense block is reduced from five-layer convolution to three-layer convolution, but the number becomes denser. And the dense blocks of each part are connected by the residual network. This method improves the retention of the original information. More dense blocks also provide a basis for more detailed information recovery, and will not excessively affect network capacity to get a better response effect.

Compared with the RRDB block containing three five-layer convolutional dense blocks, ERDB uses six three-layer convolutional dense blocks. ERDBs are used more frequently for residual blocks in order to have better resilience at different scales of detail, and hope that the recovery on the entire image can be more uniform, so that the new ERDBNet can have a better recovery effect.

Except for ERDBNet's update on the network architecture, in other aspects, the processing methods of ESRGAN is retained. The discriminator uses relative discrimination, perceptual loss, and continues to use the loss function of ESRGAN. Because the purpose of this model is to improve the richness of the original image in detail, the main improvement is to improve its network architecture. The parts that originally performed well are retained, so the original discriminator and perceptual loss are still maintained, so that the new model ERDBNet can obtain a better recovery capability.

4. Experiments Details

In training, the picture magnification is 4 times. After downsampling the original HD image, the corresponding LR image is obtained. In this experiment, the batch size is set to 36. First, a PSNR model with 11 loss function is trained, which has a new network structure. The learning rate is initialized to attenuate 2 times per small batch update. The optimization algorithm selects Adam, where the first-order moment estimate is 0.9 and the second-order moment estimate is 0.99. The model training for the paper is implemented using the PyTorch and is trained using the NVIDIA 1080Ti GPU.

In the selection of data sets, the DIV2K data set [11] is mainly used, which is a data set of LR images with 2K resolution and multiple magnifications. Before training, a larger training set is obtained by processing the original data, such as cropping and rotating, to help the trained model have better adaptability.

Not only that, in order to verify the level of the evaluation model, Set5 [12] and Set14 [13] are also introduced to evaluate the data set.

Three pictures were selected as a display in the training results. The first one is 'Face' in the Set14 data set, which is used to show the effect of general image recovery capabilities. The second and third pictures are '0804' and '0834 in the DIV2K data set.' They are two images with rich details. At the same time, put HR, SRCNN, SRGAN, EDSR, ESRGAN as a comparison chart showing the results of the paper model.

**Table 1. Compare graph model results**

| methods | face from Set14 | '0804' from DIV2K | '0834' from DIV2K |
|---------|----------------|----------------|----------------|
| PSNR    | PSNR           | PSNR           | PSNR           |
| HR      | ∞              | Bicubic 23.35  | Bicubic 22.05  |
| Bicubic | 30.00          | MSRGAN 24.90   | MSRGAN 23.52   |
| MSRGAN  | 31.82          | EDSR 24.92     | EDSR 23.24     |
| EDSR    | 25.80          | ESRGAN 22.39   | ESRGAN 22.99   |
| ESRGAN  | 24.13          | ERDBNet 22.65  | ERDBNet 21.12  |
| ERDBNet | 29.67          |                |                |

The table shows the local image PSNR for the small images in the fig 4
It can be seen from the results that, compared with ESRGAN, ERDBNet is similar in the recovery of general photos, but in other pictures with larger depth, ERDBNet has better recovery capabilities.

5. Conclusion
In this paper, the ERDBNet was proposed. This model made changes to the network structure based on the ESRGAN model to form the current ERDBNet. The ERDB block uses a three-layer convolutional layer and a more complex dense block link. The experimental results also show that the generated image is more uniform on multiple scales and has a better performance in detail recovery.
References

[1] Ledig, C., Theis, L., Husz’ar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image superresolution using a generative adversarial network. In: CVPR. (2017)

[2] Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV. (2014)

[3] 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) workshops. 2018: 0-0.

[4] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network [C]//European conference on computer vision. Springer, Cham, 2016: 391-407.

[5] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.

[6] Mao X J, Shen C, Yang Y B. Image restoration using convolutional auto-encoders with symmetric skip connections [J]. arXiv preprint arXiv:1606.08921, 2016.

[7] Tong T, Li G, Liu X, et al. Image super-resolution using dense skip connections [C]//Proceedings of the IEEE international conference on computer vision. 2017: 4799-4807.

[8] 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.

[9] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution [C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 136-144.

[10] Kingma D P, Ba J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv:1412.6980, 2014.

[11] Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: Dataset and study [C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 126-135.

[12] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [J]. 2012.

[13] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [J]. 2012.