

# Super-resolution Image Reconstruction Based on Double Regression Network Model

Jieyi Lv

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, 710021, China  
e-mail:ljyly150@163.com

Zhongsheng Wang

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, 710021, China  
e-mail:wzhsh1681@163.com

**Abstract**—By learning nonlinear mapping functions from low resolution (LR) images to high resolution (HR) images, deep neural networks show good performance in image super-resolution (SR). However, the existing SR approach has two potential limitations. First, learning the mapping function from LR to HR images is usually an ill-conditioned problem, since there exist an infinite number of HR images that can be down-sampled to the same LR image. Thus, the space of possible functions can be very large, making it difficult to find a good solution. Second, paired LR-HR data may not be available in real-world applications, and the underlying degradation method is often unknown. For this more general case, existing SR models tend to generate adaptive problems and produce poor performance. To solve the above problem, we propose a dual regression scheme that reduces the space of possible functions by introducing additional constraints on LR data.

**Keywords**—*Super Resolution; Mapping Function; Deep Neural Network; Double Regression Model*

## I. INTRODUCTION

Image super-resolution reconstruction is to reconstruct high resolution image from low resolution image. This typical problem has been widely used in astronomy, physics, medicine and other fields. In the initial super-resolution

reconstruction, interpolation methods are used, including nearest neighbor interpolation, bilinear interpolation, bicubic interpolation and so on. The interpolation method can effectively enhance the image resolution with a small amount of computation, but simply increasing the number of pixels will lead to blurred edges of the image. In addition, methods such as image super-resolution, local linear regression, dictionary learning and random forest based on sparse coding have also been widely used in many fields.

Single image super resolution is an ill-conditioned inverse problem: one LR image can correspond to multiple HR images, and the reconstructed HR images often have defects such as detail loss, edge aliasing and blurring. Deep learning technology has greatly promoted the rapid development of the field of computer vision. The vast majority of current SISR algorithms are based on end-to-end deep learning technology, that is, they directly learn the mapping between LR and HR. Although deep learning-based SISR method has made great progress, it still has problems in practical application: better results tend to rely on deeper networks, and more parameters also require more training data. This requires longer training and reasoning times, as

well as greater computing power and memory. As a result, its usefulness is greatly limited, especially in resource-constrained mobile devices.

In this paper, a new dual regression scheme is proposed to form a closed loop to improve SR performance. We introduce an additional constraint to reduce the possible space so that the super-resolved image can reconstruct the input LR image. Ideally, if the mapping from LR→HR is optimal, the super resolution image can be down-sampled to obtain the same input LR image. With such constraints, we can estimate the underlying down-sampling kernel, thereby reducing the space of possible functions and finding a good mapping from LR to HR. The specific scheme is shown in Figure 1

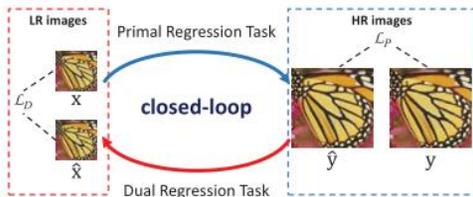


Figure 1 Composition of double regression model

## II. RELATED WORK

### A. Attention Mechanism

Attention mechanisms stem from studies of the human visual system. Human vision tends to focus on salient areas and ignore useless information, which increases the efficiency with which the brain processes information. The compression-expansion module proposed by Hu et al. using the attention mechanism can explicitly model the interdependence between feature channels, automatically obtain the importance of each feature channel through learning, and then suppress the less useful features according to the importance, so as to improve the performance of image classification network. The RCAN method proposed by Zhang et al. introduces the attention

model of Hu et al. The channel attention module can adaptively select feature channels with richer information, which improves the performance of SISR. However, the feature channels of this attention mechanism are independent of each other, which limits the flow of feature information between channels.

### B. Recursive Learning

Increasing the network depth can generally improve the performance of SISR. However, deeper networks, which generally have more parameters, require more training data. In practice, the acquisition of training data is often limited, and the risk of overfitting of network training is also increasing. One advantage of a recursive network is that it can increase the depth of the network without increasing the number of parameters. DRCN introduces recursion techniques into SISR methods for the first time: a convolutional layer is used as a recursion unit, and the weights are shared among the recursion units. DRRN improves the use of recursive techniques: residual blocks are used as recursive units, and parameters are shared among residual blocks, which also improves the performance of DRCN. Similarly, MemNet, proposed by Tai et al., also uses residual blocks as recursive units to build the network.

## III. RESEARCH CONTENTS OF THE THESIS

### A. Theoretical basis of dual regression model

The image super-resolution under the dual regression model focuses on the input LR image to achieve the purpose of restoring HR image. In principle, LR image  $I_{xLR}$  can be expressed as the output of HR image after degradation function.

$$I_{xLR} = d(I_{yHR}, \partial) \quad (1)$$

Where  $d$  is the SR degradation function responsible for converting HR image into LR image,  $I_{yHR}$  Is the input HR image, which is the reference image, and the input parameters that describe the degradation function of the image. The degradation parameters are usually the scaling coefficient, blur type, and noise. In practice, the degradation process and the dependent parameters are unknown and usually only LR images are used to obtain HR images via SR methods. The SR process is responsible for predicting the inverse of the degradation function  $d$ .

$$g(I_{xLR}, \delta) = d^{-1}(I_{xLR}) = I_{yE} \approx I_{yHR} \quad (2)$$

Where  $g$  is the SR function and describes the input parameters of the function  $g$ ,  $I_{yE}$  is corresponds to the input  $I_{xLR}$  of the estimated HR image. It is also worth noting that the super-resolution function in Eq. (2) is ill-posed because the function  $g$  is a non-mapping function. As a result,  $I_{yE}$  not the only one. There are many possibilities. The degradation process of the input LR image is unknown, and this process is affected by many factors such as sensor-induced noise, artifacts due to lossy compression, speckle noise, motion blur, and misfocused images. In most studies, a single down sampling function is used as the image degradation function.

$$d(I_{yHR}, \partial) = (I_{yHR}) \downarrow_{S_f}, \{s\} \subseteq \partial \quad (3)$$

Thereinto,  $\downarrow_{S_f}$  represents the down-sampling operation,  $S_f$  Is the sampling factor. One of the down-sampling functions often used in SR is

bicubic interpolation with antialiasing. In order to simulate a more realistic environment, researchers use more operations in the down-sampling function, and the overall down-sampling operation is:

$$d(I_{yHR}, \partial) = (I_{yHR} \otimes \kappa) \downarrow_{S_f} + n_{\sigma}, \{\kappa, s, \sigma\} \subseteq \partial \quad (4)$$

Where  $I_{yHR} \otimes \kappa$  represents HR image,  $I_{yHR}$  convolution  $\kappa$  with fuzzy kernel,  $n_{\sigma}$  Is additive white Gaussian noise with  $\sigma$  standard deviation. The degradation function defined in Eq. (4) is closer to the actual function because it considers more parameters than the simple down-sampled degradation function. The purpose of image super resolution is to minimize the loss function, as follows:

$$\hat{\phi} = [\min L(I_{yE}, I_{yHR})]_{\phi} + h\psi(\phi) \quad (5)$$

Where  $L(I_{yE}, I_{yHR})$  is the loss function between the output HR image after SR and the actual HR image,  $\psi(\phi)$  Is the regularization term. The most commonly used loss function in SR is based on the pixel mean square error, which can also be called the pixel loss. Figure 2 is a schematic diagram of the super-resolution reconstruction process.

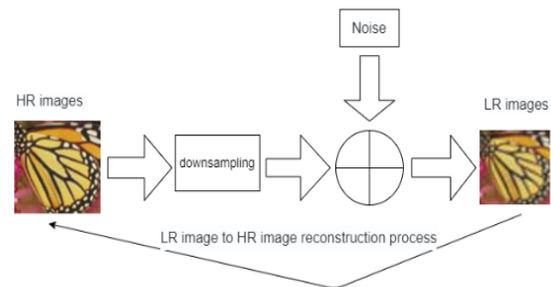


Figure 2 Super resolution reconstruction process

### B. Recursive learning

One of the learning strategies based on the basic network is to recursively learn high-level features using the same module. This approach also minimizes model parameters, since the strategy requires only one module to be recursively updated, as shown in Figure 3.

One of the most commonly used recursive networks is the deep recursive convolutional network. With one convolutional layer, the DRCN can reach a receptive field of  $41 \times 41$  without additional parameters. The deep recursive residual network uses the residual module Res Block as a part of the recursive module, which has a total of 25 recursions and performs better than the baseline Res Block. In addition to end-to-end recursion, the researchers also used a two-state recursive network, which shares signals between LR images and generated HR image states within the network. In general, while reducing parameters, recursive learning networks can learn complex representations of data at the cost of computational performance. In addition, increased computational requirements may cause gradients to explode or disappear. Therefore, recursive learning is often used in conjunction with multi-supervised or residual learning to minimize the risk of gradient explosion or disappearance.

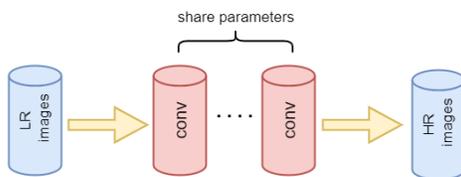


Figure 3 Recursive learning

### C. Dual regression scheme for paired data

Existing methods only focus on learning the mapping from LR to HR images. However, the

space of possible mapping functions can be very large, making training very difficult. To solve this problem, we propose a dual regression scheme by introducing an additional constraint on LR data. Specifically, in addition to learning the LR→HR mapping, we also learn the inverse/dual mapping from the super resolution image to the LR image. We learn to reconstruct HR images with the original map  $P$  and LR images with the dual map  $D$  simultaneously. Note that the dual map can be viewed as an estimate of the underlying down-sampled kernel. Formally, we formulate the SR problem as a binary regression scheme involving two regression tasks. The specific expression form is shown in Figure 1.

The original regression task: We need to find a function  $P: X \rightarrow Y$ . Make the prediction  $P(x)$  similar to its corresponding HR image  $Y$ .

Dual regression task: We seek a function  $D: Y \rightarrow X$ , so that the prediction of  $D(y)$  is similar to the original input LR image  $x$ .

Primal and dual learning tasks can form a closed loop to provide information supervision for training models  $P$  and  $D$ . If  $P(x)$  is the correct HR image, then the down-sampled image  $D(P(x))$  should be very close to the input LR image  $x$ . With this constraint, we can reduce the space of functions of possible maps, making it easier to learn better maps to reconstruct HR images. By jointly learning these two learning tasks, we propose to train the super-resolution model as follows. Given  $N$  pairs of samples  $S_p = \{(x_i, y_i)\}_{i=1}^N$

Where  $x_i$  and  $y_i$  denotes those in the paired data set  $i$  for low resolution and high resolution images. The training loss can be written as

$$\sum_{i=1}^N \zeta_p(P(x_i), y_i) + \lambda \zeta_D(D(P(x_i)), x_i) \quad (6)$$

Where  $\zeta_p$  and  $\zeta_D$  represent the loss functions of primal and dual regression tasks, respectively. Here,  $\lambda$  controls the weight of the dual regression loss.

#### D. Dual regression scheme for unpaired data

We consider a more general SR case where there is no corresponding HR data with real-world LR data. More critically, the degradation methods of LR images are often unknown, which makes this problem very challenging. In this case, existing SR models often produce serious adaptation problems. To solve this problem, we propose an efficient algorithm to adapt the SR model to the new LR data. Dual regression maps learn the underlying degradation methods and do not necessarily depend on HR images. Therefore, we can directly learn from unpaired real-world LR data for model adaptation. In order to ensure the performance of HR image reconstruction, we also add the information of pairwise synthetic data which is very easy to obtain. Given  $M$  unpaired LR samples and  $N$  paired synthetic samples, the objective function can be expressed as follows.

$$\sum_{i=1}^{M+N} I_{S_p}(x_i) \zeta_D(P(x_i), y_i) + \lambda \zeta_D(D(P(x_i)), x_i) \quad (7)$$

Where,  $I_{S_p}(x_i)$  is an index function, if  $x_i \in S_p$ , the function is equal to 1, otherwise the function is equal to 0.

#### IV. EXPERIMENT AND RESULTS

The hardware environment of this section is as follows: CPU frequency is 3.6GHz, memory is 32GB, graphics card is NVIDIA RTX 2080Ti; The software environment is: Windows operating system, Pytorch deep learning framework,

Python3.7 programming language. In the experiment, 800 high-quality training images from the public dataset DIV2K were used as the training set. In the test phase, Set5, Set14 and BSD100 were used as the standard test set, which all contained various animals and plants as well as natural scene images.

Before training, the training set is enhanced by horizontal flip and Angle rotation. Considering that putting a single image directly into the model would increase the amount of computation, which is not convenient for training, each image was randomly cropped to obtain  $256 \times 256$  image blocks. LR images in the training set and test set were obtained by bicubic interpolation in Matlab language. In order to evaluate the model performance, the PEak-to-noise ratio (PSNR) and structural similarity are calculated on the Y channel (i.e., brightness) of the YCbCr channel in this experiment. The algorithm in this paper has two models in total. The generative network model Ours P for PSNR value generates the final model Ours for adversarial network training. The experimental results of the open datasets Set5, Set14 and BSD100 are respectively compared with those of the classical algorithms FSRCNN, SRGAN and ESRGAN. In the experiment, the number of modules in the intermediate feature layer of SRGAN and ESRGAN is set to 24, which is consistent with the original paper. The number of ECB modules in the proposed algorithm is set to 8, and the number of input channels is 24. The PSNR and SSIM values and the number of parameters at  $\times 2$  and  $\times 4$  magnifications were compared respectively.

Considering that the training of generative adversarial network should go through two training processes, in order to save time, only the generative network model Ours oriented to PSNR value is trained. It can be seen that the deeper the

network, the higher the objective evaluation index, but at the same time, the longer the training time. At the same time, it can be seen from the table that the model continues to increase the number of network layers in a deeper situation, which does

not significantly improve the objective evaluation index, but increases the number of model parameters. Therefore, the algorithm in this paper is finally set as 8.

TABLE I. COMPARISON OF THE RESULTS OF PSNR RECONSTRUCTION BY TWO TIMES OF EACH ALGORITHM

Datasets	FSRCNN	SRGAN	ESRGAN	Ours
Set5	32.40	33.19	33.46	34.68
Set14	29.52	29.56	29.47	30.56
Bsd100	26.74	29.47	29.78	29.78

TABLE II. COMPARISON OF SSIM RESULTS OF FOUR TIMES RECONSTRUCTION OF EACH ALGORITHM

Datasets	FSRCNN	SRGAN	ESRGAN	Ours
Set5	0.8657	0.8657	0.8569	0.6895
Set14	0.7564	0.7369	0.8697	0.7698
Bsd100	0.7156	0.6689	0.6783	0.6856

Figure 4 and 5 shows the super-resolution image reconstructed by the proposed method and other advanced methods at  $\times 4$  magnification. The reconstruction results obtained by the proposed method are significantly better than those obtained by other methods. The first group of pictures recovered by FSRCNN method are blurred and have serious distortion; the second group of pictures recovered by SRGAN method have serious double shadow phenomenon around the branches; and the first group of pictures recovered by ESRGAN method also have serious blur phenomenon. The proposed method almost perfectly recovers the images under  $\times 4$  magnification. The recovery results of this method are similar. Compared with the current state-of-the-art, more lightweight network, the

proposed method can better recover the image contour in the Set14 dataset at  $\times 4$  magnification. It is easy to observe that the reconstruction results in this paper are also significantly better than the other two cutting-edge methods.

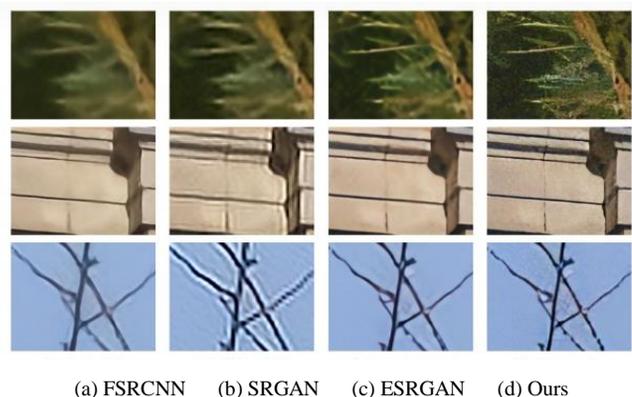
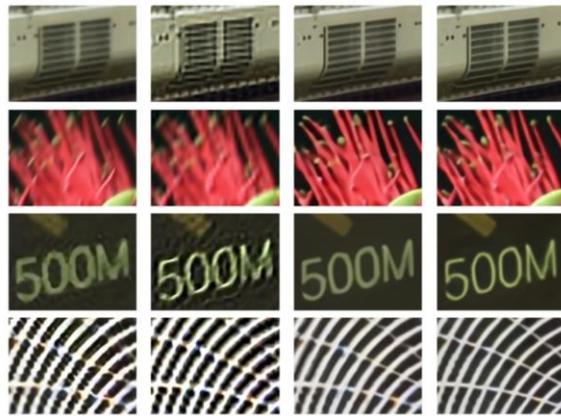


Figure 4 Double magnification reconstruction comparison of each algorithm



(a) FSRCNN (b) SRGAN (c) ESRGAN (d)Ours

Figure 5 Comparison of four times magnification reconstruction of each algorithm

## V. SUMMARY

This paper presents a new binary regression scheme for paired and unpaired data. On pairwise data, we introduce an additional constraint by reconstructing LR images to reduce the space of possible functions. As a result, we can significantly improve the performance of SR models. In addition, we focus on unpaired data and apply the dual regression scheme to real-world data. We performed ablation studies with the dual regression protocol, and the model with the dual regression protocol performed better on all datasets compared to baseline. These results show that the dual regression scheme can improve the reconstruction of HR images by introducing additional constraints to reduce the space of mapping functions. We also evaluated the influence of our dual regression scheme on other models. Compared with other classical algorithms, we compared PSNR results, SSIM results and intuitive dataset images from two levels of double and four times magnification. It can be seen that the improved algorithm is significantly better than the classical algorithm. Due to the large number of super-resolution scenes, we only selected the data sets that were biased towards buildings, trees,

numbers and other relevant directions. There may be some omission problems in the algorithm adaptation of medical and other professional fields, which still needs to be improved in the subsequent research.

## REFERENCES

- [1] Lee S, Kim JH, Heo J P. Super-Resolution of License Plate Image via Character-Based Perceptual Loss[C]//2020 IEEE International Conference on Big Data and Smart Computing. IEEE, 2020:560-563.
- [2] Yong Guo, Yin Zheng, Mingkui Tan, Qi Chen, Jian Chen, Peilin Zhao, and Junzhou Huang. Nat: Neural architecture transformer for accurate and compact architectures. In *Advances in Neural Information Processing Systems*, pages 735–747, 2019, 1.
- [3] Jiezhong Cao, Yong Guo, Qingyao Wu, Chunhua Shen, Junzhou Huang, and Mingkui Tan. Adversarial learning with local coordinate coding. In *International Conference on Machine Learning*, 2018, 2.
- [4] Tai Y, Yang J, Liu X, et al. Memnet: A Persistent Memory Network for Image Restoration[C] //Proceedings of the IEEE international conference on computer vision. 2017: 4539-4547.
- [5] Zhang Y, Tian Y, Kong Y, et al. Residual Dense Network for Image Super-Resolution[C]//Proc eedings of the IEEE conference on computer vision and pattern recognition. 2018: 2472-2481.
- [6] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, 6.
- [7] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwang-gil Jeon, and Wei Wu. Feedback network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2019, 2.
- [8] Lin Wang, Wonjune Cho, and Kuk-Jin Yoon. Deceiving image-to-image translation networks for autonomous driving with adversarial perturbations. *arXiv preprint arXiv: 2001.01506*, 2020, 2.
- [9] Li Zhang, Alok Deshpande, and Xin Chen. Denoising vs. deblurring: Hdr imaging techniques using moving cameras. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 522–529. IEEE, 2010.
- [10] Timo Stoffregen and Lindsay Kleeman. Event cameras, contrast maximization and reward functions: an analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12300–12308, 2019, 2.
- [11] Christian Reinbacher, Gottfried Munda, and Thomas Pock. Real-time panoramic tracking for event cameras. In *2017 IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2017, 1, 2.
- [12] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018, 5.