A REVIEW OF SINGLE IMAGE SUPER-RESOLUTION RECONSTRUCTION BASED ON DEEP LEARNING

Qiuyue Deng[1], Lei Yu[2] and Haicheng Hu[3]

[1]Master, Dept. of Information and Communication Eng., Harbin Engineering University
(145 Nantong street, Nangang District, Harbin150000, China)
E-mail: doreenyue@hrbeu.edu.cn

[2]Professor, Dept. of Information and Communication Eng., Harbin Engineering University
(145 Nantong street, Nangang District, Harbin150000, China)
E-mail: yulei@hrbeu.edu.cn

[3]Master, Dept. of Information and Communication Eng., Harbin Engineering University
(145 Nantong street, Nangang District, Harbin150000, China)
E-mail: czgg@hrbeu.edu.cn

Image super-resolution (SR) is an important task in the field of computer vision. The introduction of deep learning (DL) makes a breakthrough in the field of image super-resolution. In order to better grasp the current research hotspot, this paper focuses on combing the single image super-resolution (SISR) methods based on DL. Firstly, according to different baseline networks, the classic SISR models based on DL is classified. Then it summarizes several popular datasets, evaluation indicators, and the experimental results of these models. Finally, the current challenges and future trends in the field of SISR are analyzed.

Key Words : *single image super-resolution, deep learning, convolutional neural network, residual learning, recursive learning, generative adversarial net*

1. INTRODUCTION

Image resolution is used to measure the quality of the image, which reflects the ability of the image to express the detail information. In general, the higher the resolution of the image, the richer the details of the image. The purpose of image super-resolution reconstruction is to recover high-resolution (HR) image from low-resolution (LR) image and increases the details of the image. SISR is a basic task in the field of computer vision, which is not only widely used in medical imaging, image transmission and preservation, and security and monitoring, but also helps to improve other visual tasks: target tracking and detection, image segmentation and so on. It has attracted more and more scholars to study and explore. The basic idea of the early traditional methods is to extract features from LR images, then carry out feature information registration, target region fusion and redundant information elimination, and finally synthesize HR images. SISR is an ill-posed problem because multiple HR images can degenerate into the same LR images, which is very challenging. In the face of these challenges, the traditional algorithms are difficult to make a breakthrough. In recent years, DL has achieved rapid development, and its powerful feature extraction ability is very suitable for the field of computer vision. A large number of researchers have introduced DL into the field of SISR, and achieved excellent results.

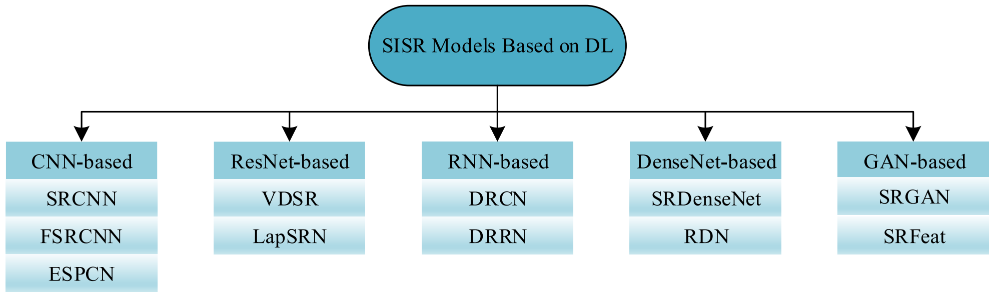

Fig.1 classification diagram of SISR models based on DL.

2. INTRODUCTION OF SISR MODELS BASED ON DL

According to baseline networks, SISR models based on DL can be divided into: convolutional neural network-based (CNN-based) models, residual network-based (ResNet-based) models, recursive neural network-based (RNN-based) models, densely connected network-based (DenseNet-based) models and generative adversarial network-based (GAN-based) models. Figure 1 shows the classification structure.

(1) CNN-based models

The pioneering work of DL in the field of SISR is the pre-upsampling SR framework (SRCNN) proposed by Dong et al.[1]. Its structure is divided into feature extraction part, nonlinear mapping part and reconstruction part. SRCNN establishes the basic structure of SISR based on DL. SRCNN achieves better reconstruction performance than traditional algorithms, but its input is preprocessed by interpolation, which increases the amount of network computation. In view of the shortcomings of SRCNN, Dong et al. improved it and proposed the post-upsampling SR framework (FSRCNN)[2]. The input of FSRCNN is the original image, which is reconstructed by deconvolution at the end of the network. The end-to-end learning mode reduces the computational complexity of the network, improves the training speed, and realizes the lightweight and efficient network performance. Shi et al. proposed efficient sub-pixel convolutional neural network (ESPCN)[3] based on sub-pixel convolution, which is also performs upsampling at the end of the network. The essence of sub-pixel convolution is pixel rearrangement, which is similar to interpolation. The difference is that image interpolation only uses the pixel values in the field of the position to be interpolated, while sub-pixel convolution is to learn an interpolation function.

(2) ResNet-based models

Kim et al. proposed that HR and LR images share most of the information, but LR images lack part of the high-frequency information. So the network only needs to learn the lost high-frequency information. Based on this idea, Kim et al. proposed a model named very deep convolutional network (VDSR)[4]. Thanks to residual learning, VDSR is the first network with a depth of 20 layers. Lai et al. also used residual learning to propose laplacian pyramid super-resolution network (LapSRN)[5]. Its structure is similar to pyramid, which is divided into feature extraction branch and reconstruction branch. The feature extraction branch is composed of convolution layer and deconvolution layer. The convolution layer learns residual features, and the deconvolution layer upsamples these residual features. The reconstruction branch uses deconvolution layer to upsample the input image. LapSRN realizes multi-scale SISR and achieves excellent results in large scaling factors.

(3) RNN-based models

In order to ensure the quality of reconstruction and reduce the complexity of the network, recursive learning is introduced into the field of SISR. Kim et al. proposed the deeply recursive convolutional network (DRCN)[6] based on recursive learning, which uses a single convolution layer as recursive unit. It can get larger receptive field and learn higher level features with less parameters. Inspired by DRCN, Tai et al. proposed the deep recursive residual network (DRRN)[7], whose main branch is a stack of recursive modules. The global residual learning and local residual learning are combined. DRRN realizes the depth and simplicity of the network.

(4) DenseNet-based models

Since Huang et al. proposed DenseNet[8], dense connection is more and more widely used in network design. Tong et al introduced dense connection into SISR domain for the first time [9]. The model uses stacked dense modules to construct a 69-layer network, and also inserts dense connections between different dense modules. Zhang et al. proposed residual dense network (RDN)[10] by combining residual connection and dense connection. Dense connection not only helps to alleviate the gradient disappearance phenomenon of very deep network, but also can integrate the characteristics of each layer of the network to provide more abundant information.

(5) GAN-based models

Although the existing SISR networks have achieved better and better reconstruction performance, and the evaluation metrics have been greatly improved, the reconstructed images still have the phenomenon of blurred local details, which is not comfortable for people's visual experience. The generative adversarial network (GAN), proposed by Goodfellow et al.[11] in 2014, is often used to solve this kind of phenomenon. Ledig et al. firstly introduced GAN into the field of SISR, and proposed the model named SRGAN[12]. Park et al. improved SRGAN and proposed the model named SRFeat[13]. Bulat et al.[14] proposed two GANs for face SR. The SISR models based on GAN can still reconstruct clear details in the case of large scaling factors.

3. EXPERIMENTAL RESULTS AND ANALYSIS

(1) Datasets

DIV2k[15] contains 1000 high-definition natural images with resolution of about 2K. Among these images, 800 images are used for training, 100 images for validation and the rest of 100 images for testing. Public test datasets are Set5[16], Set14[17], BSD100[18], Urban100[19] and Manga109[20]. Set5 and Set14 are both classic color datasets. Set5 consists of only five images. Compared with Set5, Set14 has 14 images with more categories. BSD100 contains 100 color images of different scenes. Urban100 is also composed of 100 color images. It is a relatively new dataset, which focuses on urban scenes and building structures. Manga109 is a dataset containing 109 cartoons.

(2) Image quality assessment

Image quality assessment methods are divided into objective methods and subjective methods. Objective methods mainly use some metrics to evaluate image quality quantitatively. The commonly used metrics are peak signal-to-noise ratio and structural similarity.

a) Peak signal-to-noise ratio

Peak signal-to-noise ratio (PSNR) is used to measure the similarity of two images from the perspective of pixels. It is often used to measure the reconstruction quality of lossy images. In general, the higher the PSNR, the higher the quality of the reconstructed images. It is defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}((I(i) - \hat{I}(i)))^2 \tag{1a}$$

$$PSNR = 10\log_{10}(\frac{L^2}{MSE}) \tag{1b}$$

Where MSE is the mean square error between HR image I and reconstructed image $\hat{I}$, and L represents the maximum possible pixel value.

b) Structural similarity

Structural similarity (SSIM) is to evaluate the quality of reconstructed images from the statistical relationship between two images. In general, the closer the value of SSIM is to 1, the higher the quality of reconstructed images is. It is defined as follows:

$$SSIM\left(I,\hat{I}\right)=\frac{(2\mu_I\mu_{\hat{I}}+C_1)(\sigma_{I\hat{I}}+C_2)}{(\mu_I^2+\mu_{\hat{I}}^2+C_1)(\sigma_I^2+\sigma_{\hat{I}}^2+C_2)} \tag{1c}$$

Where, μ_I and $\mu_{\hat{I}}$ represents the average image intensity of HR image I and reconstructed image $\hat{I}$ respectively. $\sigma_{I\hat{I}}$ represents the covariance between I and $\hat{I}$. σ_I^2 and $\sigma_{\hat{I}}^2$ represents the variance of I and $\hat{I}$ respectively. $C_1=(K_1L)^2$ and $C_2=(K_2L)^2$ are constants used to avoid instability. Where K_1 and K_2 are constants far less than 1, and L is the maximum possible pixel value.

(3) Comparison of experimental results

The introduction of DL has brought a breakthrough in the field of SISR, and the reconstruction performance has been continuously improved. This survey introduces the SISR models based on five classical networks. Table 1 shows the reconstruction results of each classical model on five public datasets (scaling factors are x2, x3, and x4). The results of state-of-the-art methods involved are cited from their papers. It can be seen from table 1 that the value of the evaluation metrics is constantly increasing (the highest PSNR / SSIM value has been thickened), indicating that the reconstruction performance is constantly improving. Figure 2 shows the reconstructed images of some networks. In order to better compare the results, the capture area of the original image and the reconstructed images (the red box area in the picture) are enlarged.

Table 1 The reconstruction results of each network on the public datasets.

Method	Scale	Set5	Set14	BSD100	Urban100	Manga109
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
SRCNN	x2	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663
FSRCNN		37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9710
ESPCN		37.00/0.9559	32.75/0.9098	31.51/0.8939	29.87/0.9065	36.21/0.9694
VDSR		37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9750
LapSRN		37.52/0.9591	33.08/0.9130	31.80/0.8952	30.41/0.9103	37.27/0.9740
DRCN		37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.55/0.9732
DRRN		37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.88/0.9749
RDN		**38.24/0.9614**	**34.01/0.9212**	**32.34/0.9017**	**32.89/0.9353**	**39.18/0.9780**
SRCNN	x3	32.75/0.9090	29.30/0.8215	28.41/0.7863	26.24/0.7989	30.48/0.9117
FSRCNN		33.18/0.9140	29.37/0.8240	28.53/0.7910	26.43/0.8080	31.10/0.9210
ESPCN		33.02/0.9135	29.49/0.8271	28.50/0.7937	26.41/0.8161	30.79/0.9181
VDSR		33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9340
DRCN		33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.24/0.9343
DRRN		34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.71/0.9379
RDN		**34.71/0.9296**	**30.57/0.8468**	**29.26/0.8093**	**28.80/0.8653**	**34.13/0.9484**
SRCNN	x4	30.48/0.8626	27.50/0.7513	26.90/0.7101	24.52/0.7221	27.58/0.8555
FSRCNN		30.72/0.8660	27.61/0.7550	26.98/0.7150	24.62/0.7280	27.90/0.8610
ESPCN		30.66/0.8646	27.71/0.7562	26.98/0.7124	24.60/0.7360	27.70/0.8560
VDSR		31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8870
LapSRN		31.54/0.8852	28.19/0.7720	27.32/0.7280	25.21/0.7562	29.09/0.8900
DRCN		31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.93/0.8854
DRRN		31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.45/0.8946
SRDenseNet		32.02/0.8934	28.50/0.7782	27.53/0.7337	26.05/0.7819	- / -
RDN		**32.47/0.8990**	**28.81/0.7871**	**27.72/0.7419**	**26.61/0.8028**	**31.00/0.9151**

Fig.2 Visual comparison between different models.

4. DEVELOPMENT DIRECTIONS IN FUTURE

(1) Lightweight networks

With the continuous improvement of reconstructed image quality, the structure of the networks is more and more complex, and the models have more and more parameters, which not only the requirement of hardware facilities is higher and higher, but also is very difficult to deploy on mobile devices, making the landing application of SR very difficult. In recent years, some lightweight models have been proposed. For example, Hui et al.[21] proposed the information distillation network (IDN), which effectively saves memory space and computation. Inspired by IDN[21], Hui et al.[22] further improves IDN and proposes the multi-information distillation network (IMDN). Liu et al.[23] made a deeper thinking on the basis of IMDN, introduced local residual connection, and proposed residual feature distillation mechanism. Reducing the complexity of networks as much as possible without affecting the reconstruction performance has become a major research direction in the field of SISR.

(2) Towards real-world scenarios

At present, most of the SISR networks are used to reconstruct LR images with special degradation. Usually, the existing HR images are used to get the corresponding LR images by bicubic interpolation, so as to obtain the image pairs for training. What the network learns is the reconstruction function in bicubic interpolation degradation mode. But for those images with complex or unknown degradation mode, the reconstruction effect is not ideal. However, it is more common in real life that the degradation mode is unknown. How to make the SISR models suitable for the more common degradation mode has become a research difficulty in the field of SISR.

(3) Scaling factors

Most of the SISR networks use deconvolution layer or sub-pixel convolution layer to achieve the final reconstruction. Because the deconvolution layer is easy to cause chessboard effect, most of the networks use sub-pixel convolution layer. The sub-pixel convolution layer can achieve different magnification by controlling different number of feature channels. With the increase of scaling factor, the reconstruction process becomes more and more difficult, and the metrics of reconstruction results also deteriorate. Therefore, few researches set the scaling factor higher than 8. How to keep accurate high-frequency detail in the reconstruction images is a challenge in the field of SISR under extreme upsampling conditions, such as 16 times or 32 times.

5. SUMMARY AND PROSPECT

SISR plays a very important role in real life, and also has a very important auxiliary role for other visual tasks, so SISR is a very significant research topic. In this paper, we summarize most of the existing classical models of SISR based on DL, and introduce the network structure of these classical models. The difficulties, and future directions in the field of SISR are analyzed, which provides some ideas for future research.

ACKNOWLEDGMENT: The work is supported by NSFC (61771155) and Fundamental Research Funds for the Central Universities. And thanks a lot to Miss Zheng Liying for her brilliant help to the research.

REFERENCES

1) Chao, D., Chen, C. L., He, K. and Tang, X. : Learning a deep convolutional network for image super-resolution, *European Conference on Computer Vision*, 2014.

2) Chao, D., Chen, C. L. and Tang, X. : Accelerating the super-resolution convolutional neural network, *Springer, Cham*, 2016.

3) Shi, W., Caballero, J., Huszár, F., Totz, J. and Wang, Z. : Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874–1883, 2016.

4) Kim, J., Lee, J. K. and Lee, K. M. : Accurate image super-

resolution using very deep convolutional networks, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646–1654, 2016.

5) Lai, W. S., Huang, J. B., Ahuja, N. and Yang, M. H. : Deep laplacian pyramid networks for fast and accurate super-resolution, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5835–5843, 2017.

6) Kim, J., Lee, J. K. and Lee, K. M. : Deeply-recursive convolutional network for image super-resolution, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1637–1645, 2016.

7) Ying, T., Jian, Y. and Liu, X. : Image super-resolution via deep recursive residual network, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2790–2798, 2017.

8) Huang, G., Liu, Z., Laurens, V. and Weinberger, K. Q. : Densely connected convolutional networks, *IEEE Computer Society*, pp. 4700–4708, 2016.

9) Tong, T., Li, G., Liu, X. and Gao, Q. : Image super-resolution using dense skip connections, *IEEE International Conference on Computer Vision*, pp. 4809–4817, 2017.

10) Zhang, Y., Tian, Y., Kong, Y., Zhong, B. and Fu, Y. : Residual dense network for image super-resolution, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2472–2481, 2018.

11) Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., D Warde-Farley, Ozair, S., Courville, A. and Bengio, Y. : Generative adversarial networks, *Neural Information Processing Systems*, pp. 2672-2680, 2014.

12) Ledig, C., Theis, L., F Huszar, Caballero, J., Cunningham, A. and Acosta, A. : Photo-realistic single image super-resolution using a generative adversarial network, *IEEE Computer Society*, pp. 105−114, 2016.

13) Park, S. J., Son, H., Cho, S., Hong, K. S. and Lee, S. : SRFeat: single image super-resolution with feature discrimination, *European Conference on Computer Vision*, pp. 455−471, 2018.

14) Bulat, A., Jing, Y. and Tzimiropoulos, G. : To learn image super-resolution, use a gan to learn how to do image degradation first, *European Conference on Computer Vision*, pp. 187-202, 2018.

15) Agustsson, E. and Timofte, R. : NTIRE 2017 challenge on single image super-resolution: dataset and study, *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1110–1121, 2017.

16) Bevilacqua, M., Roumy, A., Guillemot, C. and Morel, A. : Low-complexity single image super-resolution based on nonnegative neighbor embedding, *British Machine Vision Conference*, 2012.

17) Zeyde, R., Elad, M. and Protter, M. : On single image scale-up using sparse-representations, *the 7th International Conference on Curves and Surfaces. Springer*, *Berlin*, *Heidelberg*, 2010.

18) Martin, D., Fowlkes, C., Tal, D. and Malik, J. : A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, *the 8th IEEE International Conference on Computer Vision*, *Vancouver*, 2001.

19) Huang, J. B., Singh, A. and Ahuja, N. : Single image super-resolution from transformed self-exemplars, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5197–5206, 2015.

20) Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y. and Aizawa, K. : Manga109 dataset and creation of metadata, *the 1st International Workshop. ACM*, 2016.

21) Zheng, H., Wang, X. and Gao, X. : Fast and accurate single image super-resolution via information distillation network, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 723–731, 2018.

22) Hui, Z., Gao, X., Y Yang and Wang, X. : Lightweight image super-resolution with information multi-distillation network, *ACM Multimedia*, pp. 2024–2032, 2019.

23) Liu, J., Tang, J. and Wu, G. : Residual feature distillation network for lightweight image super-resolution, *European Conference on Computer Vision Workshops*, 2020.