A Light Model for Super-Resolution of Remote Sensing Images

Song Tingting*  

*School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China  

*email: songtt@std.uestc.edu.cn

Abstract. Remote sensing image super-resolution is an important research topic, which helps to improve the quality of remote sensing images. However, because remote sensing images are usually quite large while the satellite equipment often have low computing capacity and small memory, it is appealing to develop lightweight and fast models to perform the super-resolution task for the remote sensing images. In this paper, we empirically study the effectiveness of conventional super-resolution approaches for remote sensing images, and propose an effective way to reduce the model parameters and computational cost. Specifically, motivated by Res2Net, we design a new multi-scale hierarchy residual block to replace the ResBlock in EDSR to provide a more diverse receptive field for each residual block. The proposed modified method has fewer parameters and faster speed compared with the original EDSR. Moreover, we also build two benchmark super-resolution datasets (i.e., DOTA-SR and LEVIR-SR) from DOTA and LEVIR-CD, respectively, for experimental evaluation. We perform experiments on the two datasets, and results show that our method is light and have comparable performance over the existing super-resolution baselines.

1. Introduction  

With the rapid development of many infrastructures, we can easily acquire different types of satellite images with different resolutions. Currently, many works focus on understanding the remote sensing images through characterizing various objects such as land classification and scene object detection. A few research works have addressed the Super-Resolution (SR) problem for remote sensing images. SR technology aims to restore the high-resolution image from a given low-resolution image and can provide higher quality images and thus improve the performance of other downstream tasks. Single Image Super-Resolution (SISR) is a branch of SR and has attracted much attention in the past few years. Due to the low computing power and small memory of the satellite devices, we need a lightweight model to perform the super-resolution task for remote sensing images.
To this end, we explore the problem of SISR for remote sensing images and empirically study the effectiveness of conventional SISR approaches. Then, we propose an effective way to reduce the model parameters and computational cost. Specifically, motivated by Res2Net [1], we design a new multi-scale hierarchy residual block to replace the ResBlock in EDSR [2] to provide a more diverse receptive field for each residual block. The proposed modified method has fewer parameters and faster speed compared with the original EDSR [2]. We build two super-resolution datasets from DOTA [3] and LEVIR-CD [4] for evaluation (i.e., DOTA-SR and LEVIR-SR). And we evaluate the performance of existing super-resolution methods such as SRCNN [5], EDSR [2] and RCAN [6]. Finally, we conduct the experiments on DOTA-SR and LEVIR-SR, which demonstrate the effectiveness of our method.

2. Materials and Methods

2.1 Basic structures

In the mainstream super-resolution models, there usually exists a feature extraction model (backbone) and an upscale module. For the design of the backbone, people usually stack some typical structures such as residual block, dense connection, skip connection [7] and attention [6] to improve the performance. While for the upscale module, there exists interpolation, pixel shuffle [8], transpose convolution, and so on. After that, someone uses additional losses such as GAN loss and perceptual loss [9] to recover clearer texture or remove the Batch Normalization (BN) or ReLU layer for better generalization performance.

SRCNN [5] is the first deep learning model for the super-resolution task. The detail structure can be seen in figure 1, which contains only three convolutional layers and two ReLU layers.

![Fig. 1 The structure of SRCNN [5]](image1)

EDSR [2] uses a few ResBlocks followed with pixel shuffle [8] after the feature extraction. They remold the ResBlock to adapt to the SR task and remove the original BN layers to make the network more flexible, then they use the residual scaling to stable the training stage. Figure 2 shows the structure of Resblock, there are only two convolution layers and a ReLU layer in the block.

![Fig. 2 The ResBlock used in EDSR [2]](image2)

RCAN [6] is another SR model, which has more parameters than aforementioned models. It uses the channel attention (CA) to improve the power of feature representation of the network and uses the residual in residual (RIR) block to make the network deeper. The CA block gives the channels of the features different weights to increase the difference between them. The RIR block uses the short and long skip connection to better transmit characteristic information.
Fig. 3 The RCAB, RG and RIR blocks in RCAN [6]

Figure 3 shows the basic blocks of RCAN [6]. The RCAB consists of the ResBlock in EDSR [2] and a CA module, the RG consists of many RCAB blocks and uses the short skip connections to merge the features. The RIR stacks many RG blocks and uses the long skip connections to merge the features.

2.2 Res2Net block
Recently, some researchers design a Res2Net [1] module to replace the residual module which can expand the receptive fields and can represent the multi-scale features. We use this module to replace one 3x3 convolution layer in ResBlock. In this way, the overall number of parameters of the model is greatly reduced, so we can obtain a more light model for the SR task of remote sensing images. Figure 4 shows the modified Res2Net [1] block.

Fig. 4 The modified Res2Net [1] block

3. Experiments

3.1 Datasets
We select two remote sensing datasets DOTA [3] and LEVIR-CD [4] to build our benchmark datasets for evaluating the SR methods, named as DOTA-SR and LEVIR-SR. The DOTA [3] dataset contains
2,806 images and 15 categories. The LEVIR-CD [4] dataset contains 637 pairs of high-resolution remote sensing images with a size of 1024×1024. We crop these images randomly to retain images suitable for super-resolution tasks. Specifically, as for DOTA-SR, we pick 900 images for training, 101 images for validation and 111 images for testing. For LEVIR-SR, we pick 700 images for training, 101 images for validation and 89 for testing. These images are used as the high-resolution images and we use MATLAB functions to create the low-resolution parts. The bicubic downsampling scales of 2× and 4× are adopted for both datasets.

3.2 Evaluation metrics
As for comparison, we choose three SR algorithms: SRCNN [5], EDSR [2] and RCAN [6]. We use the bicubic interpolation method as the baseline. We evaluate these models on the two proposed remote sensing datasets and use Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) [10] to measure the SR results. In Equation (1), MSE means the mean square error of all pixels of two pictures while MAX means the maximum possible pixel value of the picture I.

\[
PSNR = 10 \cdot \log_{10} \left( \frac{MAX^2}{MSE} \right) \tag{1}
\]

In Equation (2), \( \mu_x, \mu_y \) and \( \sigma_x, \sigma_y \) represent the mean value and variance value of the image X, Y respectively. \( C_1, C_2 \) and \( C_3 \) are the constants.

\[
l(X,Y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad c(X,Y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad s(X,Y) = \frac{\sigma_x \sigma_y + C_3}{\sigma_x \sigma_y + C_3} \tag{2}
\]

\[
SSIM(X,Y) = l(X,Y) \cdot c(X,Y) \cdot s(X,Y) \tag{3}
\]

After that, we also compare the model parameters and the average (Avg) test time. In order to reduce the memory consumption, we use 24 ResBlock or Res2Net [1] blocks and set the channel dimension to 128. For the RCAN [6], we use the default setting in the paper. We use L1 loss to train all the models. Equation (4) shows the pixel-wise L1 loss.

\[
L_i(\hat{I}, I) = \frac{1}{h \cdot w \cdot c \cdot h \cdot w} \sum_{i,j,k} \left| \hat{I}_{i,j,k} - I_{i,j,k} \right| \tag{4}
\]

3.3 Results analysis
After running 300 epochs on the two datasets, we record the test results in Table 1. We observe that, all the deep learning models achieve better performance compared with the bicubic baseline. Our lightweight model is better than the baseline and SRCNN [5] among all the scales. Compared with EDSR [2], our proposed method achieves comparable performance but with significant lower parameters and a faster speed. Our memory usage dropped by 36.78% compared with EDSR [2] and 67.94% with RCAN [6]. Our model achieves average 5.01% increase on inference speed compared to EDSR [2] and 66.32% on RCAN [6] at scale 2 while achieves average 2.57% increase to EDSR [2] and 27.37% to RCAN [6] at scale 4.

**Table 1** Evaluation of some state-of-the-art SR algorithms
Since the image resolution is lower at scale 4, the inference speed of the model is faster than that at scale 2. The RCAN [6] runs almost 3 times than other algorithms at scale 2 and has the largest number of parameters although it has the highest PSNR and SSIM. The remote sensing tasks are often run on low computing power and small memory devices such as satellites. Therefore, in order to balance efficiency and performance, our model is more suitable for the super-resolution task of remote sensing images.

4. Conclusion

In this paper, we empirically study the effectiveness of conventional super-resolution approaches for remote sensing images. We propose a multi-scale hierarchy residual block to replace the ResBlock in EDSR. The proposed modified method has fewer parameters and faster speed compared with the original EDSR. To evaluate the effectiveness of our method, we build two remote sensing datasets DOTA-SR and LEVIR-SR. The experimental results clearly show that our model is light and has comparable performance over several existing super-resolution baselines.

References

[1] Gao S, Cheng M M, Zhao K, Zhang X Y and Torr P 2019 Res2net: a new multi-scale backbone architecture IEEE Transactions on Pattern Analysis and Machine Intelligence pp (99) 1-1

[2] Lim B, Son S, Kim H, Nah S and Lee K M 2017 Enhanced deep residual networks for single image super-resolution IEEE Conference on Computer Vision and Pattern Recognition Workshops

[3] Xia G S, Bai X, Ding J, Zhu Z, Belongie S, Luo J, Datcu M, Pelillo M and Zhang L 2018 Dota: a large-scale dataset for object detection in aerial images IEEE Conference on Computer Vision and Pattern Recognition pp 3974–3983

[4] Chen H and Shi Z 2020 A spatial-temporal attention-based method and a new dataset for remote sensing image change detection Remote Sensing 12 (10) p 1662

[5] Dong C, Loy C C, He K and Tang X 2016 Image super-resolution using deep convolutional networks IEEE Trans Pattern Anal Mach Intell 38 (2) pp 295–307
[6] Zhang Y, Li K, Wang L, Zhong B and Fu Y 2018 Image super-resolution using very deep residual channel attention networks *ECCV* pp 294–310

[7] Anwar S and Barnes N 2020 Densely residual laplacian super-resolution *IEEE Transactions on Pattern Analysis and Machine Intelligence*

[8] Shi W, Caballero J, F Huszár, Totz J and Wang Z 2016 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

[9] Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Loy CC, Qiao Y and Tang X 2018 Esrgan: enhanced super-resolution generative adversarial networks *ECCV* pp 63-79

[10] Wang Z 2004 Image quality assessment: from error visibility to structural similarity *IEEE Transactions on Image Processing* pp 600-612