SIMULTANEOUS SUPER-RESOLUTION AND SEGMENTATION FOR REMOTE SENSING IMAGES

Sen Lei[1], Zhenwei Shi[1], Xi Wu[1], Bin Pan[1], Xia Xu[1], Hongxun Hao[2]

[1]Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China
[2] The Flight Technology College, Civil Aviation University of China, Tianjin 300300, China
{senlei, shizhenwei, xiwu1000, panbin, xuxia}@buaa.edu.cn, hxhao8946@aliyun.com

ABSTRACT

In this paper, we present an algorithm to simultaneously obtain high-resolution images and segmentation maps from low-resolution inputs. Super-resolution and segmentation both are challenging task, but they may have certain relationship. Super-resolution will provide images with more details that may help to improve the segmentation accuracy, while label maps in segmentation dataset may contribute to finer edges during super-resolution process. Therefore, we aim to combine these two tasks and explore the influence for each other. For this end, we proposed a new deep neural network to simultaneously address the super-resolution and segmentation tasks for remote sensing images, which is named $\mathrm{S^2Net}$. The $\mathrm{S^2Net}$ is an integrated network composed of a super-resolution sub-network and a segmentation sub-network, which is trained in an end-to-end manner. Experimental results demonstrate that this combination can enhance the performance on these two tasks.

Index Terms— Remote sensing images, Super-resolution, Segmentation, $\mathrm{S^2Net}$

1. INTRODUCTION

Super-resolution aims to recover high-resolution images from the corresponding low-resolution ones, which has been widely used in many applications such as security and surveillance imaging, medical imaging and remote sensing image reconstruction. In the field of remote sensing, high-resolution images which contains many details are important for remote sensing applications such as image segmentation, target detection and recognition [1, 2]. Apart from developing physical imaging technologies, image super-resolution is an alternative way to obtain high-resolution remote sensing images.

Single image super-resolution generates a high-resolution image from a low-resolution input, which has received increasing attentions in recent years. Sparse coding method is used to learning compact dictionary for recovering low-resolution images [3]. Pan $et\ al.$ [4] combined compressive sensing with structural self-similarity to recover high-resolution remote sensing images from low-resolution ones.

Recently, deep convolutional neural networks (CNN) have made a breakthrough in the computer vision community and also has been widely used in many remote sensing tasks including super-resolution [5, 6, 7]. The deep CNNs can extract high-level feature representations automatically from data and learn a deep mapping function between low/high-resolution images. Dong $et\ al.$[8] introduced a three-layers CNN named SRCNN to generate high-resolution images trained in an end-to-end manner. Lei $et\ al.$[7] proposed a local-global combined network to learn multilevel representation of remote sensing images. Wang $et\ al.$[9] utilized multiple convolutional neural network,s to learn wavelet multiscale representations of remote sensing images.

However, these existed methods mainly focus on the recovery of the low-resolution inputs and little researches explore the influence of super-resolution task for some high-level applications such as segmentation, where they may have certain rela,tionship. Super-resolution will provide images with more details that may be helpful for following segmentation. Meanwhile, label maps in segmentation dataset may contribute to finer edges during super-resolution process. Therefore, more attentions should be paid on the combination of super-resolution and segmentation tasks.

In this paper, we aims to super-resolve low-resolution remote sensing images with a large upscale factor 4, and explore whether the super-resolution task would be combined with a high-level segmentation task. For this end, we here propose a new deep neural network to simultaneously perform super-resolution and segmentation for remote sensing images, which we call $\mathrm{S^2Net}$. Specifically, the $\mathrm{S^2Net}$ is composed of a super-resolution sub-network and a segmentation sub-network to handle these two issues respectively, which is trained in an end-to-end manner. The flowchart of the proposed method is shown in Fig.1.

The mainly contributions of this paper lie in that we pro-

Thanks to the National Key R&D Program of China under the Grant 2017YFC1405605, the National Natural Science Foundation of China under the Grants 61671037, and the Beijing Natural Science Foundation under the Grant 4192034. *Corresponding author: Zhenwei Shi.*

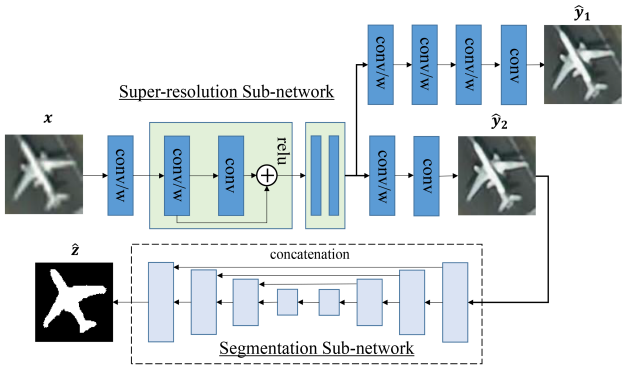

Fig. 1: Flowchart of the proposed method for remote sensing images. The 'conv/w' indicates that the convolutional layer is followed with a nonlinear function ReLU.

posed a new method named $\mathrm{S}^2\mathrm{Net}$ to simultaneously address super-resolution and segmentation tasks for remote sensing images, and experimental results show that the combination can enhance the performances of each task.

2. THE PROPOSED METHOD

2.1. Super-resolution Sub-network

Before segmentation, the low-resolution images are firstly reconstructed by the super-resolution sub-network. The architecture of the super-resolution sub-network is illustrated in Fig.1. We found that in an end-to-end training manner the reconstructions ($\hat{y}_2$ in Fig.1) followed with the segmentation part would not be pleasant for human perception. Thus we design two branches in the super-resolution network to generate high-resolution images. The first branch is composed of four convolutional layers and generate final high-resolution images for visualizations, and the second one is a a shallow network with two convolutional layer to obtain reconstructions for the following segmentation sub-network. They are both built on a same backbone and share the network weights. In this paper, the backbone consists of one convolutional layer and two residual blocks, where the residual block is constructed following [10] without batch normalization.

In order to prevent from information loss, pooling layers are avoided and only fully convolutional layers are utilized. Moreover, these convolution layers are followed a nonlinear function ReLU except output layers of the two branches, which are showed in Fig.1.

Here, we denote the low-resolution remote sensing images as $\{x^{(i)}, ..., x^{(N)}\}$ and the corresponding high-resolution ones as $\{y^{(i)}, ..., y^{(N)}\}$, respectively. The low-resolution input x is firstly upscaled to the same size of high-resolution reference y using bicubic interpolation. Furthermore, we use $\hat{y}_1$ and $\hat{y}_2$ present the outputs of these two branches, and pixel-wise loss functions are defined to optimize for the super-resolution network:

$$L_{sr1} = \frac{1}{N}\sum_{i=1}^{N} ||y^{(i)} - \hat{y}_1^{(i)}||^2$$
$$L_{sr2} = \frac{1}{N}\sum_{i=1}^{N} ||y^{(i)} - \hat{y}_2^{(i)}||^2 \quad (1)$$

where N is the total number of training samples.

2.2. Segmentation Sub-network

The outputs of the second branch of super-resolution part are taken as inputs of the following segmentation sub-network to obatin segmentation maps. In order to distinguish the object from its surroundings, strided-convolutional layers are used to enlarge the receive field of the segmentation model. And deconvolution operations are applied to get original spatial resolution. Specifically, the receive field of CNNs can be computed as follows:

$$RF^{(l+1)} = (k^{(l+1)} - 1) * \prod_{i=1}^{l} s^{(i)} + RF^{(l)} \quad (2)$$

where $RF^{(l)}$ and $s^{(i)}$ is the receive field and the stride of layer l, respectively, and $k^{(l)}$ denote the kernel size of this layer. In our experiments, images are 96×96 pixels and we thus design the network with a maximum receive field of 63 pixels, and the feature maps of bottom and top layers are concatenated for finer segmentation maps. The detailed configurations of the segmentation architecture are presented in Table 1. In order to accelerate the training phase, we add batch normalization in this model.

Following Mask R-CNN [11], in this paper, we used binary cross-entroy loss function for the segmentation sub-network which is defined as:

$$L_{seg} = \frac{1}{N}\sum_{i=1}^{N}(z^{(i)} * log(\hat{z}^{(i)}) + (1 - z^{(i)}) * log(1 - \hat{z}^{(i)})) \quad (3)$$

where $\hat{z}$ and z denote the output of the segmentation sub-network and the corresponding label.

Therefore, the overall loss function of $\mathrm{S}^2\mathrm{Net}$ model can be computed as:

$$L = \lambda_{sr1} * L_{sr1} + \lambda_{sr2} * L_{sr2} + \lambda_{seg} * L_{seg} \quad (4)$$

where λ_{sr1}, λ_{sr2} and λ_{seg} are the weights of their corresponding losses.

3. EXPERIMENTS

3.1. Dataset and Implementation Details

The experimental dataset contains three classes of remote sensing images including airplanes, ships and oiltanks with

Table 1: The Detailed Configurations of the Segmentation Sub-network

Layer names	Kernels	Sizes/Strides
seg_conv1	64	$3 \times 3/2$
seg_conv2	64	$3 \times 3/2$
seg_conv3	64	$3 \times 3/2$
seg_conv4	64	$3 \times 3/2$
seg_conv5	64	$3 \times 3/1$
seg_deconv6	64	$3 \times 3/2$
cconcat: conv3/deconv6	–	$-/-$
seg_deconv7	64	$3 \times 3/2$
cconcat: conv2/deconv7	–	$-/-$
seg_deconv8	64	$3 \times 3/2$
cconcat: conv1/deconv8	–	$-/-$
seg_logits	1	$3 \times 3/1$

the size of 96×96, and the numbers of these targets are 2609, 1732 and 2208 respectively. We randomly select 80% of the samples for training and the others for test. All the images are downsampled (scale factor = 4) as low-resolution images with the original as high-resolution references. Furthermore, the training samples are augmented by random rotation and mirroring and they all are normalized in the range of [0, 1].

In this paper, three metrics are used to evaluate the proposed method, including PSNR (peak-signal-noise-ratio), SSIM (Structural Similarity Index Measure) and IoU (Intersection of Union). PSNR and SSIM are utilized to measure the super-resolution performance, and IoU is used to evaluate the segmentation performance. Since the samples in the dataset are RGB images, PSNR and SSIM are computed by averaging among three channels.

In the training phase, λ_{sr1}, λ_{sr2} and λ_{seg} are fixed as 1, 1 and 0.01. For optimization, we use Adam to minimize with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Moreover, learning rate is 0.0001 and the number of mini-batch is 64. All these experiments are implemented with tensorflow package.

3.2. Experimental Results

Here, we evaluate the performance of the proposed method on the test set, compared with some other methods including bicubic interpolation, SRCNN [8], SR-9 and Seg-8. It should be noted that SR-9 and Seg-8 both are the baseline models of our proposed method, where SR-9 with 9 convolutional layers has the same architecture with the super-resolution sub-network without the second branch and Seg-8 is similar with the segmentation sub-network with convolutional and deconvolutional layers. These methods are trained with the same configurations with the proposed method.

Table.2 presents the results of different methods on the test set. We can see that $\mathrm{S}^2\mathrm{Net}$ obtains higher PSNR and SSIM than SR-9 and meanwhile achieves better segmentation results than Seg-8, which proves the effectiveness of the proposed method. Fig.2 and Fig.3 show the super-resolution and segmentation results for remote sensing targets, respectively. From Fig.2, it can be found that the results of the proposed method have less artifacts than other methods. Moreover, Fig.3 demonstrates that via combining with super-resolution sub-network, $\mathrm{S}^2\mathrm{Net}$ obtains more accurate segmentation results than Seg-8.

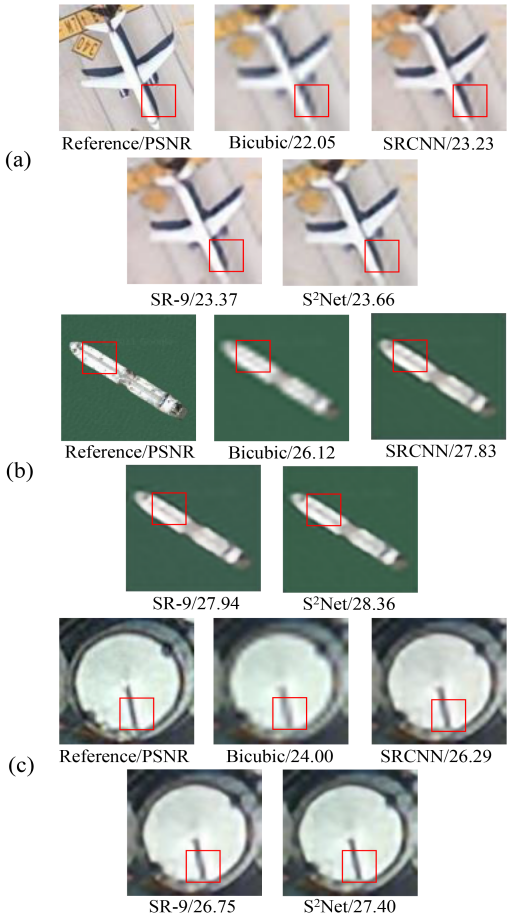

Fig. 2: Super-resolution results for remote sensing targets (PSNR/dB): (a) airplane; (b) ship; (c) oiltank (**Zoom in for best view**).

4. CONCLUSION

In this paper, we design a new network named $\mathrm{S}^2\mathrm{Net}$ to simultaneously perform super-resolution and segmentation tasks for remote sensing images. The $\mathrm{S}^2\mathrm{Net}$ is composed of a super-resolution sub-network with a two-branch structure and a segmentation sub-network with low/high-level representation concatenations. The experimental results on three kind of remote sensing targets including airplanes, ships and oiltanks show that the $\mathrm{S}^2\mathrm{Net}$ can obtain improvements on both super-resolution and segmentation tasks.

Table 2: Mean PSNR (dB), SSIM and IoU over all the test data set

class	Bicubic PSNR / SSIM	SRCNN PSNR / SSIM	SR-9 PSNR / SSIM	Seg-8 IoU	S^2Net PSNR / SSIM / IoU
airplane	23.42/0.721	24.89/0.777	25.14/0.788	0.666	**25.36/ 0.793 / 0.673**
ship	28.59/0.827	29.26/0.842	29.52/0.849	0.553	**29.61/ 0.852 / 0.632**
oiltank	28.32/0.829	29.23/0.863	29.45/0.867	**0.784**	**29.81/ 0.871** /0.765
average	26.78/0.792	27.80/0.827	28.04/0.835	0.668	**28.26/ 0.839 / 0.690**

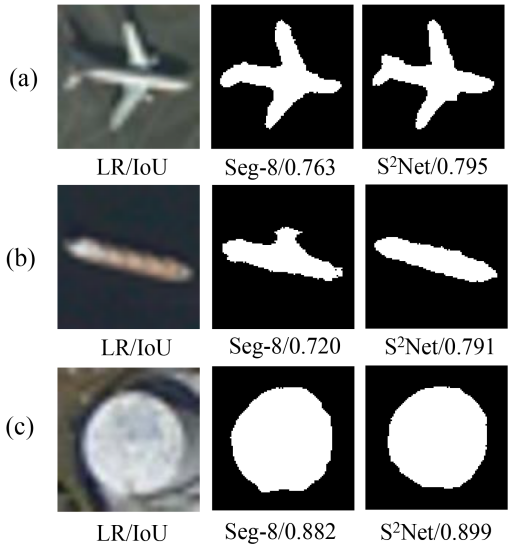

Fig. 3: Segmentation results for remote sensing targets: (a) airplane; (b) ship; (c) oiltank (LR denotes the low-resolution input).

5. REFERENCES

[1] Z. Zou and Z. Shi, "Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1100–1111, Mar. 2018.

[2] B. Pan, Z. Shi, X. Xu, T. Shi, N. Zhang, and X. Zhu, "Coinnet: Copy initialization network for multispectral imagery semantic segmentation," *IEEE Geosci. Remote Sens. Lett.*, 2018.

[3] J. Yang, J. Wright, T. S. Huang, and et al, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[4] Z. Pan, J. Yu, H. Huang, and et al, "Super-resolution based on compressive sensing and structural self-similarity for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4864–4876, 2013.

[5] Z. Shi and Z. Zou, "Can a machine generate human-like language descriptions for a remote sensing image?," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3623–3634, June. 2017.

[6] B. Pan, Z. Shi, and X. Xu, "Mugnet: Deep learning for hyperspectral image classification using limited samples," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, no. Part A, pp. 108–119, Nov. 2018.

[7] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local-global-combined network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1243–1247, Aug. 2017.

[8] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[9] T. Wang, W. Sun, H. Qi, and et al, "Aerial image super resolution via wavelet multiscale convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 769–773, 2018.

[10] K. He, X. Zhang, S. Ren, and et al, "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[11] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.