SCIENCE CHINA
Information Sciences

• LETTER •

January 2023, Vol. 66 119105:1–119105:2
https://doi.org/10.1007/s11432-021-3575-1

Unpaired remote sensing image super-resolution with content-preserving weak supervision neural network

Jie WU[1†], Runmin CONG[2†], Leyuan FANG[1*], Chunle GUO[3], Bob ZHANG[4] & Pedram GHAMISI[5,6]

[1]*College of Electrical and Information Engineering, Hunan University, Changsha* 410082, *China;*
[2]*Institute of Information Science, Beijing Jiaotong University, Beijing* 100044, *China;*
[3]*College of Computer Science, Nankai University, Tianjin* 300071, *China;*
[4]*Department of Computer and Information Science, University of Macau, Macao* 999078, *China;*
[5]*Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Freiberg* 09599, *Germany;*
[6]*Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna* 1030, *Austria*

Received 9 November 2021/Revised 5 February 2022/Accepted 12 July 2022/Published online 17 November 2022

Citation Wu J, Cong R M, Fang L Y, et al. Unpaired remote sensing image super-resolution with content-preserving weak supervision neural network. Sci China Inf Sci, 2023, 66(1): 119105, https://doi.org/10.1007/s11432-021-3575-1

Dear editor,

Image super-resolution (SR) uses image processing methods to increase the resolution of the image without increasing any hardware cost. Remote sensing images with high spatial resolution contain rich imagery information, which can be utilized for object detection, building segmentation, building extraction, change detection, and so on.

With the development of deep learning, remote sensing SR methods based on convolutional neural networks (CNNs) have achieved state-of-the-art results. However, most of the remote sensing image SR methods based on CNNs require paired HR-LR data for training. In the real world, it is difficult for a satellite to obtain paired images of different resolutions in the same scene, and few satellites can obtain HR images. Existing methods usually use simple degradation models, such as bicubic, to synthesize paired datasets. In general, these paired models cannot be generalized to real-world remote sensing SR, because such a simple bicubic degradation process is quite different from the degradation of real images that includes complex changes, such as noise, blur, and compression loss.

More recently, the generative adversarial networks (GANs) have achieved impressive results on various computer vision tasks, such as SR [1] and image domain translation. With the advent of unpaired image domain translation methods (e.g., CycleGAN [2]), we intend to synthesize real remote sensing LR images by learning the mapping function from the bicubic downsampled domain to the real LR domain to synthesize paired dataset. However, compared with natural images, remote sensing images usually contain more objects and richer content (see Figure A1(a) and (b)) so that the adversarial training in remote sensing image tasks is more likely to cause content distortion (i.e., fake edges, artificial textures, and unreal objects), due to numerous ambiguous solutions (see Figure A1(c)).

Proposed CPWSNN for unpaired remote sensing SR. To address the above issues, in this study, we proposed an unpaired remote sensing image SR method called the content-preserving weak supervision neural network (CPWSNN). Firstly, we incorporate the perceptual loss [3] into the image domain translation to synthesize pseudo-LR remote sensing images from real HR images, as shown in Figure 1(a). The perceptual loss ensure that the generated pseudo-LR images and the real LR images have the same content and semantic information. Then, the generated pseudo-LR images and the real HR images compose the paired HR-LR training data to provide supervision for the SR network which utilizes a pixel-wise loss. Furthermore, we propose the degradation consistency loss and the edge retention loss to constrain the GAN-based SR network training procedure. The degradation consistency loss is proposed to constrain the solution space of the SR model to avoid the artificial objects in its results. The edge retention loss is proposed to prevent the SR network from generating fake edges and textures by constraining the edges' information in its results. Finally, the results can have more realistic textures and detailed information with the object information unchanged. The SR reconstruction process is shown in Figure 1(b). The specific steps of the proposed method are described as follows.

Step 1. The domain translation aims to generate pseudo-LR remote sensing images that are the LR version of real HR remote sensing images by learning a translation between the bicubic domain and the real LR domain, as shown in Figure 1(a). Then, the pseudo-LR images and real HR images compose paired data to provide supervision for the SR network. Similar to the CycleGAN [2], our domain translation

* Corresponding author (email: leyuan_fang@hnu.edu.cn)
† Wu J and Cong R M have the same contribution to this work.

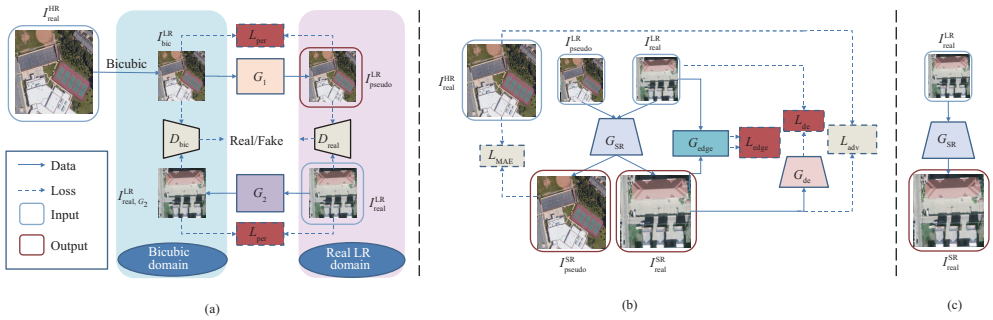

Figure 1 The framework of the proposed CPWSNN. (a) The framework for generating the pseudo-LR remote sensing image $I^{\mathrm{LR}}_{\mathrm{pseudo}}$ by domain translation with the perceptual loss L_{per}. (b) Remote sensing image SR reconstruction by the SR network with the edge retention loss L_{edge} and the degradation consistency loss L_{de}. In the testing phase (c), the well-trained G_{SR} reconstructs $I^{\mathrm{LR}}_{\mathrm{real}}$ to $I^{\mathrm{SR}}_{\mathrm{real}}$.

is composed of two generators (G_1, G_2) and two discriminators (D_{real}, D_{bic}). Given a real HR image $I^{\mathrm{HR}}_{\mathrm{real}}$, we downscale it by the bicubic interpolation operation to get $I^{\mathrm{LR}}_{\mathrm{bic}}$. G_1 is used to learn the mapping function from the bicubic domain to the real LR domain and generate a pseudo-LR image $I^{\mathrm{LR}}_{\mathrm{pseudo}}$ from $I^{\mathrm{LR}}_{\mathrm{bic}}$. G_2 is a reverse mapping network to map the real LR image $I^{\mathrm{LR}}_{\mathrm{real}}$ to its bicubic domain as $I^{\mathrm{LR}}_{\mathrm{real},G_2}$ and is used to assist the training of G_1. D_{real} is a discriminator of G_1 to distinguish $I^{\mathrm{LR}}_{\mathrm{pseudo}}$ from $I^{\mathrm{LR}}_{\mathrm{real}}$. Similarly, D_{bic} is the discriminator of G_2 to distinguish $I^{\mathrm{LR}}_{\mathrm{real},G_2}$ from $I^{\mathrm{LR}}_{\mathrm{bic}}$. The architecture of G_1 and G_2 are composed of eight residual blocks [4]. Following the loss function of CycleGAN, the full loss function of our domain translation (described by Eq. (B1)) includes the adversarial loss, the cycle consistency loss, and the identity loss. Besides, we adopt the perceptual loss to avoid object deformation and other content distortions in the pseudo-LR image. In general, since there are more objects and semantic information in the feature space than the pixel space, the perceptual loss can measure the difference between two different images to preserve the content. The perceptual loss is described in Eq. (B2).

Step 2. The SR network G_{SR} aims to reconstruct a detailed SR image $I^{\mathrm{SR}}_{\mathrm{real}}$ from a real LR image $I^{\mathrm{LR}}_{\mathrm{real}}$. After obtaining the pseudo-LR image $I^{\mathrm{LR}}_{\mathrm{pseudo}}$ by the domain translation, we utilize the G_{SR} to generate the pseudo-SR image $I^{\mathrm{SR}}_{\mathrm{pseudo}}$. The pseudo-LR image $I^{\mathrm{LR}}_{\mathrm{pseudo}}$ and real HR image $I^{\mathrm{HR}}_{\mathrm{real}}$ provide paired supervision for the training of the SR network that uses MAE loss. In addition, we adopt the GAN-based framework to train the SR network on $I^{\mathrm{SR}}_{\mathrm{real}}$ and $I^{\mathrm{HR}}_{\mathrm{real}}$, which can produce more realistic results on real LR data. D_{hr} is a discriminator (utilizing the same structure as PatchGAN [5]) to distinguish $I^{\mathrm{SR}}_{\mathrm{real}}$ from $I^{\mathrm{HR}}_{\mathrm{real}}$. However, the training process of GANs is unstable resulting in SR reconstructed results with fake textures, artificial objects, and other content distortions. So, we propose the edge retention loss L_{edge} and degradation consistency loss L_{de} to constrain the solution space of the SR network. The full loss function of the SR reconstruction is described in Eq. (B3).

Edge retention loss. The edge retention loss is proposed to prevent the SR network from generating fake edges and textures using the edge priors. In this loss, we use the DexiNed network [6] to extract the edge information from real LR and SR images and adopt the binary cross entropy (BCE) loss to measure the difference between the edge information of real LR and SR images. The edge retention loss can be expressed in Eq. (B4)

Degradation consistency loss. The degradation consistency loss is introduced to avoid the generation of unreal objects or other content distortions in SR results. The degradation network G_{de} (see Figure B1) is adopted to provide constraints for the SR network. Specifically, G_{de} degrades the SR result to $I^{\mathrm{LR}}_{\mathrm{de}}$, which is $I^{\mathrm{LR}}_{\mathrm{de}} = G_{\mathrm{de}}(I^{\mathrm{SR}}_{\mathrm{real}})$. The degradation consistency loss L_{de} is expressed in Eq. (B5).

Experimental results on real and synthetic remote sensing datasets (shown in Appendix C) demonstrate that our unpaired method achieves competitive SR results and high robustness (even under mixed degradations).

Supporting information Appendixes A–D. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant No. 61922029), Science and Technology Plan Project Fund of Hunan Province (Grant No. 2019RS2016), Key Research and Development Program of Hunan (Grant No. 2021SK2039), and Natural Science Foundation of Hunan Province (Grant No. 2021JJ30003).

References

1 Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018
2 Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, 2017. 2223–2232
3 Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of European Conference on Computer Vision, 2016. 694–711
4 He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 770–778
5 Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 1125–1134
6 Poma X S, Riba E, Sappa A. Dense extreme inception network: towards a robust CNN model for edge detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020. 1923–1932