Remote Sensing Single-Image Super-Resolution Based on a Deep Compendium Model

J. M. Haut, *Student Member, IEEE,* M. E. Paoletti, *Student Member, IEEE,* R. Fernandez-Beltran,

J. Plaza, *Senior Member, IEEE,* A. Plaza, *Fellow, IEEE,* and J. Li, *Senior Member, IEEE*

Abstract—This letter introduces a novel remote sensing single-image super-resolution (SR) architecture based on a deep efficient compendium model. The current deep learning (DL)-based SR trend stands for using deeper networks to improve the performance. However, this practice often results in a degradation of visual results. To address this issue, the proposed approach harmonizes several different improvements on the network design to achieve state-of-the-art performance when super-resolving remote sensing imagery. On the one hand, the proposal combines residual units and skip connections to extract more informative features on both local and global image areas. On the other hand, it makes use of parallelized 1×1 convolutional filters (network in network) to reconstruct the super-resolved result while reducing the information loss through the network. Our experiments, conducted using seven different SR methods over the well-known UC Merced remote sensing dataset, and two additional GaoFen-2 test images, show that the proposed model is able to provide competitive advantages.

Index Terms—Super-resolution, deep learning, remote sensing.

I. INTRODUCTION

Image super-resolution (SR) has found a fertile domain in the remote sensing field in order to overcome the optical limitations inherent to airborne and space acquisition instruments. Remote target tracking [1], land-cover mapping [2] and fine-grained image classification [3] are some of the most popular applications in which SR has shown to provide competitive advantages. In general, single-image SR aims at enhancing the resolution of a given input image from a single view of the target scene. That is, the SR technology pursues to recover spatial details not captured by the imaging sensor, which logically implies a high level of visual uncertainty and eventually demands strong image priors to effectively relieve the ill-posed nature of the problem.

Different kinds of image reconstruction and learning paradigms have been successfully applied to super-resolve remotely sensed imagery [4]. From the most traditional signal reconstruction mechanisms to the most recent machine learning algorithms, all the existing SR methodologies have their own advantages and limitations. For instance, Zhang *et al.* [5] propose a land cover

J. M. Haut, M. E. Paoletti, J. Plaza and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain.(e-mail: juanmariohaut@unex.es; mpaoletti@unex.es; jplaza@unex.es; aplaza@unex.es). R. Fernandez-Beltran is with the Institute of New Imaging Technologies, University Jaume I, 12071 Castellón de la Plana, Spain. (e-mail: rufernan@uji.es). J. Li is with the Guangdong Provincial Key Laboratory of Urbanization and Geosimulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: lijun48@mail.sysu.edu.cn).

mapping SR approach which takes advantage of the self-similarity property to learn multi-scale patch relations useful to increase the input image resolution. Despite the fact that this and other approaches, such as [6], do not require any external training set, their performances are usually rather unsatisfactory in applications where precision is important. Other authors make use of different machine learning paradigms to define more accurate SR models based on a training procedure using exemplar high-resolution data. For instance, Yang *et al.* [7] present a SR technique based on the sparse coding approach which learns a coupled dictionary from an external high-resolution training set in order to project the input low-resolution image at a higher resolution scale. Alternative works propose the use of probabilistic models to learn the mapping between low-resolution and high-resolution image domains [8]. Even though these and other recent models [9] [10] have shown to obtain a good SR performance, they are typically based on low-level visual features, which eventually limits the SR capabilities, especially under the most challenging remote sensing scenarios.

Recently, convolutional neural networks (CNN) have shown a great potential to capture high level features from optical data. Therefore, this paradigm has become one of the most important technologies to deal with the SR problem and many CNN-based learning models have been proposed in the literature. For instance, Dong *et al.* [11] propose a deep learning (DL) based approach to super-resolve low-resolution image patches. In the training stage, the method up-scales the low-resolution input images by a bi-cubic interpolation, and then a 3-layer CNN is trained to learn a mapping between these interpolated versions and their ground-truth counterparts. Conceptually, each layer represents: (1) patch extraction, (2) non-linear mapping and (3) high-resolution generation. The first two layers use a formulation based on the Rectified Linear Unit (ReLU) on the filter responses, e.i. $max(0, W*Y+B)$, where W and B represent the filters and biases, respectively, $*$ is the convolution operator, and Y is the input image. The third layer is a regular convolutional layer with filter responses $W * Y + B$. Finally, the loss function used to train the parameters W and B of the network is the Mean Square Error (MSE). Despite the remarkable performance achieved by this reference work, alternative CNN-based SR approaches have been also proposed in the literature. In particular, a relevant extension can be found in [12], where authors introduce several improvements based on three main points: (1) defining an end-to-end mapping to avoid the initial interpolation process, (2) reformulating the mapping layer by shrinking the input feature dimension, and (3) using a deeper architecture to achieve a superior restoration quality. Other authors propose different network improvements instead. For instance, Kim *et al.* [13] define a SR approach considering a deeper architecture, a data augmenting scheme together with residual and multi-scaling learning. Even though all these exhibit satisfactory SR performance, they have not been specifically designed to manage remote sensing imagery, which eventually constrains their effectiveness over this application domain. Unlike general purpose optical data, remotely sensed images exhibit special complexity because they are typically highly-detailed and fully-focused multi-band shots. As a result, other works propose CNN-based SR methods aimed at dealing with satellite and aerial data: Lei *et al.* [14] present a DL approach especially designed to super-resolve remote sensing data. Specifically, this method defines a multi-level CNN architecture in order to learn multi-scale local and global image features when introducing new spatial details. Notwithstanding the remarkable performance achieved by these and other SR approaches [15], [16], the intrinsic complexity of airborne and space

optical data, together with the specific particularities and difficulties of the DL domain, still make room for improvements focused on the development of new architectures able to obtain performance advantages in the SR task.

Taking into account this scenario, this letter proposes a novel SR approach based on a deep efficient compendium model which integrates different improvements on the network design to effectively super-resolve remotely sensed images: (i) residual units, (ii) skip connections and (iii) network in network (NIN). On the one hand, the proposed architecture uses a combination of residual units and skip connections to extract more descriptive and informative visual features on both local and global image areas. On the other hand, parallelized 1×1 convolutional filters, also NIN [17], are employed to generate the output super-resolved image by substantially reducing the dimensions of the previous layers and hence the information loss through the network. With this design, the proposed approach pursues to competently super-resolve remote sensing data while avoiding undesirable visual artifacts. Our experiments, which include seven different SR methods, the UC Merced dataset and two additional GaoFen-2 test data products, reveal very competitive performance of our newly proposed approach.

TABLE I
KERNEL SIZES AND NUMBER OF FILTERS PER LAYER

Feature extractor network (FE-net)					
Layer	$C^{(1)} - C^{(12)}$				
Kernel size	3×3				
N filters	196, 166, 148, 133, 120, 108, 97, 86, 76, 66, 57, 48				
Reconstruction network (R-net)					
Layer	A	$B^{(1)}$	$B^{(2)}$	$U^{(1)}$	$U^{(2)}*$	R
Kernel size	1×1		3×3			
N filters	64	32	32	384	384	1

*A second upsampling layer is added when factor is

4x

II. METHODOLOGY

We propose an architecture based on fully convolutional neural networks (FCNs) [18]. Our model is composed by two parts (Fig. 1): i) the *feature extractor* part (FE-net) and ii) the *reconstruction* part (R-net), which are connected through a "concatenation" layer, where all the feature maps obtained by FE-net are concatenated through skip connections before being fed to R-net. The FE-net receives the original image $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as input, which passes through a hierarchical set of 12 convolutional layers, denoted as $C^{(n)}$, in order to extract the corresponding feature maps as a linear transformation between each layer's input and the layer's kernel. The kernel size has been set to 3×3, in order to reduce the number of parameters while considering enough spatial information. Also, with the aim of learning the non-linear relationships between the data, non-linear activation functions have been allocated behind each learnable layer: $\mathbf{X}^{(l+1)} = \mathcal{H}\left(\mathbf{W}^{(l)} * \mathbf{X}^{(l)} + b^{(l)}\right)$, where $\mathcal{H}(\cdot)$ is implemented by the parametric rectifier linear unit (P-ReLU) [19] in order to deal with the decaying ReLU effect and the vanishing gradient problem. At the end, all the output feature maps are concatenated and sent to R-net.

The R-net follows the NIN architecture, implemented by an inception module with two branches (layers A and $B^{(1)}$, with kernels 1×1, and $B^{(2)}$ with kernel 3×3) in order to reduce the input volume's depth. The obtained output is upsampled by one or more upsampling layers U of 3×3 kernels, which increment the spatial dimension by a $2\times$ factor through phase shift reshaping [20]. Then, upsampled feature maps are processed by a convolutional layer (or reconstruction layer with 3×3 kernel) which obtains a single map that is combined with a bicubic-upsampled image $\mathbf{X}' \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, where
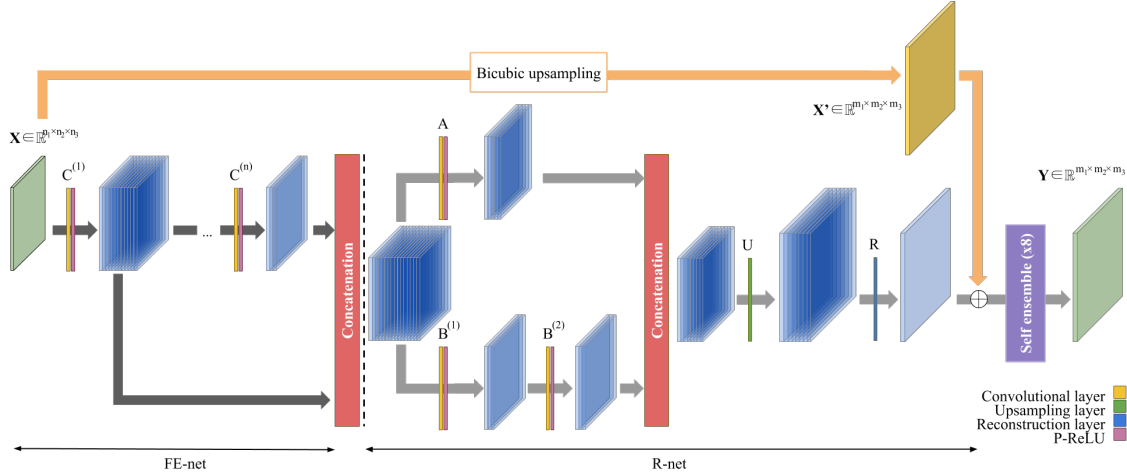
Fig. 1. Proposed model architecture, composed by FE-net (whose layer's outputs are concatenated at the end) and the R-net (composed by the inception module, whose concatenated output feeds the up-sampler, which is refined by a bicubic up-sampling). The final output is obtained as the ensemble of 8 models

$m_1 = f \cdot n_1$, $m_2 = f \cdot n_2$ and $m_3 = n_3$ being f the scaling factor. Finally, with the aim of stabilizing the performance of the model, a self-ensemble learning step has been included [21], obtaining the final super-resolved image $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ as the ensemble of 8 models. Model topology is described in Table I.

III. EXPERIMENTS

A. Datasets and methods

The remote sensing UC Merced dataset has been considered in our experiments, which contains 21 ground-truth classes with 100 RGB images per class and a spatial resolution of 0.3 m/pixel. The dataset has been randomly split into two balanced halves for training and test purposes. Besides, 20% of the training data has been reserved for validation. Regarding the experimental protocol, the original 256×256 high-resolution images have been initially down-sampled using the considered scaling factor and the bi-cubic kernel to generate the corresponding low-resolution counterparts. Then, the low-resolution test images have been super-resolved using different learning-based SR models after training and

validating the methods. It is important to highlight that all the experiments have been carried out following the standard SR procedure for RGB imagery, which is based on the uniform color space transformation YCbCr [4]. In order to assess the SR performance, two reference image quality metrics have been employed: peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [22]. Our experiments include seven different SR methods available in the literature: the bi-cubic interpolation kernel (BC), the original sparse coding approach (SR) [7] and five different DL-SR methods, i.e. SRCNN [11], FSRCNN [12], CNN-7 [14], LGCNet [14] and the proposed approach. An additional dataset has been also considered in order to assess the performance of the proposal when transferring the knowledge learned from the UC Merced dataset to a different remotely sensed image collection. The considered external test collection is made of two multi-spectral data products captured by the GaoFen-2 satellite, where only the RGB channels (3.2 m/pixel) have been selected for the experiments. This collection has been kindly provided by LGCNet authors for qualitatively assessment, since there is not

available high-resolution reference data.

B. Results and discussion

Tables II-IV provide a quantitative assessment of the experiments conducted over the UC Merced dataset. In particular, Table II presents the average PSNR (dB) and SSIM results for the considered SR methods when using three different scaling factors, i.e. $2\times$, $3\times$ and $4\times$. Table III shows the average PSNR (dB) metric per class when considering a $3\times$ upscaling factor. Additionally, Table III contains a summary of the quantitative metric results as well as the corresponding inference times for the considered SR methods. Note that the best metric value in all the tables is highlighted using bold font.

According to the average quantitative results reported in Table II, it is possible to point out several important observations. The first remarkable point is related to the effect of using different scaling factors when super-resolving the considered test images. As we can observe, all the considered methods obtain a better metric result with smaller magnification ratios because, logically, there is more available visual information to introduce high-frequency components when considering small factors over the UC Merced dataset. Nonetheless, the proposed approach consistently provides the highest metric improvement with respect to the second best method for all the tested scaling ratios and metrics. In the case of the PSNR metric, the proposed approach outperforms, on average, LGCNet in 0.20dB, CNN-7 in 0.45dB, FSRCNN in 0.40dB, SRCNN in 0.70dB, SC in 0.95dB and the bi-cubic baseline (BC) in 2.17dB. Analogously, the proposed approach average SSIM result is higher on a 0.013 for LGCNet, 0.020 for CNN-7, 0.019 for FSRCNN, 0.026 for SRCNN, 0.030 for SC and 0.068 for BC. This initial quantitative comparison reveals that the most recent DL-based SR models, i.e. LGCNet, CNN-7, FSRCNN and the proposed approach,

HR BC (27.53dB) SC (28.37dB) SRCNN (28.85dB)

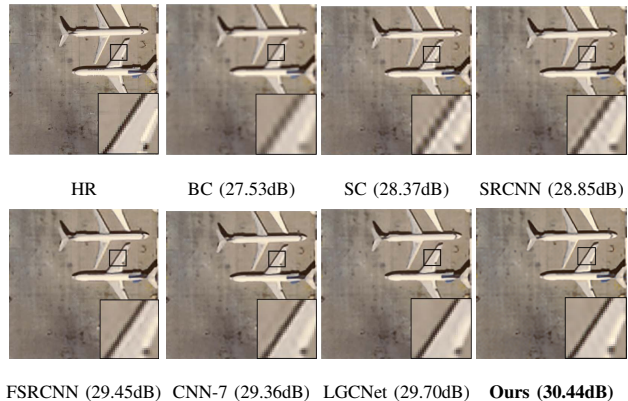FSRCNN (29.45dB) CNN-7 (29.36dB) LGCNet (29.70dB) **Ours (30.44dB)**

Fig. 2. UC Merced airplane test image qualitative assessment for a $3\times$ factor.

are the most effective to super-resolve the UC Merced remote sensing data for all the considered scaling ratios. However, the proposed approach clearly provides the best average result in terms of the PSNR and SSIM image quality metrics.

When analyzing the quantitative SR results per class, we can also observe a remarkable improvement provided by the proposed approach. Table III shows that the proposed approach obtains the best PSNR result using a $3\times$ scaling factor in 13 UC Merced categories whereas the second best SR method (LGCNet) achieves the highest value in the rest classes (8). Despite the fact that LGCNet and other recent DL competitors provide a good SR performance, the proposed approach shows a superior overall result when evaluating the individual class results in more detail. On the one hand, the proposed approach absolute PSNR class improvement over LGCNet is 13.06dB which indicates the proposed architecture effectiveness to super-resolve UC Merced remote sensing data. On the other hand, it is also possible to observe that the proposed approach is especially effective in those classes which contain a high level of spatial details, such as dense residential, harbor or parking, and therefore they require a higher amount of spatial details in the super-resolution process.

TABLE II

QUANTITATIVE ASSESSMENT FOR THE CONSIDERED SR METHODS (IN COLUMNS) USING THREE DIFFERENT SCALING FACTORS (IN ROWS).
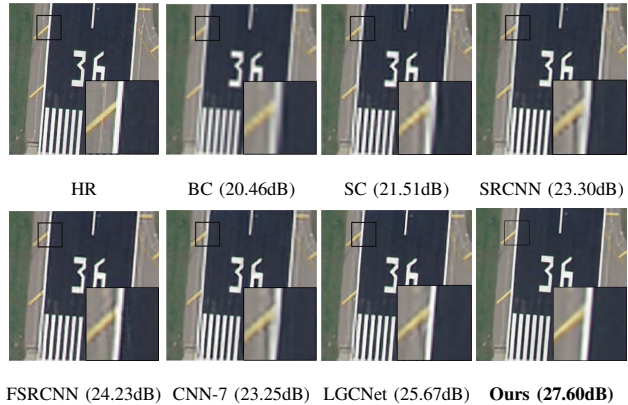
	Bicubic	SC [7]	SRCNN [11]	FSRCNN [12]	CNN-7 [14]	LGCNet [14]	Proposed
scale	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
2	30.76 / 0.8789	32.77 / 0.9166	32.84 / 0.9152	33.18 / 0.9196	33.15 / 0.9191	33.48 / 0.9235	**33.65 / 0.9274**
3	27.46 / 0.7631	28.26 / 0.7971	28.66 / 0.8038	29.09 / 0.8167	29.02 / 0.8155	29.28 / 0.8238	**29.52 / 0.8394**
4	25.65 / 0.6725	26.51 / 0.7152	26.78 / 0.7219	26.93 / 0.7267	26.86 / 0.7264	27.02 / 0.7333	**27.22 / 0.7528**

TABLE III

CLASS-BASED UC MERCED QUANTITATIVE SUPER-RESOLUTION ASSESSMENT CONSIDERING A $3\times$ SCALING FACTOR.

Class	Bicubic	SC [7]	SRCNN [11]	FSRCNN [12]	CNN-7 [14]	LGCNet [14]	Proposed
1	26.86	27.23	27.47	27.61	27.59	27.66	**29.06**
2	26.71	27.67	28.24	28.98	28.81	29.12	**30.77**
3	33.33	34.06	34.33	34.64	34.59	**34.72**	33.76
4	36.14	36.87	37.00	37.21	37.22	**37.37**	36.38
5	25.09	26.11	26.84	27.50	27.39	27.81	**28.51**
6	25.21	25.82	26.11	26.21	26.22	26.39	**26.81**
7	25.76	26.75	27.41	28.02	27.89	28.25	**28.79**
8	27.53	28.09	28.24	28.35	28.35	28.44	**28.16**
9	27.36	28.28	28.69	29.27	29.16	29.52	**30.45**
10	35.21	35.92	36.15	36.43	36.39	**36.51**	34.43
11	21.25	22.11	22.82	23.29	23.32	23.63	**26.55**
12	26.48	27.20	27.67	28.06	27.99	28.29	**29.28**
13	25.68	26.54	27.06	27.58	27.48	**27.76**	27.21
14	22.25	23.25	23.89	24.34	24.30	24.59	**26.05**
15	24.59	25.30	25.65	26.53	26.19	26.58	**27.77**
16	21.75	22.59	23.11	23.34	23.37	23.69	**24.95**
17	28.12	28.71	28.89	29.07	29.03	**29.12**	28.89
18	29.30	30.25	30.61	31.01	30.93	31.15	**32.53**
19	28.34	29.33	29.40	30.23	29.94	**30.53**	29.81
20	29.97	30.86	31.33	31.92	31.87	**32.17**	29.02
21	29.75	30.62	30.98	31.34	31.32	**31.58**	30.76

HR   BC (20.46dB)   SC (21.51dB)   SRCNN (23.30dB)

FSRCNN (24.23dB)   CNN-7 (23.25dB)   LGCNet (25.67dB)   **Ours (27.60dB)**

Fig. 3. UC Merced road test image qualitative assessment for a $4\times$ factor.

In addition to the quantitative evaluation reported by Tables II-III, some visual super-resolved results are provided in Figs. 2-5 as a qualitative evaluation for the tested methods. In particular, Figs. 2-3 show the SR results for two specific test UC Merced images, airplane and road considering $3\times$ and $4\times$ scaling factors respectively. The visual results presented in Figs. 2-3 reveal that each particular SR model tends to generate different visual features on the super-resolved output. Whereas the bi-cubic baseline (BC) together with SC and SRCNN seem very sensitive to the upscaling aliasing effect, the most recent DL-based models FSRCNN, CNN-7, LGCNet and the proposed approach provide a substantially more robust result for UC Merced remote sensing data. That is, SC and SRCNN models were designed to super-resolve images starting from their corresponding interpolated versions which logically introduces an unavoidable aliasing effect which is eventually super-resolved in the final result. The most recent DL-SR models, e.g. FSRCNN, CNN-7, LGCNet and the proposed approach, work for relieving this effect by using deeper architectures which allow them to recover cleaner high-resolution image patterns. For instance, it is easy to appreciate in Fig. 2 that BC, SC and SRCNN introduce an important aliasing effect on the airplane wing.

According to the visual result displayed in Fig. 3, it

is possible to see that the proposed approach is able to remove a large amount of the noise present in the road details of FSRCNN, CNN-7 and LGCNet. That is, the image lines of the proposed approach result are sharper and certainly the most similar to their high-resolution counterparts which eventually leads to the best visual perceived quality. Moreover, the qualitative super-resolution results provided in Figs. 4-5 show that the proposed model is able to reduce the ringing artifacts when compared to other methods which also indicates the proposed model higher robustness to transfer the knowledge acquired from the UC Merced dataset to the GaoFen-2 collection despite the existing spatial resolution differences. Finally, the computational time results reported in Table IV reveal that the proposed approach is able to achieve a high computational performance level using a GPU NVIDIA GeForce GTX 1080 whereas the considered competitors make use of a more powerful GPU device, i.e. NVIDIA Titan Z.

The current trend in DL-SR is inspired on using deeper architectures to improve the resulting performance (e.g. [13], [14]), however this practice may result in a poor propagation of activations and gradients which eventually degrades the quality of the convolutional features and hence the super-resolved result because output images can be affected by noisy artifacts. To mitigate these issues, the proposed approach makes use of a combination of residual units and skip connections in order to extract more informative features on both local and global image areas. On the other hand, parallelized 1×1 convolutional filters are used to reconstruct the resulting super-resolved image by means of a network architecture which substantially reduces the dimensions of the previous layers in order to minimize the information loss through the network. The combination of these improvements allows the proposed approach to produce a more effective remote sensing image computation while
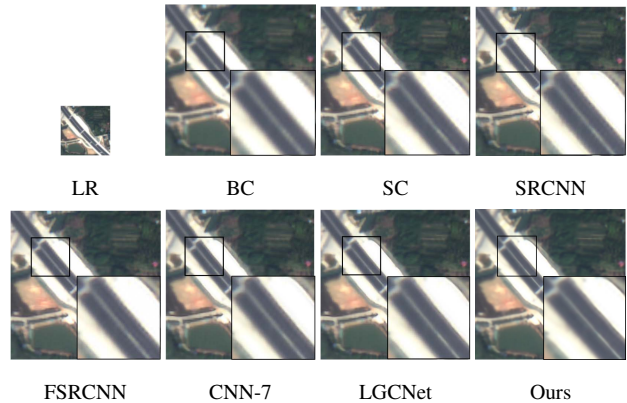

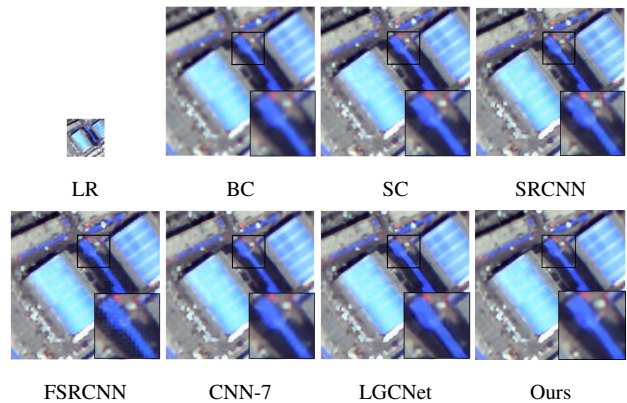Fig. 4. GaoFen-2 airport test image qualitative assessment for a $3\times$ factor.

Fig. 5. GaoFen-2 factory test image qualitative assessment for a $4\times$ factor.

achieving state-of-the-art performance.

TABLE IV

METRICS SUMMARY FOR THE BEST CONSIDERED SR METHODS

scale	VDSR [13] PSNR / SSIM / time	LGCNet [14] PSNR / SSIM / time	LGCNet+ [14] PSNR / SSIM / time	Proposed PSNR / SSIM / time
2	33.47/0.9234/0.119	33.48/0.9235/**0.063**	33.53/0.9242/0.070	**33.65/0.9274**/0.187
3	29.34/0.8263/0.118	29.28/0.8238/**0.061**	29.35/0.8251/0.069	**29.52/0.8394**/0.100
4	27.11/0.7360/0.120	27.02/0.7333/**0.061**	27.13/0.7375/0.073	**27.22/0.7528**/0.066

IV. CONCLUSIONS AND FUTURE LINES

This letter presents a novel remote sensing single-image super-resolution approach based on a deep efficient compendium model. The proposed architecture integrates different improvements on the network design to achieve state-of-the-art performance to super-resolve remote sensing data: (i) residual units, (ii) skip

connections and (iii) network in network. Our experiments, conducted over the remote sensing UC Merced dataset and GaoFen-2 tests images using seven different SR methods available in the literature, reveal that the presented approach is able to archive a state-of-the-art SR performance in the remote sensing field. Future work will be aimed at extending the proposed model to deep self-learning architectures and comprehensively analyzing the effect of considering different network modifications within the remote sensing SR domain.

REFERENCES

[1] Y. Xian, Z. I. Petrou, Y. Tian, and W. N. Meier, "Super-resolved fine-scale sea ice motion tracking," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5427–5439, 2017.

[2] Y. Zhang, Y. Zhang, W. Li, Y. Huang, and J. Yang, "Super-resolution surface mapping for scanning radar: Inverse filtering based on the fast iterative adaptive approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 127–144, 2018.

[3] S. Hao, W. Wang, Y. Ye, E. Li, and L. Bruzzone, "A deep network architecture for super-resolution-aided hyperspectral image classification with classwise loss," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14, 2018.

[4] R. Fernandez-Beltran, P. Latorre-Carmona, and F. Pla, "Single-frame super-resolution in remote sensing: a practical overview," *International Journal of Remote Sensing*, vol. 38, no. 1, pp. 314–354, 2017.

[5] Y. Zhang, F. Ling, X. Li, and Y. Du, "Super-resolution land cover mapping using multiscale self-similarity redundancy," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 11, pp. 5130–5145, 2015.

[6] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "A new deep generative network for unsupervised remote sensing single-image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, no. 99, 2018.

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

[8] R. Fernandez-Beltran, P. Latorre-Carmona, and F. Pla, "Latent topic-based super-resolution for remote sensing," *Remote Sensing Letters*, vol. 8, no. 6, pp. 498–507, 2017.

[9] B. Hou, K. Zhou, and L. Jiao, "Adaptive super-resolution for remote sensing images based on sparse representation with global joint dictionary model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2312–2327, April 2018.

[10] X. Lu, Y. Yuan, and P. Yan, "Alternatively constrained dictionary learning for image superresolution," *IEEE transactions on cybernetics*, vol. 44, no. 3, pp. 366–377, 2014.

[11] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[12] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision*. Springer, 2016, pp. 391–407.

[13] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

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

[15] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, no. 2, 2017, p. 5.

[16] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 1963–1974, 2017.

[17] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[20] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[21] R. Timofte, R. Rothe, and L. Van Gool, "Seven ways to improve example-based single image super resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1865–1873.

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.