An Efficient Deep Unsupervised Superresolution Model for Remote Sensing Images

Mohammad Moein Sheikholeslami [ID], Saeed Nadi [ID], Amin Alizadeh Naeini [ID], and Pedram Ghamisi [ID], *Senior Member, IEEE*

Abstract—**Superresolution (SR) has provided an effective solution to the increasing need for high-resolution images in remote sensing applications. Among various SR methods, deep learning-based SR (DLSR) has made a significant breakthrough. However, supervised DLSR methods require a considerable amount of training data, which is hardly available in the remote sensing field. To address this issue, some research works have recently proposed and revealed the capability of deep learning in unsupervised SR. This article presents an efficient unsupervised SR (EUSR) deep learning model using dense skip connections, which boosts the reconstruction performance in parallel with the reduction of computational burden. To do this, several blocks containing densely connected convolutional layers are implemented to increase the depth of the model. Some skip connections also concatenate feature maps of different blocks to enable better SR performance. Moreover, a bottle-neck block abstracts the feature maps in fewer feature maps to remarkably reduce the computational burden. According to our experiments, the proposed EUSR leads to better results than the state-of-the-art DLSR method in terms of reconstruction quality with less computational burden. Furthermore, results indicate that the EUSR is more robust than its rival in dealing with images of different classes and larger sizes.**

Index Terms—**Deep learning, remote sensing, superresolution (SR).**

I. INTRODUCTION

A SIZABLE number of applications in remote sensing [1]–[3] need high-spatial resolution images (HSRIs). However, HSRIs are not often accessible due to several reasons including high cost of image acquisition and technological and physical limits of sensors in capturing images with more spatial details [4]. To overcome these restrictions, data fusion or superresolution (SR) techniques have been applied. Between these two solutions, the SR has attracted more attention because it is applicable even in the case of single images where there is no additional information.

In recent years, a wide range of SR methods, including nonnegative matrix factorization based methods [5], [6] and sparse representation based methods [7], [8] have been proposed. According to the SR literature, deep learning-based SR (DLSR) methods initially proposed inspired by sparse representation based methods and led to significant improvement compared with their rivals [9], [10]. These methods are generally categorized into two main supervised and unsupervised groups. In recent years, supervised DLSR has experienced an illustrative development especially for nonsatellite images [11]–[13]. In spite of the fact that satellite images are of a more complex nature contaminated by noise, fewer research works have been carried out regarding the SR in this field [14]. In [15], a convolutional model was trained using satellite images to show the effectiveness of appropriate training datasets in yielding better SR results for remote sensing data. Lei *et al.* [16] proposed a local-global combined model to reconstruct high-resolution (HR) images through learning multilevel representations of the low-resolution (LR) counterparts. In [17], a number of convolutional models were trained using a set of Landsat and Sentinel-2 image pairs to assess the SR performance for spatial enhancement of Landsat images. In a different approach, Ma *et al.* [18] developed a DLSR scheme in 2019, which superresolves the LR image in the frequency domain. In this research, first the LR image is transformed to the frequency domain, then a deep learning model predicts the high-frequency details, and eventually an inverse transformation reconstructs the superresolved image in the original (spatial) domain.

Since supervised DLSR methods usually lead to better results than unsupervised ones, most of the research works have been dedicated to this group. However, the unavailability of high-quality training data can justify the need for unsupervised DLSR methods. To the best of our knowledge, in the remote sensing community, there are only two unsupervised DLSR methods [19], [20], each of which adopted a different approach to the problem. Wang *et al.* [20] proposed an unsupervised DLSR model, inspired by CycleGAN (cycle generative adversarial network) [21], which consists of two separate generative networks. The first network superresolves LR images and the second one downsamples HR images. The main drawback of this model is its need for unpaired training datasets. In other words, the model still needs a huge amount of LR–HR image pairs for training. This disadvantage is addressed by Haut *et al.* [19], and their proposed model only utilizes a single LR input image. In that work, a generative model was presented to reconstruct the

HR image from a random noise with the same size as the HR. Once the model is optimized, the input is replaced with the HR output for the next iteration. This procedure is repeated with a predetermined number of iterations. A challenging issue in that study is related to the loss function, where there are not any target HR images to be compared with the output. This issue has been addressed by downsampling the HR output and then computing the loss value using the LR image as the target. The results reported in the article reveals the proficiency of the model in unsupervised SR task for satellite images. However, its performance in terms of reconstruction, and especially computational burden should be still improved. For the sake of brevity, this method has been called as the unsupervised generative SR (UGSR).

It is essential to have in mind that the model has to be trained for input images in such unsupervised DLSR frameworks. This fact highlights the need of these methods for having a model with less computational burden. Thus, an efficient unsupervised SR model (EUSR) based on a DLSR method using dense skip connections (SRDenseNet) [22] is presented in this study. SRDenseNet is a superior supervised SR architecture, which combines low- and high-level feature maps by skip connections for better image reconstruction. Moreover, these types of connections enable the model to achieve deeper status in parallel with optimization time reduction. Unlike the UGSR, the input of our model is initialized by an LR image. This means the downsampling blocks are removed from our model and as a result of this, both layers and computational burden of the EUSR are substantially reduced.

The remainder of this article is organized as follows. Section II provides a description of the methodology used in the proposed model. Section III discusses the results and compares the EUSR with the competing method. Eventually, Section IV concludes the article and asserts some future research.

II. METHODOLOGY

The proposed model aims to learn a mapping function $M : \mathbb{R}^{W \times H \times 3} \rightarrow \mathbb{R}^{2W \times 2H \times 3}$ between LR ($I_{LR}^0 \in \mathbb{R}^{W \times H \times 3}$) and HR ($I_{HR} \in \mathbb{R}^{2W \times 2H \times 3}$) images, where W and H are, respectively, width and height of the LR image. In this sense, each input LR image I_{LR}^0 is fed to the model to generate the corresponding HR image $I_{HR} = M(I_{LR}^0, \theta)$, where θ is the set of model parameters. Due to the absence of the original HR counterpart, the SR output (I_{HR}) is downsampled by a scale factor equal to $1/2$ using Lanczos3 [23] downsampler $\psi : \mathbb{R}^{2W \times 2H \times 3} \rightarrow \mathbb{R}^{W \times H \times 3}$. The downsampling result $I_{LR} \in \mathbb{R}^{W \times H \times 3}$ is calculated as $I_{LR} = \psi(I_{HR})$. Then, I_{LR} and I_{LR}^0 are compared with each other using a loss function (L) based on peak signal-to-noise ratio (PSNR) [24]. PSNR is a metric used to evaluate the quality of an image with respect to another. In the PSNR (1) I and I^0 are, respectively, the target and reference images, and Peak_{sv} is the maximum signal value in the image (e.g., 255 for images with radiometric resolution of 8 bits)

$$\text{PSNR}(I, I^0) = 10 \log_{10} \frac{\text{Peak}_{sv}^2}{\text{MSE}(I, I^0)} \quad (1)$$

Algorithm 1: The EUSR Algorithm.

Input: LR image I_{LR}^0
Output: HR super-resolved image I_{HR}
repeat
   1. Input the LR image I_{LR}^0 to the model M and Estimate the HR image $I_{HR} = M(I_{LR}^0, \theta)$
   2. Downsample I_{HR} using Lanczos3 downsampler ψ to provide $I_{LR} = \psi(I_{HR})$
   3. Compute the loss value $L(\theta) = 1/PSNR(I_{LR}, I_{LR}^0)$
   4. Adam(L) optimizes the M
until satisfy stop criterion
return I_{HR}

---

where MSE stands for mean squared error. The larger PSNR value is equivalent to more similar images. Accordingly, the loss function L is defined as follows:

$$L(\theta) = \frac{1}{\text{PSNR}(I_{LR}, I_{LR}^0)}. \quad (2)$$

Considering $I_{HR} = M(I_{LR}^0, \theta)$, $I_{LR} = \psi(I_{HR})$, (1) and (2)

$$L(\theta) = \frac{1}{10} \left(\log_{10} \frac{\text{Peak}_{sv}^2}{\text{MSE}(\psi(M(I_{LR}^0, \theta)), I_{LR}^0)} \right)^{-1}. \quad (3)$$

The idea is that similar LRs result in similar HRs. To obtain more similar LRs, the model parameters θ need to be optimized. This is accomplished through minimizing the loss function $L(\theta)$ using Adam optimizer [25]. The aforementioned approach is summarized in Algorithm 1.

In the proposed model, different kinds of layers are utilized, including the followings.

1) Two-dimensional (2-D) convolution (Conv.) [26]: A Conv layer consists of a group of kernels and biases, which are applied to extract features from the input. The jth output feature map O_j at the position (x, y) can be calculated as

$$O_j^{xy} = B_j + \sum_{k} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} w_j^{mn} x_k^{(x+m)(y+n)} \quad (4)$$

where k indexes over the depth of the input $X \in \mathbb{R}^{p \times q \times r}$, and w_j^{mn} is the weight value of the 2-D kernel $W \in \mathbb{R}^{M \times N}$ at the position (m, n). Moreover, B_j is the bias matrix.

2) Rectified linear unit (ReLU) activation [27]: This activation layer applies the nonlinear ReLU function $f(x)$, as follows:

$$f(x) = \max\{0, x\}. \quad (5)$$

3) Batch normalization (BN) [28]: This layer applies a normalization on each input batch of data x, using the following equation:

$$y = \frac{x - \text{mean}(x)}{\sqrt{var(x) - \varepsilon}} \cdot \gamma + \beta \quad (6)$$

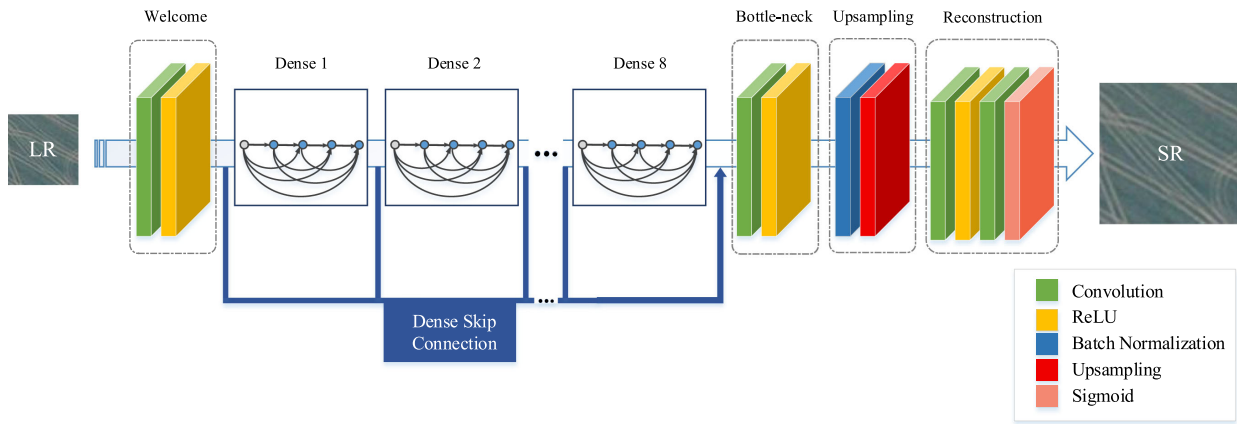where γ, β, are learned during the training, and ε is numerical stability parameter.

Fig. 1. Architecture of the proposed model for unsupervised SR.

4) Upsampling: This layer upsamples input feature maps using bilinear method and a given scale factor.
5) Sigmoid activation: Sigmoid is also an activation function, which maps each input *x* to a value between 0 and 1. The sigmoid function is provided in (7), as follows:

$$Sig(x) = (1 + \exp(-x))^{-1}. \tag{7}$$

For our model, an architecture was needed to provide both high SR performance and low computational burden. To satisfy this need, the proposed architecture was inspired by the SRDenseNet, which is based on a dense convolutional network (DenseNet) [29]. For our purpose, due to the following reasons, the DenseNet architecture has been considered. The DenseNet uses dense connections, which improve the gradient and information flow throughout the model. Moreover, the DenseNet requires fewer parameters than alternative architectures (specifically residual networks [30]) [29].

As Fig. 1 demonstrates, the model is constituted of several blocks, namely, welcome, dense, bottleneck, upsampling, and reconstruction blocks.

A. Welcome

This block is the first block encountered by the LR image and involves 2-D Conv and ReLU activation [27] layers. This block aims to extract some low-level feature maps as inputs to dense blocks.

B. Dense

Dense block is a set of Conv layers each of which is followed by ReLU and BN [28] layers; in addition, the output of each layer is fed to all subsequent layers. However, in our study, in order to both boost and accelerate the model, BN layers are eliminated from the dense blocks (Fig. 2). Dense blocks have been initially introduced in DenseNet [29]. Employing dense blocks, models with fewer parameters and less computational burden are generated because feature maps are reused for several layers. Furthermore, as it is obvious from Fig. 1, skip connections are applied to concatenate feature maps from different levels. This way, the upsampling block could benefit from the current
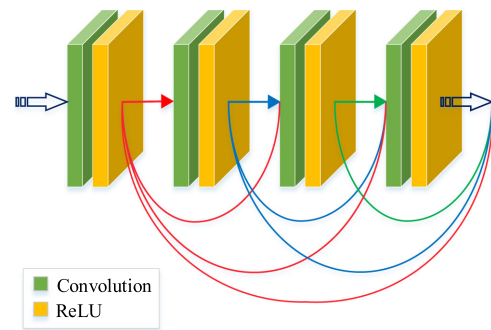

Fig. 2. Architecture of a four-layer dense block, whereas dense blocks in the proposed model consist of eight layers.

collection of the feature maps ranging from low- to high-level ones.

C. Bottleneck

As mentioned earlier, the feature concatenation results in a set of feature maps obtained by all the previous layers, and this puts a huge computational burden on the model. The solution applied here is the bottleneck block which involves a 256-kernel Conv followed by a ReLU activation. This block summarizes 1040 feature maps in only 256 feature maps, which considerably decreases the computational burden of the model.

D. Upsampling

In this block, a BN layer and a bilinear upsampling layer afterward undergo to enlarge the size of the feature maps by a scale factor of two. By doing this, feature maps are prepared to reconstruct the output SR image.

E. Reconstruction

The reconstruction block is a set of two Conv layers each of which is followed by ReLU and sigmoid activation layers, respectively. The second Conv layer in this block has three kernels to generate the RGB channels of the SR image. In

Fig. 3. Image set used in this experiment, from top left to bottom right: agricultural, chaparral, forest, intersection, and residential, residential2. First and second rows indicate vegetation and urban classes, respectively.

addition, the application of the sigmoid is to map the values between zero and one. By doing so, the values of the output image become similar to the target LR image. Note that, unlike the common approach using BN along with Conv layers, we have reduced the use of BN in a way that it is just used in upsampling block, and this task increases the speed of the convergence.

III. EXPERIMENTS

In this section, the proposed method is compared with the UGSR, as the most state-of-the-art method, and three other non-DLSR methods. The experiments have been divided into four subsections, namely, datasets, parameter settings, accuracy assessment, convergence speed analysis, and sensitivity analysis to the image size.

A. Datasets

We have used two datasets, namely UC Merced [31] and RSCNN7 [32]. UC Merced includes 2100 remote sensing images in 21 various land use classes, i.e., 100 images per each class. The image size in the dataset is 256×256 pixels. RSCNN7 is composed of seven different classes, including grass, field, industry, lake, forest, residential, and parking. Each class has 400 images of size 400×400 pixels. For our experiments, we have chosen six different images which could be classified into vegetation and urban categories (see Fig. 3). These two classes have been selected due to their importance and generality in a wide range of applications. We then cropped them into 128×128 pixels to accelerate the experiments. However, image crops with the resolution of 256×256 pixels have been applied in sensitivity analysis to the image size. To generate an LR version of these HR images, we have followed a downsampling procedure using the Lanczos3 kernel. This way, the acquisition condition of the sensor has been preserved, whereas the spatial resolution has been reduced. Thus, the HR act as test reference data.

TABLE I
COMPREHENDED DETAILS OF THE PROPOSED MODEL ARCHITECTURE

Block	Layers	Kernel Size	No. Kernels
Welcome	Conv+ReLU	(1, 1)	16
Dense Block×8	[Conv+ReLU]×8	(3, 3)	16
Bottle-neck	Conv+ReLU	(1, 1)	256
Upsampling	BN	-	-
	Upsampling(2x)	-	-
Reconstruction	Conv+ReLU	(5, 5)	128
	Conv+Sigmoid	(1, 1)	3

B. Parameter Settings

SRDenseNet [22] was an inspiration for our model architecture. Consequently, some parameters, including the number of the dense blocks, and the growth rate have been set according to this article. The comprehended configuration of the proposed model is available in Table I. Furthermore, all parameters of the UGSR and other non-DLSR competing methods were set as suggested in [19] and [4], respectively.

C. Accuracy Assessment

In order to evaluate the performance of the EUSR, we selected the UGSR as the most state-of-the-art unsupervised DLSR method in the remote sensing community. Furthermore, three unsupervised non-DLSR methods, namely, blind deblurring (BDB) [33], fast superresolution (FSR) [34], and gradient profile prior (GPP) [35] were selected. The reason why we chose these three methods is that they conduct unsupervised SR without using any data except for input LR image, just like the proposed method. The UGSR was suggested in [19] to be trained for 2000 epochs ($P = 2000$). On the other hand, for the EUSR, an early stopping technique was implemented with the patience set to 60 epochs ($P = 60$), which was determined by [22]. This arrangement for stopping criteria has been made in all subsections except for Subsection III-D, where both criteria have been separately considered to analyze the convergence speed of each method. In addition, the structural similarity (SSIM) [36] and the PSNR have been utilized as frequently used metrics for assessment of the SR results.

According to Table II, in all images except for intersection and residential2 images, the EUSR outperforms the competing methods in terms of PSNR. In addition, the proposed method obtains equal or higher SSIM values than the competing methods in all images. Moreover, DLSR methods (the EUSR and UGSR) prove their superior performance with respect to non-DLSR methods (BDB, FSR, and GPP) according to Table II and Fig. 4; thus, in the next experiments only DLSR methods, i.e., the EUSR and UGSR are compared. As it is clear from Fig. 5, apart from the agricultural image, the results of the DLSR methods are almost close to each other. However, the EUSR shows an illustrative enhancement in terms of convergence speed and computational burden (Subsection III-D).
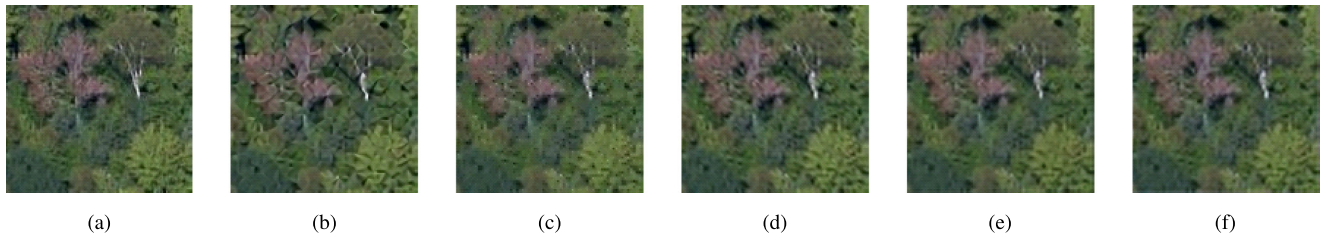
Fig. 4. Results of the competing and the proposed methods for the image forest: (a) HR(PSNR(dB)\SSIM), (b) BDB(22.15\0.59), (c) FSR(28.95\0.89), (d) GPP(28.70\0.89), (e) UGSR(29.69\0.91), and (f) EUSR(29.90\0.91).

TABLE II
COMPARISON OF THE PROPOSED METHOD AND THE COMPETING METHODS IN TERMS OF THE PSNR(dB) AND SSIM USING IMAGES WITH THE SIZE EQUAL TO 128 × 128 PIXELS

Images	Metrics	BDB	FSR	GPP	UGSR	EUSR
agricultural	PSNR	27.73	34.91	35.67	30.33	**36.36**
	SSIM	0.79	0.94	**0.95**	0.83	**0.95**
chaparral	PSNR	19.87	27.42	26.52	28.53	**28.58**
	SSIM	0.71	0.92	0.92	**0.94**	**0.94**
forest	PSNR	22.15	28.95	28.70	29.69	**29.90**
	SSIM	0.59	0.89	0.89	**0.91**	**0.91**
intersection	PSNR	23.07	29.02	28.48	**29.76**	29.72
	SSIM	0.63	0.88	0.86	0.88	**0.89**
residential	PSNR	17.72	23.69	23.01	24.08	**24.21**
	SSIM	0.62	**0.86**	0.84	0.85	**0.86**
residential2	PSNR	17.95	26.32	25.02	**27.94**	27.68
	SSIM	0.56	0.90	0.88	**0.92**	**0.92**

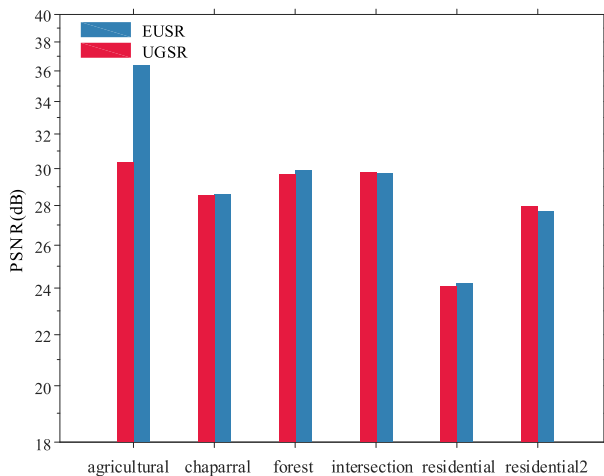The best result for each image and metric is highlighted in bold font.

Fig. 5. PSNR(dB) results for the UGSR and the proposed method.

D. Convergence Speed Analysis

By convergence speed, we mean the ability of a model in converging to a final desired answer as soon as possible. This proves the efficiency of a method in terms of the computational burden. Note that methods with lower computational burden need lower computational resources. Computational burden has

been always an important criterion to rank the proficiency of methods. Considering this, we have designed an experiment to compare the time and epochs needed for model convergence and the PSNR of the results. To do so, we have run the models and got the results in two different ways: first, after 2000 epochs of training ($P = 2000$), second, after 60 epochs without any decrease in the loss value ($P = 60$). We have selected 2000 epochs because our rival has introduced this as the optimum value [19]. We have also employed 60 epochs as the early stopping patience according to [22].

It is worth mentioning that all the experiments were conducted on an ASUS FX-553 laptop that has a NVIDIA GeForce GTX-1050Ti GPU (4 GB GDDR5). Additionally, all the DLSR models have been implemented using Keras [37] and TensorFlow [38].

The results show the prominence of the proposed method in the convergence (see Table III).

In addition to the impressive performance in terms of PSNR, as can be seen from Table III, the proposed method robustly converges before 250 training epochs to the results even better than ones achieved by the UGSR after 2000 epochs. High convergence speed along with high accuracy is the factors indicating the efficiency of a method [39]. In our proposed method, convergence speed has increased by 90.24% while accuracy has been better than that of the rival that means, the EUSR is much more efficient than the UGSR.

Our efficiency in the computational burden, can be explained from two aspects: removing BNs, and applying an LR image as an input instead of a random noise. Applying an LR image, in comparison to the UGSR input of which is a random noise with the same size as HR, significantly decreases the number of layers. Unlike the hourglass architecture [40] used in the UGSR, in the EUSR there is no need to downsampling blocks, which constitutes almost half of the layers. Therefore, the model is optimized faster. Moreover, as Fig. 6 presents, using the LR image as the model input, we apply all the data in hand and start from higher PSNR values. This obviously requires fewer optimization epochs, compared with the random noise input.

E. Sensitivity Analysis to the Image Size

Regarding the image size, it is to say that the research works suffer from limitations of data, time, and hardware resources especially in the DLSR field. However, practically

TABLE III
PSNR (DB), NUMBER OF EPOCHS, AND TIME (S) PROVIDED FOR THE UGSR AND THE PROPOSED METHOD USING
TWO PATIENCE VALUES ($P = 60$ AND $P = 2000$)

Image	Method	PSNR(dB)		Epoch		Time(s)	
		$P = 60$	$P = 2000$	$P = 60$	$P = 2000$	$P = 60$	$P = 2000$
agricultural	UGSR	26.18	30.33	**159**	2000	33	241
	EUSR	**36.36**	**36.47**	193	2000	**25**	**142**
chaparral	UGSR	16.96	28.53	**33**	2000	**19**	241
	EUSR	**28.58**	**28.61**	137	2000	22	**144**
forest	UGSR	21.90	29.69	**54**	2000	**21**	240
	EUSR	**29.90**	**30.00**	154	2000	23	**143**
intersection	UGSR	21.48	29.76	210	2000	39	239
	EUSR	**29.72**	**29.77**	**182**	2000	**25**	**141**
residential	UGSR	14.38	24.08	**66**	2000	**23**	244
	EUSR	**24.21**	**24.27**	200	2000	26	**143**
residential2	UGSR	14.68	**27.94**	**57**	2000	21	240
	EUSR	**27.68**	27.71	118	2000	**20**	**142**

The best result for each image, patience value and metric is highlighted in bold font.
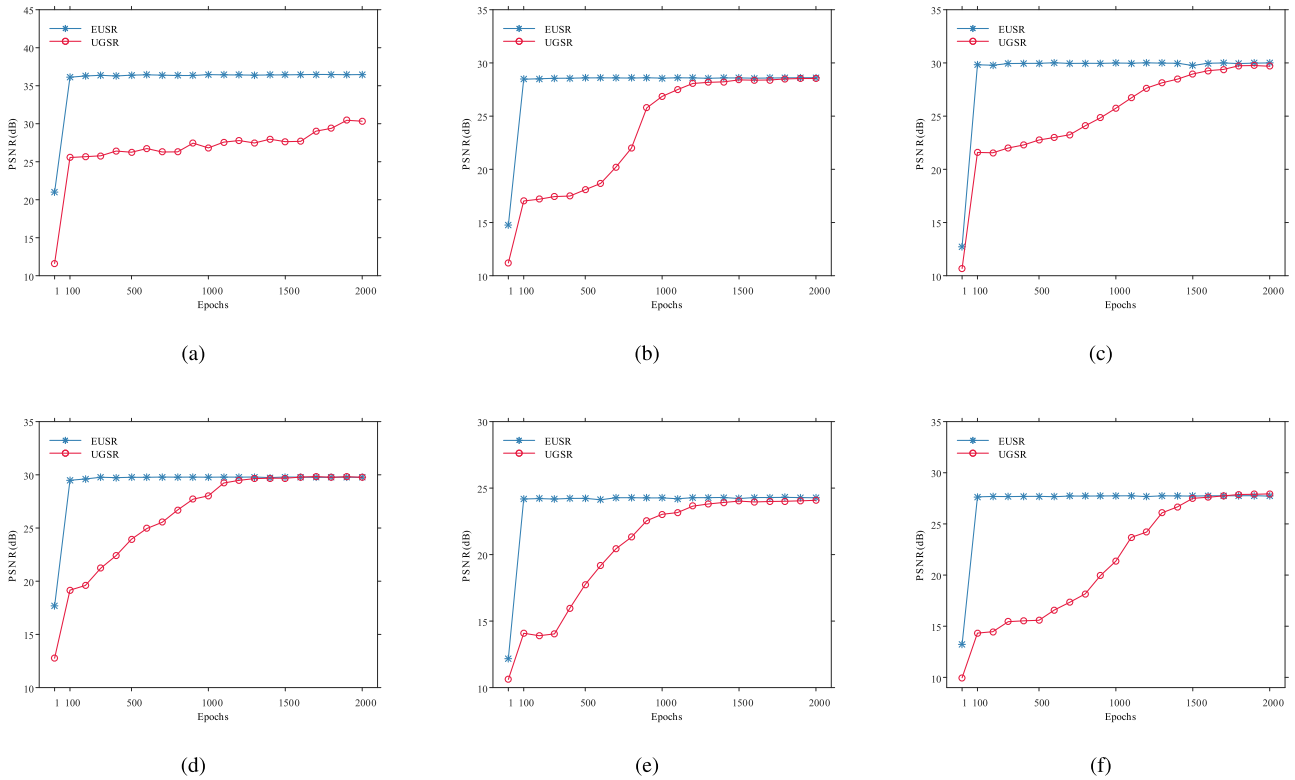

Fig. 6. PSNR evolution of the UGSR and the proposed method for the images: (a) agricultural, (b) chaparral, (c) forest, (d) intersection, (e) residential, and (f) residential2.

it is essential and inevitable to work on images with larger sizes. Accordingly, this analysis has been designed to assess the performance and applicability of our method in case of larger sizes. In this experiment, we have used larger crops (256×256 pixels) of the image set. As it can be seen from Table IV, the proposed method achieves a remarkable performance with respect to the UGSR (a 49.91% improvement on average of six images). The results are brought for $P = 60$

and $P = 2000$ for the EUSR and the UGSR, respectively. It is worth mentioning that the UGSR method might obtain more acceptable results in case of more training epochs. Nonetheless, the UGSR needs much more epochs for training in case of larger images, which proves less efficiency of this method for practical applications.

In Figs. 7 and 8, the sensitivity of both methods to the image size can be observed. Increasing the image size, not only raises

(a)   (b)   (c)
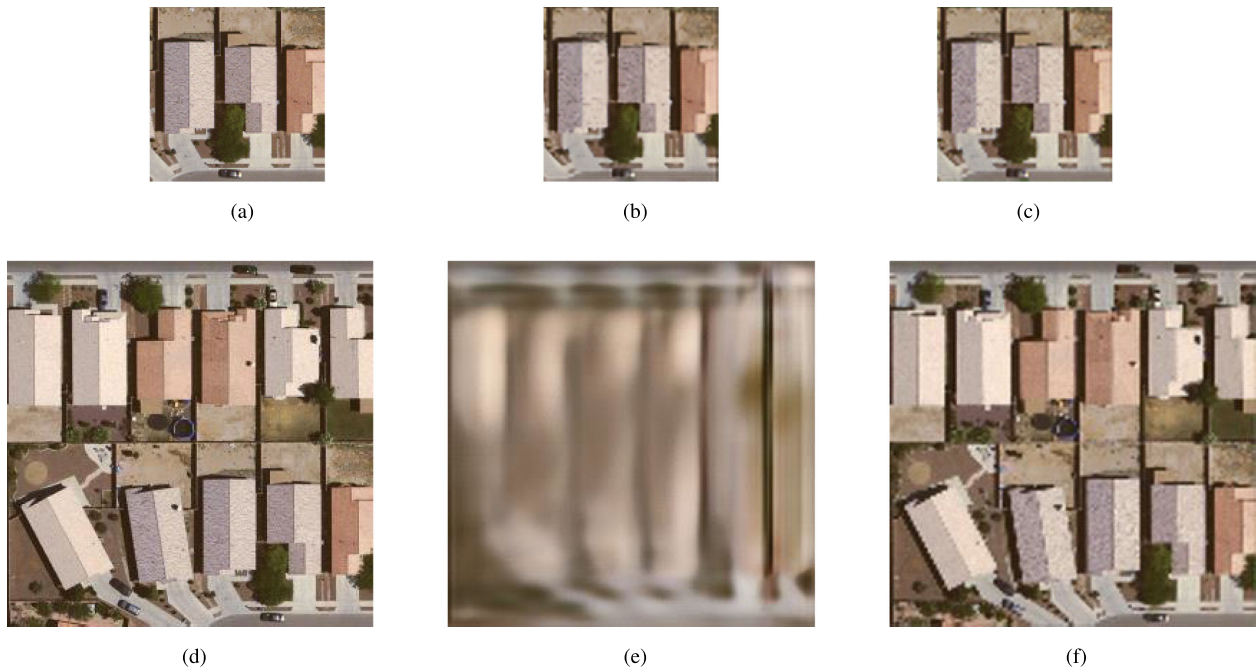
(d)   (e)   (f)

Fig. 7. HR and superresolved results obtained by the UGSR ($P = 2000$) and the EUSR ($P = 60$) over the residential image. The first and second rows are, respectively, specified to 128×128 and 256×256 pixel resolutions. (a) HR (128×128 pixels), (b) UGSR(128×128 pixels), (c) EUSR(128×128 pixels), (d) HR (256×256 pixels), (e) UGSR(256×256 pixels), and (f) EUSR(256×256 pixels).

(a)   (b)   (c)
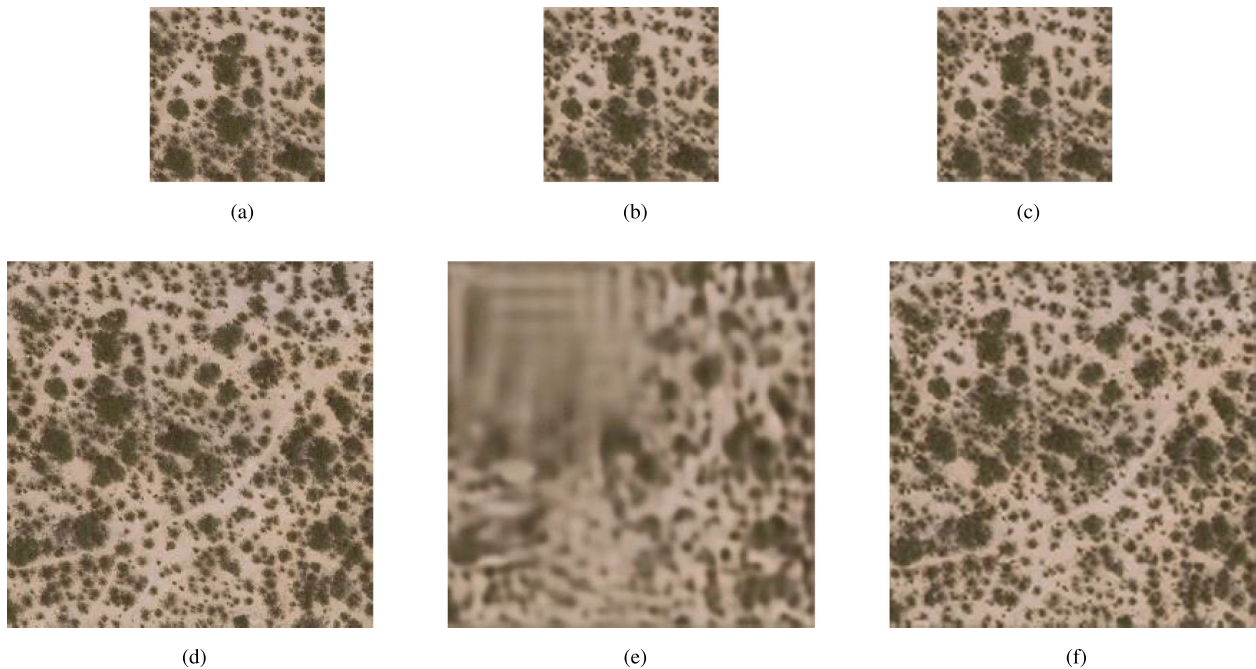
(d)   (e)   (f)

Fig. 8. HR and superresolved results obtained by the UGSR ($P = 2000$) and the EUSR ($P = 60$) over the chaparral image. The first and second rows are, respectively, specified to 128×128 and 256×256 pixel resolutions. (a) HR(128×128 pixels), (b) UGSR(128×128 pixels), (c) EUSR(128×128 pixels), (d) HR(256×256 pixels), (e) UGSR(256×256 pixels), and (f) EUSR(256×256 pixels).

the computational burden but also could cause some artifacts such as watering effect (the case in Figs. 7(e) and 8(e)). To be more likely, as the image size is increased by utilizing larger crops, the information content of the image is also increased, which is equivalent to more spatial details. Reconstruction of HR images which have high spatial details requires a highly competent SR performance. Therefore, our proposed method is more practical than its rival, because in the real world we deal with HR images.

TABLE IV
PSNR (dB) OF 256 × 256 PIXEL IMAGES RESULTED OF THE PROPOSED
($P = 60$) AND THE UGSR ($P = 2000$) METHODS

Image	Method	PSNR(dB)
agricultural	UGSR	26.62
	EUSR	**36.49**
chaparral	UGSR	19.71
	EUSR	**28.58**
forest	UGSR	23.22
	EUSR	**30.38**
intersection	UGSR	17.52
	EUSR	**28.00**
residential	UGSR	14.93
	EUSR	**25.11**
residential2	UGSR	16.57
	EUSR	**28.00**

The best result for each image is highlighted in bold
font.

IV. CONCLUSION

This article presents an efficient unsupervised DLSR method for remote sensing images. The proposed architecture takes advantage of densely connected layers to simultaneously provide robust proficient SR performance and efficiency in the computational burden. Our experiments, including accuracy assessment, convergence speed analysis, and sensitivity analysis to the image size, prove the ability of the proposed model to efficiently achieve a state-of-the-art performance. Future research can be concentrated on implementing perceptual loss in the proposed model to obtain visually more pleasant HR images.

REFERENCES

[1] M.-T. Pham, E. Aptoula, and S. Lefèvre, "Feature profiles from attribute filtering for classification of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 249–256, Jan. 2018.

[2] G. Sumbul, R. G. Cinbis, and S. Aksoy, "Fine-grained object recognition and zero-shot learning in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 770–779, Feb. 2018.

[3] T. Wu, J. Luo, J. Fang, J. Ma, and X. Song, "Unsupervised object-based change detection via a Weibull mixture model-based binarization for high-resolution remote sensing images," *IEEE Geosci. Remote Sens. Letters*, vol. 15, no. 1, pp. 63–67, Jan. 2018.

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

[5] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. A. Morel, "Neighbor embedding based single-image super-resolution using semi-nonnegative matrix factorization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2012, pp. 1289–1292.

[6] X. Li, L. Wang, Q. Cheng, P. Wu, W. Gan, and L. Fang, "Cloud removal in remote sensing images using nonnegative matrix factorization and error correction," *ISPRS J. Photogrammetry Remote Sens.*, vol. 148, pp. 103–113, 2019.

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

[8] X. Li, H. Shen, L. Zhang, H. Zhang, Q. Yuan, and G. Yang, "Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7086–7098, 2014.

[9] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Nov. 2015.

[10] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.

[11] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.

[12] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1637–1645.

[13] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1874–1883.

[14] J. Yang, Y.-Q. Zhao, J. C.-W. Chan, and L. Xiao, "A multi-scale wavelet 3D-CNN for hyperspectral image super-resolution," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1557.

[15] L. Liebel and M. Körner, "Single-image super resolution for multispectral remote sensing data using convolutional neural networks," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 41, pp. 883–890, 2016.

[16] 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.

[17] D. Pouliot, R. Latifovic, J. Pasher, and J. Duffe, "Landsat super-resolution enhancement using convolution neural networks and sentinel-2 for training," *Remote Sens.*, vol. 10, no. 3, 2018, Art. no. 394.

[18] W. Ma, Z. Pan, J. Guo, and B. Lei, "Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3512–3527, Jun. 2019.

[19] 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 Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6792–6810, Nov. 2018.

[20] P. Wang, H. Zhang, F. Zhou, and Z. Jiang, "Unsupervised remote sensing image super-resolution using cycle CNN," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3117–3120.

[21] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.

[22] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4799–4807.

[23] K. Turkowski, "Filters for common resampling tasks," in *Graphics Gems*. New York, NY, USA: Academic, 1990, pp. 147–165.

[24] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, 2008.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[26] S. Ji, W. Xu, M. Yang, and K. Yu, "3-D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[27] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*.

[28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.

[29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[31] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.

[32] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, Nov. 2015.

[33] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 783–798.

[34] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouamé, and J.-Y. Tourneret, "Fast single image super-resolution using a new analytical solution for $l_2 - l_2$ problems," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3683–3697, Aug. 2016.

[35] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[37] F. Chollet *et al.*, "Keras," 2015. [Online]. Available: https://keras.io
[38] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: http://tensorflow.org/
[39] A. A. Naeini, M. Babadi, S. M. J. Mirzadeh, and S. Amini, "Particle swarm optimization for object-based feature selection of VHSR satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 379–383, Mar. 2018.
[40] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 483–499.

Amin Alizadeh Naeini received the M.Sc. and Ph.D. degrees in photogrammetry from the University of Tehran, in 2011 and 2015, respectively.

He was working as an Assistant Professor with the Department of Geomatics Engineering, University of Isfahan from 2015 to 2019. Since then, he has been being with the Department of Earth and Space Science, York University, Canada. His primary research interests include hyperspectral image processing, deep learning based image matching and geometric processing of satellite images.

Mohammad Moein Sheikholeslami was born in 1995. He received the B.S. degree in 2017 in survey engineering from the University of Isfahan, Isfahan, Iran, where he is currently working toward the M.Sc. degree in remote sensing.

His current research interests include SR, geometric processing of satellite images, and application of machine (deep) learning in remote sensing.

Pedram Ghamisi (Senior Member, IEEE) received the B.Sc. degree in civil (survey) engineering from the Tehran South Campus of Azad University, Tehran, Iran, in 2008, the M.Sc. degree (Hons.) in remote sensing from the K. N. Toosi University of Technology, Tehran, Iran, in 2012, and the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2015.

In 2013 and 2014, he was with the School of Geography, Planning, and Environmental Management, University of Queensland, Brisbane, QLD, Australia. In 2015, he was an Alexander von Humboldt Research Fellow with the Technical University of Munich, Munich, Germany, and Heidelberg University, Heidelberg, Germany. From 2015 to 2018, he was a Research Scientist with the German Aerospace Center, Oberpfaffenhofen, Germany. Since 2018, he has been working as the Head of the Machine Learning Group, Helmholtz-Zentrum Dresden-Rossendorf (HZDR). He is also the CTO and the Co-Founder of VasoGnosis Inc., Milwaukee, WI, USA, where he is involved in the development of advanced diagnostic and analysis tools for brain diseases using cloud computing and deep learning algorithms. He is also the Vice-Chair of the IEEE Image Analysis and Data Fusion Committee. His research interests include interdisciplinary research on remote sensing and machine (deep) learning, image and signal processing, and multisensor data fusion.

Dr. Ghamisi was the recipient of the Best Researcher Award for M.Sc. students at the K. N. Toosi University of Technology in the academic year of 2010–2011, the IEEE Mikio Takagi Prize for winning the Student Paper Competition at IEEE International Geoscience and Remote Sensing Symposium in 2013, the Talented International Researcher by Iran's National Elites Foundation in 2016, the first prize of the data fusion contest organized by the Image Analysis and Data Fusion Technical Committee of IEEE-GRSS in 2017, the Best Reviewer Prize of IEEE Geoscience and Remote Sensing Letters in 2017, the Alexander von Humboldt Fellowship from the Technical University of Munich, and the High Potential Program Award from HZDR. He is also an Associate Editor for MDPI-*Remote Sensing* and IEEE Geoscience and Remote Sensing Letters.

Saeed Nadi received the Ph.D. degree in geospatial information systems from the University of Tehran, Tehran, Iran.

He is currently an Assistant Professor with the Department of Geomatics Engineering, University of Isfahan, Isfahan, Iran. His research interests include the use of advanced machine learning techniques for spatial information extraction from remote sensing images, especially for environmental applications.