Exploiting Digital Surface Models for Inferring Super-Resolution for Remotely Sensed Images

Savvas Karatsiolis, Chirag Padubidri, and Andreas Kamilaris

***Abstract*— Despite the plethora of successful super-resolution (SR) reconstruction (SRR) models applied to natural images, their application to remote sensing imagery tends to produce poor results. Remote sensing imagery is often more complicated than natural images, has its peculiarities such as being of lower resolution, contains noise, and often depicts large textured surfaces. As a result, applying nonspecialized SRR models like the enhanced SR generative adversarial network (ESRGAN) on remote sensing imagery results in artifacts and poor reconstructions. To address these problems, we propose a novel strategy for enabling an SRR model to output realistic remote sensing images: instead of relying on feature-space similarities as a perceptual loss, the model considers pixel-level information inferred from the normalized digital surface model (nDSM) of the image. This allows the application of better-informed updates during the training of the model which sources from a task (elevation map inference) that is closely related to remote sensing. Nonetheless, the nDSM auxiliary information is not required during production, i.e., the model infers an SR image without additional data. We assess our model on two remotely sensed datasets of different spatial resolutions that also contain the DSMs of the images: the Data Fusion 2018 Contest (DFC2018) dataset and the dataset containing the national LiDAR flyby of Luxembourg. We compare our model with ESRGAN, and we show that it achieves better performance and does not introduce any artifacts in the results. In particular, the results for the high-resolution DFC2018 dataset are realistic and almost indistinguishable from the ground-truth images.**

***Index Terms*— Deep learning (DL), normalized digital surface model (nDSM), perceptual loss, remote sensing, super-resolution (SR) reconstruction (SRR).**

I. Introduction

HIGH-QUALITY aerial photography and satellite imagery facilitate the development of interesting remote sensing applications for large-scale monitoring and Earth observation, including land monitoring, urban planning, and surveillance. However, the severe weakness of remotely sensed imagery, in general, is its low spatial resolution, i.e., the detailed level is insufficient for detecting certain objects of interest like tree types, solar panels on rooftops, and cars. Remotely sensed imagery usually has low spatial resolution due to the cost and time required to collect high-quality/low-noise images and the vulnerability of such images to environmental variations during acquisition like atmospheric and light variations. Excessive costs may diminish the advantages of using high-quality imagery in remote sensing applications. A common alternative to address this limitation is the use of low-quality (low spatial resolution/high noise) images to reconstruct scene information as much as possible and then perform inference based on the information-enriched data. This strategy maintains the lower image acquisition cost and improves the quality of the final output.

Traditional upsampling methods such as nearest-neighbor and bicubic interpolation [1] rely on surrounding pixels to add a small amount of information to an image and tend to produce blurry and distorted results mainly because they fail to recover high-frequency information. Inevitably, demanding applications like small-object detection tasks do not generally benefit much from image interpolation by methods that rely on neighboring pixels to add some level of detail to the image. Some early attempts to produce better results than the traditional interpolation methods involved the learning of degradation models [2] and the feature matching of low-resolution (LR)/high-resolution (HR) patches to facilitate the recovery of HR images [3], [4], [5]. Slightly more sophisticated methods built sparse representations that comprised of a dictionary used to reconstruct the HR counterpart of an LR patch [6], [7]. While image upsampling using sparse representations tends to slightly improve the recovery of high-frequency information, it is a very computationally intensive technique [7]. The limited performance of these approaches in effectively converting an LR image to its realistic HR counterpart originates from their inability to learn.

One of the reasons that deep learning (DL) and convolutional neural networks (CNNs) have become extremely popular is their ability to supply end-to-end models that perform inference based on raw data without relying on hand-engineered features or extensive incorporation of task-related knowledge into the model. These characteristics positioned DL as the mainstream approach nowadays to solve challenging remote sensing tasks. As such, DL is widely used for tackling the super-resolution (SR) reconstruction (SRR) task, i.e., converting a single LR image to an HR one. The output of the SRR task is called an SR image, and its goal is

Manuscript received 8 April 2022; revised 8 August 2022 and 10 September 2022; accepted 16 September 2022. Date of publication 26 September 2022; date of current version 11 October 2022. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant 739578 and in part by the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy. *(Corresponding author: Savvas Karatsiolis.)*

Savvas Karatsiolis and Chirag Padubidri are with the CYENS Center of Excellence, Nicosia 1016, Cyprus (e-mail: s.karatsiolis@cyens.org.cy; c.padubidri@cyens.org.cy).

Andreas Kamilaris is with the Department of Computer Science, University of Twente, 7522NB Enschede, The Netherlands, and also with the CYENS Center of Excellence, Nicosia 1016, Cyprus (e-mail: a.kamilaris@cyens.org.cy).

Digital Object Identifier 10.1109/TGRS.2022.3209340

to learn how to produce SR images from LR images that are indistinguishable from the ground truth, i.e., the HR images. Furthermore, the development of efficient SSR models will greatly benefit DL models performing a plethora of remote sensing tasks since high-quality imagery is especially beneficial for DL models [8], [9], [10], [11]. With the ever-increasing usage of DL methodologies for developing remote sensing applications [12], [13], [14], [15], [16], training and inferring on HR images greatly increase the chances of obtaining good results on notoriously difficult tasks.

II. Related Work

The first attempts of using DL for the SSR task used pixel loss between the SR output and the HR image (ground truth). Pixel losses are straightforward to implement. Specifically, minimizing the mean squared error (MSE) conveniently maximizes the peak-signal-to-noise ratio (PSNR), which is a commonly used measure for evaluating SRR models. However, PSNR is not a good measure of perceptual similarity because it fails to capture perceptually relevant differences [17]. In particular, the textual detailed level is not reflected in the magnitude of the measured PSNR. Pixel losses tend to produce overly smoothed outputs that constitute candidate HR reconstructions: the model calculates a statistical average of the plausible HR reconstructions introduced to it during training. SRCNN [18] was one of the early attempts that used a DL model trained on a pixel loss for the SRR task. Many following attempts experimented with various advanced architectural features in the DL model to mitigate the effects of pixel losses. Kim *et al.* [19] applied residual learning [20] into a very deep CNN, Zhang *et al.* [21] applied deep residual channel attention mechanisms, and Lai *et al.* [22] proposed the Laplacian super-resolution network (LapSRN), which supported high up-sampling factors with the use of residual skip connections. Despite the extensive focus on identifying novel architectural features that improve SRR models, the gap between the quality of the HR images and the SR outputs remained. To overcome the limitations created by applying a pixel loss between the ground truth and the SR image, Johnson *et al.* [23] introduced a perceptual loss to measure semantic similarity between the two images. They specifically used a Visual Geometry Group (VGG)-16 [24] model trained on ImageNet [25] and minimized the Euclidean distance between the features of the HR images and the features of the SR images (i.e., a perceptual loss). They showed that this strategy allowed the model to reconstruct fine details and edges. These results are in line with Mahendran and Vedaldi [26] who also showed that matching the features of higher layers in the pretrained model preserves the image content and the spatial structure of an image. Johnson *et al.* [23] trained two SRR models: one that did not use any pixel loss during training and relied solely on the perceptual loss and one that only used a pixel loss. The outputs produced by the two models confirmed that while the perceptual loss is better at reconstructing minute details and producing visually appealing results, the pixel loss gives much fewer artifacts mainly because of its smoothing effect on the pixel values. This result suggests that both losses are useful for the SRR task.

Further, generative adversarial networks (GANs) [27] are highly effective generative models for producing realistic images. The GAN learns a mapping from one manifold to another via an adversarial game between a generator model and a real/generated image discriminator model. GAN's ability to produce sharp images by learning the actual data distribution [27] suggests that the adversarial loss might be a good fit for the SRR task. Indeed, Ledig *et al.* [17] proposed a GAN-based model for the SR task (SRGAN), combining three losses: a content loss (MSE pixel loss), a perceptual loss (VGG feature matching like in [23]), and an adversarial loss that encourages the network to favor solutions that reside on the manifold of natural images. Wang *et al.* [28] proposed some improvements to SRGAN including: 1) the implementation of residual-in-residual dense blocks (RRDBs), which constitute an extension to densely connected networks [29]; 2) the use of relativistic adversarial loss [30] which stabilizes the GAN's training and improves its performance; and 3) the application of the perceptual loss before the activations of the VGG layers. Wang *et al.* [28] called their improved model enhanced SR GAN (ESRGAN). While the ESRGAN's performance on natural images is quite impressive, it tends to create artifacts in remotely sensed imagery [31]. This may emanate from the complexity and variability of the scenes depicted in remote sensing [31] or from the images' lower spatial resolution and the higher noise they usually exhibit. Furthermore, a huge portion of remotely sensed images often includes textured surfaces, in contrast to the images contained in the ImageNet dataset that have more high-frequency components spread throughout the image area.

These peculiarities of remotely sensed images are better managed by models that are oriented to work with such data. In this direction, Gong *et al.* [31] proposed the enlighten-GAN model that uses a self-supervised hierarchical perceptual loss. Liu *et al.* [32] exploited the salient maps of images to learn additional structure priors and to make the model focus more on the salient objects. Huan *et al.* [33] proposed a multiscale residual network with hierarchical feature fusion and multiscale dilation residual blocks. Courtrai *et al.* [34] used a cycle-GAN [35] to convert LR images to HR images as well as HR images to LR images, which is a process that seems to help the model learn the mapping between the two domains. Courtrai *et al.* [34] also integrated a YOLOv3 [36] model into their architecture to conduct small object detection. The integrated object detection model, together with the cycle-GAN, trains the generator synergically. Despite small object detection being the model's main aim, the generator produces upsampled images to facilitate the task.

Summing up, previous works on the SRR task for remote sensing imagery focus either on the architecture of the model or on small training procedure differentiations that potentially improve the results to a certain extent. In this article, the authors exploit the best practices derived from the state-of-the-art experimentation up to date, suggesting partly keeping the training principles (content and adversarial losses) of the highly successful ESRGAN while modifying the way the perceptual loss is conceived in the context of SR in general and in remote sensing specifically. The key idea of the proposed

approach lies in the observation that single-image SR is an ill-posed problem in the sense that for any LR image exists numerous HR images that could correspond to it [23], [31], [32]. Thus, for any successful model to achieve superior performance, it must derive significant pixel-level knowledge during the training. Up to date, most promising models use a perceptual loss that is based on feature matching, i.e., matching the similarity of two images in feature space [17], [23], [28]. Alternatively, the authors propose the replacement of the perceptual loss with a pixel-level loss which is more appropriate for SRR models operating on remotely sensed imagery. Specifically, the authors exploit the normalized digital surface model (nDSM), defined as the difference between the DSM and the digital elevation model (DEM), i.e., nDSM = DSM − DEM. The nDSM holds a great amount of pixel-level information that can restrain the model's flexibility in outputting a statistical average of viable solutions: a candidate solution must have the same nDSM as the ground-truth HR image. The main difference between the proposed approach and a feature matching perceptual loss is that the nDSM contains most of the spatial relations within an image while feature space similarities may be misleading: semantically unrelated images can have a similar subset of features. Johnson *et al.* [23] also note this while discussing which VGG layers (lower level or higher level) to choose when constructing the perceptual loss. The results of this article suggest that the gradients flowing from the nDSM back to the SRR model during training improve the quality of the latter. Some task-specific training techniques have also been applied, which stabilize training and improve the results.

III. Using DSMs to Apply SR

As mentioned in the related work, the ESRGAN model's performance in upsampling natural images is impressive in part due to the perceptual loss used during training. The perceptual loss of the ESRGAN is calculated using a second model pretrained on a second task relevant to the primary task of interest, e.g., a classification task conducted via a VGG model. Regarding the ESRGAN's training, the VGG model used for calculating the perceptual loss was trained on the 1000 classes of the ImageNet dataset. This substantial number of classes, in combination with the millions of images contained in the dataset and the effectiveness of the VGG model, facilitated the training of the SRR task. However, remote sensing imagery differs from the images contained in the ImageNet dataset in several aspects: they have lower spatial resolution and level of detail, they have higher noise, and they depict larger textured surfaces instead of individual objects dominating the image. Thus, a pretrained VGG model on the ImageNet dataset might not be the best choice for training an SRR model that takes as input aerial photography or satellite imagery. Even if an ESRGAN-like model is trained on a remote sensing task from scratch, using its learned features for building a perceptual loss, it will most probably not be exposed to hundreds of classes or have access to millions of HR images. Such limitations are quite common when dealing with remote sensing tasks. Furthermore, both large textured surfaces and image variability in remote sensing imagery tend to reduce the effectiveness of feature space similarity metrics. In Sections III-A–III-D, we describe the loss functions used in the proposed SRR model.

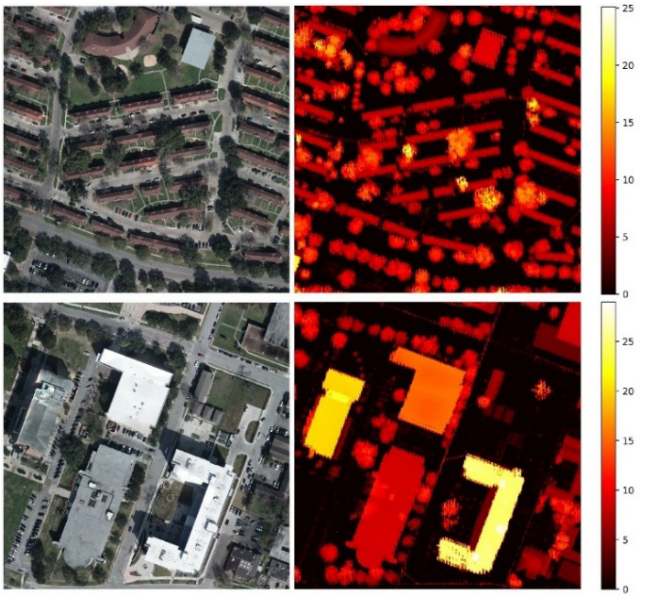

Fig. 1. nDSM [38] infers the height map of the objects depicted in an aerial image. (Left) RGB aerial images are shown. (Right) Predicted elevation heat maps are shown. The color bars indicate the color-coded height in meters.

A. nDSM-Based Loss

The authors address the problems imposed on the SRR task of remote sensing imagery by applying a pixel-level loss based on information that is closely related to the domain, harnessing the nDSM. An nDSM inferring model captures the spatial relations in remotely sensed images to infer the heights of the depicted objects. Interestingly, neural networks (NNs) predicting depth from single images also use object interactions, like shadows, to identify objects in the scene [37]. To infer the nDSMs, we use a model developed in previous work [38] that converts single RGB images to nDSMs and we pretrain it with the data used for the SRR task. The model inferring the nDSMs from RGB images uses a U-Net architecture [39] to compress an image into smaller representations that the model then decodes to form the elevation map. For a detailed model architecture and details regarding its training, we kindly refer the readers to [38]. Fig. 1 shows examples of inferring the elevation map of an aerial image via the nDSM.

The pretrained nDSM provides the means for defining a loss that closely relates to the domain data. Besides the nDSM-based loss, our proposed SRR methodology also uses a content loss (pixel-loss) and an adversarial loss. The content loss forces the model to output images that maintain the content of the LR image while the adversarial loss drives the model to infer images that are sharper and more realistic.

Ideally, an SR image calculated from its LR counterpart should result in the same nDSM as the ground-truth HR image corresponding to the LR image. The closer an SR image is to the ground-truth HR image, the closer their inferred nDSMs

should be. This is reflected in the following loss:

$$L_{\text{nDSM}} = \| f_{\text{nDSM}}(f_{\text{SR}}(x_{\text{LR}})) - f_{\text{nDSM}}(x_{\text{HR}}) \|_2 \tag{1}$$

where $f_{\text{nDSM}}(\cdot)$ is the nDSM-inferring model, $f_{\text{SR}}(\cdot)$ is the SRR model, x_{LR} is the LR image, and x_{HR} is the HR image. During training, the LR image is used as the input of the SRR model and its SR reconstruction is passed through the pretrained nDSM. Then, the ground-truth HR image is also passed through the nDSM, and the Euclidean distance of the inferred height maps is calculated. During production, the nDSM is no longer required, since its sole purpose is to facilitate the model training by forcing the parameters' update operation to favor weights that output SR images that are similar to the HR images.

B. Adversarial Loss

In addition to the content loss and the nDSM loss, we also use an adversarial loss to bias the model toward images that reside on the HR images' manifold. We adopt the original GAN methodology proposed by SRGAN [17] and not the relativistic GAN suggested in ESRGAN [28] because we did not observe any improvements in our results when the latter was used. However, we noticed that the proposed nDSM loss stabilizes the training and facilitates learning, which makes the use of a vanilla GAN sufficient. The adversarial loss is defined as

$$L_{\text{adversarial}} = \sum_{n} -\log f_{\text{DGAN}}(f_{\text{SR}}(x_{\text{LR}})) \tag{2}$$

with f_{DGAN} being the discriminator of the GAN, f_{SR} is the SRR model, and x_{LR} is the LR image. We use this formulation instead of minimizing $\log(1 - f_{\text{DGAN}}(f_{\text{SR}}(x_{\text{LR}}))$ to avoid the saturating gradient issue [27]. The GAN discriminator is trained on predicting whether input images source from the HR images' distribution or the SR images' distribution.

C. Proposed SR Model Architecture

ESRGAN employs the basic architecture of SRGAN [17], incorporating modifications such as the removal of batch normalization [40] everywhere in the model and the use of RRDBs as the basic block of the model. The specific architecture performs most computations in the LR feature space and uses up-scaling units located near the output which increases the resolution of the feature maps calculated by the RRDBs. We apply some further modifications to the ESRGAN architecture that enhance its performance on remote sensing imagery.

1) Each subpixel up-sampling (x2) layer is followed by two convolutional layers with parametric rectified linear units (PReLUs) [41].
2) After the upsampling units, we use two additional convolutional layers with PReLU activations.
3) The output convolutional layer applies a hyperbolic tangent activation function followed by an operation that converts the resulting values in the range $[0, 1]$. Specifically, the rescaling operation applies the function $0.5 \times (x - 1) + 1$ to an input x.

Fig. 2 shows the proposed modified architecture for the SR network. The modifications made to the original ESRGAN architecture are noted in Fig. 2. The model implements several blocks, each consisting of three RRDBs that contain residual nodes and dense connections. Each residual node applies a scaling parameter β to the output of each RRDB before adding it to the residual path. A similar scaling is applied at the output of each RRDB and specifically at the residual node that merges the input of the block with its output. Residual scaling prevents instability during the training and allows for smoother updates [28].

D. Content Loss

Most SRR models use either the MSE or mean absolute error (MAE) to implement the content loss. These error functions tend to be a good fit for close-range photography, but this might not be the case for remotely sensed imagery. As mentioned before, aerial images usually have low spatial resolution and are noisy. Large texture surfaces of a wide variety make the SRR task even harder. Images of rocky areas, random soil formations, dumping fields with randomly disposed waste, and varied objects' orientations make it extremely hard for a DL model to learn the data distribution. Furthermore, trees with entangled branches and an infinite number of leaves configurations render the SRR task on remote sensing imagery extremely hard even for the state-of-the-art models like the ESRGAN. In particular, soil and leaves' configurations are very complex and thus very hard to model in the SR context. MSE penalizes large prediction errors which makes MAE more suitable when the dataset contains several outliers. In the case of aerial imagery-based SRR, a wise strategy is to avoid high penalization on the reconstruction error of entities whose distribution is a priori difficult to learn (e.g., soil and trees) and to penalize large errors on easier-to-learn reconstructions like cars and houses. Thus, we propose the use of the Huber loss [42] instead of MAE or MSE because it applies either of the two losses, depending on the error magnitude. The proposed content loss is a Huber function with a transition point ϵ and $a = f_{\text{SR}}(x_{\text{LR}}) - x_{\text{HR}}$ defined as

$$L_{\text{content}} = \begin{cases} \frac{1}{2}(a)^2, & \text{if } |a| \leq \epsilon \\ \epsilon(|a|) - \frac{1}{2}\epsilon, & \text{otherwise.} \end{cases} \tag{3}$$

We propose this design choice after a series of observations made during trial-and-error experimentation with various content losses. Using MAE or MSE, the SRR model tends to predict tree and soil reconstructions with relatively low errors, but reconstructions are overly smooth and blurry. By not heavily penalizing errors of such depictions (trees, soil, or other complicated structures), we shift the burden of generating realistic reconstructions to the other losses (i.e., the nDSM and the adversarial loss). Accordingly, by greatly penalizing the reconstruction error of structures like houses and cars, the influence of the nDSM and the adversarial loss is reduced, and the model avoids the generation of high-frequency artifacts. Meyer [43] proposed an alternative probabilistic interpretation of the Huber loss which justifies its

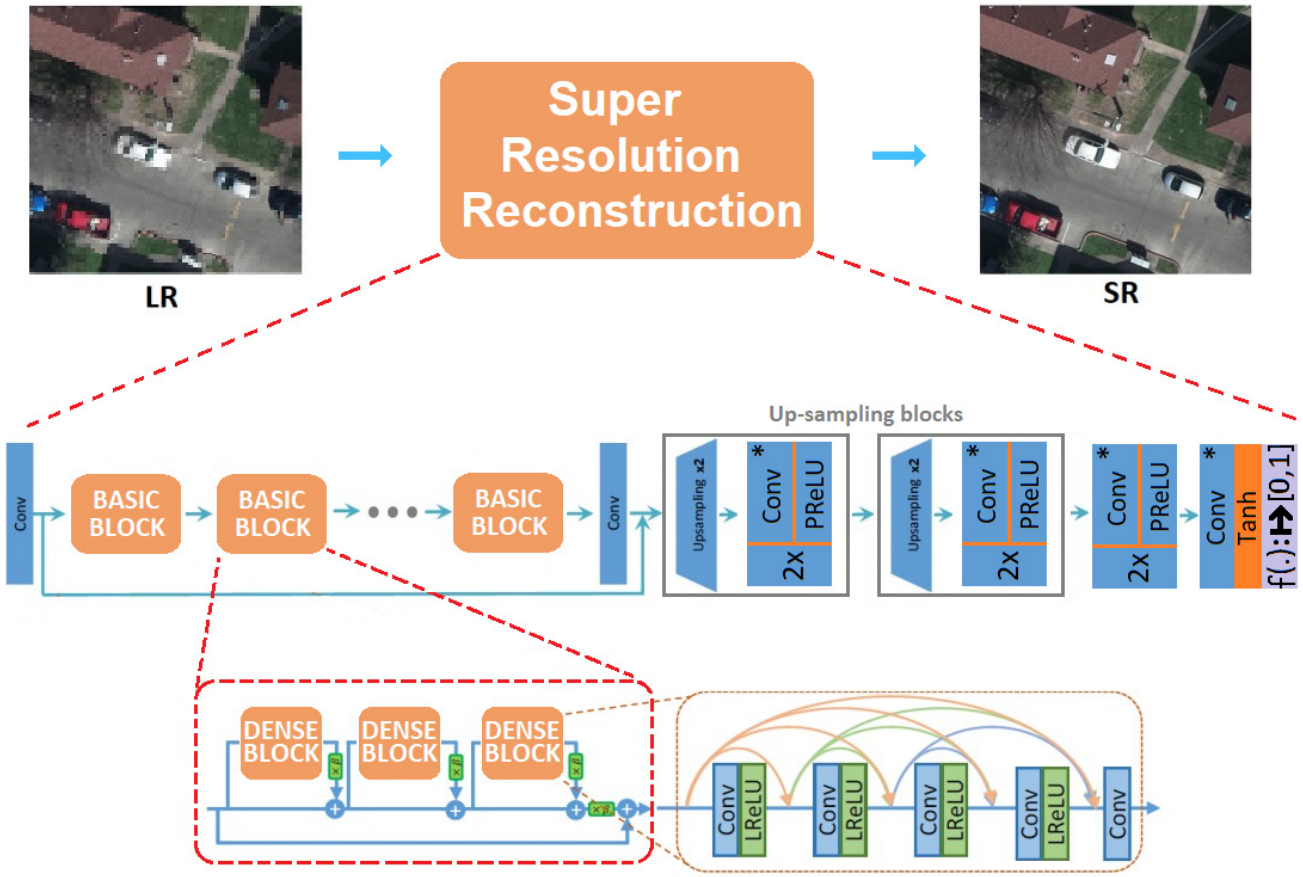

Fig. 2. Modified ESRGAN architecture is used in the proposed SR model. Several RRDBs process the image while maintaining its LR dimensions, and upsampling is applied near the output of the model. We introduce two convolutional layers with PReLU activations between the upsampling units as well as two additional convolutional layers with PReLU activations after the last upsampling unit. Furthermore, we use a hyperbolic tangent activation function at the output and a rescaling layer just before the inferred SR image to adjust the values in the range [0,1]. Our modifications on the basic ESRGAN are shown with an asterisk on the top-right area of the network components.

use on tasks dealing with aerial photography since these tasks generally contain significant noise and often have low quality: Huber loss minimization relates to minimizing an upper bound on the Kullback–Leibler divergence between the Laplacian distribution of noise in the ground-truth data and the Laplacian distributed prediction noise. Meyer [43] further showed that the optimal transition point of the Huber function is closely related to the noise in the ground-truth data. Taking the above into account, we train the proposed models for minimizing the following combination of the three losses, as described above (content, nDSM, and adversarial loss):

$$L = \alpha L_{\text{nDSM}} + L_{\text{content}} + \beta L_{\text{adversarial}} \tag{4}$$

with α and β being the weighting factors of the losses.

IV. Experiments and Results

We evaluate our methodology with two datasets, one containing images mainly of an urban area taken by an aircraft equipped with image and ranging laser scanner (LiDAR) sensors and one dataset containing aerial images mainly of rural areas. Both datasets contain the corresponding DSMs and DEMs. This variety in landscapes enables us to assess the developed models' performance in different imagery dataset contexts and spatial resolutions. As mentioned before, the training of the models requires the nDSM of the area used in the training data while an nDSM is not required during inference. Still, this limitation only allows the use of datasets that include DSMs and DEMs such as the Data Fusion 2018 Contest (DFC2018) dataset [44], [45] and the dataset containing the national LiDAR flyby of Luxembourg, conducted in 2019 by the country's administration for cadaster and topography [46], [47]. The DFC2018 dataset is part of a set of community data provided by the IEEE Geoscience and Remote Sensing Society (GRSS). In this article, we specifically use the multispectral LiDAR classification challenge data. The RGB images of the DFC2018 dataset have a 5-cm/pixel spatial resolution while the LiDAR resolution is 50 cm/pixel. The data belong to a 4172 $\times$ 1202 m^2 area and includes the University of Houston, Houston, TX, USA, and its surroundings. The Luxembourg dataset contains RGB images of 20-cm/pixel spatial resolution, and the LiDAR resolution is 50 cm/pixel with a density of 15 points/m^2. The Luxembourg dataset is in georeferenced raster format and uses the Luxembourg Reference Frame (LUREF) (EPSG 2169) coordinate system and projection. We use the datasets to train two models for the SRR $\times$4 task (one model for each dataset) and the DFC2018 dataset to train a model for the SRR $\times$8 task. We do not use

the Luxembourg dataset to train a model on the SRR $\times 8$ task because of the poor quality of the downscaled images.

A. Training Details

We train our models with the Adam optimizer and a learning rate of 0.0001, scaling the nDSM and the adversarial losses with factors $\alpha = 0.01$ and $\beta = 0.001$, respectively, as shown in (4). We also apply label smoothing of 0.2 to the GAN training, and we pretrain the SRR models with MAE. This puts the weights in an appropriate configuration to avoid local minima and stabilize the GAN training [17]. The MAE of the pretrained models also provides an indication of what constitutes a suitable value for the transition point of the Huber loss (content loss). Our experiments showed that a Huber loss transition point that is twice the MAE of the pretrained models gives better results. In our experiments, we use $\times 4$ and $\times 8$ upsampling factors. For the $\times 4$ upsampling experiments, we train the models with randomly cropped patches of size 520×520 pixels, which are downscaled with bicubic interpolation to LR inputs of size 130×130 (the models apply $\times 4$ upsampling, and thus, the SR outputs match the size of the original HR images). For the $\times 8$ upsampling experiment on the DFC2018 dataset, the 520×520 random patches are downscaled via bicubic interpolation to LR inputs of size 65×65. The LR images in all experiments are created by downscaling the HR images via bicubic interpolation.

B. Results

Figs. 3 and 4 show the results for the SRR $\times 4$ models for the DFC2018 and the Luxembourg datasets, respectively. Fig. 5 shows the results of the SRR $\times 8$ model for the DFC2018 dataset. The SRR $\times 4$ model dealing with the DFC2018 dataset reconstructs finer image details and more high-frequency components in comparison to the model trained on the Luxembourg dataset. This is not surprising since the resolution of the images in the DFC2018 dataset is four times higher than the resolution of the images in the Luxembourg dataset, and thus, the model learns the data distribution of a more detailed scenery. This is also reflected in the performance metrics shown in Table I, i.e., the SRR model trained with the DFC2018 dataset achieves a higher structural similarity index measure (SSIM) [48] and a PSNR than the model trained with the Luxembourg dataset for the SRR $\times 4$ task. The results also suggest that the effectiveness of the proposed SRR approach depends on the quality of the nDSM used. This is one of the reasons why our approach works better on the DFC2018 dataset, as it has a more accurate nDSM. Since the quality of the nDSM relates to the quality of the images contained in the dataset, the effectiveness of the proposed approach inherently relates to the quality of the images in the dataset. Hence, the results of the SSR $\times 4$ model trained with the high-quality DFC2018 dataset are often indistinguishable from the ground-truth HR images (a more detailed analysis of this is provided in Section IV-D regarding the limitations of the model). Table I also shows the values of PSNR and SSIM achieved by the ESRGAN trained on the datasets and the corresponding values achieved when bicubic interpolation is applied to the LR images. The overall findings indicate that the proposed model achieves higher scores on the PSNR and SSIM metrics than when bicubic interpolation or the ESRGAN model is applied. An analysis of the comparison between our model and the ESRGAN model is held in Section IV-C. We must note the difficulty in comparing the performance of our model with the performance of previous studies on SRR models applied to remote sensing imagery because the datasets involved in the comparison must contain DSMs.

TABLE I
PERFORMANCE OF THE PROPOSED APPROACH

		SSIM	PSNR (DB)
X 4 DFC2018	**OUR**	**0.92**	**31.17**
	ESRGAN	0.91	30.43
	BICUBIC	0.83	28.41
X 4 LUXEMBOURG	**OUR**	**0.83**	**26.46**
	ESRGAN	0.81	26.1
	BICUBIC	0.68	23.3
X 8 DFC2018	**OUR**	**0.88**	**28.62**
	ESRGAN	0.85	26.9
	BICUBIC	0.66	24.55

The proposed SRR model performing $\times 4$ upsampling recovers significant information content which was lost during the downscaling of the images (Figs. 3 and 4). Various objects like cars and street poles are properly reconstructed, and in many cases, some fine details like shadows and street lines are almost identical to their HR counterparts. Buildings are also properly reconstructed with high-level details, and large surfaces like rooftops are depicted with their original texture which was severely degraded during downscaling. The reconstructions with the least fidelity are those of trees and soil, which is something expected given the distribution complexity of their surfaces and their extreme diversity in visual representations. As expected, the results of the SRR model performing $\times 8$ upsampling show lower fidelity because of the high information loss during the downsampling of the ground-truth images. Regardless of the lower quality, the resulting images (Fig. 5) recover a lot of details like difficult-to-identify train rails, rooftop textures, car shapes, road details, street poles, and shadows. The significant image quality improvement observed at the output of the model in comparison to the LR input image reflects the quality metrics' improvement shown in Table I.

The results shown in Figs. 3–5 may encourage the use of the proposed model for upsampling LR images to feed other remote sensing models performing challenging tasks that may require inputs of a higher level of detail. Some examples of such tasks are humans counting, tree identification, car type identification, dumping detection, chimney detection, detection of parking spots for disabled people, power cable identification, etc. Such tasks are supported by our results as shown in Figs. 3–5, i.e., the objects involved in these tasks are enhanced by our model and they become more evident (their detailed level is enhanced).

Fig. 3. SRR ×4 results for the DFC2018 dataset. (Columns from left to right) LR, HR (ground truth), and SR images (inferred by our model).

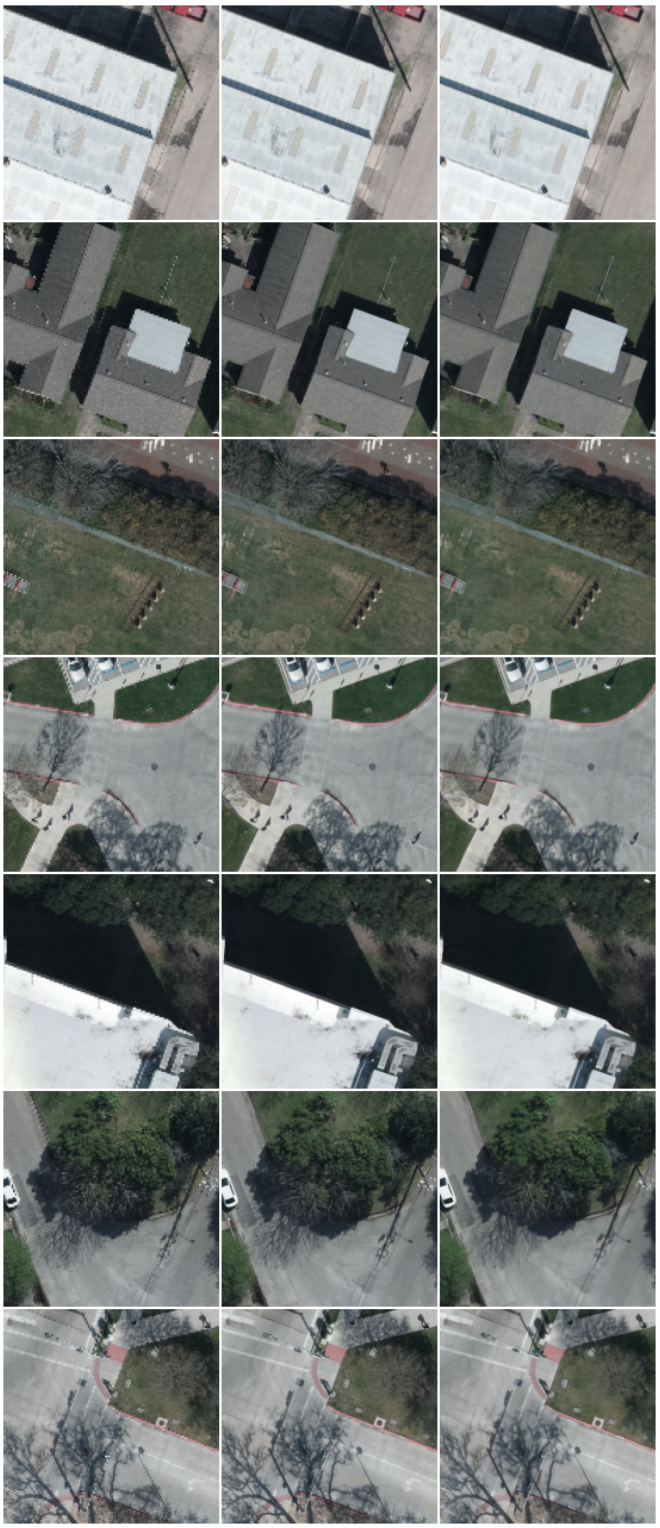

Fig. 3. *(Continued.)* SRR ×4 results for the DFC2018 dataset. (Columns from left to right) LR, HR (ground truth), and SR images (inferred by our model).

C. Comparison With the ESRGAN

We give special importance to comparing our model with the ESRGAN model because the ESRGAN outperforms most SRR methods including the SRGAN (of which it is an improvement). The ESRGAN model is very popular and

Fig. 4. SRR $\times 4$ results for the Luxembourg dataset. (Columns from left to right) LR, HR (ground truth), and SR images (inferred by our model).

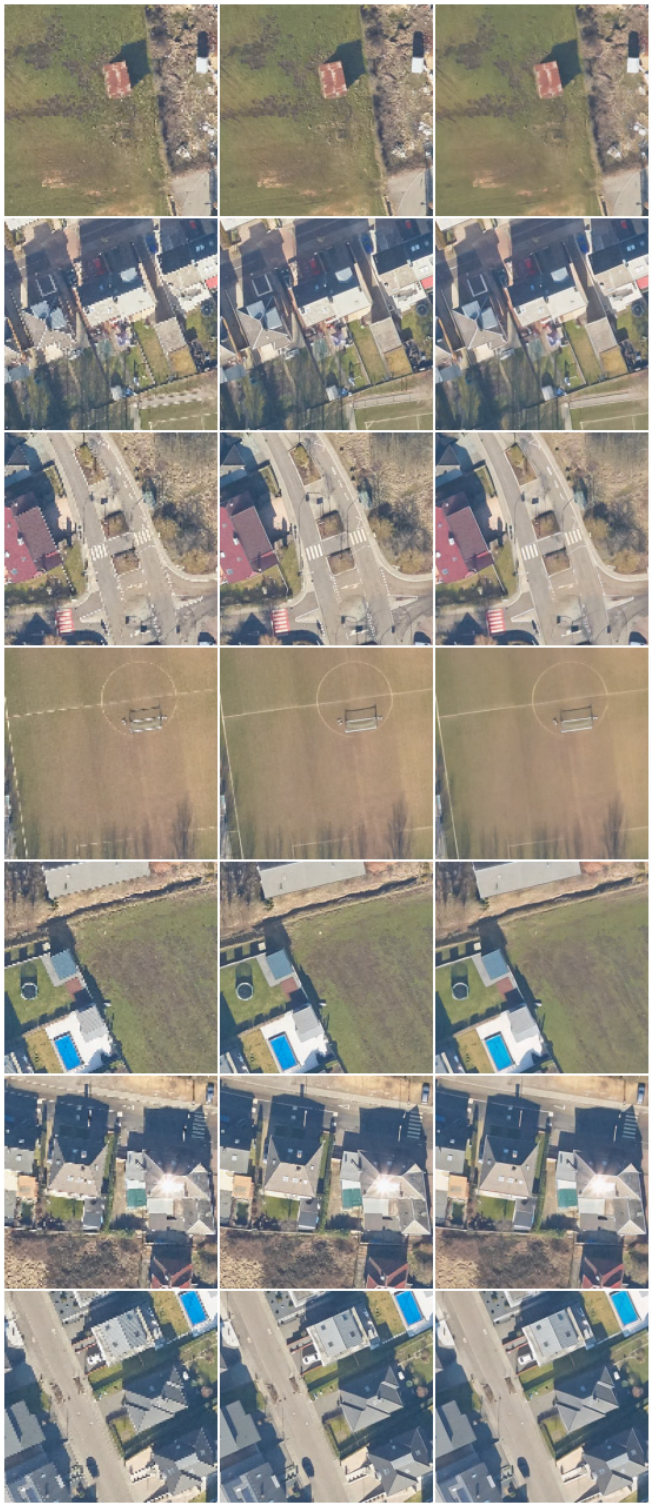

Fig. 4. *(Continued.)* SRR $\times 4$ results for the Luxembourg dataset. (Columns from left to right) LR, HR (ground truth), and SR images (inferred by our model).

constitutes a suitable and established model for applying the SRR task to various applications. Furthermore, our model is developed based on similar concepts used by the ESRGAN (architecture and loss functions) while introducing important and necessary modifications to make it suitable for applying the SRR task on remote sensing imagery.

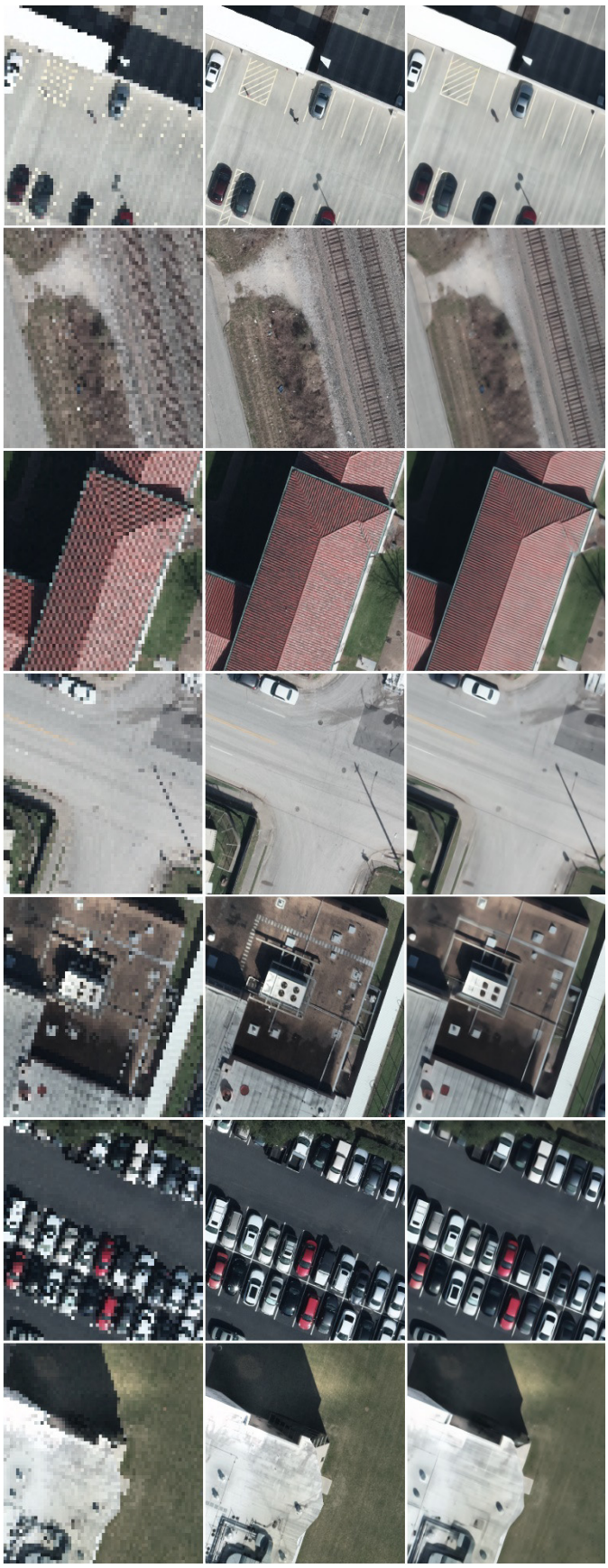

Fig. 5. SRR $\times 8$ results for the DFC2018 dataset. Columns from left to right: LR, HR (ground truth), and SR images (inferred by our model).

Fig. 5. *(Continued.)* SRR $\times 8$ results for the DFC2018 dataset. Columns from left to right: LR, HR (ground truth), and SR images (inferred by our model).

Despite being one of the best performing models for applying the SR task on natural images, ESRGAN's performance on aerial and remotely sensed imagery has a significant drawback: ESRGAN creates artifacts, especially on large textured

surfaces that are very common in remotely sensed imagery [31]. We trained the ESRGAN model [49] on our datasets to compare its results with ours and assess the performance of our approach. We specifically used SSIM and PSNR as the comparison metrics. The resulting metrics' values are shown in Table I. Our approach achieves better performance on both metrics for both datasets and upsampling factors ($\times 4$ DFC2018, $\times 4$ Luxembourg, and $\times 8$ DFC2018). Visual inspection of our approach and ESRGAN's results reveals no significant differences in the quality of the generated images except for some annoying artifacts created by the ESRGAN, especially on large flat surfaces. Some examples of these artifacts when applying the $\times 4$ SR task on remote sensing imagery with the ESRGAN are shown in Fig. 6. Occasionally (on about 5% of the images produced by the ESRGAN), these artifacts are so acute that distort a portion of the SR image significantly. The images produced by the ESRGAN reveal its eagerness to reconstruct high-frequency components, a property that proves to be productive when working with natural images, but it is problematic when working with remotely sensed imagery. The main reason ESRGAN achieves lower scores on the PSNR and SSIM evaluation metric values than our model is the artifacts it produces. We confirmed this by calculating the evaluation metric values only for the images produced by ESRGAN that do not contain artifacts (the screening was conducted with visual examination). The metric values scored by the ESRGAN's results after excluding the images containing artifacts were closer to the scores achieved by our method than when including the distorted images. The generated artifacts are even more evident when we use the ESRGAN for the $\times 8$ SR task. Fig. 7 shows some comparative examples of using the ESRGAN and our method for the $\times 8$ SR task.

Fig. 7 shows that there are more artifacts (both in number and intensity) and more distortion on large surfaces like building roofs and roads compared to the artifacts observed in the application of the $\times 4$ SR task with the ESRGAN as shown in Fig. 6.

We believe that ESRGAN produces these artifacts because of the nature of its perceptual loss: the pretrained VGG19 model that is used for obtaining the perceptual loss of the inferred SR images is not a good option for remotely sensed imagery. VGG19 is pretrained on the ImageNet dataset that has a very different data distribution than remotely sensed imagery.

This difference in data distribution renders the feature mapping layer used for the perceptual loss incapable of calculating appropriate remotely sensed image representations. Therefore, the essence of what constitutes a high-quality remotely sensed image cannot be captured by the perceptual loss calculated with a model pretrained on a different data distribution, e.g., the VGG19 used by the ESRGAN model. Our approach proposes a solution to this problem by replacing the ImageNet classifier with a pretrained nDSM prediction model. The nDSM is trained on predicting the height maps of remotely sensed images, and thus, it is domain-specific and relevant to the images used in the SR task.

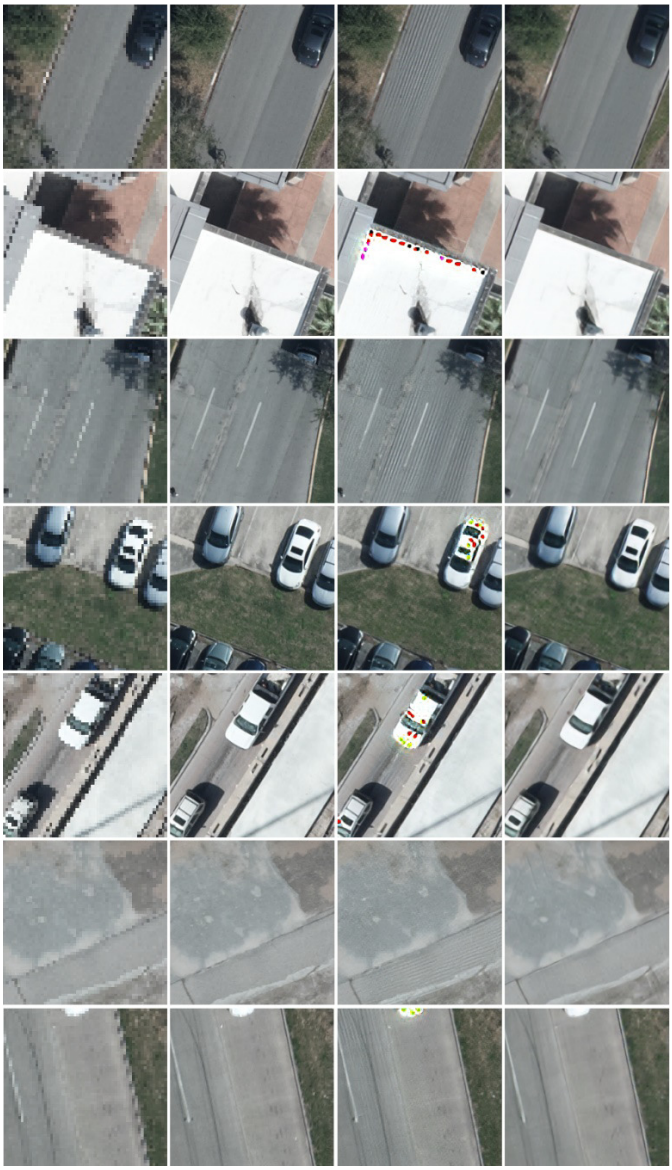

Fig. 6. Examples of the artifacts created by the ESRGAN on large surfaces depicted in remotely sensed imagery for $\times 4$ SR. (From left to right) LR images, the original HR images, the SR images generated by the ESRGAN, and the SR images generated with our approach. Besides its occasional difficulty to reconstruct some objects (e.g., cars), ESRGAN adds artifacts on large surfaces (e.g., roads and roofs). On the contrary, our model does not.

D. Limitations

The main limitations of our work have to do with the upsampling factor targeted, e.g., $\times 4$ and $\times 8$, the spatial resolution of the HR images (training images) and the proper alignment between the HR images and their DSMs. We tested our model on up to an $\times 8$ upsampling factor, and the results are decent (Fig. 5). We do not show the results from experiments with larger upsampling factors because beyond $\times 8$ upsampling, the SR task becomes overwhelmingly ill-posed [29], [37], [38]: it is extremely difficult for LR images (downscaled images) to contain information regarding fine details of the original HR images. Similarly, SRGAN and ESRGAN focus solely on a $\times 4$ resolution upsampling factor [17], [28]. In contrast to the case of natural images, the benefits of applying the SR task on

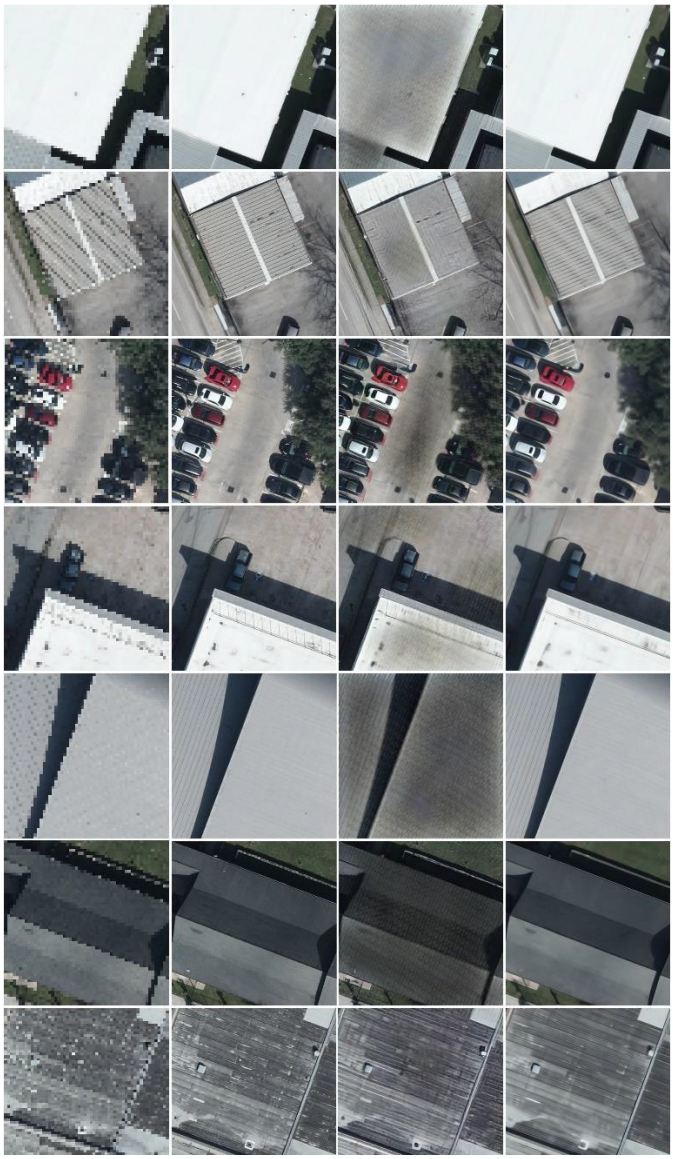

Fig. 7. Examples of the artifacts created by the ESRGAN on large surfaces depicted in remotely sensed imagery for $\times 8$ SR. (From left to right) LR images, the original HR images, the SR images generated by the ESRGAN, and the SR images generated with our approach. The ESRGAN tends to add a significant amount of texture and lines to the roads and building roofs.

remotely sensed imagery for $\times 8$ upsampling are de facto more profound because of the high cost of HR imagery. This makes our approach more suitable for remotely sensed imagery since SRGAN and ESRGAN are optimized for $\times 4$ spatial resolution upsampling. However, the experiments we conducted revealed that our model does not perform well for upsampling factors greater than $\times 8$.

We also experimented with the spatial resolution of the HR images used for training the model. We used a dataset available from the national Open Government Data (OGD) initiative of Austria [50]. The dataset contains RGB images of 40-cm/pixel spatial resolution, and the LiDAR resolution is 40 cm/pixel (the acquired RGB images were of 20-cm/pixel resolution, but the publicly available data are downscaled to 40-cm/pixel resolution). At this lower spatial resolution, the nDSM that provides the perceptual loss does not generalize well, probably because it loses its ability to detect low-level features like corners, lines, shadow edges, or other features that allow it to predict the heights of the objects in the image. Thus, the advantages of using an nDSM-based perceptual loss are lost which explains why our model operating on images of lower spatial resolution performs similar to the case of using a combination of an MAE and a GAN loss alone, i.e., without the nDSM-based perceptual loss. In other words, while the nDSM-based perceptual loss has a significant contribution to the performance of the model when the spatial resolution of the HR images is adequate for the nDSM to learn how to infer the height maps of the images, this contribution ceases to exist when the resolution of the HR images (and the quality of the data) prevents the nDSM from generalizing well.

Another limitation of our model related to the nDSMs of the HR images is the proper alignment between them. For example, the nDSMs of the DFC2018 dataset are properly aligned with the RGB images and this allows the model to achieve very good results for both upsampling factors ($\times 4$ and $\times 8$). The nDSMs of the Luxembourg dataset are not perfectly aligned, and we suspect that this plays a significant role in the model's inability to perform well on the $\times 8$ SR task for the specific dataset (the results are not better than using an MAE and a GAN loss only). The nDSM trained on the Luxembourg dataset is performing much worse than the nDSM trained for the DFC2018 dataset. Specifically, the nDSM for the DFC2018 dataset has an MAE of 0.54 m on its test set while the nDSM for the Luxembourg dataset has an MAE of 0.83 m. This great difference is attributed partly to the lower resolution of the Luxembourg dataset and mostly to the misalignment between its nDSMs and RGB images. To investigate the effects of the misalignment problem, we trained the model on the DFC2018 dataset with the nDSMs shifted by a constant value of two pixels in random directions (up, down, left, and right) for each image and the nDSMs distorted with a random transformation (small rotation and slight skew). In the case of the constant shifting of the nDSMs, the results were very similar to the unaltered nDSMs. On the contrary, in the case of random small transformations, we observed a significant reduction in the performance metrics (in the range of 5%–10%). This suggests that the quality of the nDSMs is also important for our method to achieve its potential.

V. Conclusion

This article proposes an SRR model that works with remotely sensed images, addressing the limitations of existing state-of-the-art models by including auxiliary information that is important during the model's training phase, i.e., including the corresponding nDSM of the remote sensing imagery dataset. In other words, the proposed SRR model, during training, considers the difference between the nDSM inferred by the calculated SR image and the ground-truth HR image, instead of using a perceptual loss. Furthermore, the contribution of this article includes applying some architectural changes to the ESRGAN model and employing a Huber loss as a content loss to mitigate the difficulties imposed by

remotely sensed images. Visual inspection, together with the significant improvement of the SSIM and PSNR metrics of the inferred SR images obtained, suggests that the proposed model is suitable for the SRR task and can cope with popular and notorious remote sensing imagery limitations such as big surface textures and lower spatial resolution. Summing up, this article shows that an nDSM-based loss seems to be suitable for the SRR task on remote sensing imagery, supplying the model with enriched pixel-level information. This approach allows us to detect objects of interest that were otherwise impossible to identify such as car types, powerlines, parking spots, chimneys, and tree types. This information could be very useful to city planners, policymakers, operators of municipalities, and local communities.

REFERENCES

[1] J. Allebach and P. W. Wong, "Edge-directed interpolation," in *Proc. 3rd IEEE Int. Conf. Image Process.*, vol. 3, Sep. 1996, pp. 707–710.

[2] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays using convex projections," *J. Opt. Soc. Amer.*, vol. 6, no. 11, pp. 1715–1726, Nov. 1989, doi: 10.1364/JOSAA.6.001715.

[3] R. Achanta, F. J. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Proc. Int. Conf. Comput. Vis. Syst.*, May 2008, pp. 66–75.

[4] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 30, no. 2, p. 12, 2011, doi: 10.1145/1944846.1944852.

[5] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[6] Y. Zhang, Q. Fan, F. Bao, Y. Liu, and C. Zhang, "Single-image super-resolution based on rational fractal interpolation," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3782–3797, Aug. 2018, doi: 10.1109/TIP.2018.2826139.

[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, doi: 10.1109/TIP.2010.2050625.

[8] V. Thambawita, I. Strümke, S. A. Hicks, P. Halvorsen, S. Parasa, and M. A. Riegler, "Impact of image resolution on deep learning performance in endoscopy image classification: An experimental study using a large dataset of endoscopic images," *Diagnostics*, vol. 11, no. 12, p. 2183, Nov. 2021, doi: 10.3390/diagnostics11122183.

[9] C. F. Sabottke and B. M. Spieler, "The effect of image resolution on deep learning in radiography," *Radiol., Artif. Intell.*, vol. 2, no. 1, Jan. 2020, Art. no. e190015.

[10] M. Koziarski and B. Cyganek, "Impact of low resolution on image recognition with deep neural networks: An experimental study," *Int. J. Appl. Math. Comput. Sci.*, vol. 28, no. 4, pp. 735–744, Dec. 2018, doi: 10.2478/amcs-2018-0056.

[11] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *Proc. 8th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2016, pp. 1–6, doi: 10.1109/QoMEX.2016.7498955.

[12] L. Khelifi and M. Mignotte, "Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis," *IEEE Access*, vol. 8, pp. 126385–126400, 2020, doi: 10.1109/ACCESS.2020.3008036.

[13] Q. Yuan *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, May 2020, Art. no. 111716, doi: 10.1016/j.rse.2020.111716.

[14] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114417, doi: 10.1016/j.eswa.2020.114417.

[15] L. P. Osco *et al.*, "A review on deep learning in UAV remote sensing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 102, Oct. 2021, Art. no. 102456, doi: 10.1016/j.jag.2021.102456.

[16] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019, doi: 10.1016/j.isprsjprs.2019.04.015.

[17] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114, doi: 10.1109/CVPR.2017.19.

[18] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 184–199.

[19] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654, doi: 10.1109/CVPR.2016.182.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[21] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. 15th Eur. Conf. Comput. Vis. (ECCV)*, vol. 11211. Munich, Germany, Sep. 2018, pp. 294–310, doi: 10.1007/978-3-030-01234-2_18.

[22] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5835–5843, doi: 10.1109/CVPR.2017.618.

[23] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 694–711, doi: 10.1007/978-3-319-46475-6_43.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*. San Diego, CA, USA, May 2015, pp. 1–14.

[25] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[26] A. Mahendran and A. Vedaldi, "Understanding deep image representations by inverting them," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5188–5196, doi: 10.1109/CVPR.2015.7299155.

[27] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 2. Cambridge, MA, USA: MIT Press, 2014, pp. 2672–2680.

[28] N. C. Rakotonirina and A. Rasoanaivo, "ESRGAN+: Further improving enhanced super-resolution generative adversarial network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 63–79, doi: 10.1109/icassp40776.2020.9054071.

[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.*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[30] A. Jolicoeur-Martineau, "The relativistic discriminator: A key element missing from standard GAN," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, New Orleans, LA, USA, May 2019.

[31] Y. Gong *et al.*, "Enlighten-GAN for super resolution reconstruction in mid-resolution remote sensing images," *Remote Sens.*, vol. 13, no. 6, p. 1104, Mar. 2021, doi: 10.3390/rs13061104.

[32] B. Liu *et al.*, "Saliency-guided remote sensing image super-resolution," *Remote Sens.*, vol. 13, no. 24, p. 5144, Dec. 2021, doi: 10.3390/rs13245144.

[33] H. Huan *et al.*, "End-to-end super-resolution for remote-sensing images using an improved multi-scale residual network," *Remote Sens.*, vol. 13, no. 4, p. 666, Feb. 2021, doi: 10.3390/rs13040666.

[34] L. Courtrai, M.-T. Pham, and S. Lefèvre, "Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks," *Remote Sens.*, vol. 12, no. 19, p. 3152, Sep. 2020, doi: 10.3390/rs12193152.

[35] 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. (ICCV)*, Oct. 2017, pp. 2242–2251, doi: 10.1109/ICCV.2017.244.

[36] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[37] T. Van Dijk and G. D. Croon, "How do neural networks see depth in single images?" in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2183–2191, doi: 10.1109/ICCV.2019.00227.

[38] S. Karatsiolis, A. Kamilaris, and I. Cole, "IMG2nDSM: Height estimation from single airborne RGB images with deep learning," *Remote Sens.*, vol. 13, no. 12, p. 2417, Jun. 2021, doi: 10.3390/rs13122417.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 37. Lille, France, Jul. 2015, pp. 448–456.
[41] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034, doi: 10.1109/ICCV.2015.123.
[42] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 1964, doi: 10.1214/aoms/1177703732.
[43] G. P. Meyer, "An alternative probabilistic interpretation of the Huber loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5261–5269.
[44] *2018 IEEE GRSS Data Fusion Contest*. Accessed: Feb. 26, 2022. [Online]. Available: http://dase.grss-ieee.org/index.php
[45] *IEEE France GRSS Chapter*. Accessed: Feb. 26, 2022. [Online]. Available: https://site.ieee.org/france-grss/2018/
[46] Government of the Grand Duchy of Luxembourg. *The Luxembourg Data Platform, LiDAR 2019 Digital Surface Model (DSM)*. Accessed: Feb. 26, 2022. [Online]. Available: https://data.public.lu/fr/datasets/lidar-2019-modele-numerique-de-la-surface-mns
[47] Government of the Grand Duchy of Luxembourg. *The Luxembourg Data Platform, Technical Orthophoto of the Grand Duchy of Luxembourg, Winter 2019 Edition*. Accessed: Feb. 26, 2022. [Online]. Available: https://data.public.lu/fr/datasets/orthophoto-technique-du-grand-duche-de-luxembourg-edition-2019-hiver
[48] 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, doi: 10.1109/TIP.2003.819861.
[49] X. Wang *et al.* (2018). *ESRGAN's Official Implementation*. Accessed: Jul. 15, 2022. [Online]. Available: https://github.com/xinntao/ESRGAN
[50] Austria National Open Government Data (OGD) Initiative. (2019). *GeoData Burgenland*. Accessed: Jul. 16, 2022. [Online]. Available: https://geodaten.bgld.gv.at/de/home.html

Savvas Karatsiolis received the HND degree in electrical engineering, the B.Sc. degree in computer engineering, and the M.Sc. degree in information systems, and the Ph.D. degree in computer science from the University of Cyprus, Nicosia, Cyprus, in 1998, 2003, 2011, and 2019, respectively, with a focus on deep learning and Artificial Intelligence (AI).

He is currently a Research Associate with the CYENS Research Center of Excellence, Nicosia. His areas of research are machine-learning theory, computer vision, generative models, and unsupervised/self-supervised learning.

Chirag Padubidri, photograph and biography not available at the time of publication.

Andreas Kamilaris received the B.A. degree in computer science from the University of Cyprus, Nicosia, Cyprus, in 2007, the M.S. degree in distributed systems from the ETH University of Zurich, Zürich, Switzerland, in 2009, and the Ph.D. degree from the Computer Science Department, University of Cyprus, focusing on the topic of enabling smart homes using web technologies, in 2013.

He received the Marie Skłodowska-Curie Fellowship in 2016, working at the Institute of Agriculture and Food Research and Technology (IRTA Barcelona), Barcelona, Spain, performing research on big data analysis and applications in the agri-food sector. In April 2018, he joined the Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, Enschede, The Netherlands, as an Assistant Professor. In October 2019, he joined the CYENS Center of Excellence, Nicosia, as the Leader of the SuPerWorld MRG.