

Multispectral Satellite Image Generation Using StyleGAN3

Michael Alibani , Member, IEEE, Nicola Acito , and Giovanni Corsini , Life Member, IEEE

Abstract—Satellite-based remote sensing images are essential for Earth surface analysis, serving diverse purposes in both civilian and military domains. Satellite images are used for analysis and decision making and are considered a reliable source of information. Recently, the field of image generation has developed increasingly sophisticated techniques, such as generative neural models, usually known as generative adversarial networks (GANs), to create synthetic images from scratch that appear almost real. Generative models have traditionally been applied to RGB or grayscale images and have been used for generating fake images of faces, animals, or objects. Currently, there are few studies regarding the application of GAN to multispectral satellite images. This work aims to test GAN models against the generation of multispectral satellite images, and in particular, the work explores the ability of the state-of-the-art StyleGAN3 model to produce high-quality synthetic Sentinel-2 images. The work delves into the configuration, training process, and evaluation of StyleGAN3 using custom Sentinel-2 datasets. StyleGAN3 results are compared with those provided by the proGAN model, the only GAN model tested so far on multispectral satellite data. Evaluation methods include visual inspection, spectral signature analysis, and a modified Fréchet inception distance for quantitative assessment. Results show that StyleGAN3 outperforms proGAN model exhibiting visually pleasing images. The quantitative comparison shows that StyleGAN3 provides the best results in terms of FID scores, in particular the improvement compared to proGAN increases as the spatial extent and spectral dimension of the generated images increases.

Index Terms—Generative adversarial network (GAN), GeoFake, Sentinel-2, StyleGAN3, synthetic image generation.

I. INTRODUCTION

THE rapid advancement of artificial intelligence (AI) algorithms, coupled with the abundance of available data for training, has sparked significant interest within the scientific community in the field of image generation. Generative methods have proven effective in synthesizing various types of images, including medical images [1], realistic photographs of objects, scenes, and human faces [2]. These generated images have been utilized to augment existing datasets, enhancing the training process and improving the performance of machine learning models. Furthermore, generative models have found applications in creating avatars for online gaming and social

media profiles. In addition, artists are exploring the possibilities of merging human and machine-generated content, redefining the boundaries of art itself [3].

Existing research works mainly concern the generation of captured photographs from conventional red green blue (RGB) cameras with a focus on human face images [4], [5]. In this work we concentrate on the generation of multispectral remotely sensed images from satellite platforms, which still remains a relatively unexplored topic within the scientific community. The generation of satellite (and aerial) imagery presents unique challenges compared to the generation of images of human faces or objects, as they require the reconstruction of both local and global features to obtain realistic results.

In the context of remote sensing, generative models for satellite images could be a useful data augmentation tool for enriching existing datasets adopted to train machine learning and deep learning models [6], [7]. Data augmentation through generative models can also be used for evaluating the performance of a certain algorithm (not necessarily based on AI) by providing statistically more reliable values for the performance indicators adopted. The generative approach is particularly useful in application scenarios where limited images are available or when there is significant disparity between the different categories of interest [1], [8] as well as for reproducing unique features that are unusual or challenging to find in nature, such as camouflaged objects or rare materials [9]. The generation of remote sensing images also finds application in the field of defense and security. Fake satellite images can be generated specifically to hide important military infrastructure and/or to create false scenarios in order to deceive opposing analysts [10].

Furthermore, the impressive results obtained by AI-based generative models in the creation of “fake” images acquired by conventional RGB cameras have stimulated the analysis of their impact on real life [11] and the development of methods to evaluate the authenticity of a visual content and discriminate between false and real images [12], [13], [14]. This problem also arises in the remote sensing field [15], [10], where, on the one hand, it is crucial to investigate the potential of AI-based generative models in order to assess the implications and risks associated with such strategies. On the other hand, it is important to have large datasets of generated images in order to analyze and design the possible strategies for the detection of false satellite data [16].

The most successful and widely adopted approaches for image manipulation and synthesis are based on the generative adversarial networks (GANs) framework introduced by

Manuscript received 12 October 2023; revised 2 January 2024; accepted 16 January 2024. Date of publication 22 January 2024; date of current version 12 February 2024. This work was supported by the European Union - FSE-REACT-EU, PON Research and Innovation 2014-2020 under Grant DM1062/2021. (Corresponding author: Michael Alibani.)

The authors are with the Department of Information Engineering, University of Pisa, 56122 Pisa, Italy (e-mail: michael.alibani@unipi.it; nicola.acito@unipi.it; giovanni.corsini@unipi.it).

Digital Object Identifier 10.1109/JSTARS.2024.3356618

Goodfellow et al. [17]. GANs have been employed in various computer vision applications, including style transfer [18], super-resolution [19], and image-to-image translation [20]. While early GAN models had limitations in terms of image resolution, variation, and visual quality, current models have significantly improved in these aspects.

The first works that combined GAN models with remote sensing data were mainly focused on image-to-image translation for mapping purposes, where photos were synthesized from edge maps by means of the pix2pix algorithm [20]. Then, researchers have explored the application of GANs to remote sensing for tasks, such as super-resolution [21] and cloud coverage removal [22], [23]. In recent years, GAN models have been proposed to produce images from scratch that closely mimic the spatial distribution observed in the training set [24]. However, the application of GAN methods, from the original Deep Convolutional GAN (DCGAN) [25] to more recent models, such as StyleGAN2 [26], has primarily been restricted to RGB data acquired by aerial platforms [27]. The first study that generalizes to the multispectral domain the generation of remote sensing images is that in [28] where the architecture named Progressive GAN (ProGAN, [29] is adopted to generate multispectral images from the Sentinel-2 satellite platform [30].

Following recent advances in the generative models, in this article, we explore the capability of StyleGAN3 [31], one of the latest and most promising GAN architectures, to generate tiles of Sentinel-2 multispectral images. Sentinel-2 data have 13 spectral bands covering the visible and near infrared (VNIR) and short wave infrared (SWIR) ranges with ground sampling distance (GSD) of 10, 20, and 60 m. Particularly, we focus on the four bands in the VNIR spectral range with 10 m GSD (referred to as high resolution (HR) images, hereinafter) and the six bands in the NIR/SWIR spectral range having 20 m GSD (referred to as low resolution (LR) images, hereinafter).

In the paper, we detail: a) the modifications made to the original StyleGAN3 architecture in order to extend its functionality to multispectral Sentinel-2 data, and b) the training strategy adopted in terms of training set involved and training parameters. We also discuss the results obtained by comparing them with those provided by the ProGAN model. For this purpose, we propose a qualitative analysis based on visual inspection of examples of generated images and a quantitative analysis obtained by exploiting the well-known Fréchet inception distance (FID, [32].

The rest of this article is organized as follows. After a general introduction to the GAN architecture (see Section II), in Section III we describe the StyleGAN3 model and the modifications made to handle Sentinel-2 data. In the same section we also provide details about the training strategy adopted. Results are discussed in Section IV. Finally, Section V concludes this article.

II. GENERATIVE ADVERSARIAL NETWORKS

The general architecture of a GAN consists of two neural networks, named the generator and the discriminator, which compete in a zero-sum game to generate data that are

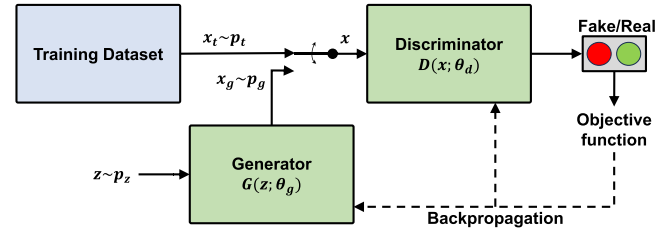


Fig. 1. GAN architecture.

indistinguishable from real ones. This adversarial competition enables the generator to learn to synthesize new data that follows the same statistical distribution as the training set. The conceptual GAN architecture is depicted in Fig. 1. The generator is a differentiable function $G(z; \theta_g)$, depending on the parameters θ_g , that takes input from a latent space z with a prior distribution p_z and generates data samples x_g according to a distribution p_g . The discriminator can be represented as a differentiable function $D(x; \theta_d)$, with parameters θ_d , that outputs the probability that sample x originates from the training data distribution p_t . During the training phase the parameters θ_d are updated so as to maximize $D(x; \theta_d)$, while the parameters θ_g are modified in order to make the statistical distribution of x_g more similar as possible to p_t . This is accomplished by solving the following minmax optimization problem:

$$\min_{\theta_g} \max_{\theta_d} V(\theta_g, \theta_d) \quad (1)$$

where

$$V(\theta_g, \theta_d) = \mathbb{E}_x[\log D(x; \theta_d)] + \mathbb{E}_z[\log(1 - D(G(z; \theta_g); \theta_d))] \quad (2)$$

with \mathbb{E} denoting the expectation. Equation (1) defines a competitive game where the discriminator is trained to best distinguish between true and generated data, while the generator learns to “fool” the discriminator. GANs were originally proposed as generative models for unsupervised learning, but they can also be trained using semisupervised or fully supervised learning methods [20], [33]. In recent years, there has been significant advancement in GAN research, leading to substantial changes in cost function used for training and network architecture. As to the cost function, various metrics have been proposed to enhance and expedite convergence, including Jensen–Shannon Divergence, least squares distance, and Wasserstein distance, which have demonstrated notable improvements [34]. The network architectures of GANs have also evolved. DCGAN [25] is one of the first proposed architecture and employs cascading 2-D convolutional layers in both the generator and discriminator. A major advance was introduced by the multiscale processing architecture resulting in ProGAN, [29] which incorporate a progressive growth mechanism into the generator, enabling the generation of higher resolution images with superior quality compared to previous approaches. In 2019, the ProGAN hierarchical processing chain was modified by introducing the innovative idea of style modulation, which resulted in the first version of the StyleGAN architecture [2]. The original StyleGAN has been modified in recent years to reduce droplet-like artefacts

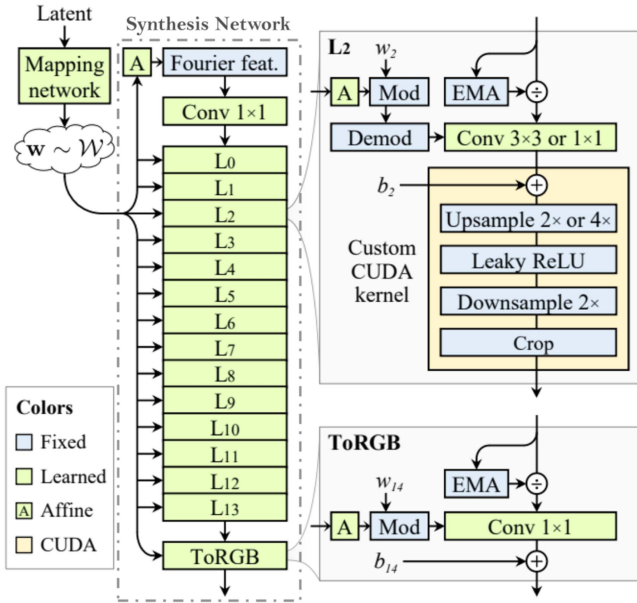


Fig. 2. Architecture of the StyleGAN3 generator. Figure from [35].

(StyleGAN2 [26] and aliasing arising in video synthesis (StyleGAN3, [35]. In parallel with the development of StyleGAN, conditional coordinate GAN [36] was developed which allows for the generation of images larger than those included in the training set.

III. GENERATING SENTINEL-2 IMAGES

In this section, first, we briefly summarize the StyleGAN3 architecture and we describe the modifications made to adapt the original model to Sentinel-2 image generation (see Section III-A). Then we describe the learning strategy by detailing the training set and the adopted training parameters (see Section III-B).

A. StyleGAN3

It is worth noting that the primary objective of this study is to assess the effectiveness of the StyleGAN3 model in generating unconditional multispectral remote sensing images, with a specific focus on Sentinel-2 satellite data. Here, StyleGAN3 is used as a tool and the complete description of the network architecture is out of the scope of this work. Therefore, in the following we briefly summarize the philosophy behind the StyleGAN3 architecture without going into the details, which can be found in [35]. StyleGAN introduced an innovative architecture for the generator part that, by using latent codes injected in each intermediate layer, enables dynamic adjustment of image style [2]. As to the discriminator part, all the versions of StyleGAN use the same architecture as ProGAN. The architecture of the StyleGAN3 generator is sketched in Fig. 2. It basically consists of two components: mapping network and synthesis network. The latter enables a progressive generation process that starts from low-resolution images and gradually refines them to higher resolutions. Regardless of the resolution

of the final output, synthesis network encompasses 14 blocks (denoted as L_i with $i = 0, \dots, 13$ in Fig. 2). Each block basically includes a convolutional layer and a nonlinearity (leakyRelu) wrapped between two data resize layers. One of the innovations of StyleGAN3 concerns the data resize layers. Specifically, the upsampling task which was accomplished by bilinear interpolation in the previous versions, in StyleGAN3 is performed by the more rigorous Whittaker–Shannon interpolation formula. It is implemented through a low-pass finite impulse response filter that approximates the “sinc” impulse response by truncation. In order to control the transition band and the ringing artifacts in the approximation of the low-pass filter, truncation is performed by using the Kaiser window. The blocks L_i s extract feature maps at different scales that are modulated by the style coefficients. The latter are in turn obtained by applying an affine transformation to the intermediate latent code w , which is the output of the mapping network. This latter ensembles two fully connected layers whose coefficients are learned during the training phase and transforms the 512×1 input latent vector z in the 512×1 intermediate latent vector w . z is the input of the StyleGAN3 generator and is randomly drawn from a multivariate normal distribution.

In our experiments we adopt the implementation of StyleGAN3 proposed in [37] based on the original software released by NVIDIA research group [38]. The adopted implementation supports standard raster image formats, such as JPG and PNG. These data formats restrict the system to handle images with three (or four, in the case of PNG) channels and having a radiometric resolution of 8-bit per channel. As stated in Section I, we analyze the generation process with respect to the HR Sentinel-2 data having four bands (B2, B3, B4, and B8) with 10 m GSD and the LR Sentinel-2 data with six bands (B5, B6, B7, B8A, B11, B12) and 20 m GSD named. Each band of the considered data has radiometric resolution of 12 bit. Therefore, different StyleGAN3 models are considered and to cope with the specific characteristics of Sentinel-2 data two main changes are applied to the implementation in [37]: a) the data read and write modules are modified to handle standard ENVI format [39] and floating-point precision; b) the number of channels of the 1×1 convolutional layer in the last block of the synthesis network (to RGB in Fig. 2) and that of the input layer of the discriminator are changed to handle the four bands and the six bands of HR and LR images, respectively.

In addition, for the purpose of comparison, we consider the official implementation of ProGAN proposed in [40], which we modified according to the aforementioned considerations in order to make the network capable of processing and generating multispectral floating point images. These steps enable the possibility of training the state-of-the-art StyleGAN3 and ProGAN on the same Sentinel-2 dataset, consequently allowing for a qualitative and quantitative comparison, as will be evident in Section IV.

B. Training Strategy

It is worth noting that, despite the StyleGAN3 network is able to manage images up to 1024×1024 pixels we decided to train

TABLE I
REAL DATASETS DESCRIPTION

Name of the dataset	Sensing date (dd/mm/yy)	File Identifier
Grosseto	07/07/2020	<i>S2A_MSIL2A_20200707T101031_N0214_R022_T32TPN_20200707T135316</i>
Pisa	23/04/2021	<i>S2A_MSIL2A_20210423T101021_N0300_R022_T32TP2_20210423T125530</i>
San Diego	19/11/2021	<i>S2A_MSIL2A_20211119T183801_N0301_R027_T11SMS_20211119T210846</i>
San Diego	03/01/2022	<i>S2B_MSIL2A_20220103T183749_N0301_R027_T11SMS_20220103T205342</i>
Melbourne	13/12/2021	<i>S2A_MSIL2A_20211213T001111_N0301_R073_T55HCU_20211213T015920</i>
Melbourne	31/12/2021	<i>S2B_MSIL2A_20211231T002049_N0301_R116_T55HCU_20211231T020637</i>
Rome	22/09/2021	<i>S2B_MSIL2A_20210922T32TQM_N0301_R116_T55HCU_20220922T020637</i>
La Paz	31/01/2022	<i>S2B_MSIL2A_20220131T175549_N0400_R141_T12QWM_20220131T204342</i>

TABLE II
TRAINED STYLEGAN3 MODELS

Name of the StyleGAN3	Sentinel-2 bands	Image size (pixels)
SGAN3_HR_64	B2, B3, B4, B8	64×64
SGAN3_LR_64	B5, B6, B7, B8A, B11, B12	64×64
SGAN3_HR_128	B2, B3, B4, B8	128×128
SGAN3_LR_128	B5, B6, B7, B8A, B11, B12	128×128

The first column contains the name of the model, the second and the third columns refer to the Sentinel-2 bands and the size of the images generated by the model, respectively.

models for reduced size images in order to limit the training time. Specifically, we consider the two data formats corresponding to 64×64 pixels and 128×128 pixels images. The unconditional generation task we are addressing via generative models requires huge and sufficiently diverse datasets of images in order to be properly trained. In this work we use datasets extracted from Sentinel-2 products that are free of charge and can be obtained directly from the online Copernicus Open Access HUB [41]. The online portal is provided by the European Space Agency to download the Sentinel's images acquired since the launch date.

We selected eight data acquired by both Sentinel-2A and Sentinel-2B sensors in different regions of the world (i.e., Italy, Australia, USA, Mexico) from 2020 to 2022. Specifically, three images cover regions that include the towns of Grosseto, Pisa (Tuscany, Italy), and Rome, two images refer to the area of San Diego (California, USA), two images were acquired on the area of Melbourne (Victoria, Australia) and one image refers to the area of La Paz (Baja California Sur, Mexico). Each dataset is composed by a $10\,980 \times 10\,980$ pixels HR image and a 5490×5490 pixels LR image. The considered images are representative of various application scenarios including urban, rural, coastal, and mountain scenarios. For each of the real dataset, Table I provides a) the assigned name, b) the sensing date, and c) the file identifier for its download.

We trained four different StyleGAN3 models, one for each combination of spectral range and image size. Table II

summarizes the four models indicating, for each of them, the name assigned, the type of image it refers to (HR with four bands or LR with six bands) and the size of the image in pixel units.

From the above-mentioned dataset, for each of the StyleGAN3 model considered, according to the spectral range and the images size of interest, we randomly picked 10^5 image tiles to define the training set. In this tiles extraction process we discarded areas covered by clouds, corrupted data and areas completely covered by sea water. Each image tile contains atmospherically corrected spectral reflectances with a 12 bit radiometric resolution that have been properly scaled within the dynamic range [0, 1] with floating point precision.

To further enrich the obtained training sets we applied to each of them the data augmentation techniques suggested in [42] and consisting in geometric and blitting transformations.

Geometric transformations encompass a range of techniques including isotropic scaling, arbitrary rotation, anisotropic scaling, and fractional translation. These transformations contribute to the augmentation process and assist in preserving the high-frequency details that may be lost during geometric transformations. In addition, pixel blitting techniques, such as x-flips, 90° rotations, and integer translations, are employed. These blitting transformations prove useful in recovering high-frequency details and further improving the quality of the training process [42].

Augmentation operations are applied sequentially to the training set during the training procedure following by means of an updating pipeline enabled by the NvLabs StyleGAN3 implementation.

Similar to StyleGAN3, we trained four different ProGAN models, one for each combination of spectral range and image size. Table III summarizes the four models, indicating the core dataset used for each of them. The dataset used to train the ProGAN models was identical to that used for StyleGAN3.

The training process took place on a single NVIDIA GeForce 3090 with 24 GB RAM. The models were trained until the generated images reached a satisfactory level of quality. The training duration was approximately seven days for each of the StyleGAN3 and ProGAN model considered. The StyleGAN3

TABLE III
TRAINED PROGAN MODELS

Name of the ProGAN	Sentinel-2 bands	Image size (pixels)
PGAN_HR_64	B2, B3, B4, B8	64×64
PGAN_LR_64	B5, B6, B7, B8A, B11, B12	64×64
PGAN_HR_128	B2, B3, B4, B8	128×128
PGAN_LR_128	B5, B6, B7, B8A, B11, B12	128×128

The first column contains the name of the model, the second and the third columns refer to the Sentinel-2 bands and the size of the images generated by the model, respectively.

networks were configured using the StyleGAN3-t configuration, and specific optimization parameters were chosen for the Adam optimizer. The values selected for the optimization parameters were $\beta_1 = 0$ and $\beta_2 = 0.99$. The learning rate was set to $d\ell r = 0.002$, the R1 regularization weight was $\gamma = 1$, and the batch size was set to 32.

The ProGAN networks were trained using Adam optimizer. The values selected for the optimization parameters were $\beta_1 = 0$ and $\beta_2 = 0.99$. The learning rate was set to $d\ell r = 0.001$, the R1 regularization weight was $\gamma = 1$, and the batch size was set to 16 [29].

IV. RESULTS

In this section, we discuss the results obtained by the networks trained on the dataset described in Section III-B. We present the generated multispectral images for both the high-resolution and low-resolution datasets, and we discuss their performance in terms of visual quality and selected analytical indicators which will comprehensively be described in Section IV-A. As stated in Section III ProGAN are considered for comparison.

A. Performance Metrics

The evaluation of generative models is a subject of ongoing debate, lacking consensus on a metric that effectively captures the models' strengths and limitations for comparison purposes [43]. While some approaches focus on quantitative assessment, others emphasize qualitative evaluation.

Currently, there is no universally accepted metric that comprehensively evaluates all aspects of generative models, such as quality, diversity, overfitting, and mode dropping [44].

Nonetheless, certain metrics are highly regarded. A widely accepted score is the FID, which compares generated images with real ones using a pretrained deep neural network classifier, based on inception models [45].

FID measures the statistical distance between the distributions of inception features extracted from both generated and real images [32]. These features are derived from the final convolutional layer of the inception model and are assumed to be Gaussian random vectors. Specifically, denoted as r and g the 2048-D features vectors obtained as output of the Inception model for real and generated data, respectively, FID measures the Fréchet

distance between the two Gaussian distributions of those vectors

$$\text{FID}(r, g) = \|\mu_r - \mu_g\|_2^2 + \text{Tr} \left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}} \right) \quad (3)$$

where (μ_r, Σ_r) and (μ_g, Σ_g) are the mean vectors and the covariance matrices of r and g , respectively [46].

FID ranges from 0 to ∞ , where lower values indicate greater similarity among groups of images.

In this work, to compute r and g we utilize a standard InceptionV3 model pretrained on the ImageNet dataset. The latter is a three-bands (RGB) dataset, so in order to take into account the multispectral nature of the HR and LR Sentinel-2 data we consider an augmented feature space for r and g . Specifically, for the HR four-bands images we calculate the FID₄ score using augmented feature vectors that result from concatenating the features extracted by InceptionV3 from bands $B2, B3, B4$, and bands $B3, B4, B8$. This approach ensures that the evaluation takes into account all the bands, facilitating a comprehensive assessment of the generated image. The same methodology is applied to evaluate the generated LR six-bands images, where the FID₆ score is defined using bands $B5, B6, B7$, and bands $B8A, B11, B12$, respectively.

It is worth noting that, while FID is not particularly informative for absolute evaluation of the performance of a given generative model, it holds significance for the relative scoring of generative models trained on the same dataset.

B. Image Generation and Evaluation

This section discusses examples of images obtained by the StyleGAN3 and ProGAN generators after the training process. We provide examples for both urban and rural scenarios and for both HR and LR images. For the sake of clarity and space, we are limiting ourselves to reporting only the figures associated with the 128×128 resolution, while the other case study (64×64) can be found in the Appendix.

To facilitate visual comparison, we have split the presented multispectral images into three-band subsets suitable for false color representation. The HR four-band images are represented by grouping bands $B2, B3$, and $B4$ to obtain an RGB image, and grouping bands $B3, B4$, and $B8$ to obtain a color infrared (CIR) image. The LR six-band images are represented by means of two false-color images, each grouping the first three and the second three bands ($B5, B6, B7$ and $B8A, B11, B12$, respectively).

Let us start the discussion focusing on the results in Fig. 3, which contains six HR image tiles organized in a two-rows and three-columns grid. The first row contains the RGB representations, while the second row contains the CIR representations of the selected images. In the first column, you can find an image randomly extracted from the training dataset. In the second column, you can find an image extracted from a pool of 50 000 images generated by StyleGAN3, and finally, in the third column, you can find an image extracted from a pool of 50 000 images generated by ProGAN. For each generative model the image shown was selected according to a specific criterion. Such a criterion is based on the consideration that by applying the Inception model to images with similar content, similar feature vectors are obtained. Therefore, first, we compute the

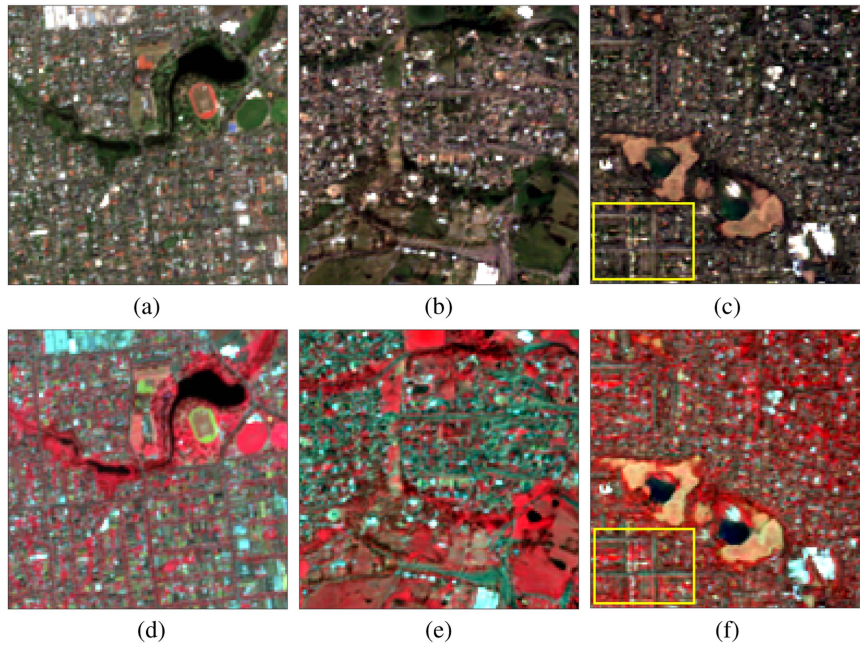


Fig. 3. Real and generated HR tiles, each measuring 128×128 pixels, of an urban scenario. (a), (b), and (c) depict the RGB representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands B_2 , B_3 , and B_4 . (d), (e), and (f) depict the corresponding false-color CIR representations obtained by combining bands B_3 , B_4 , and B_8 . (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

augmented inception feature vector for the real image, then we obtain the augmented Inception feature vectors for all the images produced by each generative model, and, finally we select the generated image closest to the real one in terms of Euclidean distance between the corresponding augmented inception feature vectors. It is therefore no coincidence that the images generated by both StyleGAN3 and ProGAN shown in Fig. 3 appear related in terms of content and describe a very similar scenario.

Specifically, Fig. 3 depicts HR tiles referring to a typical urban scenario and we can notice that the generated images are visually quite similar to the real one (first column) and faithfully reproduce the typical urban pattern, characterized by roofs of various sizes and compositions, as well as the presence of roads, green spaces, and industrial buildings. Both generated images are visually good in quality; however, the ProGAN image appears a little more blurred and with some evident artifacts, such as the straight lines highlighted in the yellow box within Fig. 3(c) and (f).

Fig. 4 shows examples concerning the generation of LR images with reference to an urban scenario. The organization of the results in Fig. 4 follows the format detailed for Fig. 3. Also in this case, the presented images generated by StyleGAN3 and ProGAN were selected according to the criterion based on the minimum Euclidean distance between augmented inception feature vectors. Of course, in this case we used the augmented inception feature vectors defined for the LR images in Section IV-A. The selected images are visually quite similar and reproduce typical urban pattern as the previous HR case. The quality of the StyleGAN3 image appears better than ProGAN one. The latter contains some visible artifacts, as highlighted by the yellow boxes within Fig. 4(c) and (f). The checkerboard

pattern in the image is not natural but is probably induced by the generation process.

Figs. 5 and 6 follow the structure described for the previous images but present rural scenario examples referring to the HR and LR cases, respectively. Also in this case the extraction approach based on Inception distance allowed to select images that are very similar in context. Both the networks successfully reproduce the geometry and composition of the fields, characterized by grassland juxtaposed with woods and bare soil. It can be said that, there is high quality in both the generated images, but the artifacts of ProGAN seem less pronounced.

In general, we can appreciate that the generative models are capable of producing high quality images. Both StyleGAN3 and ProGAN generate images that are hardly distinguishable to the naked eye, although it is evident that StyleGAN3 produces an overall better result.

In addition we underline that, for both the LR and HR cases, the generated images exhibit similarity in context but are not identical to the closest real image (in terms of Euclidean distance between the feature vectors). This fact could be considered as an indirect proof of the capability of the networks to generate images with a high degree of generalization without incurring in overfitting issues.

To further test the quality of the generated images, we compared the spectral profiles of selected pixels extracted from StyleGAN3 images, ProGAN images, and real ones, respectively. An effective generation process should produce similar spectral signatures for pixels that belong to the same natural/material class. Fig. 7 displays graphs comparing the spectral signatures of pixels highlighted by colored dots in Figs. 5 and 6 (the lines follow the color notation of the points). The first graph [see Fig. 7(a)] shows the signatures of the three

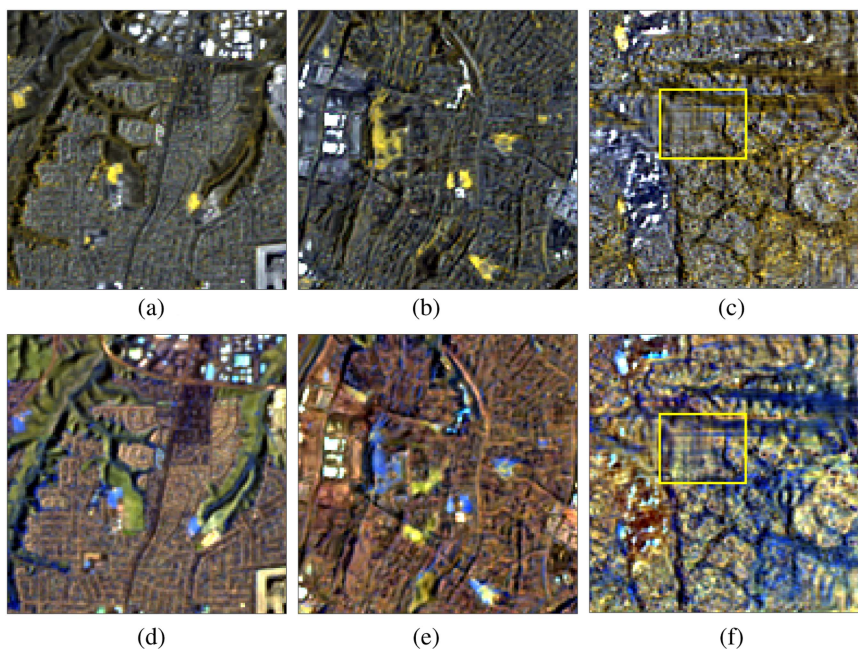


Fig. 4. Real and generated LR tiles, each measuring 128×128 pixels, of an urban scenario. (a), (b), and (c) depict the false color representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands $B5$, $B6$, and $B7$. (d), (e), and (f) depict the corresponding false-color representations obtained by combining bands $B8A$, $B11$, and $B12$. (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

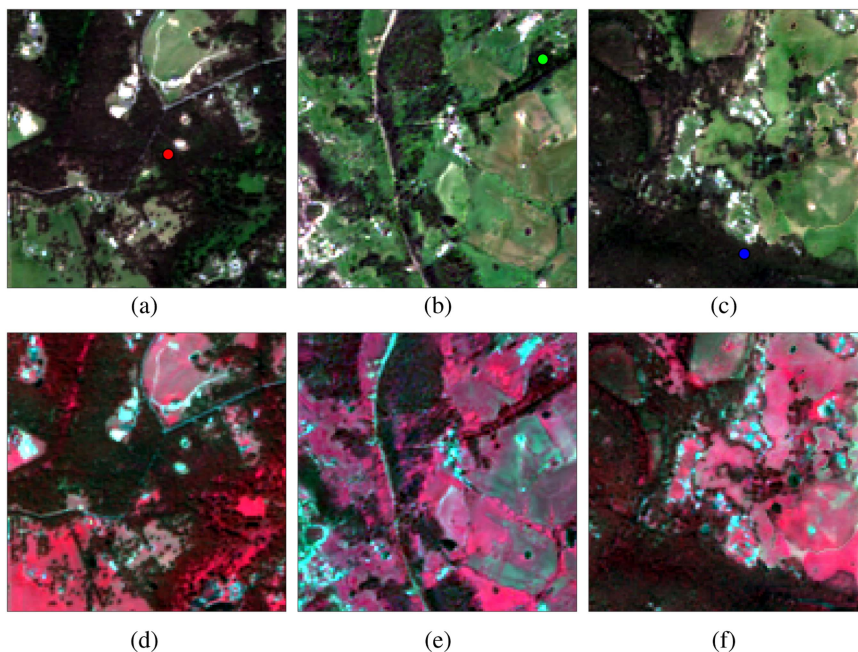


Fig. 5. Real and generated HR tiles, each measuring 128×128 pixels, of a rural scenario. (a), (b), and (c) depict the RGB representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands $B2$, $B3$, and $B4$. (d), (e), and (f) depict the corresponding false-color CIR representations obtained by combining bands $B3$, $B4$, and $B8$. (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

vegetation pixels marked by the dots in Fig. 5 (HR images), while the second graph [see Fig. 7(b)] displays the signatures of three pixels marked by the dots in Fig. 6 (LR images). This analysis is useful to evaluate the spectral signatures of pixels that belong to similar classes in order to verify if they exhibit consistent behavior. For instance, vegetation pixels [see

Fig. 7(a)] exhibit the characteristic red-edge profile between the bands $B3$ and $B8$. This is the first quantitative result that is useful to enforce the quality and consistency of the generation process.

Finally, to enforce the results we quantitatively compare the performance of StyleGAN3 and ProGAN by means of the

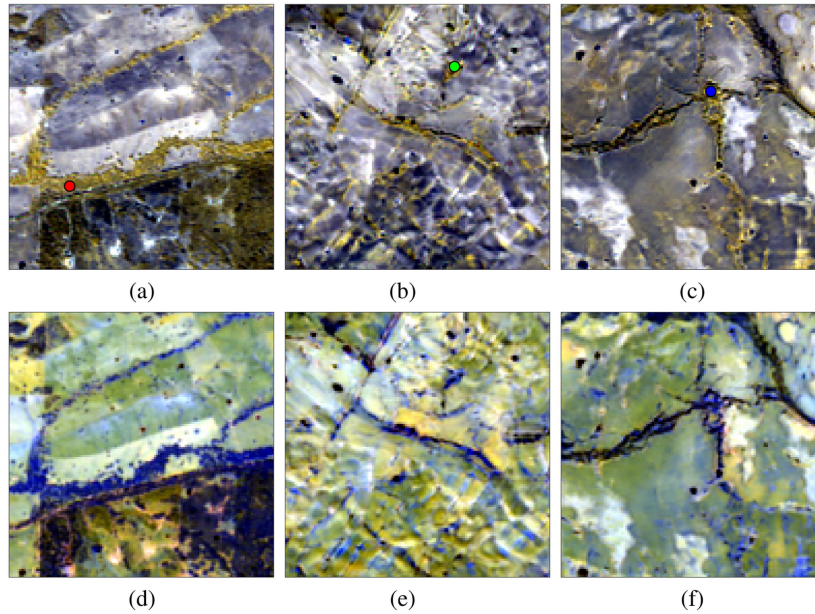


Fig. 6. Real and generated LR tiles, each measuring 128×128 pixels, of a rural scenario. (a), (b), and (c) depict the false color representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands B_5 , B_6 , and B_7 . (d), (e), and (f) depict the corresponding false-color representations obtained by combining bands B_{8A} , B_{11} , and B_{12} . (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

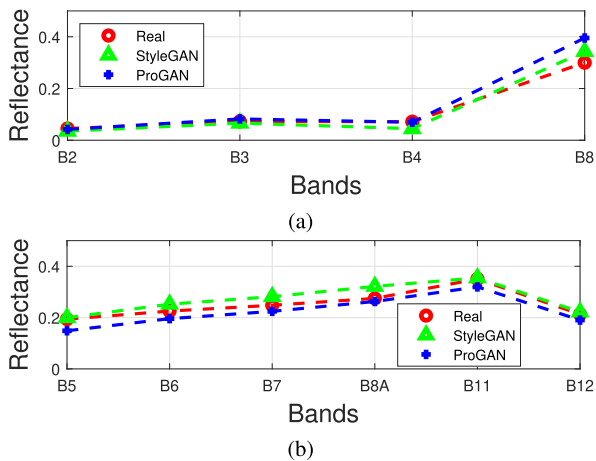


Fig. 7. Representation of the spectral reflectance of some pixels extracted from the images in Figs. 5 and 6; the colors in the graphs indicate the points they refer to. (a) Graphs associated with the points in Fig. 5(a) (red), Fig. 5(b) (green), and Fig. 5(c) (blue). (b) Graphs associated with the points in Fig. 6(a) (red), Fig. 6(b) (green), and Fig. 6(c) (blue).

augmented FID scores for all the considered spectral ranges (FID_4 for HR and FID_6 for LR). In order to obtain FID scores for each network (StyleGAN3 and ProGAN) we generated 50 000 images that are compared in terms of FID with 50 000 samples extracted from the training dataset. This procedure is repeated several times to have a statistical evaluation of the scores. This approach allows us to assess the variability and consistency of the results obtained from different runs. By the same procedure we have also evaluated FID scores involving only the real data, i.e., by extracting both the groups of 50 000 images from the real dataset. Values obtained in this way establish a lower bound for FID (the network cannot do better than this).

TABLE IV
FID SCORES OBTAINED USING THE AUGMENTED FEATURE VECTORS
GENERATED BY THE INCEPTIONV3 NETWORK

Data	FID_4		FID_6	
	64×64	128×128	64×64	128×128
Real	$2.10 \pm 7.71e-5$	$2.28 \pm 1.40e-4$	$1.93 \pm 2.21e-4$	$1.87 \pm 2.95e-4$
ProGAN	$14.65 \pm 1.85e-2$	$36.38 \pm 2.31e-2$	$72.73 \pm 2.47e-2$	$112.35 \pm 2.62e-2$
StyleGAN3	$13.41 \pm 1.16e-2$	$22.40 \pm 3.91e-2$	$18.97 \pm 1.39e-2$	$35.97 \pm 8.52e-2$

These scores were applied to both real data and the generated images from ProGAN and StyleGAN3. The table displays the FID scores for all the combinations of HR and LR, as well as tile sizes, that were considered.

The bold indicates that the FID scores obtained with StyleGAN3 are better (i.e. lower) than the scores obtained with respect to the ProGAN trained on the same dataset.

Table IV displays the mean and variance of the augmented FID scores obtained after 100 runs. It is worth noting that, although the visual comparison only involved the 128×128 case, Table IV also includes the numerical results obtained in the 64×64 case for completeness.

The results in Table IV clearly highlight that StyleGAN3 outperforms ProGAN in all the considered configurations. The first column of Table IV displays the means and standard deviations of FID_4 applied to the 64×64 HR case, where the differences in means and variances between ProGAN and StyleGAN3 are limited to 1.24 and $0.69e-2$. The variance values remain small in all cases; thus, we do not discuss them further because they are less relevant. When we consider the 128×128 HR case (second column of Table IV) the difference between FID_4 mean values provided by the two generative models, increase to 13.98 in favor of StyleGAN3.

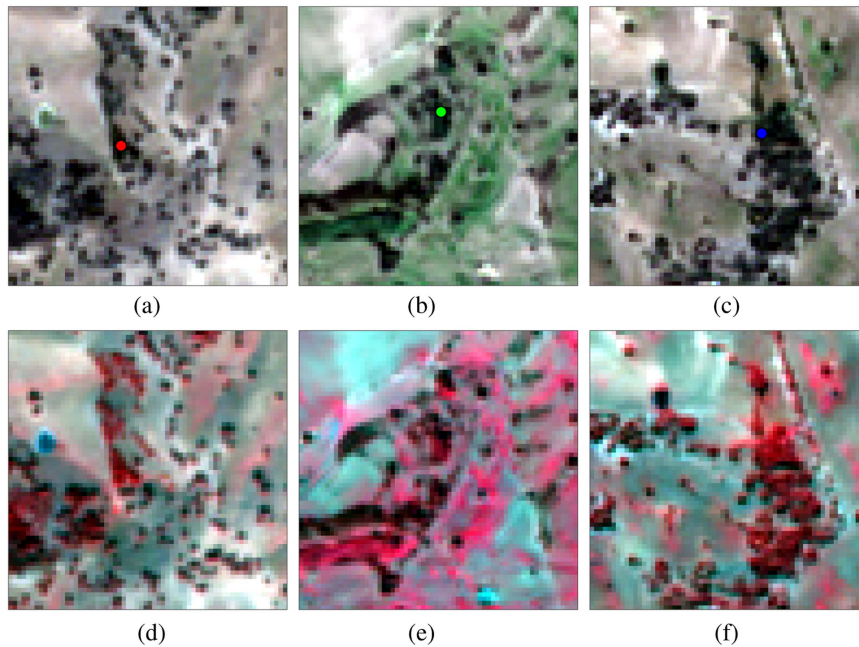


Fig. 8. Real and generated HR tiles, each measuring 64×64 pixels, of a nonurban scenario. (a), (b), and (c) depict the RGB representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands B_2 , B_3 , and B_4 . (d), (e), and (f) depict the corresponding false-color representations obtained by combining bands B_3 , B_4 , and B_8 . (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

The third column of Table IV illustrates the means and standard deviations of FID_6 applied to the 64×64 HR case. The difference in terms of mean values of FID_6 between StyleGAN3 and ProGAN grows to 53.76. This value further increases when considering the 128×128 LR case (fourth column of Table IV), where the difference becomes 73.38. We can conclude that FIDs scores provided by StyleGAN3 networks are consistently lower than those obtained by ProGAN networks in all cases. Interestingly, as the number of bands and image size increase, the difference between the scores of the two networks becomes more pronounced. This observation suggests that, as resolution and the number of bands increase; StyleGAN3 becomes more effective than ProGAN and, to some extent, confirms the qualitative results obtained from visual inspection.

V. CONCLUSION

In this article, we present a study focused on generating synthetic satellite images using unconditional generative models. Specifically, we adapted an instance of the state-of-the-art StyleGAN3 model and applied it to Sentinel-2 multispectral satellite imagery. After proper adaptations and training, the StyleGAN3 model demonstrated excellent performance in the new context of multispectral satellite imagery, proving effective in faithfully reproducing the spectral and spatial characteristics of both four-bands and six-bands images under examination. The images generated by StyleGAN3 were compared with those generated by ProGAN, both trained on the same dataset. The comparison was initially conducted through visual inspection and subsequently through quantitative analysis, involving the observation of spectral signatures in selected pixels and

evaluation using the commonly used FID metric. The use of FID required an augmentation procedure to handle the multispectral case. StyleGAN3 exhibited superior performance over ProGAN in all cases considered. When comparing StyleGAN3 and ProGAN for 64×64 tiles, FID_4 values of 13.41 and 14.65 were observed for the HR four-band images, and FID_6 values of 18.97 and 72.73 for LR six-band images. In addition, when comparing StyleGAN3 and ProGAN for larger 128×128 tiles, FID_4 values of 22.40 and 36.38 were observed for four-band images, and FID_6 values of 35.97 and 112.35 for six-band images. However, while distinguishing images generated by StyleGAN3 from real ones through visual inspection alone is challenging, and the FID values are better than those of ProGAN, StyleGAN3's FID values still fall far from the lower bound established by comparing groups of real images

This can be interpreted as a sign that there is still room to define automatic methods able to distinguish fake images from real ones. The latter is an important conclusion considering the serious implications of false data in terms of security and should stimulate researchers to study new algorithms for fake image detection. In the near future we intend to extend our study to images with larger dimensions (beyond 128×128) and more spectral bands (e.g., hyperspectral images).

APPENDIX

This appendix presents results for the 64×64 case study that were omitted from this article to maintain conciseness and clarity. These results are reported here to provide the reader with an overall view of the findings. The organization of the results is the same as that adopted in Section IV-B.

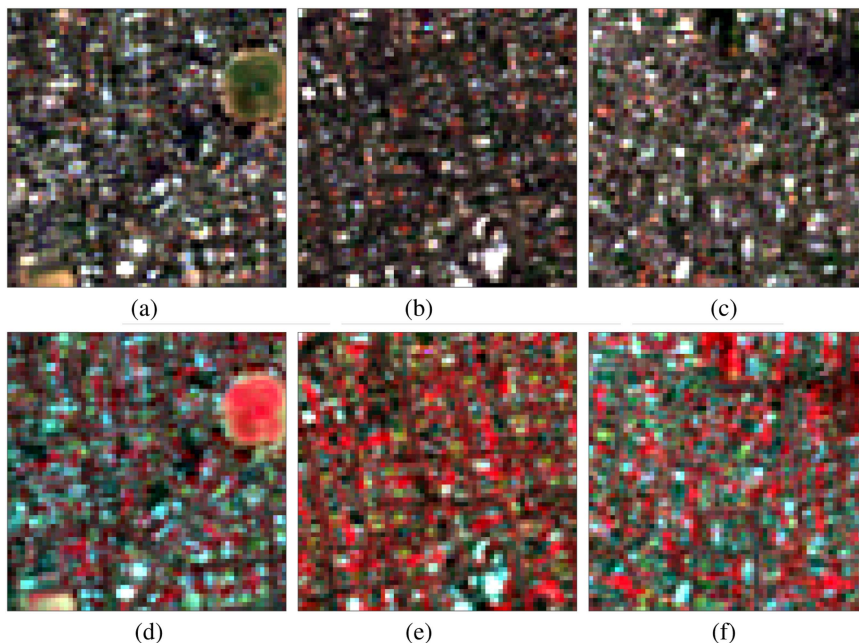


Fig. 9. Real and generated HR tiles, each measuring 64×64 pixels, of an urban scenario. (a), (b), and (c) depict the RGB representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands $B2$, $B3$, and $B4$. (d), (e), and (f) depict the corresponding false-color representations obtained by combining bands $B3$, $B4$, and $B8$. (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

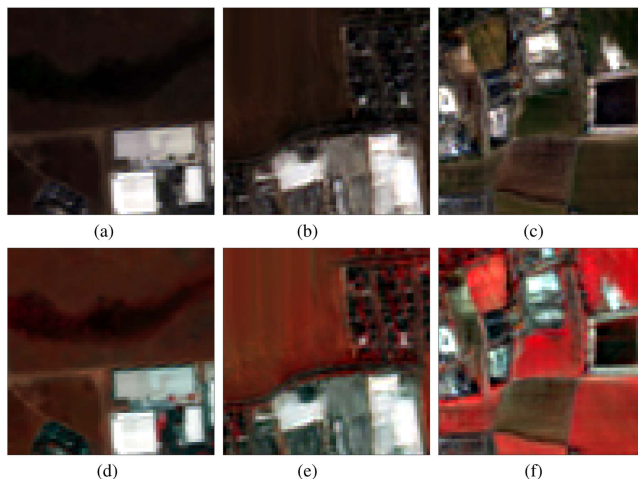


Fig. 10. Real and generated HR tiles, each measuring 64×64 pixels, of an urban scenario. (a), (b), and (c) depict the RGB representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands $B2$, $B3$, and $B4$. (d), (e), and (f) depict the corresponding false-color representations obtained by combining bands $B3$, $B4$, and $B8$. (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

Particularly, Figs. 8–10 depict the RGB and corresponding CIR representations of HR images extracted from the training, StyleGAN3, and ProGAN datasets. Conversely, Figs. 11–13 show the false-color representations of LR image datasets.

We provide some examples from both urban (see Figs. 9–11) and nonurban scenarios (Figs. 8, 12, and 13) with reference to both HR and LR models.

Fig. 8 shows the capability of the networks to reproduce the pattern of nonurban scenarios with alternating bare soil,

vegetation, and woods. Fig. 12 shows a coastal area with rural lands and several water bodies, while Fig. 13 depicts the pattern of a typical mountainous region. Whereas, for urban scenarios, the networks exhibit the capability to reproduce low-density (see Fig. 10), middensity (see Fig. 9), and high-density (see Fig. 11) urban images.

Finally, in order to give examples of the capability of the generative networks to reproduce the spectral behavior of real data, we compare the spectral signatures of pixels extracted from

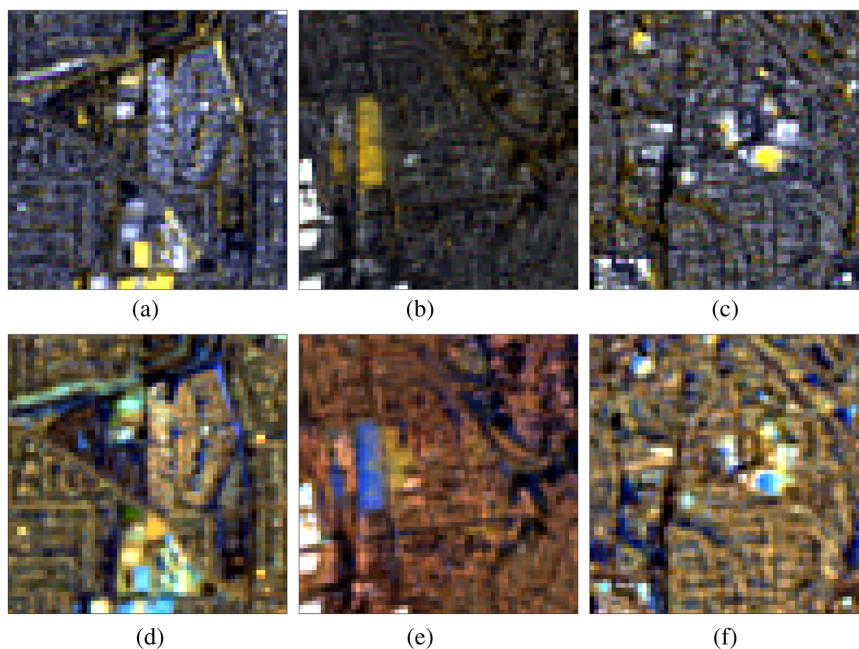


Fig. 11. Real and generated LR tiles, each measuring 64×64 pixels, of an urban scenario (a), (b), and (c) depict the RGB representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands B_5 , B_6 , and B_7 . (d), (e), and (f) depict the corresponding false-color representations obtained by combining bands B_{8A} , B_{11} , and B_{12} . (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

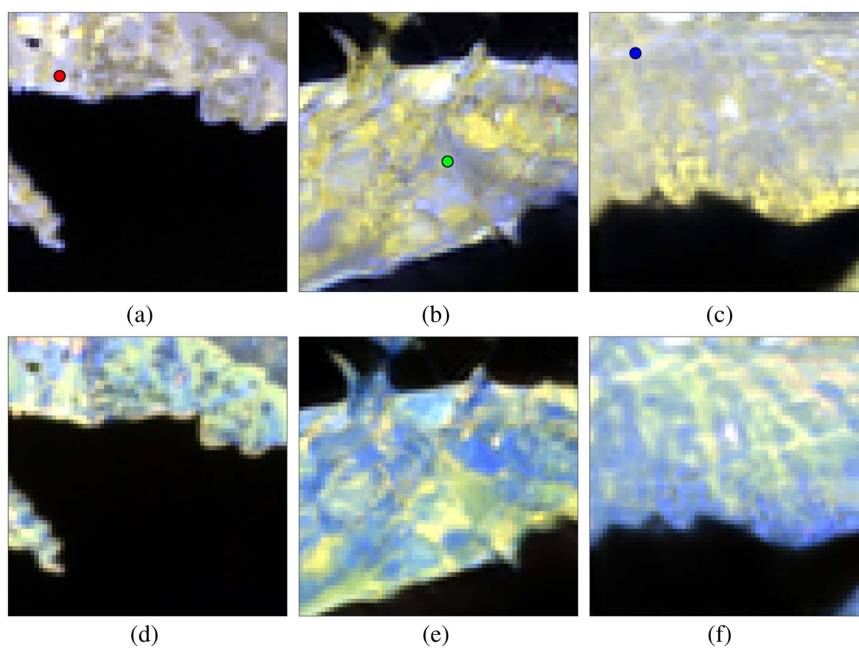


Fig. 12. Real and generated LR tiles, each measuring 64×64 pixels, of a nonurban scenario. (a), (b), and (c) depict the RGB representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands B_5 , B_6 , and B_7 . (d), (e), and (f) depict the corresponding false-color representations obtained by combining bands B_{8A} , B_{11} , and B_{12} . (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

similar classes. Fig. 14 displays the spectral signatures of the pixels that are highlighted by colored dots in Figs. 8 and 12 (the lines follow the color notation of the points). The first graph [see Fig. 14(a)] shows the signatures of two vegetation pixels from HR images, while the second graph [see Fig. 14(b)] displays

the signatures of two soil pixels from LR images. The spectral signatures extracted from the same classes exhibit consistent behavior. For example, also in this case, the vegetation pixels have the typical red-edge profile in all the considered images [see Fig. 14(a)].

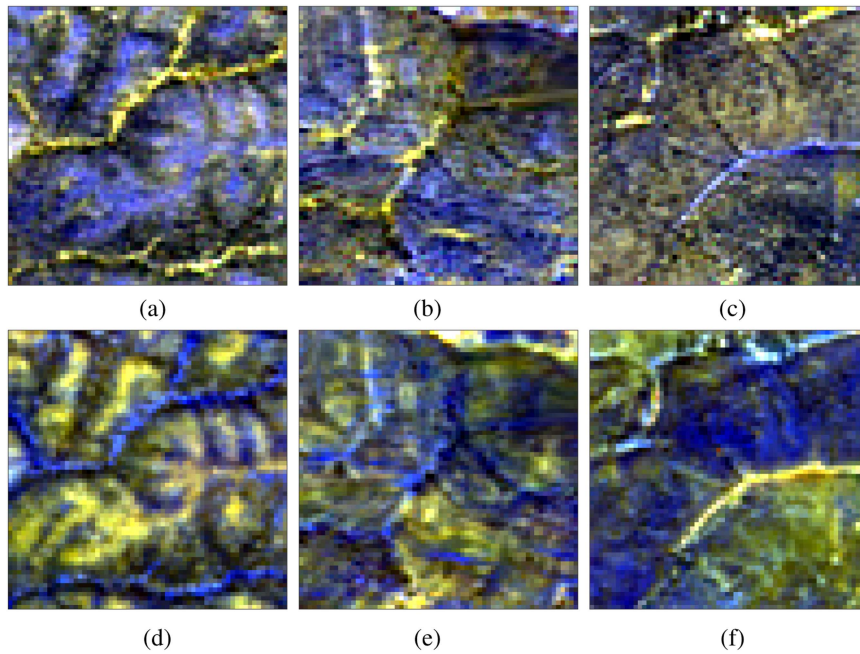


Fig. 13. Real and generated LR tiles, each measuring 64x64 pixels, of a nonurban scenario. (a), (b), and (c) depict the RGB representations of images belonging to the Training, StyleGAN3, and ProGAN datasets, respectively, obtained by combining bands B_5 , B_6 , and B_7 . (d), (e), and (f) depict the corresponding false-color representations obtained by combining bands B_{8A} , B_{11} , and B_{12} . (a) Training. (b) StyleGAN3. (c) ProGAN. (d) Training. (e) StyleGAN3. (f) ProGAN.

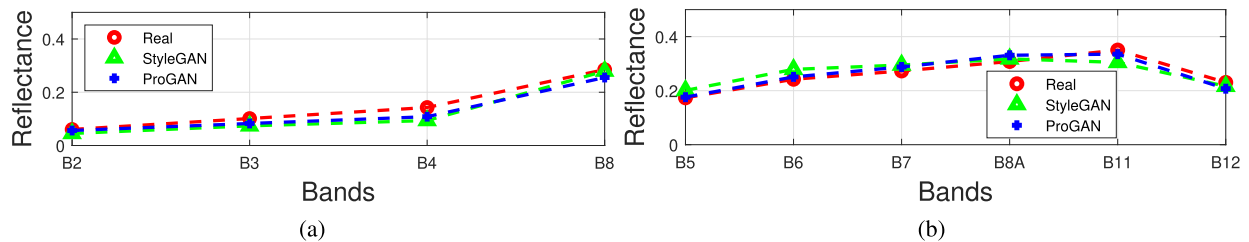


Fig. 14. Representation of the spectral reflectance of some pixels extracted from the images in Figs. 8 and 12; the colors in the graphs indicate the points they refer to. (a) Graphs associated with the points in Fig. 8(a) (red), Fig. 8(b) (green), and Fig. 8(c) (blue). (b) Graphs associated with the points in Fig. 12(a) (red), Fig. 12(b) (green), and Fig. 12(c) (blue).

REFERENCES

- [1] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.
- [2] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4396–4405.
- [3] S. Shahriar, "GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network," *Displays*, vol. 73, 2022, Art. no. 102237.
- [4] Y. Hu, L. Manikonda, and S. Kambhampati, "What we Instagram: A first analysis of Instagram photo content and user types," in *Proc. Int. AAAI Conf. Web Social Media*, 2014, vol. 8, no. 1, pp. 595–598.
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [6] N. Audebert, B. Le Saux, and S. Lefèvre, "Generative adversarial networks for realistic synthesis of hyperspectral samples," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4359–4362.
- [7] D. Lee, S. Liu, J. Gu, M.-Y. Liu, M.-H. Yang, and J. Kautz, "Context-aware synthesis and placement of object instances," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, vol. 31, pp. 10393–10403.
- [8] O. Adedeji, P. Owoade, O. Ajayi, and O. Arowolo, "Image augmentation for satellite images," 2022, *arXiv:2207.14580*.
- [9] J. Huang, S. Liu, Y. Tang, and X. Zhang, "Object-level remote sensing image augmentation using U-Net-based generative adversarial networks," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–12, 2021.
- [10] H.-S. Chen, K. Zhang, S. Hu, S. You, and C.-C. J. Kuo, "Geo-DefakeHop: High-performance geographic fake image detection," 2021, *arXiv:2110.09795*.
- [11] J. Kietzmann, L. W. Lee, I. P. McCarthy, and T. C. Kietzmann, "Deepfakes: Trick or treat?," *Bus. Horiz.*, vol. 63, no. 2, pp. 135–146, 2020.
- [12] X. Zhang, S. Karaman, and S.-F. Chang, "Detecting and simulating artifacts in GAN fake images," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, 2019, pp. 1–6.
- [13] Z. Wang, Y. Guo, and W. Zuo, "Deepfake forensics via an adversarial game," *IEEE Trans. Image Process.*, vol. 31, pp. 3541–3552, 2022.
- [14] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "CNN-generated images are surprisingly easy to spot... for now," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8692–8701.
- [15] B. Zhao, S. Zhang, C. Xu, Y. Sun, and C. Deng, "Deep fake geography? when geospatial data encounter artificial intelligence," *Cartogr. Geographic Inf. Sci.*, vol. 48, no. 4, pp. 338–352, 2021.
- [16] N. Yuchen, D. Xiangling, Z. Wenyi, and Z. Yulin, "A systematic review on detection of manipulated satellite images," in *Proc. IEEE 10th Int. Conf. Cyber Secur. Cloud Comput./IEEE 9th Int. Conf. Edge Comput. Scalable Cloud*, 2023, pp. 6–11.

- [17] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [18] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 172–189.
- [19] 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. 105–114.
- [20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5967–5976.
- [21] L. Salgueiro Romero, J. Marcello, and V. Vilaplana, "Super-resolution of sentinel-2 imagery using generative adversarial networks," *Remote Sens.*, vol. 12, no. 15, 2020, Art. no. 2424.
- [22] H. Pan, "Cloud removal for remote sensing imagery via spatial attention generative adversarial network," 2020, *arXiv:2009.13015*.
- [23] X. Wen, Z. Pan, Y. Hu, and J. Liu, "Generative adversarial learning in YUV color space for thin cloud removal on satellite imagery," *Remote Sens.*, vol. 13, no. 6, 2021, Art. no. 1079.
- [24] L. Abady et al., "An overview on the generation and detection of synthetic and manipulated satellite images," *APSIPA Trans. Signal Inf. Process.*, vol. 11, no. 1, pp. 1–56, 2022.
- [25] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. 4th Int. Conf. Learn. Representations*, 2016, San Juan, Puerto Rico, May 2–4, 2016, pp. 1–16. [Online]. Available: <http://arxiv.org/abs/1511.06434>
- [26] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of styleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8107–8116.
- [27] M. Yates, G. Hart, R. Houghton, M. T. Torres, and M. Pound, "Evaluation of synthetic aerial imagery using unconditional generative adversarial networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 190, pp. 231–251, 2022.
- [28] L. Abady, M. Barni, A. Garzelli, and B. Tondi, "Gan generation of synthetic multispectral satellite images," in *Proc. SPIE*, vol. 11533, pp. 122–133, 2020.
- [29] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–26. [Online]. Available: <https://openreview.net/forum?id=Hk99zCeAb>
- [30] J.-N. Thépaut, D. Dee, R. Engelen, and B. Pinty, "The copernicus programme and its climate change service," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 1591–1593.
- [31] T. Karras et al., "Alias-free generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 852–863.
- [32] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 6626–6637.
- [33] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 2234–2242.
- [34] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, vol. 70, pp. 214–223. [Online]. Available: <https://proceedings.mlr.press/v70/arjovsky17a.html>
- [35] T. Karras et al., "Alias-free generative adversarial networks," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 852–863, 2021.
- [36] C. H. Lin, C.-C. Chang, Y.-S. Chen, D.-C. Juan, W. Wei, and H.-T. Chen, "COCO-GAN: Generation by parts via conditional coordinating," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 4511–4520.
- [37] D. Porres, "Stylegan3-fun," [Online]. Available: <https://github.com/PDillis/stylegan3-fun>
- [38] NVlabs, "Alias-free generative adversarial networks (stylegan3)," [Online]. Available: <https://github.com/NVLabs/stylegan3>
- [39] N. Geospatial, "Envi image files," [Online]. Available: <https://www.l3harrisgeospatial.com/docs/ENVIImageFiles.html>
- [40] Karras, "Progressive growing of GANs for improved quality, stability, and variation," [Online]. Available: https://github.com/tkarras/progressive_growing_of_gans
- [41] ESA, "Copernicus hub," [Online]. Available: <https://scihub.copernicus.eu/>
- [42] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 12104–12114, 2020.
- [43] A. Borji, "Pros and cons of GAN evaluation measures," *Comput. Vis. Image Understanding*, vol. 179, pp. 41–65, 2019.
- [44] M. Kang, J. Shin, and J. Park, "StudioGAN: A taxonomy and benchmark of GANs for image synthesis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 15725–15742, 2023, doi: [10.1109/TPAMI.2023.3306436](https://doi.org/10.1109/TPAMI.2023.3306436).
- [45] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
- [46] D. Dowson and B. Landau, "The Fréchet distance between multivariate normal distributions," *J. Multivariate Anal.*, vol. 12, no. 3, pp. 450–455, 1982.



Michael Alibani (Member, IEEE) received the B.S. degree in biomedical engineering, the M.S. degree in automation and robotics engineering, and the Ph.D. degree in information engineering from the University of Pisa, Pisa, Italy, in 2012, 2017, and 2020, respectively.

Since 2022, he has been a temporary Researcher with the Department of Information Engineering, University of Pisa. His research interests include autonomous vehicles, control systems, and remote sensing. His current activity has been focused on deep

learning and machine learning techniques for remotely sensed data.



Nicola Acito received the Laurea degree (*cum laude*) in telecommunication engineering and the Ph.D. degree in "methods and technologies for environmental monitoring" from the University of Pisa, Italy, in 2001 and 2005, respectively.

From 2004 to 2008, he was a temporary Researcher with the Department of Information Engineering, University of Pisa. From 2009 to 2020, he was Researcher/Lecturer with the Italian Navy Academy, Livorno, Italy. Since 2021, he has been an Associate Professor with the University of Pisa. His research

interests include multidimensional signal and image processing in remote sensing applications, deep learning and machine learning techniques for remotely sensed data. His current activity has been focused on atmospheric compensation and superresolution of hyperspectral images.

Dr. Acito is a Reviewer of many international journals, i.e., IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS.



Giovanni Corsini (Life Member, IEEE) received the Laurea degree in electronic engineering from University of Pisa, Pisa, Italy, in 1979.

He is currently a Full Professor with the Department of Information Engineering, University of Pisa, where he teaches "image processing." His research interests include signal detection and processing with emphasis on image and multidimensional signal analysis in remote sensing applications, object detection and parameter estimation from remotely sensed data with particular emphasis on hyperspectral and multi-

spectral images.

Prof. Corsini has been a Member of technical committees at international conferences. He is a Reviewer of the most important international journals in this field, i.e., IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS, and the *International Journal on Remote Sensing*.