Enhancing Very High-Resolution Satellite Images At 15 cm: A Novel Pipeline With Synthetic Data and Single Image Superresolution

Sandeep Kumar Jangir and Reza Bahmanyar

***Abstract*—Employing a two-step pipeline that encompasses an image-to-image translation and a superresolution (SR) network, we significantly enhance satellite images with a ground sample distance (GSD) of 30 cm to a superior 15 cm GSD. Our translation network learns from the characteristics of satellite images and replicates these onto aerial images with a GSD of 15 cm, creating a tailored training dataset. The SR network then uses this dataset to train and render enhanced satellite images at a GSD of 15 cm. This innovative approach can provide a cost-effective substitute for commercial high-definition images, and broadens the usage of high-quality data across various applications.**

***Index Terms*—15 cm ground sample distance (GSD), aerial and satellite image enhancement, image-to-image (I2I) translation, superresolution (SR).**

I. Introduction

In RECENT years, very high-resolution (HR) satellite images have been widely used for several high-level computer-vision (CV) tasks, such as object detection, semantic segmentation, and more in various applications, such as traffic monitoring [1], autonomous driving [2], [3], and urban planning [4], [5]. Due to their acquisition conditions and objectives, satellite images can suffer from various technical and physical problems, such as sensor noise, atmospheric distortion, cloud cover, and irregular illumination. In addition, the quality and the spatial resolution of commercially available satellite imagery is not always high enough for many high-level CV tasks. To deal with these limitations, many image enhancement techniques, such as color adjustment [6], [7], [8], noise reduction [9], [10], [11], [12], and superresolution (SR) [13], [14], [15], [16], [17] techniques have been developed to cost-effectively improve the quality of images acquired by currently operational and older satellites. Common image enhancement tasks, such as denoising, deblurring, and SR are low-level and ill-posed inverse problems, and mapping the degraded images to their high-quality counterparts can be supervised or unsupervised. Recently, image enhancement methods based on deep learning (DL) approaches have been shown to outperform traditional methods [18], [19]. Among them, the most widely applied methods to satellite imagery are DL-based SR [20] methods, which increase the spatial resolution of images by learning and modeling complex patterns for more accurate subpixel information synthesis. These methods are categorized into single-image SR (SISR) [21] and multi-image SR (MISR) [22]. While SISR methods use a single image, MISR methods combine multiple images of an area, taken either by the same or different satellites at different orbital cycles, to recover a higher resolution image.

SISR methods are preferred over MISR methods due to the complexity of acquiring and processing large amounts of data. They learn a mapping from low-resolution (LR) to HR images [also known as ground truth (GT)] using paired training datasets. Ground sampling distance (GSD), ground area covered by a pixel, refers to the spatial resolution of aerial or satellite imagery, with smaller GSD indicating higher resolution. Traditionally, HR images are artificially downsampled to generate LR counterparts [16], [23], [24], which restricts the enhancement capability of SR networks to the original HR resolution [25], [26], [27], [28], [29]. For example, a model trained on pairs of 60 cm (LR) and 30 cm GSD (HR) images cannot accurately enhance 30 cm GSD images to 15 cm. Obtaining 15 cm GSD GT for accurate SR training is a challenge. Consequently, SR enhancement is limited by the highest resolution of the training data (currently, commercially available 30 cm GSD images), resulting in less accurate or representative 15 cm GSD images. Satellite image enhancement to better than 30 cm GSD has only been attempted in a few previous works [26], [30], [31]. Zhu et al. [30] proposed an approach to generate realistic training dataset by estimating a degradation model from commercial satellite images at 50 cm GSD and applying it to Google-owned aerial images for generating LR images. With this dataset, they train an SR network (similar to SRGAN [15]) to enhance the GeoEye-1 satellite images from 50 to 25 cm GSD. Shermeyer et al. [26] trained an SR networks (VDSR [32], RFSR [26]) to enhance satellite images from 60 to 30 cm and 120 to 30 cm GSDs. Then, they applied the trained model to enhance WorldView-3 (WV-3) satellite images from 30 to 15 cm GSD. They show that enhanced images significantly improve object detection performance, especially for small objects. Currently, organizations such as Maxar, European Space Imaging (EUSI) and Airbus offer commercial 15 cm GSD imagery based on their proprietary

Manuscript received 13 September 2023; revised 4 December 2023; accepted 18 December 2023. Date of publication 22 December 2023; date of current version 10 January 2024. This work of Sandeep Kumar Jangir was supported by a DAAD-DLR Research Fellowship under Grant ID-57540125 as a part of his Ph.D. studies. *(Corresponding author: Sandeep Kumar Jangir.)*

The authors are with the Department of Photogrammetry and Image Analysis, Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany (e-mail: sandeep.jangir@dlr.de; reza.bahmanyar@dlr.de).

Digital Object Identifier 10.1109/JSTARS.2023.3346181

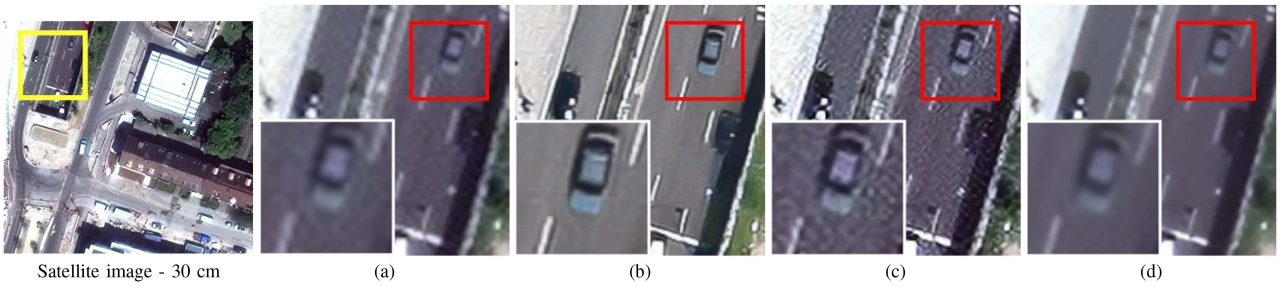

Fig. 1. Enhanced 15 cm GSD satellite images. (a) Bicubic Interpolation, Real-ESRGAN [17] trained on (b) SynthSat dataset, (c) traditional approach (HR/GT-15 cm (aerial); LR-30 cm), and (d) ground image dataset DIV2K [35].

high-definition (HD) technologies. A recent publication from Chouteau et al. at Airbus [31] explains an approach for generating 15 cm HD satellite images. They apply their proprietary measured degradation parameters from their satellite to their extremely high resolution (3–7.5 cm GSD) aerial imagery to generate LR images. By this dataset, they train an SR network and use it to the enhancement of the Pléiades Neo satellite images from 30 to 15 cm GSD.

In this article, we propose an image-to-image superresolution (I2I-SR) pipeline to i) generate HR (15-cm GT and 30-cm LR) synthetic dataset with paired images, and ii) use it to train an SISR network to enhance the RGB channels of multispectral satellite images from 30 to 15 cm GSD. To this end, we develop a novel I2I network, called SynthSat-I2I, for synthesizing 30-cm satellite images from 15-cm aerial imagery. The synthesized images contain the content of the aerial images with the degradation and style of the satellite imagery. The synthesized images and the aerial images form the SynthSat dataset. Using this dataset, we train an SR network to enhance the synthesized 30 cm GSD satellite image to its paired 15 cm GSD aerial image. Since aerial images are acquired from much lower altitudes than satellite images, they suffer less from various distortions in addition to their better spatial resolutions. The trained SR network using our dataset can provide 15-cm satellite images, which exhibit the increased sharpness, detail, and clarity typical of aerial imagery, alongside the style of aerial images. In our experiments, we train state-of-the-art (SOTA) SR networks that model real-world degradation for real-world or blind SR, such as Real-ESRGAN [17], BSRGAN [33], SwinIR [34], and simple GAN-based ESRGAN [16], on SynthSat dataset. Since there is no 15 cm GSD GT for the 30-cm satellite image, we evaluate the SR networks based on the quality of the images produced. We select Real-ESRGAN as the SR network for our I2I-SR pipeline as it produces better 15 cm GSD images which are sharper, with perceptual quality and visual appearance closer to aerial imagery compared to other SR methods.

Fig. 1 presents results of enhancing a satellite image from 30 to 15 cm GSD using different techniques: bicubic interpolation, Real-ESRGAN trained on SynthSat dataset, Real-ESRGAN trained with traditional dataset creation approach (downsampling 15 cm GSD GT aerial images using bicubic interpolation to 30 cm GSD LR images), and Real-ESRGAN trained on ground image dataset (DIV2K) [35].

Visual comparisons with existing HD products show the competitive quality of our enhanced images. For example, Fig. 2 compares our enhancement result from our I2I-SR pipeline and HD image from EUSI for the same WV-4 satellite image. Our result is sharper with fewer artifacts, especially in the high frequency components. With satellites, such as WV-3, WV-4, Pléiades Neo, and SuperView, providing multispectral images at 30 cm GSD from around the world, our enhancement approach can provide community low-cost SR images for various downstream applications. Fig. 3 gives an overview of our approach. We also perform an ablation study to verify the importance of different parts of the SynthSat-I2I network. We also believe that our proposed pipeline can be used to create the synthetic dataset for denoising, dehazing, and other enhancement algorithms that require pairwise datasets.

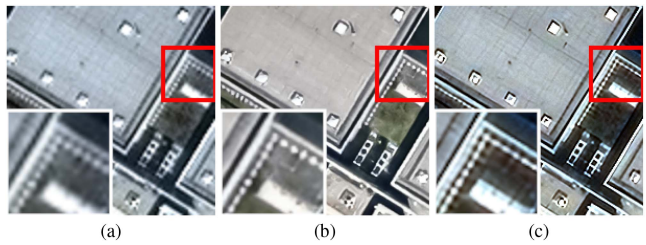

Fig. 2. (a) 30 cm GSD satellite. (b) I2I-SR 15 cm GSD. (c) 15 cm HD from EUSI.

The contributions of this article can be summarized as follows.

1) *SynthSat-I2I network:* This article introduces a novel I2I-SR pipeline, featuring the SynthSat-I2I network. This network is designed to synthesize 30 cm GSD satellite images from HR aerial imagery with a resolution of 15 cm GSD. The synthesized images aim to retain the content of aerial images while incorporating the degradation and style characteristics of satellite imagery.
2) *Creation of SynthSat dataset:* The proposed pipeline contributes to the generation of a synthetic dataset named SynthSat. This dataset consists of paired images, including 15 cm GSD aerial images and their corresponding synthesized 30 cm GSD satellite images. The dataset serves as a valuable resource for training SR networks to enhance satellite images beyond their native resolution.
3) *Training SR networks:* This article demonstrates the application of the SynthSat dataset to train SOTA SR networks to reach a higher resolution of 15 cm GSD.
4) *Overcoming resolution limitations:* This article addresses the limitation in training data resolution by proposing a method to enhance satellite images beyond the highest

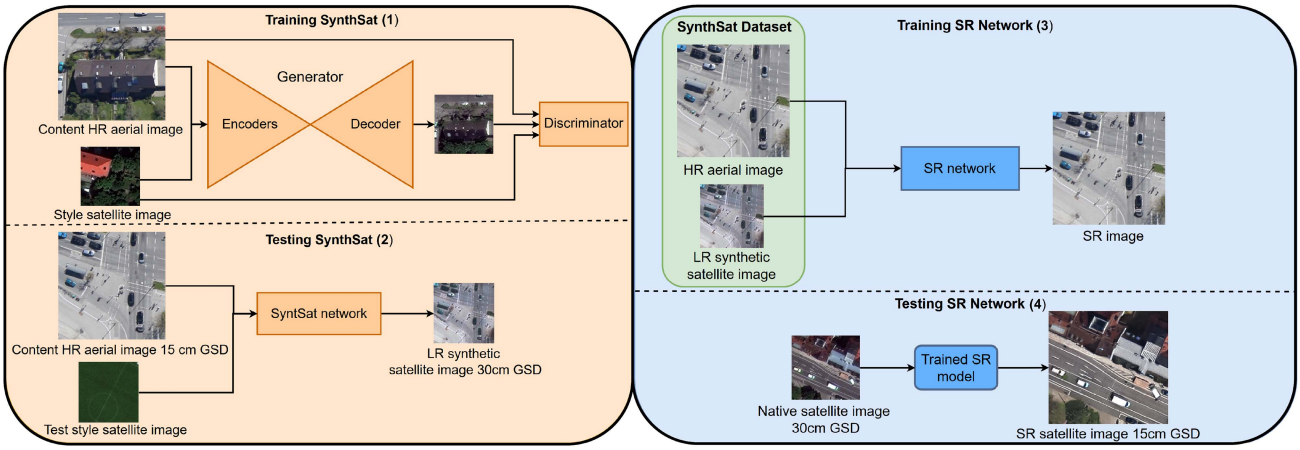

Fig. 3. Overview of the I2I-SR pipeline, which involves training SynthSat-I2I to generate the SynthSat dataset, followed by training an SR algorithm, and finally testing it on real 30 cm GSD satellite images.

resolution commercially available (30 cm GSD). This is crucial for tasks requiring finer details, such as those in traffic monitoring, autonomous driving, and urban planning.

5) *Application-based analysis:* This article showcases the practical effectiveness of enhanced 15 cm GSD satellite images in tasks, such as lane prediction and road segmentation, outperforming native 30 cm GSD images. This underscores the tangible benefits of I2I-SR pipeline for applications requiring sub-30 cm GSD resolution and pristine satellite imagery.

The rest of this article is organized as follows. Section II describes related work regarding I2I translation and recent developments in the field. Section III provides details about various parts of the proposed SynthSat method, such as network modules and loss functions. Section IV presents experiments and discussions for the individual parts of the I2I-SR pipeline, i.e., i) creation of the SynthSat dataset using the SynthSat-I2I network and ii) training SR networks using the SynthSat dataset. Section V offers an application-based analysis of the enhanced satellite images on downstream applications, such as lane prediction and road segmentation. Section VI provides an in-depth analysis of the issues and limitations of the I2I-SR pipeline. Finally, Section VII concludes this article.

II. Related Work

The first I2I translation algorithm was proposed by Gatys et al. [36], [37] which performs in an iterative procedure. Several feedforward approaches [38], [39], [40] were introduced later to speed up the translation process. With the advancement in generative adversarial networks (GANs) [41], most of the I2I translation methods rely on GANs to model the distribution of the target domain. These methods are mainly categorized into the supervised and unsupervised methods. Supervised methods require aligned image pairs from the source and target domains, which can be difficult, expensive, and sometimes impossible to obtain. Unsupervised methods do not require paired datasets, but may require additional constraints, such as labels, text, or semantic maps. They assume a shared latent space between the source and target domains, allowing them to translate source images into the target domain by sampling from that space. Variational autoencoders (VAEs) are commonly used for I2I translation [42], [43], [44]. While VAEs are more stable during training than GAN-based methods, they often produce results with artifacts and structural distortions. Arbitrary style transfer [45], [46], [47], [48] is an I2I method that translates a content image into the domain of an arbitrary style image. These algorithms fall into two categories: nonphotorealistic rendering, which produces artistic-looking images, and photorealistic rendering, which synthesizes realistic images for tasks, such as day to night, summer to winter, and damaged to inpainted. Style transfer techniques originated texture synthesis and transfer methods [49], [50], [51] for nonphotorealistic rendering. Later Gatys et al. [37] demonstrated the ability of CNNs to independently process and manipulate content and style images, enabling iterative optimization for style transfer tasks. Huang et al. [52] proposed adaptive instance normalization that can compute affine parameters from the style input and use them for normalizing the content images and producing the translated output image. Most of the recent style transfer methods especially for photorealistic rendering are based on GANs. Methods, such as [45], [53], [54] provide outstanding arbitrary style transfer results. Anokhin et al. [55] proposed a multidomain I2I translation method that produces images with different style transformations and at higher resolutions using an integrated SR block.

Recent works on style transfer, such as [56], [57], [58], [59], [60], focus on employing different approaches or introducing new modules, loss functions, and evaluation metrics to enhance the quality of style transfer. For instance, Wang et al. [56] utilized diffusion models to achieve high-quality style transfer while preserving content information effectively. In their content-style disentangled framework, Wang et al. [56] employed a diffusion-based style removal module to eliminate style information from both the content and style images to extract domain-aligned content information. The authors then use a

diffusion-based style transfer module to disentangle style information from the style image and transfer it to the content image. Style disentanglement and transfer are facilitated by the style disentanglement loss, aligning the transfer mapping of the content from the content image to its stylization with that of the style image. AesPA-Net [57] focuses on improving the attention mechanism and introduces a novel metric to quantify the frequency of repetition of local patterns in style images. This metric is integrated into their style transfer framework to enhance style transfer. Meanwhile, Huang et al. [58] proposed a novel artistic style transfer framework using separate encoders for content and style features. They employ vector quantization and a specially designed style-guided attention module for codebook-based stylization. This allows users to balance style similarity, visual fidelity, and content preservation in style transfer results by fusing continuous and quantized stylized features before decoding. RAST [59] adopts an iterative approach to style transfer through multirestorations. The method utilizes multirestoration loss and style difference loss to control the content-style balance in stylized images. Wen et al. [60] proposed a photorealistic style transfer using a reversible residual network, which includes a channel refinement module and an unbiased linear transform module and perform style transfer in the feature space. All the previously discussed I2I methodologies have been developed for ground images, and their significance is subjective from an artistic point of view. Despite most methods aiming to maintain a good balance of content retention and stylization, they face challenges when performing realistic style transfer between satellite and aerial images, as they are not designed for such tasks.

Previous works, such as [61], [62], [63] have shown potential for satellite $\rightleftharpoons$ maps. Schenkel et al. [64] proposed a cycle-consistent adversarial domain adaptation method to transfer style for aerial $\rightleftharpoons$ satellite, aerial $\rightleftharpoons$ aerial images in the near-IR and RGB bands at low spatial resolutions. I2I have also been used for converting between optical and synthetic aperture radar (SAR) images. Works such as [65], [66], [67] have been employed to translate SAR images to optical images or vice versa, aiming to enhance understanding, facilitate visualization, and for downstream applications. Li et al. [67] utilized I2I translation to convert optical images into SAR images. The translated image, along with its corresponding reference image in the same domain, is then inputted into a change detection network to generate the change map. Reyes et al. [66] conducted SAR-to-optical I2I and analyzed the impact of the translated images in subsequent applications, such as binary road segmentation. Wang et al. [66] also performed SAR-to-optical I2I for scenarios where optical data are unavailable. In such cases, this translation aids in land cover visual recognition for untrained individuals. In this article, we focus on the translation for the optical data. We propose an arbitrary style transfer method for generating realistic satellite images from content aerial images and arbitrary satellite style images. It implicitly learns degradation from satellite imagery and applies it to aerial imagery, resulting in natural-looking LR synthetic satellite imagery that, along with aerial imagery as GT, is used to train SR algorithms to produce 15-cm enhanced images from native 30-cm satellite imagery.

III. Methodology

SynthSat-I2I network aims to apply the style of a satellite image while preserving the information of the content aerial image, addressing the distortions and visual artifacts often encountered in I2I style transfer algorithms. Despite differences in quality, area coverage, sensor calibration, and other aspects, aerial and satellite imagery describe the same physical objects and share a similar latent space structure. Translated images are sampled from this latent space, representing the local structures of the content image in the feature space of the satellite image. The SynthSat-I2I network, inspired by the works of [45], [68], [69], includes a generator network G, shown in Fig. 4, with an encoder–decoder structure to extract features at different scales. A single decoder fuses image features from two encoders at i different levels. A multiscale discriminator is used to discriminate between the generated synthetic satellite image and the real satellite image at three scales. Let x_c be the content image, x_s the style image, and x_t the translated synthetic satellite image. Images x_c and x_s are fed to content encoder (CE) and style encoder (SE), respectively. The content images are downsampled using bicubic interpolation to the spatial resolution of the style images.

A. Content and Style Encoders

The CE and SE share the same architecture. The encoder extracts high- and low-level features from the images and projects them into their latent space. At the input of the encoder residual block (erb), a downsampling layer (down) with bicubic interpolation scales down the features. The encoder residual block, shown in Fig. 5, consists of three modules, each with convolution–normalization–LReLU layers and dense connections to improve information flow. The modules are followed by a concatenation layer and a convolution layer. In addition, a skip connection with convolution–normalization–LRelu layer preserve important features and promote local residual learning of meaningful information. The residual layers use a combination of spectral-instance normalization, which has been shown in previous works [68], [70], [71] to contribute to overall stability in training GANs. The outputs for each encoder are given as $\text{CE} = \{c\_\text{erb}_1, c\_\text{erb}_2, c\_\text{erb}_3, \ldots, c\_\text{erb}_i\}$ and $\text{SE} = \{s\_\text{erb}_1, s\_\text{erb}_2, s\_\text{erb}_3, \ldots, s\_\text{erb}_i\}$. Thus, the output of a encoder residual block can be generalized as the following:

$$\text{erb}_1 = \text{conv}_{3\times3}\left(\text{erb}_i\right) \text{for } i = 1 \tag{1}$$

where $\text{conv}_{3\times3}$ is convolution used for feature expansion and

$$\text{erb}_i = \text{down}_i\left(\text{erb}_{i-1}\right); \text{for } i \neq 1. \tag{2}$$

B. Feature Decoder

The decoder fuses content and style features at different levels to create synthetic satellite images. The structure of the decoder mirrors the encoder network. It fuses style and reconstructs features from the ith to the first level. For optimal style transfer, we use the attentional manifold alignment (AMA) block [69]. It consists of an attention and a space-aware interpolation module to fuse and transfer the style using the content and style features

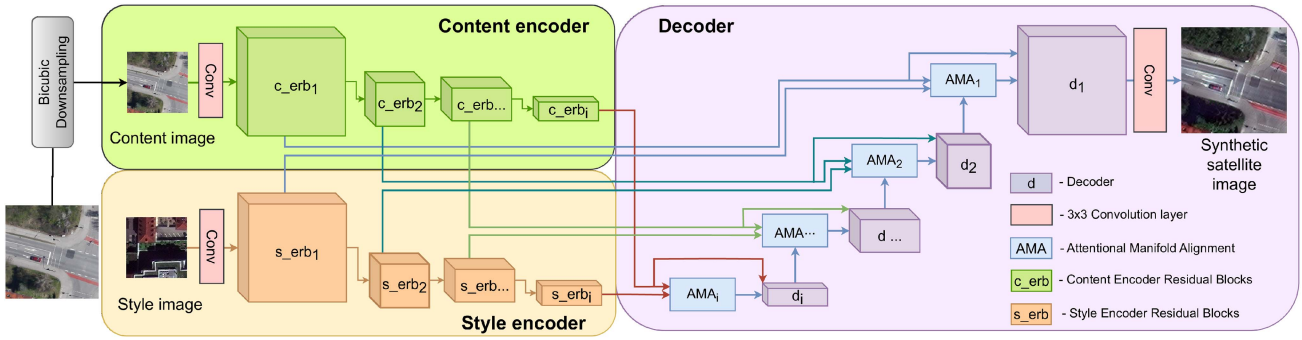

Fig. 4. Generator of SynthSat-I2I. The CE and SE share the same design and extract features at different scales. The features from the encoder are fused in the decoder network using the AMA block [69] and decoder blocks.

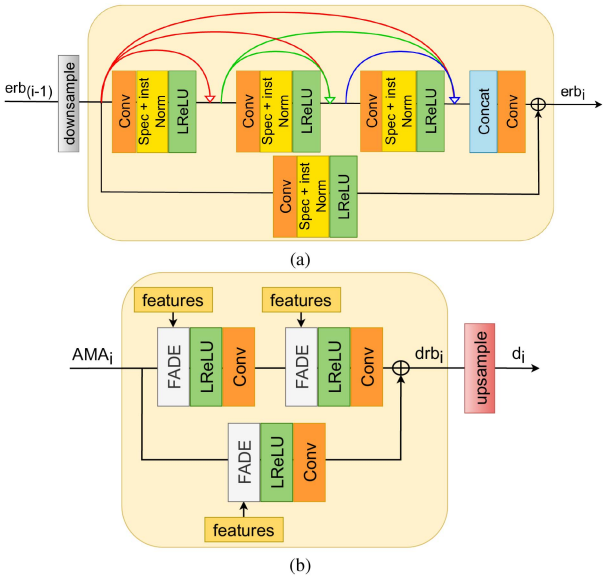

Fig. 5. (a) CE and SE use the same encoder residual block design. (b) Input to the decoder block is the output from the AMA block [69] and the CE at each level.

from the encoders at each level. The output from the AMA block is input to the decoder block (d) consisting of a decoder residual block (drb), which has a similar structure to the FADE residual block [45]. The decoder residual block (shown in Fig. 5) consists of two modules, each consisting of feature adaptive denormalization (FADE) layer [45] followed by LeakyReLU activation and a convolution block. A skip connection consisting of a decoder residual module is added for the residual learning. FADE uses multiscale feature representation from the CEs. This layer provides additional control in preserving information present in the content image. An upsampling layer (up) with bicubic interpolation is present at the end of the decoder residual block. $\text{Dec} = \{d_i, d_{i-1}, d_{i-2}, \ldots, d_1\}$ are the output features of the decoder after upsampling layer and the output at each level is given as

$$d_1 = \text{conv}_{3\times3}\left(\text{drb}_1\left(\text{AMA}_1 + d_2\right)\right)$$
$$d_6 = \text{up}_6\left(\text{drb}_6\left(\text{AMA}_6\right)\right)$$
$$d_i = \text{up}_i\left(\text{drb}_i\left(\text{AMA}_i + d_{i+1}\right)\right) \text{ for } i = \{2,3,4,5\} \quad (3)$$

where $\text{conv}_{3\times3}$ is used for condensing the features to produce RGB output. AMA_i is AMA block at ith level.

C. Discriminator

Multiscale discriminator [68], [72] consist of three discriminators (D_1, D_2, D_3) with identical network architectures. It compares x_t and x_s at three different scales created by downsampling x_t and x_s by factors of two and four. The discriminators operate at different scales, each with its own receptive field, allowing coarse-to-fine generator training and ensuring consistent style.

D. Loss Functions

Various loss functions are employed to ensure consistent style transfer and maximum content preservation. The output of the generator is $x_t = G(x_c, x_s)$. The generator and multiscale discriminators are trained alternately using a hinge-based adversarial loss [70], [73]. The adversarial loss is given as

$$\mathcal{L}_{\text{Adv}-G} = -\mathbb{E}[D(x_t)] \quad (4)$$
$$\mathcal{L}_{\text{Adv}-D} = -\mathbb{E}\left[\min\left(0, -1 + D\left(x_s\right)\right)\right] - \mathbb{E}[\min(0, -1 - D(x_t))]. \quad (5)$$

To improve the GAN loss and ensure stable training, a feature matching loss is used [72], which is computed by matching extracted features from x_t and x_s at different scales in the discriminator. It promotes the similarity between x_t and x_s in terms of their natural appearance. The feature loss is described as

$$\mathcal{L}_{\text{fm}}\left(x_t, x_s\right) = \mathbb{E}\sum_{j=1}^{T}\frac{1}{N_j}\left[\left\|D_k^{(j)}(x_t) - D_k^{(j)}(x_s)\right\|_1\right] \quad (6)$$

where D_k^j denotes the features from the jth layer of the D_k discriminator with T number of layers and N_j denotes the number of features in each layer.

Kornia [74] is an open-source library that implements differentiable CV functions for DL models. We use its edge loss for content preservation by comparing the canny edge detection

Algorithm 1: Training SynthSat-I2I.

procedure INITIALIZATION
    Initialize the parameters of the generator G and discriminators D using Xavier initialization.
end procedure
Total number of steps: T.
procedure TRAINING LOOP (T)
    for $t = 1$ to T **do**
        Generator Update:
        (i) Sample a mini-batch of content images $\{x_c^{(1)}, x_c^{(2)}, \ldots, x_c^{(N)}\}$ and style images $\{x_s^{(1)}, x_s^{(2)}, \ldots, x_s^{(N)}\}$.
        (ii) Using bicubic interpolation, downsample content images to $128 \times 128 \times 3$ to match the spatial resolution of style images.
        (iii) Generate synthetic satellite images $\{x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(N)}\}$ using x_c and x_s with the generator: $x_t^{(i)} = G(x_c^{(i)}, x_s^{(i)})$.
        (vi) Compute adversarial loss: $\mathcal{L}_{Adv-G} = -\mathbb{E}[D(x_t)]$.
        (v) Compute feature matching loss: $\mathcal{L}_{\text{fm}}(x_t, x_s)$.
        (vi) Compute edge loss: $\mathcal{L}_{\text{edge}}(x_t, x_c)$.
        (vii) Compute style loss using REMD: $\mathcal{L}_{remd}(x_t, x_s)$.
        (viii) Compute the overall generator loss:

$$\begin{aligned}\mathcal{L}_G = \mathcal{L}_{Adv-G} &+ \lambda_{fm}\mathcal{L}_{\text{fm}}(x_t, x_s) \\ &+ \lambda_{edge}\mathcal{L}_{\text{edge}}(x_t, x_c) \\ &+ \lambda_{remd}\mathcal{L}_{\text{remd}}(x_t, x_s)\end{aligned}$$

        (ix) Update the generator parameters using gradient descent.
    Discriminator Update:
        (i) Using the mini-batch of real satellite images $\{x_s^{(1)}, x_s^{(2)}, \ldots, x_s^{(N)}\}$ and generated synthetic images $\{x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(N)}\}$.
        (ii) Compute adversarial loss for the discriminators:

$$\begin{aligned}\mathcal{L}_{Adv-D} = &-\mathbb{E}\left[\min\left(0, -1 + D(x_s)\right)\right] \\ &-\mathbb{E}\left[\min(0, -1 - D(x_t))\right]\end{aligned}$$

        (iii) Update the discriminator parameters using gradient descent.
    end for
end procedure

filter output for x_c and x_t using $L1$

$$\mathcal{L}_{\text{edge}}(x_t, x_c) = \left\|\text{CannyEdge}(x_t) - \text{CannyEdge}(x_c)\right\|_1. \quad (7)$$

Following [69], [75], [76], relaxed Earth mover distance (REMD) is adapted as the style loss between x_t and x_s, guaranteeing the style similarity of x_t and x_s

$$L_{\text{REMD}} = \max\left(\frac{1}{H_s W_s}\sum_i \min_j C_{ij}, \frac{1}{H_c W_c}\sum_j \min_i C_{ij}\right) \quad (8)$$

where the pairwise cosine distance matrix C_{ij} measures the dissimilarity between high-level features in x_t and x_s. H_s, W_s and H_c, W_c are height and width of style and content images, respectively. The overall loss function is given as

$$\begin{aligned}\min_G \max_D \mathcal{L}\left(G(x_c, x_s), D(x_t, x_c, x_s)\right) = \mathcal{L}_{\text{adv}} + \lambda_{\text{fm}}\mathcal{L}_{\text{fm}} \\ (x_t, x_s) + \lambda_{\text{edge}}\mathcal{L}_{\text{edge}}(x_t, x_c) + \lambda_{\text{REMD}}\mathcal{L}_{\text{REMD}}(x_t, x_s) \quad (9)\end{aligned}$$

where $\lambda_{\text{fm}}, \lambda_{\text{edge}}$, and λ_{REMD} are the hyperparameters to define the importance of their respective loss functions.

IV. Experiment and Discussions

We train the SynthSat-I2I network to create the SynthSat dataset for training the SR algorithms. Using this dataset, we then train SR algorithms and evaluate them for enhancing the GSD of satellite images from 30 to 15 cm. We use 2× NVIDIA Titan RTX GPUs with 24 GB of VRAM for the experiments.

A. SynthSat-I2I : Creation of SynthSat Dataset

The aerial image dataset consists of 16 nonoverlapping RGB aerial images of 4866×3244 pixels at 15 cm GSD from the SkyScapes dataset [3] over the city of Munich. We tile the images into 98 674 patches of 256×256 pixels with an overlap of 80%. We use these image patches as content input images for Synthsat-I2I and also as GT for training the SR algorithms. For our satellite image dataset, we use a single WV4 pan-sharpened RGB image of $45\,386 \times 33\,753$ pixels at 30-cm GSD over the main city of Munich. We take half of the image for training SynthSat-I2I as the style input, and the rest for testing the SynthSat-I2I and the SR algorithms. We tile the train and test satellite image parts into patches of 128×128 pixels with overlap of 60%, resulting in 98 674 image patches for each of the training and test sets. We input the aerial and test satellite image patches to SynthSat-I2I resulting in 98 674 photorealistic synthetic satellite images at 30 cm GSD. The resulting image patches and their paired aerial images at 15 cm form the SynthSat dataset, which we use for training SR algorithms. We then test the SR algorithm using the test satellite image patches. Since the aerial and satellite images in our dataset are from the same city, they overlap at few points. We specifically isolated these areas to draw a comparative analysis between the original aerial (15 cm), satellite (30 cm) images, the synthetic satellite (30 cm) image derived through SynthSat-I2I, and the SR-enhanced image (15 cm) generated by SR networks trained on the SynthSat dataset. This methodical comparison underlines the potential of our I2I-SR pipeline, a demonstration of which can be observed in Fig. 6.

For training the SynthSat-I2I network, we downsample the aerial image patches using bicubic interpolation to 128×128

Fig. 6. Comparison of an aerial image, its SynthSat-I2I synthetic counterpart, the original satellite, and its SR variants.

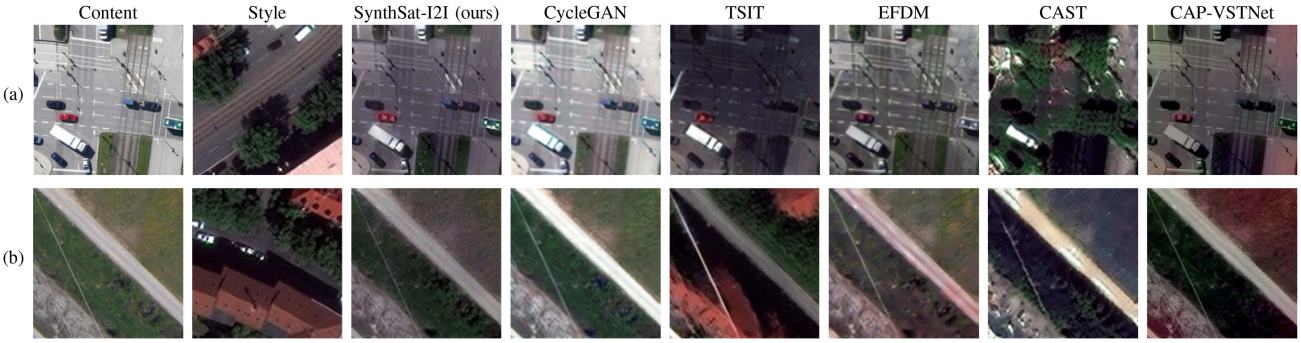

Fig. 7. Figure compares our novel style transfer method with others, in generating synthetic satellite images from aerial and satellite images.

pixels and feed them to the CE. The input to the SE are randomly sampled images from the training set of the satellite image patches We initialize the weights using Xavier initialization. Moreover, we employ ADAM optimizer [77] with $\beta_1 = 0.1$ and $\beta_2 = 0.999$, and with the learning rates of 0.0002 and 0.0004 for the generator and discriminator, respectively. Furthermore, we apply a combination of spectral [70] and instance [78] normalization to all layers of the generator and discriminator networks. We set the losses hyperparameters as $\lambda_{\text{fm}} = 0.6$, $\lambda_{\text{edge}} = 1$, $\lambda_{\text{REMD}} = 5$. We visually select the iteration that produces the most natural looking satellite images with maximum content preservation as the best performing method.

Fig. 7 shows examples of images translated from aerial to satellite domain using our SynthSat-I2I and other I2I networks, such as SOTA CycleGAN [46], TSIT [45], and latest arbitrary style I2I methods, such as EFDM [47], CAST [48], and CAP-VSTNet [60]. As figure shows, SynthSat-I2I outperforms the other methods in transferring the satellite style while preserving the image content. Most arbitrary style transfer I2I methods, designed primarily for artistic style transfer, occasionally struggle with accurate style replication and content preservation. This can be observed with CycleGAN, TSIT, and style transfer from EFDM and CAP-VSTNet and content retention from CAST in Fig. 7. SynthSat-I2I does more than just apply the style of satellite images to aerial ones. During training, the network implicitly learns the degradation model from the satellite images, which is then applied to the aerial images during testing. We quantitatively assess I2I methods by comparing synthetic and actual satellite images using nonreference metrics, such as FID [79] and LPIPS [80]. We evaluate content preservation via PSNR and SSIM comparisons of synthetic satellite and their corresponding aerial images. As per Table I, our SynthSat-I2I outperforms in both style transfer and content preservation. We also evaluate the parameters and inference times for various I2I methods, as presented in Table II. The table indicates that our method exhibits the highest number of parameters and the longest inference time compared to other methods. Despite these metrics, our approach produces superior synthetic images when compared to alternative methods. Although our inference time is relatively higher than that of other methods, it still demonstrates impressive speed, processing 18 patches per second. This slight increase in inference time serves as a deliberate tradeoff to

TABLE I
COMPARISON OF SYNTHETIC SATELLITE IMAGES GENERATED BY SYNTHSAT-I2I AND OTHER I2I METHODS

	FID ↓	LPIPS ↓	PSNR ↑	SSIM ↑
CycleGAN	38.22	0.6846	27.90	**0.88**
TSIT	29.06	0.6578	27.87	0.62
EFDM	51.90	0.6367	28.56	0.81
CAST	78.48	0.6368	27.93	0.46
CAP-VSTNet	108.23	0.75	27.86	0.68
SynthSat-I2I (ours)	**19.23**	**0.6296**	**28.63**	0.70

The bold values indicate the best scores for the respective quantitative metrics.

TABLE II
COMPARISON OF NETWORK PARAMETERS AND INFERENCE TIME FOR SYNTHSAT-I2I AND OTHER I2I METHODS

I2I networks	Parameters (million)	Inference time (in milli seconds)
CycleGAN	11	20.67
TSIT	398	22.83
EFDM	3.5	4.88
CAST	3.5	9.89
CAP-VSTNet	4	46.15
SynthSat-I2I (ours)	538	54.84

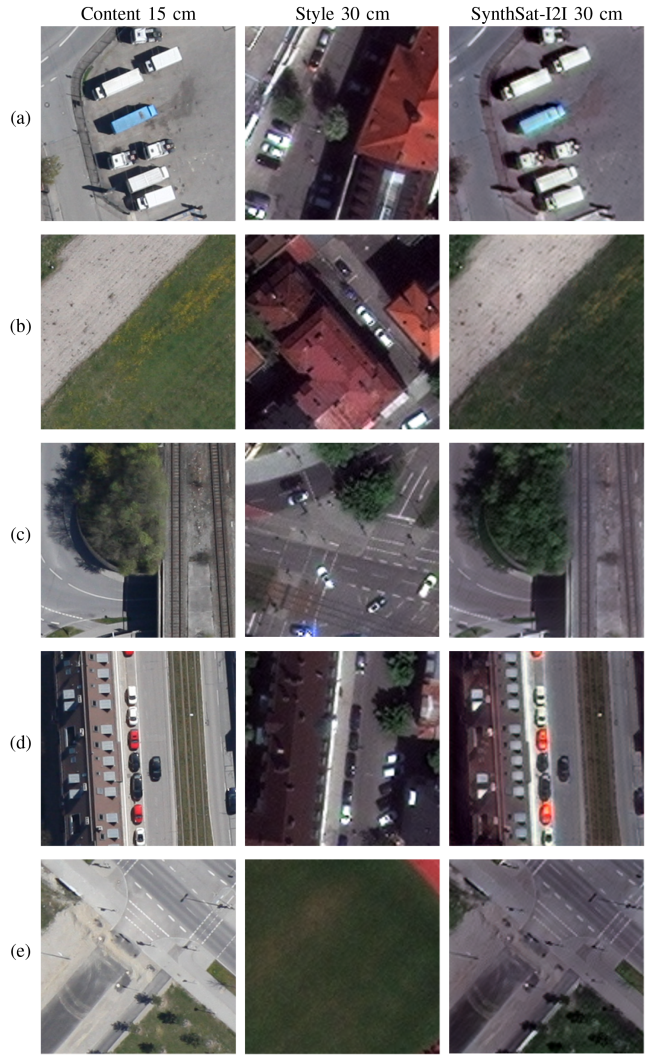

Fig. 8. Synthetic satellite image at 30 cm GSD generated by the SynthSat-I2I network using the content aerial image at 15 cm GSD and style satellite image at 30 cm GSD as the input to the network.

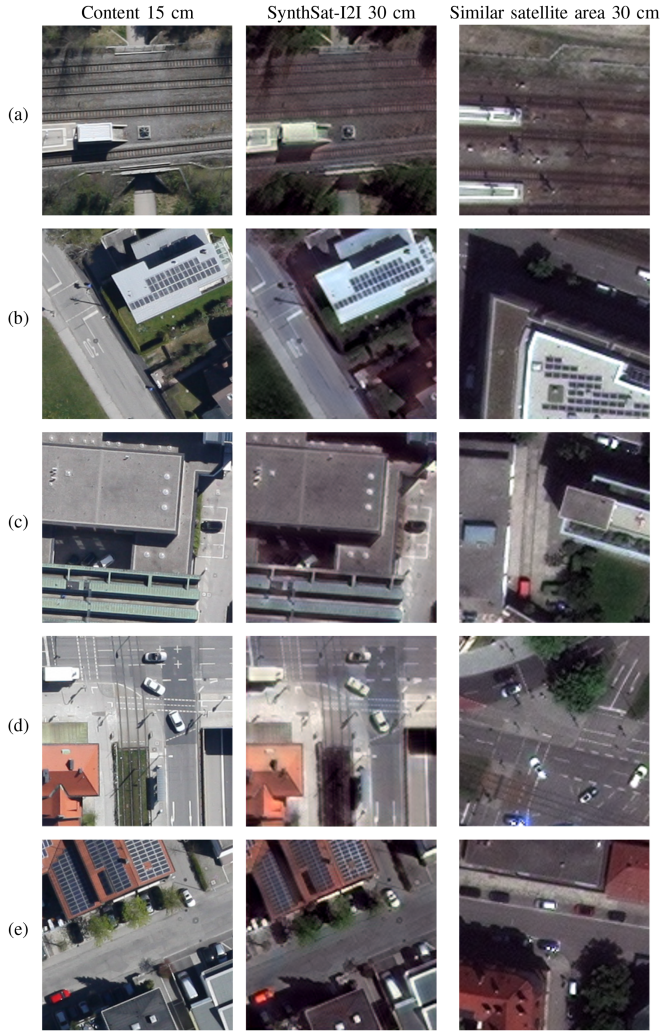

Fig. 9. Comparing synthetic satellite images from SynthSat-I2I and real satellite images at 30 cm GSD over a similar area.

generate synthetic images that preserve most content and achieve realistic style transfer.

Figs. 8 and 9 present additional examples of synthetic satellite images generated by the SynthSat-I2I network. During training, the SynthSat network effectively learns the general characteristics of satellite images, which enables it to produce natural-looking synthetic images even when presented with input-style images from significantly different geographic areas. Notably, this capability is demonstrated by the convincing results shown in images (c) and (e) of Fig. 8.

The synthetic satellite images generated by the SynthSat-I2I network exhibit degradation and style characteristics that are typical of satellite imagery, while effectively preserving the content of the original aerial images. This can be observed by comparing the synthetic satellite images at 30 cm GSD with the original 30 cm GSD satellite images in Fig. 9. The comparison focuses on similar geographical areas, including (a) train tracks, (b) solar panels on rooftops, (d) road intersections, and residential areas in (e). It is important to note that there is a one-to-one correspondence between the resulting synthetic satellite images at 30 cm GSD and the high-resolution aerial images at 15 cm GSD. Consequently, we can construct a paired dataset known as the SynthSat dataset, which is valuable for training various image enhancement tasks, such as image singe image SR.

Even though SynthSat-I2I exhibits robust translation capabilities, it may still encounter challenges with significant geographic and seasonal variations in images, angle of capture, and object distribution. These factors can hinder color translation learning and affect the quality of synthesized images.

1) Ablation Studies: We evaluate the impact from each module on the generator network of the SynthSat-I2I network using FID, LPIPS, and SSIM scores. In Table III, FID and LPIPS scores compare the style images with the generated images, while SSIM is used to check content retention. Adding each module improves overall performance. The baseline generator is a simple encoder–decoder network without AMA, FADE layer (replaced by a convolution layer), REMD, and edge losses. Initially, using AdaIN for style transfer yields unsatisfactory results [see Fig. 10(c)]. To address this, we replace AdaIN with the

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

Fig. 10. Effects of various components added to the baseline generator of SynthSat-I2I. Content (a) and style (b) images, AdaIn (c), AMA (d), $d+$ FADE (e), $d+e+\mathcal{L}_{\text{REMD}}$ (f), $d+e+f+\mathcal{L}_{\text{edge}}$ (g).

TABLE III
ABLATION STUDY ON VARIOUS COMPONENTS OF THE SYNTHSAT-I2I NETWORK

AdaIN	AMA	FADE	$\mathcal{L}_{\text{REMD}}$	$\mathcal{L}_{\text{edge}}$	FID ↓	LPIPS ↓	SSIM ↑
✓					0.6548	37.31	0.6223
×	✓				0.6500	37.01	0.7061
×	✓	✓			0.6395	33.61	0.7895
×	✓	✓	✓		0.6365	30.41	0.7948
×	✓	✓	✓	✓	0.6308	30.06	0.8194

TABLE IV
COMPARISON OF ENHANCED SATELLITE IMAGES AT 15 CM GSD USING VARIOUS SR METHODS TRAINED ON THE SYNTHSAT DATASET

Metrics	ESRGAN	BSRGAN	SwinIR	Real-ESRGAN
FID ↓	84.80	106.28	97.95	**82.90**
LPIPS ↓	0.6828	0.6829	0.6845	**0.6728**

The bold values indicate the best scores for the respective quantitative metrics.

AMA block, aligning feature maps in a common space, resulting in significant visual quality improvement [see Fig. 10(d)]. The addition of the FADE layer enables better adaptation to input image style and produces more natural-looking synthesized images, as seen in the results [see Fig. 10(e)]. Introducing the REMD loss ($\mathcal{L}_{\text{REMD}}$), an REMD loss, further improves style by comparing gram matrices of style and synthesized feature maps [see Fig. 10(f)]. To preserve edge information, we apply the edge loss ($\mathcal{L}_{\text{edge}}$), ensuring the network retains the structural details from the input image. Results demonstrate the importance of ($\mathcal{L}_{\text{edge}}$) in preserving important structural information [see Fig. 10(g)].

B. Image SR

To determine the best performing method in the selected SR algorithms for our I2I-SR pipeline, we evaluated four GAN-based SR methods. These included ESRGAN [16], and others incorporating real-world degradation modeling, such as Real-ESRGAN [17], BSRGAN [33], and SwinIR [34]. Each SR algorithm was freshly trained on the SynthSat dataset under comparable degradation conditions. We allocated 10% of the SynthSat dataset for validation, utilizing PSNR and SSIM metrics for selection of testing iterations. The trained SR models were then applied to real satellite image patches. With no available GT for test images at 15 cm GSD, visual comparison (see Fig. 11) was used to assess the SR methods. ESRGAN, lacking a degradation modeling component during training, encounters challenges in mitigating noise and pixel distortions prevalent in satellite images. Further exacerbated by the domain discrepancies between GT aerial and LR synthetic satellite images, the produced SR images display imperfect object boundaries. Notably, the lane markings in Fig. 11(a) appear fuzzy, and a noticeable color bleeding effect can be seen on the building rooftops in Fig. 11(b). Conversely, other SR methods that utilize degradation modeling can simulate multiple image degradations during training. Even minor pixel displacements between GT and LR images in the SynthSat dataset have minimal impact on the SR images, contributing to superior SR results. On comparing the outputs from Real-ESRGAN, BSRGAN, and SwinIR, it is apparent that while BSRGAN and SwinIR produce similar image quality with smooth textures and a somewhat blurred appearance, Real-ESRGAN generates images that are strikingly vibrant, sharp, and superior in texture detail. For instance, Real-ESRGAN effectively reconstructs elements, such as the slightly faded lane marking in the original satellite image [refer to Fig. 11(a)], or the area adjacent to the shadow of the building in the bottom left [as shown in Fig. 11(b)]. This superior reconstruction by Real-ESRGAN holds consistent, as illustrated in example Fig. 6(b). Fig. 12 showcases additional samples from the city of Munich, that compare the 15-cm SR results obtained from the SR algorithms. It is evident from the samples that our I2I-SR pipeline produces images with superior overall image quality, characterized by brightness, sharpness, and improved texture information. We also conducted quantitative evaluations using nonreferenced metrics FID and LPIPS, comparing the SR images at 15 cm GSD and aerial images at the same GSD. Results, as presented in Table IV, indicated that images from Real-ESRGAN have comparable quality to aerial images.

1) Testing on More Cities: We conducted additional inference tests using the trained SR models on the pan-sharpened RGB satellite image with a 30 cm GSD captured over the city of Berlin. We then compared the performance of the SR images at 15-cm resolution obtained from our pipeline with that of other SR methods. Notably, despite both Munich and Berlin satellite images being captured by the same satellite (WV4) at the same

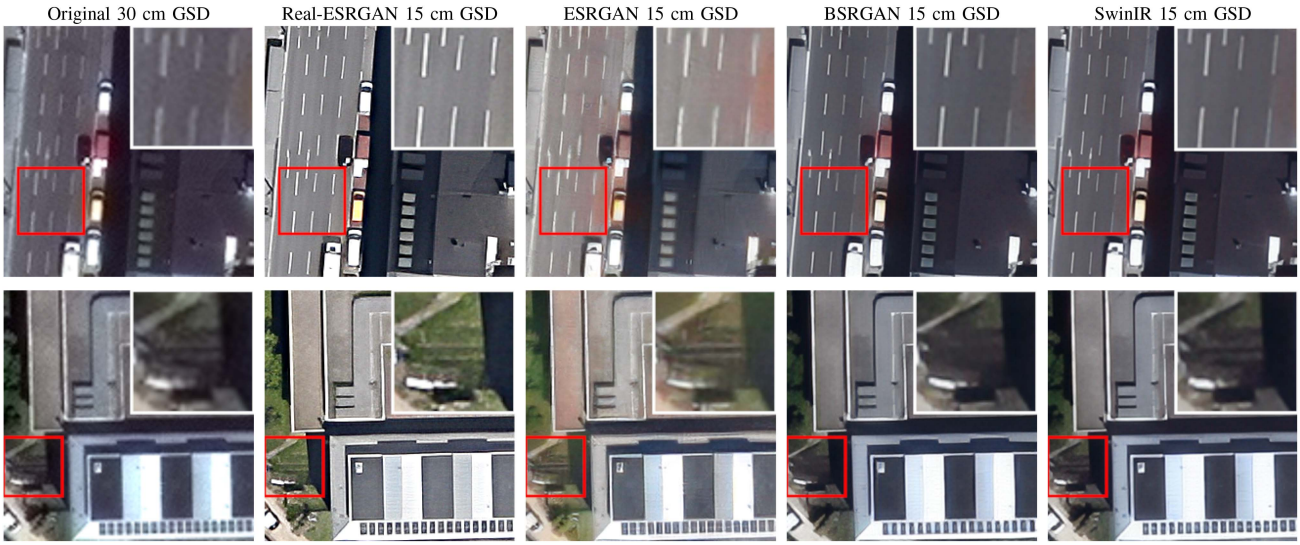

Fig. 11. Figure showcases the effectiveness of various SR algorithms, trained on SynthSat, in enhancing satellite image resolution from 30 to 15 cm GSD.

resolution (30 cm GSD), they were processed differently, resulting in varying visual qualities. To emphasize this point, Fig. 13 presents a comparison between an aerial image of Munich and the WV4 satellite images taken over Munich and Berlin. The presented Fig. 14 illustrates that the SR images obtained from our pipeline outperform the other methods, even though the SynthSat dataset was generated based on the style of the Munich satellite image. The SR images are sharper, contain more details, and exhibit fewer artifacts compared to their counterparts. The similarities between the city of Munich and Berlin, including architectural styles, road networks, and house layouts, contribute to the ability of the SR model to generalize well on the satellite image of Berlin.

Furthermore, we have conducted additional testing using the SR models on the images obtained from the Airbus Pléiades Neo satellite, which features a GSD of 30 cm. The RGB images are made available by Airbus Intelligence website on their satellite image gallery web-page [81], but their usage is restricted to personal purposes only. To carry out the evaluation, we randomly selected images from various cities and enhanced them from their native 30 cm GSD to 15 cm GSD. As there is no available 15 cm GT data, we relied on qualitative comparisons for assessment. The cities subjected to SR evaluation include Clisson-France (see Fig. 15), Pisa-Italy (see Fig. 16), Pittsburg-USA (see Fig. 17), and Detroid-USA (see Fig. 18). Examining Figs. 15–18, it becomes evident that the images produced by the SR models surpass the quality of simple bicubic interpolation operation by a significant margin. Even though the SR models are trained on a dataset comprising nadir view images, captured by a satellite sensor looking directly downward, they perform remarkably well even for oblique images, as demonstrated in Figs. 16–18. Among the SR models, our I2I-SR pipeline featuring Real-ESRGAN stands out with superior image quality, boasting brightness, sharpness, and proper grid reconstruction from the original image. On the other hand, methods such as BSRGAN and SwinIR tend to produce oversmoothed images that lack texture and high-frequency information, resulting in a blurry appearance and loss of critical details.

Notably, the Real-ESRGAN model excels in enhancing various image examples, such as the solar-roof (b) and road-crossing (c) in Fig. 15, the textures on the dome (a) and building structures (c) in Fig. 16, and the grid pattern on the building (b) in Fig. 17, as well as the highway (a) and building (d) in Fig. 18, showcasing superior texture preservation and image enhancement compared to alternative methods.

2) Comparison With Industry Products: Organizations such as Maxar, EUSI, and Airbus are leaders in the field of commercial satellite imagery. They offer high-resolution imagery with a GSD as small as 15 cm, based on their proprietary HD technologies. Using our pipeline, we believe that our method can achieve comparable results to the HD products. Figs. 19 and 20 present additional samples that compare the SR images generated by our proposed pipeline with HD product from EUSI, both of which have a GSD of 15 cm and were derived from the same WV4 image with a 30 cm GSD. SR images produced by our pipeline exhibit image quality that is comparable to that of image from EUSI. Furthermore, in certain instances, our images demonstrate superior sharpness and better reconstruction, such as in (a), (b), and (d) in Fig. 19(c) and (d) in Fig. 20.

However, it is crucial to acknowledge its inherent flaw, namely the occasional color bleeding problem. This issue occurs in some dark regions/patterns of the images, where green tints replace black colors. Given the restricted volume and diversity of the SynthSat-I2I training dataset, it may not be able to encapsulate the nuanced degradation patterns of the satellite images fully. This limitation also impacts the SR network during training, potentially impeding its capacity to accurately model the complex degradation characteristics present in the test satellite images at the original 30 cm GSD. This insufficiency can manifest as anomalies during testing, such as the color bleeding issue evident in ESRGAN [refer to Fig. 11(b)]. To mitigate these challenges, a comprehensive dataset of aerial and satellite images from diverse

Fig. 12. Figure showcases the effectiveness of Real-ESRGAN in our I2I-SR pipeline and other SR algorithms, trained on SynthSat, in enhancing satellite image resolution from 30 to 15 cm GSD over the city of Munich.

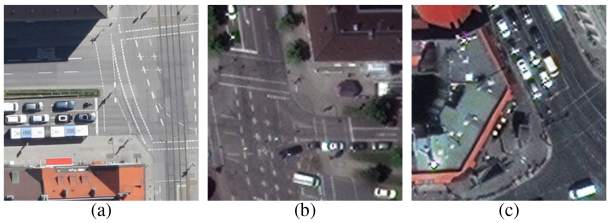

Fig. 13. (a) Aerial image over Munich at 15 cm GSD, WV4 satellite images at 30 cm GSD over (b) Munich and (c) Berlin.

regions and seasons could be used to enhance the SynthSat dataset. Moreover, incorporating advanced degradation modeling for the training of the SR algorithm within the I2I-SR pipeline might offer substantial improvements. More in-depth analysis of the color bleeding issues are discussed section VI and propose preventive measures to mitigate it effectively.

V. Application Based Analysis

Due to the absence of GT data for the enhanced 15-cm images, accurately evaluating the quantitative performance of the SR images and their accuracy through qualitative measures poses a significant challenge. To demonstrate the practical effectiveness of the SR images, we have employed two high-level application tasks: lane prediction from [3] and road segmentation from [82]. Remarkably, the enhanced 15-cm images consistently outperform the native 30-cm images in these tasks without requiring retraining of the respective networks. This compelling result emphasizes the considerable benefits of using SR satellite images at 15 cm GSD, derived from their native 30 cm GSD resolution. The enhanced images prove highly advantageous for downstream applications that excel with sub-30 cm GSD resolution or demand noise-free, sharp, and pristine satellite imagery.

A. Lane Prediction

In this study, we are employing the lane detection algorithm proposed in paper "SkyScapes: Fine-Grained Semantic Understanding of Aerial Scenes" [3], without any modifications to its default settings. Our primary objective is to evaluate the performance of the algorithm on SR images, which were enhanced from their native resolution of 30 cm GSD to 15 cm GSD. Importantly, we are conducting this evaluation without retraining the original network used in this article. The authors of "SkyScapes" trained their lane detection network on a comprehensive dataset, known as the SkyScapes dataset, which includes high-resolution aerial images at 13 cm GSD. The dataset covers various categories relevant to aerial scenes, such as urban, suburban, and rural areas. The network was trained to identify road lanes within these aerial scenes and subsequently segment them into 12 distinct categories of lane markings. These categories encompass long line (LL), dash line, tiny dash line (TDL), zebra zone, turn sign (TS), stop line, other signs, the rest of lane-markings, parking zone (PZ), no PZ, crosswalk, and plus sign. For testing the lane detection algorithm, we utilized a WV4 satellite image captured over the city of Munich, which was enhanced from its native 30 cm GSD to 15 cm GSD using various methods. For fair and consistent comparisons, we are testing the enhanced images using the checkpoints provided by the authors, ensuring that the same trained model used in their study is applied to our enhanced imagery. Our aim is to demonstrate that these enhanced satellite images perform effectively with algorithms designed for lower GSD resolutions. We want to showcase that the finer resolution SR images maintain their compatibility with and yield satisfactory results when processed by algorithms tailored for low GSD aerial imagery. This assessment will affirm the practical viability of utilizing very high resolution SR satellite images (15 cm GSD) for tasks that traditionally require lower GSD data, potentially opening new opportunities for high-resolution remote sensing applications.

We conducted a comparison between the enhanced satellite image at 15 cm GSD obtained through bicubic interpolation and 15 cm GSD SR images generated using Real-ESRGAN, BSRGAN, and SwinIR. In Fig. 21, we present the enhanced images alongside their respective lane segmentation masks for each sample. Notably, the 15-cm bicubic interpolated image shows minimal to no lane markings detected. However, upon comparing the lane segmentation from the SR methods, it becomes evident that the Real-ESRGAN, as part of the I2I-SR pipeline, consistently detects more lane markings compared to the other methods. Through visual inspection, it is apparent that Real-ESRGAN excels in identifying more LLs and TDLs, with moderate performance on TSs when compared to BSRGAN and SwinIR. This outcome aligns with expectations, given that the SR images produced by Real-ESRGAN are sharp and faithfully reconstruct all the information from the original image without losing any details. Despite the absence of GT prediction masks, the visual evidence clearly indicates that enhancing native 30 cm GSD satellite imagery to a finer resolution significantly improves the detection of features that are otherwise undetectable at the native resolution, particularly benefiting tasks that require lower GSD data.

B. Road Segmentation

We conducted an additional experiment to verify the semantic consistency of the SR images in comparison to the original images. For this purpose, we employed a Dense-U-Net-121 model [82], previously trained on the DeepGlobe18 road segmentation dataset [83], which comprises over 6000 images with a 50 cm GSD from Southeast Asia. The selected network architecture is particularly effective at capturing fine-grained details in images, making it well-suited for detecting roads of varying widths, ranging from broad four-lane motorways to narrow streets. Since the SR images have a finer GSD of 15 cm, we downsampled both the SR images and the original 30 cm GSD image to 50 cm GSD using bicubic interpolation before feeding them to the model. While this downsampling step prevents us from directly assessing the suitability of SR images for HD road mapping applications, the downsampled images still retain the artifacts introduced by the various enhancement methods. This enables us to determine whether the information in the SR images is preserved or enhanced at a high level. To ensure

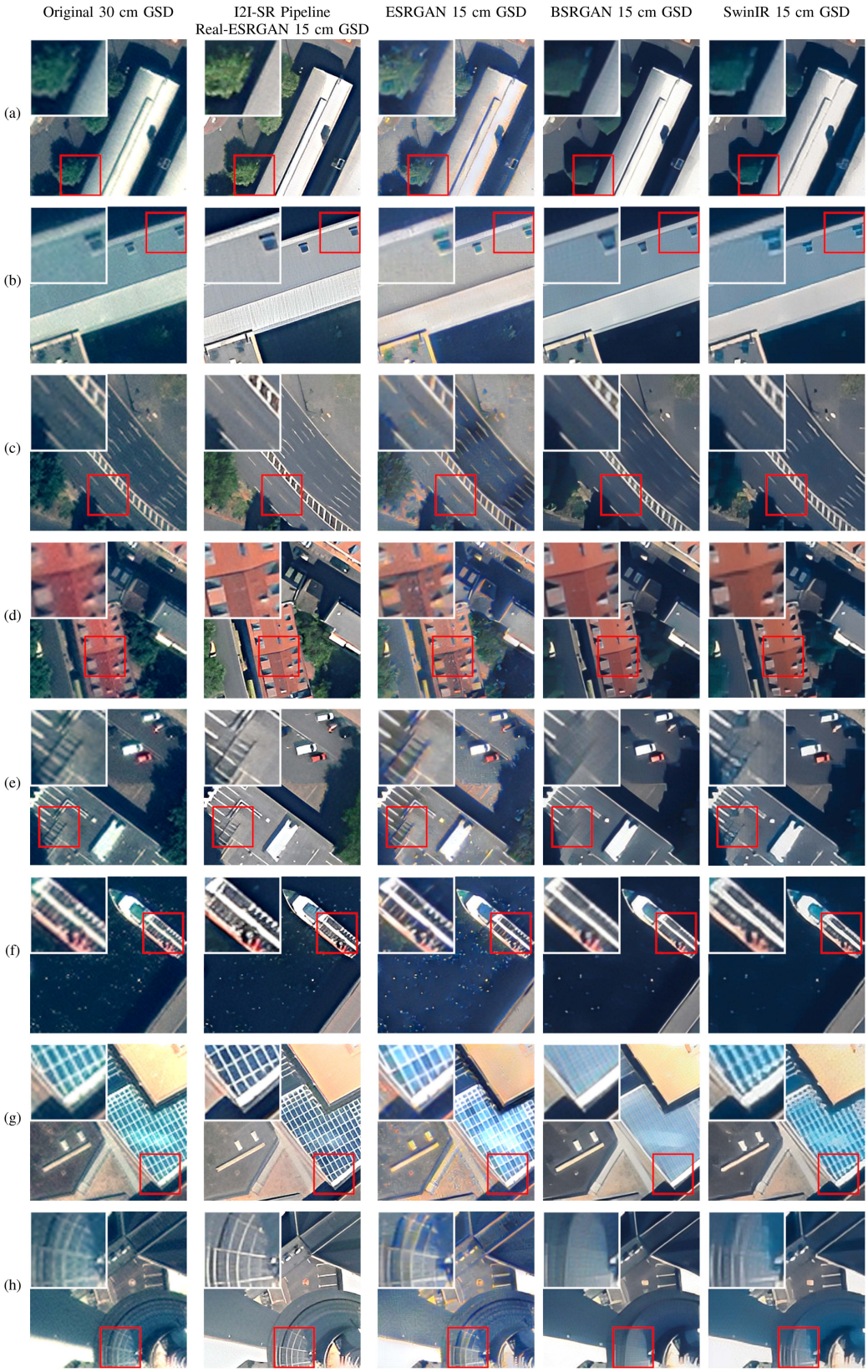

Fig. 14. Figure showcases the effectiveness of Real-ESRGAN in our I2I-SR pipeline and other SR algorithms, trained on SynthSat, in enhancing satellite image resolution from 30 to 15 cm GSD over the city of Berlin.

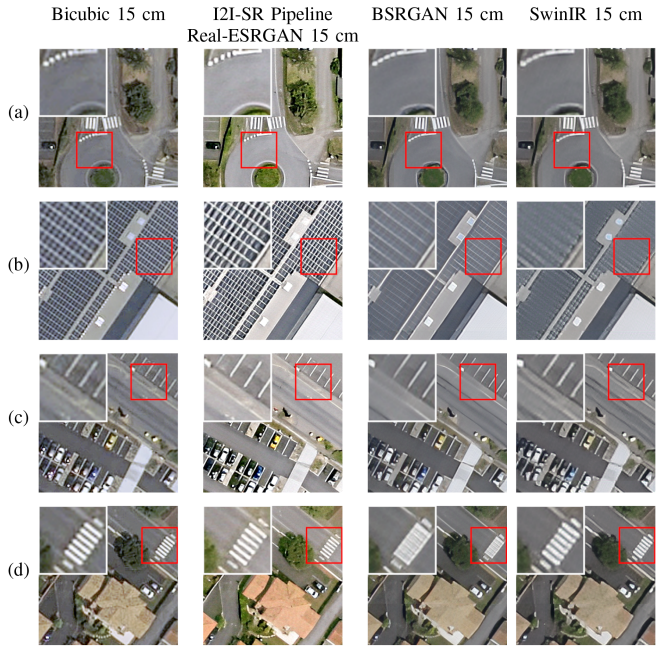

Fig. 15. Airbus Pléiades Neo image over Clisson-France enhanced from 30 to 15 cm using Bicubic interpolation, Real-ESRGAN (I2I-SR pipeline), BSRGAN, and SwinIR.

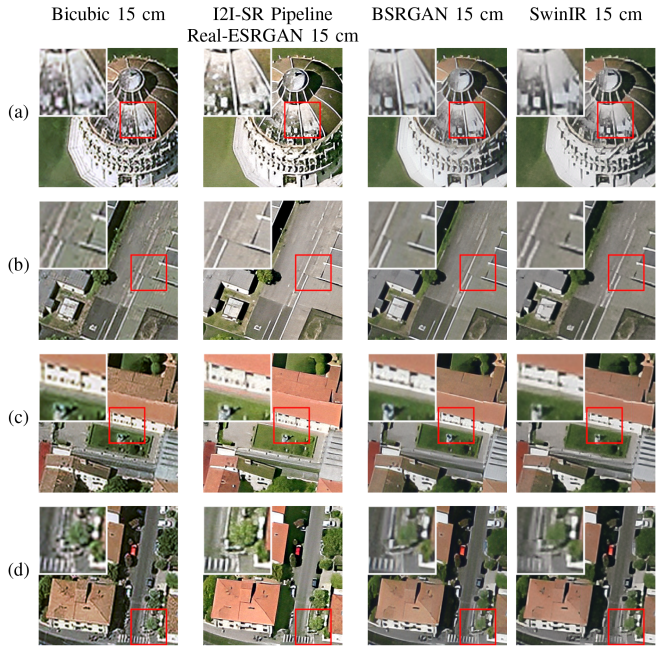

Fig. 16. Airbus Pléiades Neo image over Pisa-Italy enhanced from 30 to 15 cm using Bicubic interpolation, Real-ESRGAN (I2I-SR pipeline), BSRGAN, and SwinIR.

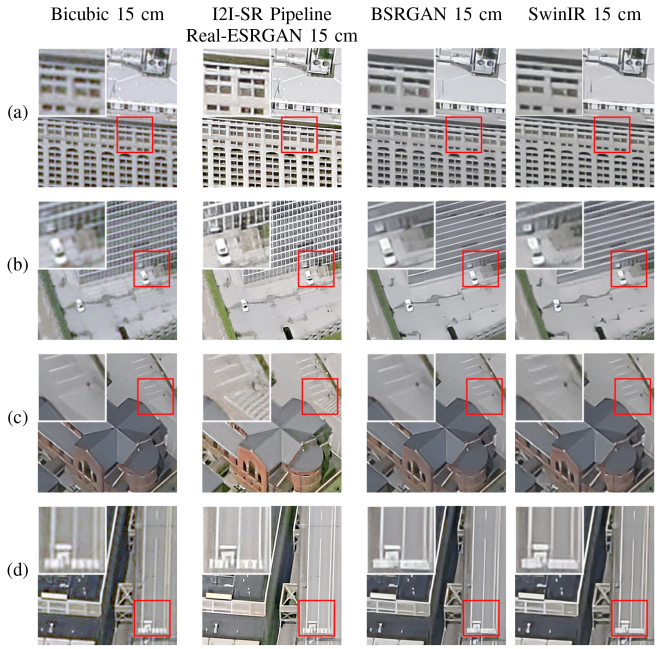

Fig. 17. Airbus Pléiades Neo image over Pittsburg-USA enhanced from 30 to 15 cm using Bicubic interpolation, Real-ESRGAN (I2I-SR pipeline), BSRGAN, and SwinIR.

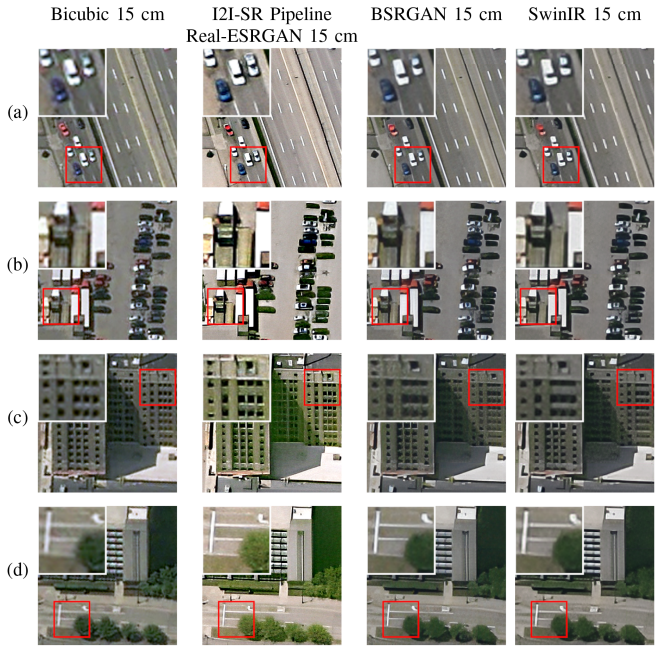

Fig. 18. Airbus Pléiades Neo image over Detroid-USA enhanced from 30 to 15 cm using Bicubic interpolation, Real-ESRGAN (I2I-SR pipeline), BSRGAN, and SwinIR.

consistency and avoid inconsistencies along patch borders after mosaicking the individual predictions, we processed the entire images at once without slicing them. The patch size was set to 3552×3488 pixels.

In this section, we present a qualitative analysis of our results, as we lack GT data for our test scene—a WV4 image over Munich city with a native 30 cm GSD. To provide visual support, we include the corresponding OpenStreetMap [84], where roads are represented with an approximate width of 5 m. Our results, shown in Fig. 22, reveal that the model extracted more road samples (a)–(c) from SR images generated by Real-ESRGAN using the I2I-SR pipeline compared to other methods. Despite dealing

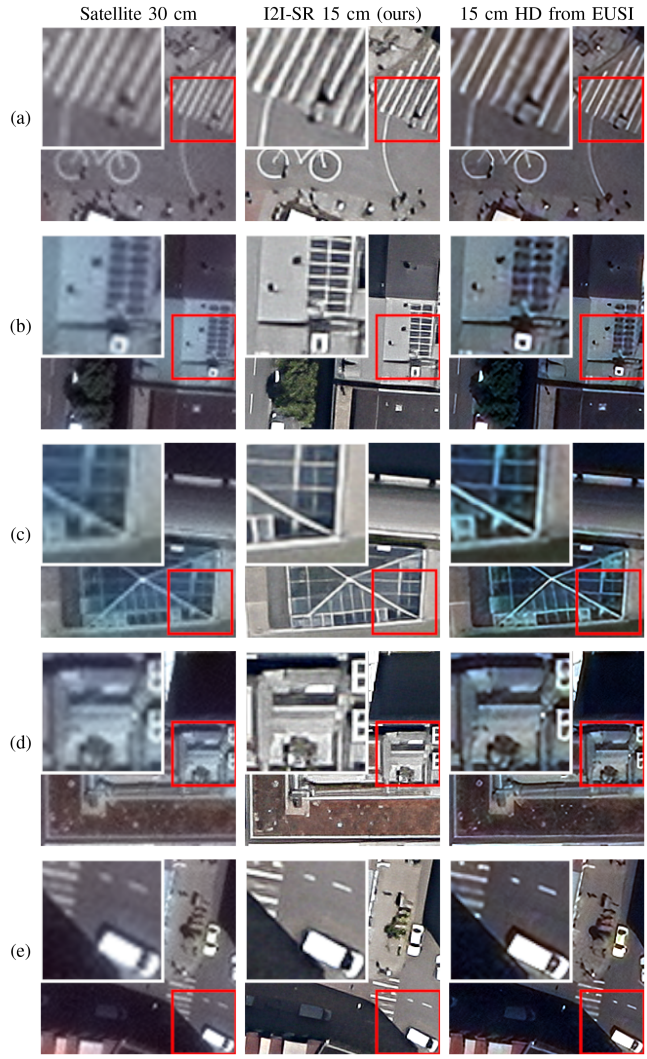

Fig. 19. (a) Satellite image at 30 cm GSD. (b) SR image at 15 cm GSD from our I2I-SR approach. (c) 15 cm HD from EUSI, over the city of Munich.

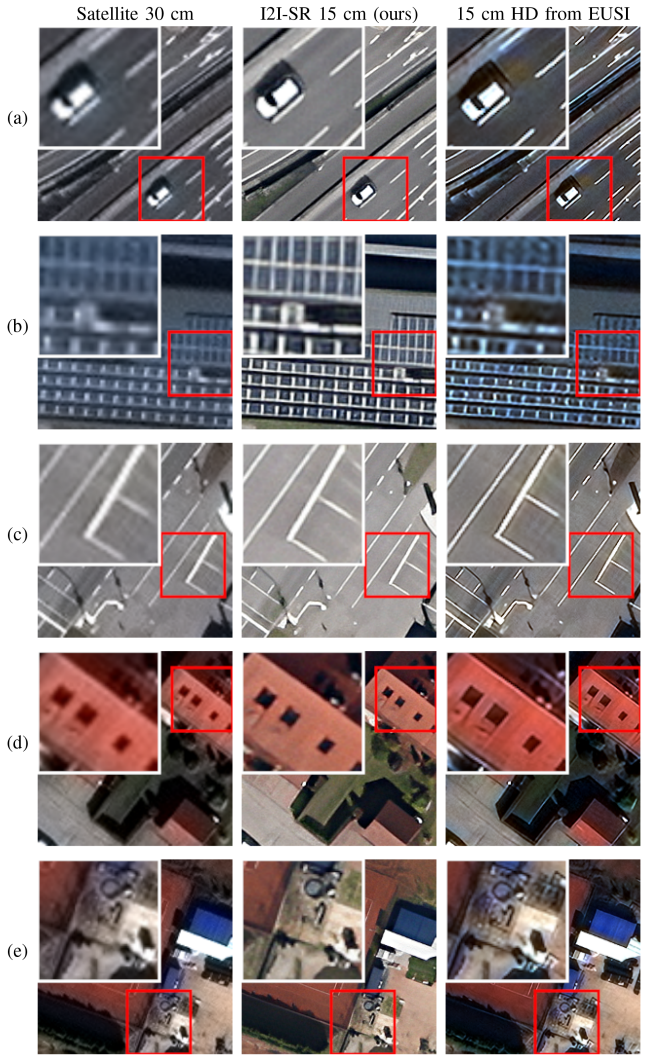

Fig. 20. (a) Satellite image at 30 cm GSD. (b) SR image at 15 cm GSD from our I2I-SR approach. (c) 15 cm HD from EUSI, over a town south of Munich.

with challenging road layouts and complex backgrounds, the segmentation model successfully captures most of the roads. Notably, single-lane roads (a), (c) are more consistently extracted than larger roads (b), (c). We attribute this disparity to two factors: first, the SR downsampled images at 50 cm GSD are significantly less blurry than those from the DeepGlobe18 dataset, as they are not native 50 cm GSD images; second, our training set lacks urban scenarios, especially dense urban scenes, such as the city center region from Munich. Furthermore, all the SR methods effectively reduce the noise present in the original image, leading to improved segmentation [e.g., the removal of false positives inside the roundabout in sample (e)]. However, we observed that BSRGAN and SwinIR tend to oversmooth the images, creating spurious connections between nearby objects, such as roads leaving the roundabout in sample (e). Overall, the images produced by our I2I-SR pipeline demonstrate its remarkable capacity for image enhancement, as roads appear more clearly to our model. However, there are still some obstacles that need to be addressed to extract the remaining roads successfully. Two significant challenges are 1) shadow occlusion and 2) color bleeding. Shadow occlusion is a common issue in remote sensing, requiring dedicated training sets and methods to overcome. This complexity is evident in sample (d) where no single method managed to fully extract the two streets leaving the roundabout along the South-Western and North-Eastern axes. The success of road recognition in shadow areas depends on the amount of retained information and, most crucially, the contrast within the shadow regions in the SR images. When color values are close to 0 or only retain some gradient, certain methods may struggle to correctly identify the roads. On the other hand, color bleeding is specific to the Real-ESRGAN images in sample (d), which can confuse the segmentation model. In these images, roads may no longer be composed of shades of gray and white, but instead, they contain large spots of diffuse green and red from nearby vegetation and building roofs, respectively. This phenomenon is responsible for the total or partial absence of roads along the Northern, Southern, and South-Eastern axes in sample (d). Addressing these challenges will be crucial for achieving

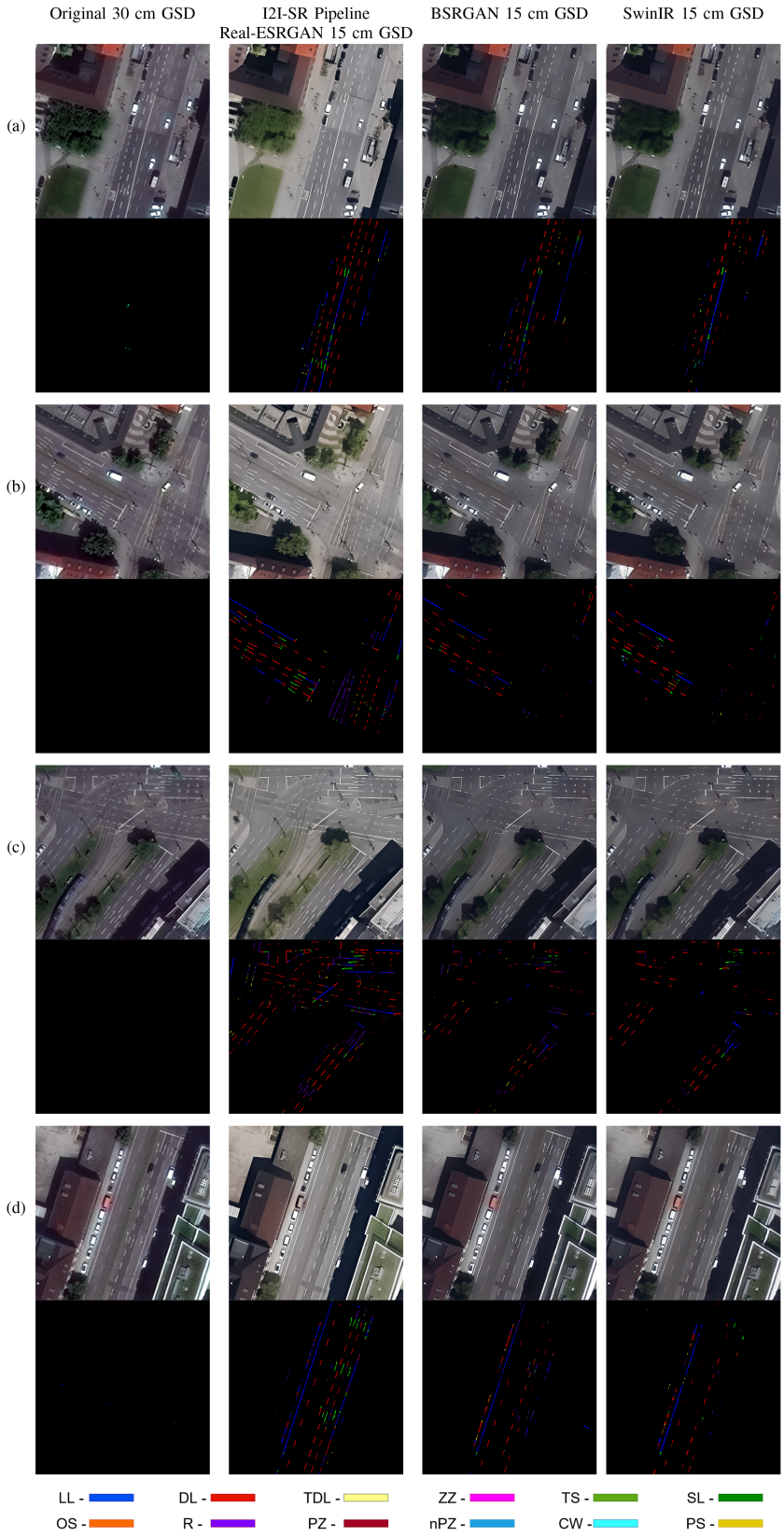

Fig. 21. Enhanced satellite images at 15 cm GSD via different methods, accompanied by lane prediction masks from [3] using 15 cm GSD images as input.

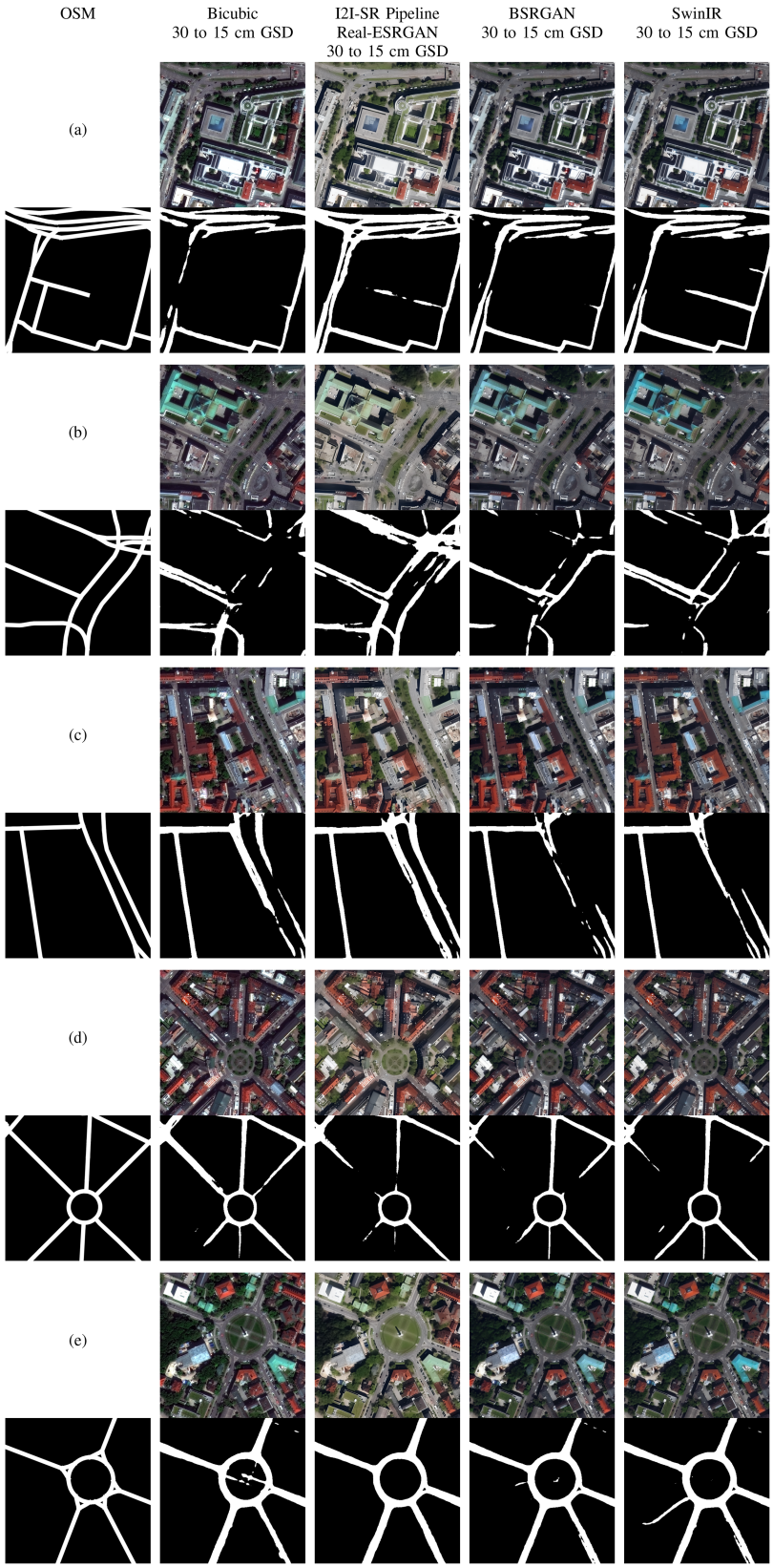

Fig. 22. Comparative analysis of RGB satellite images at 50 cm GSD from various methods and OpenStreetMaps' road segmentation mask at the same location.

complete and accurate road extraction in such complex scenes. Nevertheless, the noise-free nature and superior quality of the SR image at 15 cm GSD, produced by the I2I-SR pipeline, resulted in a significantly improved downsampled image at 50 cm GSD. As a result, it successfully detected more roads compared to the original image with similar GSD. This impressive outcome highlights the immense potential of very high-resolution SR satellite images.

VI. Evaluation of Color Bleeding

In this section, we conduct an in-depth analysis of the color bleeding issue, exploring its causes and potential solutions for mitigation. The severity of color bleeding varies among different algorithms: ESRGAN exhibits a significant problem, while Real-ESRGAN images show a moderate presence in dark/shadow regions. However, BSRGAN and SwinIR, all trained on the SynthSat dataset, do not exhibit this issue. This observation may be attributed to how real-world/blind SR algorithms, such as Real-ESRGAN, BSRGAN, and SwinIR, handle degradation. The bleeding effect appears to be more pronounced in SR methods lacking proper degradation modeling, such as ESRGAN as seen from (b), (d), (f), and (h) in Fig. 12. RGB satellite images, despite extensive processing, can still retain various artifacts, such as halo artifacts around moving vehicles, color and black and white noise artifacts, reflection artifacts, and aliasing around objects at certain angles. Real-world/blind SR algorithms attempt to model a wide range of degradations during their training process. This includes operations, such as adding diverse types of noise, using blurring functions, introducing compression artifacts, and employing various up- and downsampling techniques to simulate real-world degradations. Such comprehensive degradation modeling contributes to the effectiveness of these algorithms in handling color bleeding and other artifacts. BSRGAN and SwinIR remove degradations more aggressively from the native images, which allows them to suppress color noise/artifacts from the original 30-cm image and produce cleaner images without color bleeding. However, this approach comes with a drawback—their SR images tend to be blurry with smoothed textures, leading to the loss of some important information that might be overlaid by noise in the satellite images. Consequently, sometimes the noise artifacts might contain valuable information (example: noise present near or on a lane marking), are considered noise and are removed during the process, sacrificing SR image quality. This can clearly be seen in (a) and (c) samples in Fig. 21, where few lane marking are not even reconstructed or a part of the lane marking is missing for BSRGAN and SwinIR. Real-ESRGAN, despite being trained with comparable degradation modeling, such as BSRGAN and SwinIR, adopts a less aggressive approach, aiming to retain maximum information in the base image while removing as much degradation as possible (as observed in Figs. 12 and 14). However, in certain regions of the satellite image where the network cannot completely remove color artifacts during processing, these artifacts are amplified in the SR images, leading to the color bleeding problem. To test this theory, we selected a few images exhibiting color bleeding issues, namely,

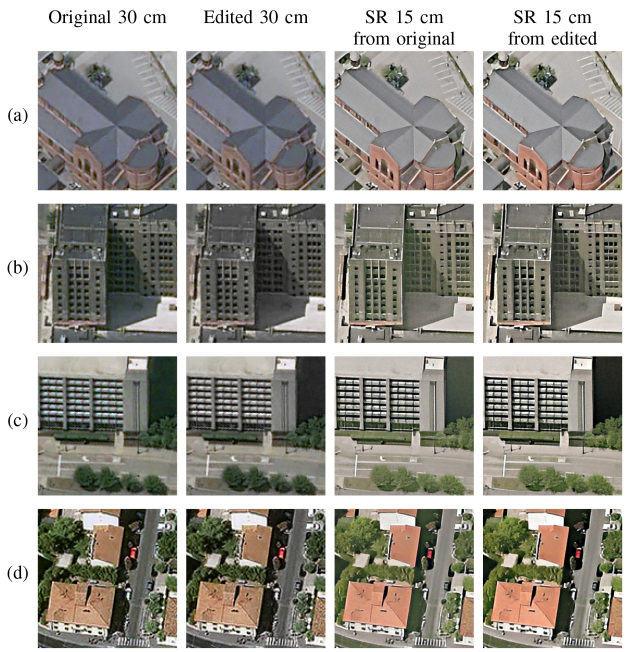

Fig. 23. Performance of the images from the Real-ESRGAN in I2I-SR pipeline when applied on original 30-cm image versus edited 30-cm image (after removing color artifact in the shadow region).

Figs. 16(d), 17(c), and 18(c) and (d). For each sample in Fig. 23, we can see that the SR images from Real-ESRGAN have a green color bleeding issue in the dark/shadow regions. We use GIMP (a free and open-source raster graphics editor used for image manipulation and image editing) to mask such regions and lower down the saturation of the green channel, thus rendering the green to gray color. Now when we performed inference on the edited images, we could see that the SR images no longer have any color bleeding issues present in the dark/shadow regions, as speculated earlier. Indeed, while this method may not be practical for removing the color artifact for every satellite image, it provided valuable insight into a potential solution. We believe that constructing a better SynthSat dataset, comprising various satellite images with diverse degradation artifacts during the training of the SynthSat-I2I network, could lead to a more robust dataset for training the SR algorithm for real-world inference.

VII. Conclusion

In this study, we introduce a novel I2I-SR pipeline composed of the SynthSat-I2I network and an SR algorithm. The SynthSat-I2I network constructs the SynthSat dataset, comprising aerial images at 15 cm GSD as GT and synthetic satellite images at 30 cm GSD for training the SR algorithm. Applying the trained SR model to actual 30 cm GSD satellite images, we successfully enhance them to 15 cm GSD. These enhanced images are brighter, sharper, and have well-reconstructed textures. We also show that selected SR algorithm is able to generate better images when compared to other methods. In addition, the limited size and diversity of the data for training the SynthSat-I2I network and the SynthSat dataset, the results can suffer from some artifacts such as color bleeding. Our future work aims to

utilize larger, more diverse datasets and explore enhancements without domain transfer.

ACKNOWLEDGMENT

The authors would like to thank European Space Imaging for the World View data and their colleague C. Henry for his contribution on the road segmentation.

REFERENCES

[1] S. Ø. Larsen, H. Koren, and R. Solberg, "Traffic monitoring using very high resolution satellite imagery," *Photogrammetric Eng. Remote Sens.*, vol. 75, no. 7, pp. 859–869, 2009.

[2] Y. Shi and H. Li, "Beyond cross-view image retrieval: Highly accurate vehicle localization using satellite image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16989–16999.

[3] S. M. Azimi, C. Henry, L. Sommer, A. Schumann, and E. Vig, "Skyscapes fine-grained semantic understanding of aerial scenes," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 7392–7402.

[4] M. Janalipour and M. Taleai, "Building change detection after earthquake using multi-criteria decision analysis based on extracted information from high spatial resolution satellite images," *Int. J. Remote Sens.*, vol. 38, no. 1, pp. 82–99, 2017.

[5] M. Ghanea, P. Moallem, and M. Momeni, "Building extraction from high-resolution satellite images in urban areas: Recent methods and strategies against significant challenges," *Int. J. Remote Sens.*, vol. 37, no. 21, pp. 5234–5248, 2016.

[6] D. J. Bora, "AERSCIEA: An efficient and robust satellite color image enhancement approach," in *Proc. Int. Conf. Res. Intell. Comput. Eng.*, 2017, pp. 3–13.

[7] H. Demirel, C. Ozcinar, and G. Anbarjafari, "Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 333–337, Apr. 2010.

[8] J. Park, J.-Y. Lee, D. Yoo, and I. S. Kweon, "Distort-and-recover: Color enhancement using deep reinforcement learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5928–5936.

[9] J. Song, J.-H. Jeong, D.-S. Park, H.-H. Kim, D.-C. Seo, and J. C. Ye, "Unsupervised denoising for satellite imagery using wavelet directional cycle-GAN," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6823–6839, Aug. 2021.

[10] S. Suresh, S. Lal, C. Chen, and T. Celik, "Multispectral satellite image denoising via adaptive cuckoo search-based wiener filter," *IEEE Trans. Geosc. Remote Sens.*, vol. 56, no. 8, pp. 4334–4345, Aug. 2018.

[11] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," *Neural Netw.*, vol. 131, pp. 251–275, 2020.

[12] Z. Wang, M. K. Ng, J. Michalski, and L. Zhuang, "A self-supervised deep denoiser for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5520414.

[13] J. Li, Z. Pei, and T. Zeng, "From beginner to master: A survey for deep learning-based single-image super-resolution," vol. abs/2109.14335, 2021. [Online]. Available: https://arxiv.org/abs/2109.14335

[14] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2472–2481.

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

[16] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 63–79.

[17] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1905–1914.

[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.*, 2014, pp. 184–199.

[19] C. Dong, C. C. Loy, X. Tang, and K. He, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[20] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2021.

[21] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 372–386.

[22] F. Salvetti, V. Mazzia, A. Khaliq, and M. Chiaberge, "Multi-image super resolution of remotely sensed images using residual attention deep neural networks," *Remote Sens.*, vol. 12, no. 14, 2020, Art. no. 2207.

[23] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1132–1140.

[24] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11057–11066.

[25] K. Jiang, Z. Wang, P. Yi, J. Jiang, J. Xiao, and Y. Yao, "Deep distillation recursive network for remote sensing imagery super-resolution," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1700.

[26] J. Shermeyer and A. Van Etten, "The effects of super-resolution on object detection performance in satellite imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 1432–1441.

[27] T. Lu, J. Wang, Y. Zhang, Z. Wang, and J. Jiang, "Satellite image super-resolution via multi-scale residual deep neural network," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1588.

[28] M. Kawulok, S. Piechaczek, K. Hrynczenko, P. Benecki, D. Kostrzewa, and J. Nalepa, "On training deep networks for satellite image super-resolution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3125–3128.

[29] Z. Wang, K. Jiang, P. Yi, Z. Han, and Z. He, "Ultra-dense GAN for satellite imagery super-resolution," *Neurocomputing*, vol. 398, pp. 328–337, 2020.

[30] X. Zhu, H. Talebi, X. Shi, F. Yang, and P. Milanfar, "Super-resolving commercial satellite imagery using realistic training data," in *Proc. IEEE Int. Conf. Image Process.*, 2020, pp. 498–502.

[31] F. Chouteau et al., "Joint super-resolution and image restoration for Pléiades Neo imagery," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 9–15, 2022.

[32] 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.*, 2016, pp. 1646–1654.

[33] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 4771–4780.

[34] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1833–1844.

[35] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1110–1121.

[36] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 262–270.

[37] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2414–2423.

[38] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[39] D. Kotovenko, A. Sanakoyeu, S. Lang, and B. Ommer, "Content and style disentanglement for artistic style transfer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 4421–4430.

[40] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images," in *Proc. 33rd Int. Conf. Mach. Learn.*, M. F. Balcan and K. Q. Weinberger, Eds. 2016, vol. 48, pp. 1349–1357.

[41] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[42] J.-Y. Zhu et al., "Toward multimodal image-to-image translation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 465–476.

[43] L. Ma, X. Jia, S. Georgoulis, T. Tuytelaars, and L. Van Gool, "Exemplar guided unsupervised image-to-image translation with semantic consistency," 2019, *arXiv:1805.11145*.

[44] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 700–708.

[45] L. Jiang, C. Zhang, M. Huang, C. Liu, J. Shi, and C. C. Loy, "TSIT: A simple and versatile framework for image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 206–222.

[46] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.
[47] Y. Zhang, M. Li, R. Li, K. Jia, and L. Zhang, "Exact feature distribution matching for arbitrary style transfer and domain generalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8025–8035.
[48] Y. Zhang et al., "Domain enhanced arbitrary image style transfer via contrastive learning," in *Proc. ACM SIGGRAPH Conf. Proc.*, 2022, pp. 1–8.
[49] M. Elad and P. Milanfar, "Style transfer via texture synthesis," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2338–2351, May 2017.
[50] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Proc. 28th Annu. Conf. Comput. Graph. Interactive Techn.*, 2001, pp. 341–346.
[51] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. IEEE 7th Int. Conf. Comput. Vis.*, 1999, pp. 1033–1038.
[52] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1510–1519.
[53] J. Liang, H. Zeng, and L. Zhang, "High-resolution photorealistic image translation in real-time: A Laplacian pyramid translation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9387–9395.
[54] Y. Jing et al., "Dynamic instance normalization for arbitrary style transfer," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 4369–4376.
[55] I. Anokhin et al., "High-resolution daytime translation without domain labels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7485–7494.
[56] Z. Wang, L. Zhao, and W. Xing, "Stylediffusion: Controllable disentangled style transfer via diffusion models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 7677–7689.
[57] K. Hong et al., "AesPA-Net: Aesthetic pattern-aware style transfer networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 22758–22767.
[58] S. Huang, J. An, D. Wei, J. Luo, and H. Pfister, "Quantart: Quantizing image style transfer towards high visual fidelity," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 5947–5956.
[59] Y. Ma, C. Zhao, X. Li, and A. Basu, "RAST: Restorable arbitrary style transfer via multi-restoration," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2023, pp. 331–340.
[60] L. Wen, C. Gao, and C. Zou, "CAP-VSTNet: Content affinity preserved versatile style transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 18300–18309.
[61] C. Xu and B. Zhao, "Satellite image spoofing: Creating remote sensing dataset with generative adversarial networks (short paper)," in *Proc. 10th Int. Conf. Geographic Inf. Sci.*, 2018, pp. 67:1–67:6.
[62] J. Marín and S. Escalera, "SSSGAN: Satellite style and structure generative adversarial networks," *Remote Sens.*, vol. 13, no. 19, 2021, Art. no. 3984.
[63] Y. Kang, S. Gao, and R. E. Roth, "Transferring multiscale map styles using generative adversarial networks," *Int. J. Cartogr.*, vol. 5, no. 2/3, pp. 115–141, 2019.
[64] F. Schenkel, S. Hinz, and W. Middelmann, "Style transfer-based domain adaptation for vegetation segmentation with optical imagery," *Appl. Opt.*, vol. 60, no. 22, pp. F109–F117, 2021.
[65] L. Wang et al., "SAR-to-optical image translation using supervised cycle-consistent adversarial networks," *IEEE Access*, vol. 7, pp. 129136–129149, 2019.
[66] M. Fuentes Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt, "SAR-to-optical image translation based on conditional generative adversarial networks–optimization, opportunities and limits," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 2067.
[67] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 179, pp. 14–34, 2021.
[68] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2332–2341.
[69] X. Luo, Z. Han, L. Yang, and L. Yang, "Progressive attentional manifold alignment for arbitrary style transfer," in *Proc. Asian Conf. Comput. Vis.*, 2022, pp. 3206–3222.
[70] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. 6th Int. Conf. Learn. Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=B1QRgziT-
[71] Z. Lin, V. Sekar, and G. Fanti, "Why spectral normalization stabilizes GANs: Analysis and improvements," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 9625–9638.
[72] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8798–8807.
[73] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.
[74] E. Riba, D. Mishkin, D. Ponsa, E. Rublee, and G. Bradski, "Kornia: An open source differentiable computer vision library for pytorch," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2020, pp. 3663–3672.
[75] T. Qiu, B. Ni, Z. Liu, and X. Chen, "Fast optimal transport artistic style transfer," in *Proc. Int. Conf. Multimedia Model.*, 2021, pp. 37–49.
[76] N. Kolkin, J. Salavon, and G. Shakhnarovich, "Style transfer by relaxed optimal transport and self-similarity," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10043–10052.
[77] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Representations*, 2015. [Online]. Available: http://arxiv.org/abs/1412.6980
[78] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," vol. abs/1607.08022, 2016. [Online]. Available: http://arxiv.org/abs/1607.08022
[79] 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, pp. 9625–9638.
[80] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595.
[81] A. intelligence, "Satellite image gallery," Accessed: Jul. 27, 2023. [Online]. Available: https://www.intelligence-airbusds.com/newsroom/satellite-image-gallery
[82] C. Henry, F. Fraundorfer, and E. Vig, "Aerial road segmentation in the presence of topological label noise," in *Proc. Int. Conf. Pattern Recognit.*, 2021, pp. 2336–2343.
[83] I. Demir et al., "DeepGlobe 2018: A challenge to parse the Earth through satellite images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 172–17209.
[84] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org," 2017. [Online]. Available: https://www.openstreetmap.org

Sandeep Kumar Jangir received the bachelor's degree in computer science and engineering from Anna University, Chennai, India, in 2017, and the master's degree in computer science from the Technical University of Munich, Munich, Germany, in 2020. He is currently working toward the Ph.D. degree in aerial and satellite image Enchantment using deep learning with the Department of Photogrammetry and Image Analysis, German Aerospace Center (DLR), Wessling, Germany.

His research interests include high- and low-level computer vision tasks, deep learning techniques, synthetic datasets, and computer graphics.

Reza Bahmanyar received the B.Sc. degree in electrical engineering from Mazandaran University, Babol, Iran, in 2009, and the M.Sc. degree in computer science from Saarland University, Saarbrücken, Germany, in 2012. He wrote the master's thesis in the Max Plank Institute of Computer Science, Saarbrücken, Germany. He received the Ph.D. degree in computer science from the Technical University of Munich, Munich, Germany, in 2016, in a joint program with the Munich Aerospace organization and the German Aerospace Center, DLR, Wessling, Germany.

Since 2016, he has been a Research Scientist with Remote Sensing Institute, DLR. His research interests include computer vision, image processing, data mining, and artificial intelligence.