Leveraging Vision Transformers and Adversarial Unsupervised Domain Adaptation to Generate the NIR Band on Out-of-Domain VHR RGB Data

Viktoria Kristollari, Vassilia Karathanassi

***Abstract*— The accessibility of very-high-resolution Remote Sensing (RS) near-infrared (NIR) imagery before 2010s is limited. In addition, public annotated image datasets often used as benchmarks contain only RGB information. Thus, RGB-to-NIR translation could greatly benefit the RS field. Generative adversarial networks (GANs) and more recently vision transformers (ViT) have shown promising results in generating spectrally super-resolved data. However, the RGB-to-hyperspectral studies scatter their attention to multiple bands, while the RGB-to-NIR studies have tackled only vegetation. In addition, prior research regarding NIR prediction on out-of-domain data is extremely limited, and attention models have not been previously tested in the RGB-to-NIR translation. Beyond that, unsupervised domain adaptation (UDA) has been overlooked in the enhancement of cross-domain band generation. This study, at first evaluates a conditional GAN and two attention networks on predicting NIR on out-of-domain RGB data. The properties of the out-of-domain data are those typically required in RS NIR prediction tasks (different regions/sensors/dates). In a following step, the study attempts to increase the source/target RGB radiometric similarity through CycleGAN-based UDA on unpaired bi-temporal data (lack of geographic correspondence in RGB source/target patches), and thus improve the initial NIR prediction. In the experiments, the vegetation, impervious and ground classes were assessed. It was shown that MST++ (ViT) produced the most satisfactory out-of-domain NIR predictions and that UDA through a CycleGAN version employing batch normalization managed to significantly enhance the NIR prediction when there was a substantial RGB radiometric domain gap.**

***Index Terms*— Prediction methods, Image generation, Infrared imaging, Neural network applications, Spectral domain analysis, Image resolution, Optical image processing**

I. Introduction

IMAGE-to-image translation (ITIT) is an image processing method whose basic idea is to learn the mapping functions between an input and an output image [1]. ITIT can be either performed for paired data (pixel co-respondence between input and output image) or unpaired data. Multiple Remote Sensing (RS) studies have investigated paired ITIT through deep learning (DL) and most often conditional generative adversarial networks (cGANs) to generate missing information. The applications are various and among others include Synthetic Aperture Radar (SAR)↔optical [1]–[9], visible (VIS)→map [10]–[14], optical→elevation (digital terrain/surface model (DTM/DSM)) [15]–[18], thermal infrared (TIR)↔VIS [19]–[22], VIS-VIS [23][24], grayscale→RGB [25][26], spectral super-resolution (SSR) (RGB/multispectral (MS)→hyperspectral (HS)) [27]–[35] [36]–[38], and RGB→near-infrared (NIR) [39]–[43]. Concerning unpaired ITIT, so far, it has been used in Remote Sensing (RS) applications exclusively for unsupervised domain adaptation (UDA) purposes as an intermediate step to enhance the semantic segmentation output [44]–[52]. It is noted that the arrow symbol ("→") in the text refers to the direction of ITIT.

In the following Introduction Subsections, we present the related literature works (I.A-I.C), and then we describe our motivations and contributions (I-D).

A. Paired ITIT– Broadly Related Work

The vast majority of the RS ITIT research has been focused on SAR↔optical paired deep ITIT because SAR data are unaffected by atmospheric conditions. Due to higher accessibility, most of the studies process medium-resolution (>=5 m) data [2]–[5]. However, recently a few studies have been conducted on very high-resolution (VHR) data [1], [6]–[9]. In [1], aiming at enhancing change detection performance, NICE-GAN [53], an introspective network based on CycleGAN [54] with multi-scale formulation in the discriminator and residual attention, was used for SAR↔optical translation. With the same goal but in a non-adversarial setting, the authors in [6] implemented an optical-SAR domain adaptation-based change detection network where distribution discrepancies in Hilbert space were included. An ITIT adaptation-based change detection technique (based on NICE-GAN) was also proposed in [7] where the features of optical images were transferred to SAR. In [8], the authors took advantage of both Pix2Pix (cGAN) [55] and CycleGAN, and performed SAR↔optical mapping by incorporating an additional network called the distortion-adaptive module in both directions. Finally, in [9], a Parallel-GAN was proposed for SAR→optical translation consisting of a backbone ITIT subnetwork and an adjoint optical image reconstruction subnetwork.

Interest has also been shown in the deep translation of VHR

The authors are with the Laboratory of Remote Sensing, School of Rural, Surveying, and Geoinformatics Engineering, National Technical University of Athens, 15780 Zografou, Greece (e-mail: vkristoll@central.ntua.gr; karathan@survey.ntua.gr).

VIS satellite images to maps because it could prove very beneficial when timely updates are required [10]–[14]. Several methods have been applied in a paired setting. In [10], a scale-consistent cGAN was proposed to simultaneously generate multi-level tile maps from multi-scale RS images. In addition, in [12], adversarial deep transfer training schemes were combined with attention-based network designs to generate maps over various regions, and in [13], a level-aware fusion network for multilevel map generation was introduced. Due to the scarcity of paired data, unpaired samples have also been proven to be useful. In [11], the authors designed a semi-supervised learning strategy based on training GANs on rich unpaired samples and then applied fine-tuning on limited paired samples. In addition, in [14], Semi-MapGen was proposed, a network based on semi-supervised GANs, which requires only a small set of accurate and complete matched data and plenty of unpaired data.

Mapping optical data to elevation information in a paired fashion is another application that has been studied in the DL literature as an affordable alternative to approaches that require Light detection and ranging (Lidar), Interferometric SAR (InSAR) or stereo pairs [15]–[18]. The study described in [15] was among the first to apply a cGAN to translate VHR optical data to elevation. Adversarial learning was also proposed in [16] where Pix2Pix was implemented to map optical to elevation data for Sentinel-2 and unmanned aerial vehicle (UAV) imagery. In a non-adversarial setting, the authors of [17] applied an encoder-decoder model with skip connections [56] and residual blocks [57] for DSM generation. Several strategies were investigated and showed that the performance can vary according to the dataset morphology. Similarly, in [18], a U-Net [56] with residual blocks was applied to create elevation information for airborne images.

Other DL studies of paired data have investigated the TIR→VIS ITIT in data collected by geostationary meteorological satellites by use of the Pix2Pix model to enrich the collected information [19]–[21]. In [19], Pix2Pix was trained on daytime pairs of the 10.8 μm longwave radiance band and the 0.675 μm visible band of the meteoritical imager (MI) onboard the Communication, Ocean and Meteorological Satellite (COMS), to create the non-existent nighttime visible reflectance band. In a later study for the above-mentioned satellite [20], Pix2Pix was applied for virtual nighttime visible imagery generation using multiband infrared observations and a brightness temperature difference. In a similar concept, in [21], Pix2Pix was trained on thermal band differences of the Advanced Meteorological Imager (AMI) sensor, of the GK-2A Geostationary Korea Multi-Purpose Satellite, to provide virtual RGB bands during day and night. Besides the above-mentioned low-resolution meteorological applications, VIS→TIR VHR mapping by Pix2Pix has also been explored [22] as an intermediate step to achieve thermal geolocation in low illumination environments, motivated by the limited availability of satellite thermal data.

Data collected by geostationary meteorological satellites have also been used for cross-satellite deep paired VIS-VIS ITIT with the Pix2Pix network to generate missing bands [23] [24]. In [23], Pix2Pix was trained on blue band radiance images of the Advanced Himawari Imager (AHI) to generate a simulated green band (useful for monitoring water and vegetation) for the Geostationary Operational Environmental Satellite (GOES-16) Advanced Baseline Imager (ABI) sensor. In addition, in [24] Pix2Pix was trained on blue band radiance images of the GK-2A/AMI sensor to generate simulated green and red bands (useful for monitoring atmospheric environments) for the Geostationary Environment Monitoring Spectrometer (GEMS) of the GK-2B satellite.

B. Paired ITIT– Closely Related Work

1). SSR:

Hyperspectral images carry valuable spectral information. However, the fact that until very recently satellites that provide global and public HS data were inexistent and the high cost of airborne sensors hinders the exploitation of HS information. Thus, hyperspectral image reconstruction from low-cost RGB cameras and MS sensors (spectral super-resolution (SSR)) has attracted recent attention. In the past, the problem was approached in a non-DL manner with linear unmixing [58], regression [59], and methods that exploit sparse representations and dictionary learning [60][61]. More recently, numerous DL studies have been published [27]–[35] [36]–[38]. All SSR literature methods so far, have evaluated their paired methodologies in the output as a whole without isolating particular bands (e.g. NIR).

In [27], a variant of the dense convolutional neural network (CNN) called "Tiramisu" [62] was trained on MS data collected from the Advanced Land Imager (ALI) on board the Earth Observing One (EO-1) satellite to predict the HS Hyperion bands. The qualitative evaluation showed that the predicted output was less noisy than the ground-truth data. In addition, the quantitative evaluation employed abundance estimation. In [28], CNN regression models were investigated to produce the Hyperion HS bands from the Landsat 7/8 MS bands. The authors showed that the CNN regression produced better performance compared to conventional regression methods. The model outputs were also evaluated by classification with support vector machines (SVM) over principal components (PCs). In [29], the authors proposed a cGAN with an additional spectral discriminator to map RGB to HS information in GF-5 data. The quantitative spectral and spatial scores showed that the proposed network was more robust than alternative non-adversarial approaches. In [30] an encoder-decoder model with attention to semantic similarity was implemented to spectrally super-resolve RGB to HS images in 1 m spatial resolution (SR). MS→HS SSR (Hyperion images) by employing semantic information was also proposed in [31] in the form of a change detection subnetwork. Finally, an encoder-decoder model was trained in [32] to spectrally super-resolve UAV and GF-5 MS images. Several input band combinations were explored and it was shown that the inclusion of the NIR band can enhance the performance of the final HS output.

In other studies [33]–[35], SSR attempts have been made to enhance the exploitation of both spatial and spectral information in high/medium SR data. In [33], the authors proposed a progressive spatial-spectral joint network to reconstruct satellite and airborne HS data from MS. In addition,

in [34], a spatial-spectral residual attention was exploited for MS→HS mapping, and in [35], a spatial-spectral feature attention module was introduced in a GAN for both synthetic and real data scenarios.

More recently, attention modules have been also exploited [36]–[38]. In [36], a hybrid transformer architecture was explored to achieve SSR on two aerial datasets through the integration of intra- and cross inter- row/column attention mechanisms. In [37], a reflectance/shading decomposition SSR framework that incorporated spatial and channel attention modules was tested on AVIRIS and GF5 data. Finally, in [38], the authors employed an integrated network of 3D CNN and U-shaped Transformer in a coarse-to-fine scheme to utilize both local and global information in aerial imagery.

2). RGB-to-NIR Translation:

The RS studies published so far in RGB→NIR paired translation have focused exclusively on vegetation applications [39]–[43], [63]. The significance of the NIR band in providing rich information for the determination of vegetation parameters has since long been established [64]–[67]. In the last years, MS sensors onboard UAVs have been extensively used in precision agriculture. However, small producers have difficulty in affording the required equipment. Low-cost RGB cameras on board lightweight UAVs are a more affordable option [41]. Thus, RGB→NIR translation could prove very useful for vegetation monitoring.

Before the broad application of DL, this problem had been approached by regression analysis on conventional cameras depicting crops [63], where a green-NIR correlation had been indicated. In recent years, several DL studies have been published for the RGB→NIR translation in vegetation areas and the vast majority have employed the Pix2Pix model [39]–[43]. In [39], Pix2Pix was trained on UAV RGB crop data to generate the NIR band. The L1 loss was replaced with the Charbonnier penalty function and both in- (same crop type) and out-of-domain (different crop types) experiments were performed. In [40], RGB→NIR translation was performed on Worldview-2 data by Pix2Pix with residual blocks in the generator. The authors focused on forest areas and evaluated the performance in a cross-domain setting (SPOT, Planet with finetuning). The Planet data translation showed lower evaluation scores due to higher heterogeneity compared to the training inputs. It was also stated that adversarial training increased the performance and that the inclusion of NIR in the classification task lowered the size of the needed annotations. One of the research motivations was the fact that public RGB databases do not contain the NIR band. However, ITIT could generate the missing information. In [41], Pix2Pix was trained on low-cost UAV RGB cameras to estimate the NIR band for agricultural purposes. The network outperformed a previously proposed endmember-based method [68]. The authors also investigated combinations of the original L1 loss with the structural similarity index (SSIM) and a perceptual loss to achieve slightly better results. In [42] Pix2Pix was employed for RGB→NIR field imagery translation (agricultural areas) with a DenseNet architecture [62] as the generator. Comparison with the original U-Net generator showed slight improvement. Finally, in [43], RGB→NIR translation was performed on Sentinel-2 data collected all year round. It was observed that the model was unaffected by corrupted pixels but did not show satisfactory generalization ability to Landsat-8 data. The failure in the out-of-domain performance could be attributed to differences in illumination, as well as atmospheric and sensor conditions.

C. UDA

Unpaired ITIT has recently widely been used in the form of UDA in VHR VIS RS when annotations are available in the source domain but not in the target domain [44]–[52]. The goal is to enhance cross-domain object detection (CDOD) or semantic segmentation (CDSS) tasks by decreasing domain shifts caused by the rich structure diversity, the variability in atmospheric/lighting conditions and viewing angles, as well as the different sensor characteristics.

Adversarial approaches and particularly CycleGAN-based models, which employ a cycle-consistency loss, have often been proposed [44]–[48]. In [44], the segmentation models were optimized in two opposite directions by implementing bidirectional adversarial domain adaptation through CycleGAN, which takes advantage of the information from both domains. In [45], a CycleGAN-based model with residual connections was proposed followed by a semantic segmentation stage. An in-network resizer module was included to address the scale discrepancy. In [46], a two-stage cross-domain self-training object detection framework was investigated. The first stage introduced a CycleGAN-based strategy to mitigate the domain-shift and increase the quality of pseudo-labels. In [47], the two-stage CDSS task was approached by a dual space CycleGAN-based model where frequency information through discrete wavelet transform was integrated. Finally, in [48], CycleGAN-based UDA was explored to overcome the limitations of a coastal marine debris estimation model.

In other studies, adversarial learning has been applied, without the cyclic-consistency loss, or alternative non-adversarial learning strategies have been employed [49]–[52]. In [49], CDSS was performed by a curriculum-style local-to-global cross-domain adaptation framework. The adaptation process was conducted in an easy-to-hard way using an entropy-based score and adversarial learning. In [50], a two-stage framework was applied, which performs fine-grained local and category-level alignment on top of global alignment. The framework used adversarial learning and knowledge distillation. In [51], the authors proposed a deep covariance alignment model to align category features. Finally, in [52], the authors implemented a lightweight UDA model relying on latent representation separation and mixing across domains which can be used in an one-shot setting.

D. Motivations and Contributions

Before 2010s the accessibility of VHR RS NIR imagery was limited since most of the airborne data collected until that time contained only RGB information, and satellite VHR data were not yet available to the public. In addition, public image datasets like Google Earth and several annotated datasets (e.g. [69][70][71]) which are often used as benchmarks contain only

RGB information. Thus, artificially generating the missing NIR band could greatly benefit RS applications when: a) requiring the use of RGB airborne data collected before 2010s (e.g. change detection) or publicly available in Google Earth; and b) exploring the performance of novel methods on public annotated RGB benchmark datasets after enriching them with the NIR band.

As mentioned in the above literature review, the SSR-published studies have evaluated their paired data methodologies in the output as a whole without isolating particular bands like the NIR. In addition, to the best of our knowledge, the RGB→NIR literature in total, has exclusively explored only the vegetation category. However, the NIR information has also among others proven useful for general scene recognition [72], mineral mapping [73], plastic litter detection [74][75], and nighttime image generation for the monitoring of environmental and socio-economic dynamics [76]. Thus, the generation of the NIR band would be significant to be explored in more detail and for more land cover categories. In addition, the evaluation of the methods on out-of-domain data is crucial since in typical RS applications the NIR prediction is required on data collected on dates different than those of the training set or referring to different regions with similar spectral characteristics. In both cases, atmospheric optical depth and sun/sensor viewing angle may differ, resulting in radiometric differences (radiometric domain gap).

Concerning UDA, since it has been irrefutably recognized as capable of decreasing the domain discrepancies in the CDSS task, it should be at least logical to test it for the enhancement of cross-domain band generation (e.g. NIR). The objective is to increase the radiometric similarity between the training set (source data) and the out-of-domain set (target data) through UDA, thereby improving the NIR prediction of the out-of-domain set. A practical common scenario would be training a model with recent satellite RGB (input), NIR (output) data, and then applying UDA on airborne RGB past data as a pre-step before using the pretrained model to generate the unavailable past NIR. Unlike the CDSS/CDOD tasks, where spectral similarity is not required when applying UDA (e.g. applying an SS model trained on green roofs (labeled source data) to detect red→green roofs (UDA) (unlabeled target data)), in the cross-domain band generation spectral similarity between the source and target data when applying UDA is significant to predict a reliable missing band.

Attempting to contribute to the previous VHR RGB→NIR prediction literature, this paper evaluates at first a conditional GAN (cGAN) and two attention networks trained on paired data (pixel correspondence between RGB and NIR) on predicting NIR on out-of-domain RGB data, and then attempts to improve the NIR prediction by employing a CycleGAN-based UDA on unpaired RGB bi-temporal data. The unpaired process follows a more general scenario where the source and target RGB patches used in the CycleGAN training do not geographically correspond. In summary, the main contributions can be summarized as follows:

1) We explore the performance of a cGAN and two attention models on predicting NIR on out-of-domain VHR RGB data referring to different regions/sensors/dates than the training set. Prior research regarding NIR prediction on out-of-domain data is extremely limited. In addition, attention models have not been previously tested in the RGB→NIR translation.

2) We explore the possibility of UDA through CycleGAN in improving the NIR prediction on out-of-domain imagery. In three configurations, the effects of batch size and normalization techniques are examined. Through this study, UDA is employed for the first time in the enhancement of cross-domain band generation.

3) We implement the models on three main thematic categories: a) impervious surfaces/urban fabric (manmade objects); b) vegetation (forest, crops); and c) ground. In earlier work, the impervious and ground categories have been completely out-of-focus.

In the following sections, at first the data and methodology are described (section II) and then the results are presented and discussed (section III). Finally, the conclusions are summarized (section IV).

II. Data and Methodology

A. Datasets – Pre-processing

For the implementation of the methodology, Geoeye-1 (GE01) and Worldview-2/3 (WV-2/3) satellite VHR images were employed. The procured images were pan-sharpened by the vendor and contained four bands (RGB-NIR). The spatial resolution for GE01 and WV-2 images was 0.5 m, whereas for WV-3 was 0.3 m. They were collected from four European areas (Granada/Spain (G), Tønsberg/Norway (T), Rhodes/Greece (R), Venice/Italy (V)) in a bi-temporal fashion. The areas were heterogeneous since the morphology differed (G: dense high urban fabric/red-tiled roofs/steep mountains/few agricultural fields, T: sparse low buildings/grey-tiled roofs/flat terrain/high presence of agriculture and forest, R: dense urban fabric with terraces/few crops, V: very dense homogenous buildings/red-tiled roofs/limited vegetation). More info can be found in [77] where this dataset was used for the first time. In the text "1" refers to the earliest date and "2" to the latest (e.g. G1, T2).

The pre-processing steps included: a) the creation of mosaics for WV-3 since the area of interest was depicted in

TABLE I
Dataset Information

Area	Collection date	Satellite	Spatial Resolution (m)	Size (km^2)
Granada	19/7/2013	GE01	0.5	21
	2/7/2018	WV-3	0.3	
Tønsberg	20/9/2013	WV-2	0.5	25
	12/7/2019	GE01	0.5	
Rhodes	23/4/2013	WV-2	0.5	33
	5/6/2019	WV-3	0.3	
Venice	4/5/2013	GE01	0.5	17
	13/5/2018	WV-2	0.5	

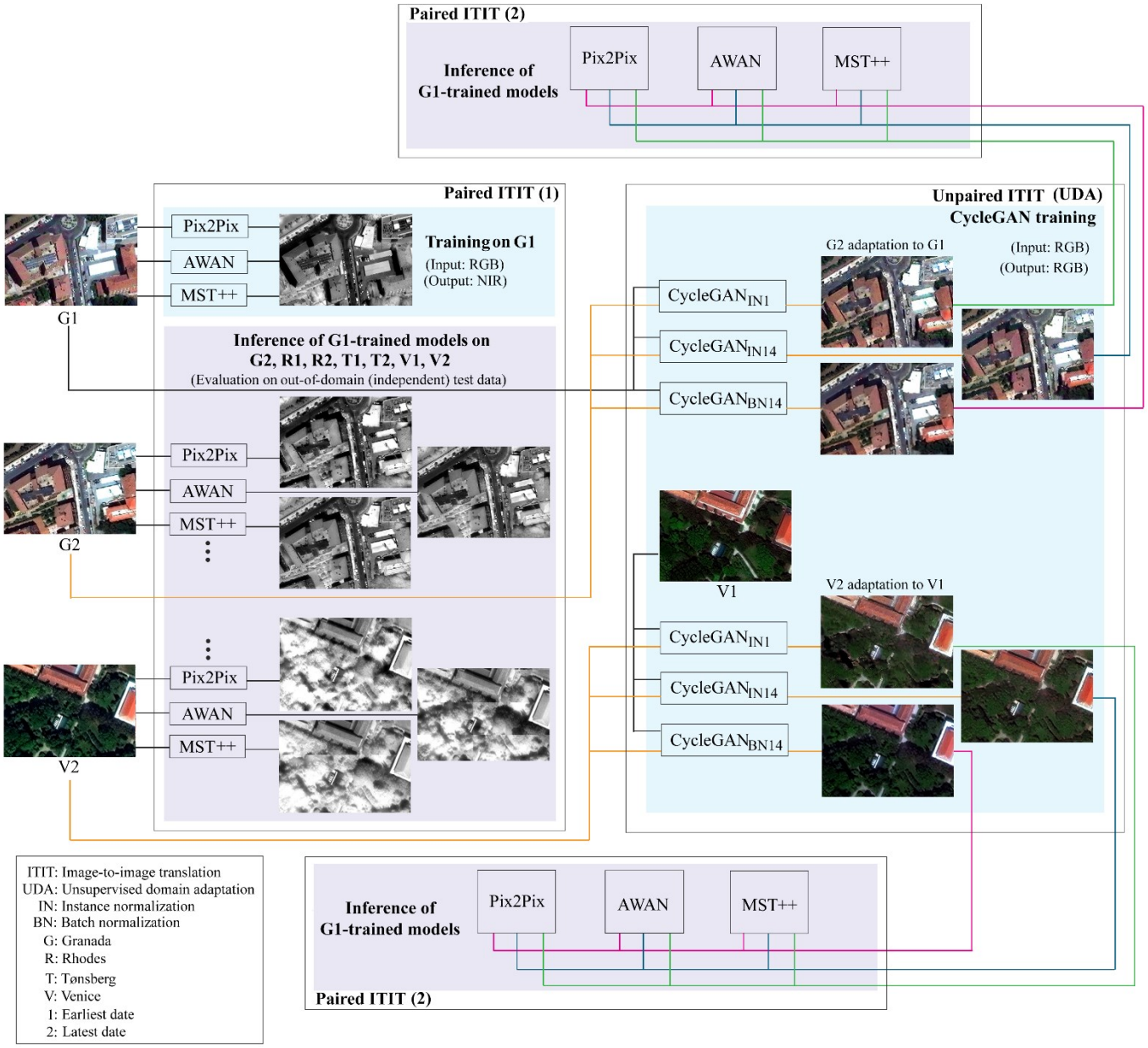

Fig. 1. Flowchart of the methodology.

multiple tiles; b) resampling of the WV-3 images to 0.5 m SR (same as GE01, WV2); c) conversion of image values to 8-bit radiometric resolution (often encountered on airborne images before 2010s and public datasets); and d) normalization. TABLE I shows details about the images. Water areas were masked and not taken into account because they were out of the focus of this study.

B. Method

In this paper, after evaluating a cGAN (Pix2Pix), an adaptive weighted attention network (AWAN) [78], and a ViT (MST++) [79] trained on paired data (each pixel in the RGB source data corresponds to a pixel in the target NIR data) on predicting NIR on out-of-domain data, a CycleGAN-based UDA was employed on unpaired data to improve the NIR prediction on out-of-domain imagery. Pix2Pix and CycleGAN were selected because they have been widely used in the ITIT literature, and AWAN and MST because they won the NTIRE Spectral Reconstruction Challenge in 2020 and 2022 respectively. The models were implemented in Pytorch [80], in a machine with i7-8700K CPU and NVIDIA 1070 Ti GPU. A flowchart of the methodology is provided in Fig. 1.

In more detail, after the data pre-processing, the paired ITIT was applied (Fig. 1/Paired ITIT (1)) where Pix2Pix, AWAN and MST++ were trained on RGB data to predict NIR. All models were trained on the G1 image alone. The performance of the G1-trained models was assessed on seven out-of-domain images (G2, R1, R2, T1, T2, V1, V2). This evaluation permits to give a performance estimation in scenarios where the unavailable NIR image corresponds to a region spectrally similar to a region with available NIR. A possible application could be in the NIR enrichment of annotated benchmark public

RGB datasets.

Following the paired ITIT, the unpaired ITIT was applied (UDA) by training three CycleGAN versions (IN1, IN14, BN14) on the RGB images of Granada and Venice. The effect of different normalization techniques (batch normalization (BN), instance normalization (IN)) and batch sizes (1, 14) was evaluated. The process is unpaired because the G1 RGB patches during the model training did not correspond geographically to the G2 RGB patches (random selection). The same goes for V1/V2. This approach covers a more challenging general scenario.

The above pairs were selected because they were collected in the same month to avoid seasonal changes. The UDA aimed at adapting G2 data to G1, and V2 data to V1, and thus increasing RGB radiometric similarity based: a) on the logical assumption that the G1-trained models (Fig. 1/Paired ITIT (1)) should perform better on G1 compared to G2, since G1 consisted the training set; and b) based on the fact that the G1-trained models performed better on V1 compared to V2 (section III). G2 and V2 are considered out-of-domain data compared to G1 because G2 is collected by a different sensor and on a different date, and V2 refers to a different region. Since CycleGAN uses a cyclic loss function, the reverse adaptation is also implemented during training but it is not of interest in this paper.

The ultimate objective of UDA was to enhance the NIR prediction of G2 and V2 in the paired ITIT. Thus, after UDA, a second inference was performed on the G1-trained Pix2Pix, AWAN, and MST++ models (Fig 1. /Paired ITIT (2)) where the NIR G2 and V2 predictions were compared with the initial predictions (before UDA). It is noted that besides the CycleGAN-based UDA, histogram matching (HM) was also performed for comparison.

1) Paired ITIT – Pix2Pix:

GANs are generative models that learn a mapping from random noise vector z to output image y, $G: z \rightarrow y$ [81]. GANs consist of a generator G and a discriminator D. In image generation applications, the goal of the generator is to produce synthetic (fake) images that challenge the ability of the discriminator to differentiate them from real images. The training is described by the objective function shown in (1) where G aims at minimizing L_{GAN} against an adversarial D that aims at maximizing it.

$$\min_G \max_D L_{\mathrm{GAN}}(G, D) \tag{1}$$

$$L_{\mathrm{GAN}}(G, D) = \mathbb{E}_y [\log D(y)] + \mathbb{E}_z [\log (1 - D(G(z))] \tag{2}$$

Conditional GANs learn a mapping from the observed image x and random noise vector z, to y, $G: \{x, z\} \rightarrow y$ [55]. In this case, the minmax two-player training is performed on (3).

$$L_{\mathrm{cGAN}}(G, D) = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log (1 - D(x, G(x, z))] \tag{3}$$

In our study, Pix2Pix which is a cGAN, was implemented according to the suggestions in [55] with a modification in the number of convolutional layers. It was trained in an alternating way according to (4) which includes the addition of the $L1$ loss (5) to (3) so that the generator except for antagonizing the discriminator is also tasked to produce outputs similar to the ground-truth (reconstruction loss). In (5) λ is a trade-off parameter between $L_{cGAN}(G, D)$ and $L_{L1}(G)$ and was set to 100.

$$\min_G \max_D L_{\mathrm{cGAN}}(G, D) + \lambda L_{L1}(G) \tag{4}$$

$$L_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1] \tag{5}$$

The generator followed a U-Net [56] architecture. It consisted of eight convolutional layers in the encoder part and eight transposed convolutional layers in the decoder part. The respective layers in the original Pix2Pix implementation were four. In our study the number was increased in an attempt to improve the prediction accuracy. Skip connections through concatenation were applied between the encoder and the decoder layers. Dropout was used in three decoder layers with value 0.5 (randomly zeroes some of the input tensor) to provide stochasticity both in training and inference. Concerning the discriminator, it consisted of five convolutional layers and was in a PatchGAN form [55]. BN followed the convolutional layers in both the generator and discriminator. Finally, the dimensions of the generator input patch were 256×256×3 (RGB) and of the output patch were 256×256×1 (NIR prediction).

The training details are shown in TABLE II. The learning rate was set to 2×10^{-4}. The number of epochs was set to 50 for the G1-trained models. The inference time for a 17 km^2 image was 11 s. The selection of the final weights was performed by an empirical process based on the lowest $L1$ loss values in combination with the observation of the performance of the predictions on image samples.

2) Paired ITIT – AWAN:

AWAN contains a backbone with multiple dual residual attention blocks (DRAB). Long and short skip connections form the dual residual learning. An adaptive weighted channel attention (AWCA) module and a patch-level second-order

TABLE II
TRAINING DETAILS

Training details	Pix2Pix	AWAN	MST++
Epochs	50	50	50
Batch size	32	32	10
Patch size	256	64	128
Training steps	1312	350	300
Trainable params G	3,404,801	4,354,045	1,619,625
Trainable params D	175,793	___-	-
Training time (h) (G1/all)	3	3	3
GPU memory usage (GB)	1.6	8	7

non-local (PSNL) module were investigated to capture channel correlations and long-range spatial context.

In our study, the network was implemented according to the suggestions in [78]. The selected DRAB number was 8 and the selected output channel number for the DRAB was 100. The dimensions of the input RGB patch of the network were 64×64×3 and of the output patch were 64×64×1. It is noted that the original output bands for AWAN were 31 but, in our case, it is only the NIR band. To retain consistency with Pix2Pix, $L1$ loss was used. For fair comparison, it was decided to train for a similar amount of time as Pix2Pix, thus, the number of training steps was set accordingly. The training details are shown in TABLE II. The learning rate was set to 1×10^{-4}. The inference time for a 17 km^2 image was 144 s.

3) Paired ITIT – MST++:

MST++ (multi-stage spectral-wise transformer) employs spectral-wise multi-head self-attention (S-MSA) in its basic unit: spectral-wise attention block (SAB). Each spectral feature map is treated as in S-MSA as a token to calculate the spectral self-attention. SABs form a U-shaped single-stage spectral wise transformer (SST) to extract multi-resolution contextual information. MST++ is cascaded by several SSTs to enhance the reconstruction quality from coarse to fine.

In our study, the network was implemented according to the suggestions in [79]. The dimensions of the input RGB patch of the network were 128×128×3 and of the output patch were 128×128×1. It is noted that the original output bands for AWAN were 31 but, in our case, it is only the NIR band. Thus, a convolutional layer was added after the output layer to produce 1 band instead of 31. To retain consistency with Pix2Pix, $L1$ loss was used. For fair comparison, it was decided to train for a similar amount of time as Pix2Pix, thus, the batch size and the number of training steps were set accordingly. The training details are shown in TABLE II. The learning rate was set to 1×10^{-4}. The inference time for a 17 km^2 image was 92 s.

4) Unpaired ITIT (UDA) – CycleGAN:

CycleGAN performs unpaired ITIT with adversarial training. Since the $G: x \rightarrow y$ mapping is under-constrained, an inverse mapping $F: y \rightarrow x$ is also employed and a Cycle-

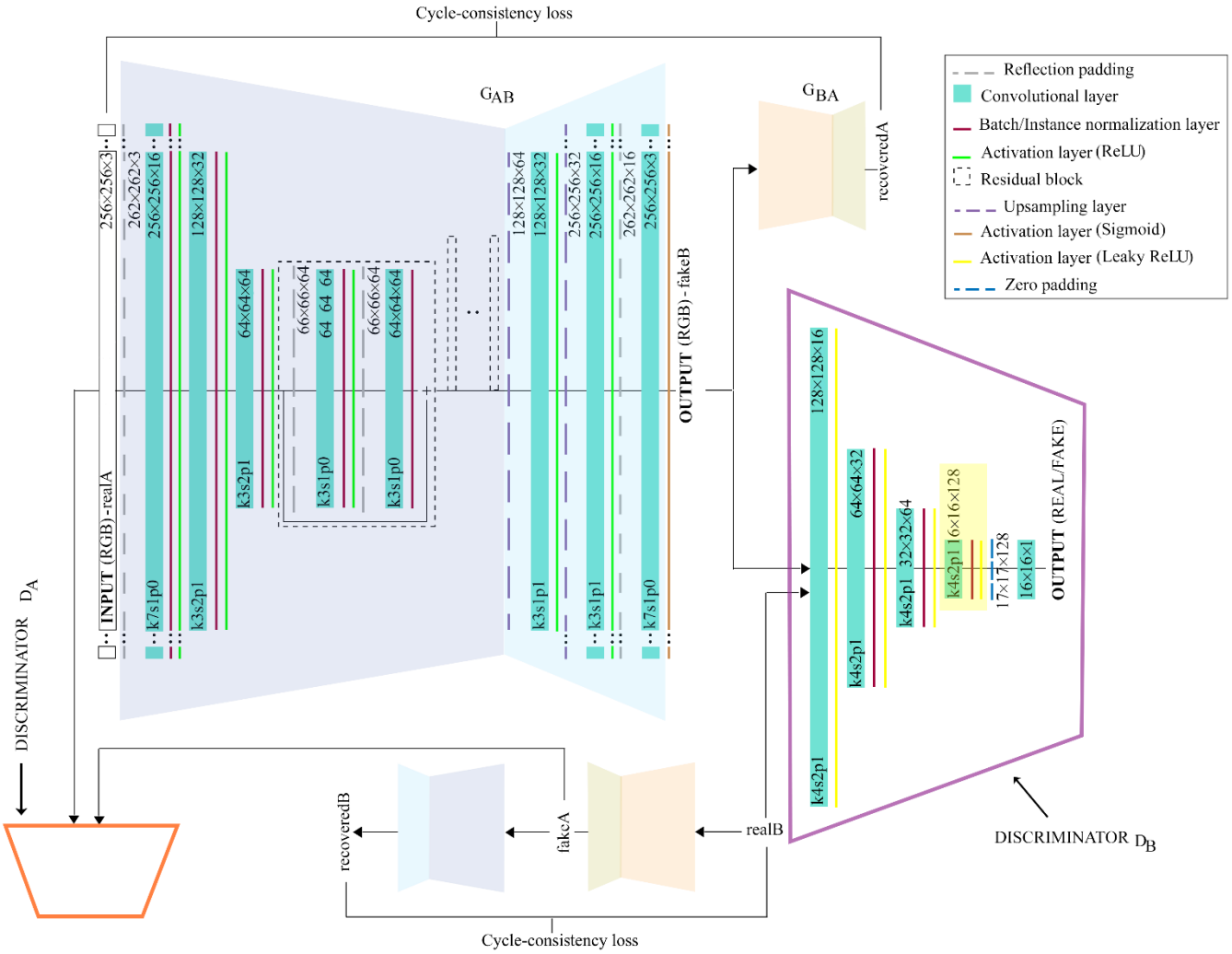

Fig. 2. Architecture of the implemented CycleGAN model. In the first version, nine residual blocks were used in the generator, while in the second and third, three. The yellow highlighted layers in the discriminator were removed in the second and third versions.

consistency loss is introduced to impose $F(G(x)) \approx x$ and $G(F(y)) \approx y$ (6).

$$L_{\text{cyc}}(G,F) = \mathbb{E}_x\left[\|F(G(x)) - x\|_1\right] + \mathbb{E}_y\left[\|G(F(y)) - y\|_1\right] \quad (6)$$

Both mappings are simultaneously trained. The objective function is a combination of the cycle-consistency loss and the adversarial losses. Two discriminators (D_x , D_y) are includedin CycleGAN, one for each mapping. D_x aims at differentiating x from $F(y)$, and D_y aims at distinguishing y from $G(x)$. Except for the adversarial losses for the $G: x \rightarrow y$ and $F: y \rightarrow x$ mapping and the corresponding discriminators (D_y , D_x) , as suggested in [54] an identity loss (7) [82] is additionally utilized to retain color fidelity between the input and the output. Thus, the full objective is expressed in 8.

$$L_{\text{identity}}(G,F) = \mathbb{E}_y[\|G(y) - y\|_1] + \mathbb{E}_x[\|F(x) - x\|_1] \quad (7)$$

$$L(G,F,D_x,D_y) = L_{\text{GAN}}(G,D_y) + L_{\text{GAN}}(F,D_x) + \lambda_1 L_{\text{cyc}}(G,F) + \lambda_2 L_{\text{identity}}(G,F) \quad (8)$$

where λ_1:10 and λ_2:5

In our study, CycleGAN was implemented according to the suggestions in [54]. It is noted that three versions, which differed in the use of normalization layers and batch size, were explored. In the original paper IN was used in the generator and BN in the discriminator. In the first and second versions, IN was employed in the generator and the discriminator with batch sizes 1 and 14 (maximum capacity of the available computer memory) respectively. In the third version, BN was employed with batch size 14. All three versions were trained on the G1/G2 (GIN1, GIN14, GBN14) and the V1/V2 (VIN1, VIN14, VBN14) RGB pair of images.

The architecture of the three versions is shown in Fig. 2. Three convolutional layers and nine residual blocks formed the first version generator encoder, and two Upsampling and three convolutional layers formed the generator decoder. The PatchGAN discriminator consisted of five convolutional layers. In the second and third versions, the difference compared to the first version architecture was the use of three residual blocks instead of nine in the generator, and the removal of a convolutional layer (shown in yellow highlight) in the discriminator to alleviate the computational load. The learning rate for the first version was set constant to $2\text{x}10^{-4}$ for the first 100 epochs and then linear decay to zero was implemented. For the second and third versions, the learning rate decay was implemented for the last ten epochs. Finally, the dimensions of the generator input and output patch were 256×256×3 (RGB). It is noted that since CycleGAN aims at unpaired ITIT, realA and realB patches (Fig. 2) did not geographically match during training since they were selected randomly.

The training details for the three versions are shown in

TABLE III
UDA TRAINING DETAILS

Training details	GIN1	GIN14	GBN14	VIN1	VIN14	VBN14
Epochs	200	80	80	200	80	80
Batch size	1	14	14	1	14	14
Patch size	256	256	256	256	256	256
Training steps	1312	1312	1312	1092	1092	1092
Trainable params G	715,651	272,515	272,515	715,651	272,515	272,515
Trainable params D	175,089	43,057	43,057	175,089	43,057	43,057
Training time	8	26	24	7	21	20
GPU memory usage (GB)	0.5	6	6	0.5	6	6

TABLE III. The inference time for a batch of 32 patches for the first version was 0.113 s and for the second and third was 0.073 s. The selection of the final weights was based on the lowest cycle-consistency loss along with observing image samples.

III. RESULTS AND DISCUSSION

A. Paired ITIT – First stage

The results of the paired ITIT were evaluated quantitatively and qualitatively. For the quantitative evaluation, the root mean square error (RMSE) (14) and the structural similarity (SSIM) (15) [83] were calculated between the predicted NIR and the ground-truth with kernel size 7x7. It is noted that RMSE is sensitive to spectral information, while SSIM is sensitive to geometry. In this paper more importance is assigned to the accuracy of spectral representation.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (14)$$

$$\text{SSIM} = \frac{(2\mu_y\mu_{\hat{y}} + c_1)(2\sigma_{y\hat{y}} + c_2)}{(\mu_y{}^2 + \mu_{\hat{y}}{}^2 + c_1)(\sigma_y{}^2 + \sigma_{\hat{y}}{}^2 + c_2)} \quad (15)$$

where $c_1 = (k_1 L)^3$, $c_2 = (k_2 L)^3$, L: the dynamic range of the pixel values, k_1=0.01, and k_2=0.03.

1) Quantitative evaluation – Evaluation scores:

RMSE and SSIM were estimated for the predicted images produced by Pix2Pix, AWAN, and MST++. It is noted that in contrast to the suggestion in [55], for Pix2Pix, BN was retained in the inference and was not changed to IN because it performed better in preliminary experiments. The scores were calculated for three separate categories: impervious (e.g. buildings, roads), vegetation (forest, crops) and ground. The three categories in the images were delineated by masks that were created in a graphics editor [84]. At first, vegetation was detected by the normalized difference vegetation index (NDVI) [85] and then the ground category was manually masked. The remaining area constituted the impervious category. The evaluation scores are presented in TABLE IV. The bold font indicates the best scores. The mean score values are depicted for each image. Total mean

TABLE IIV
NIR EVALUATION SCORES IN PAIRED ITIT – FIRST STAGE

		impervious			vegetation			ground		
		Pix2Pix	AWAN	MST++	Pix2Pix	AWAN	MST++	Pix2Pix	AWAN	MST++
RMSE→	G2	0.069	0.049	**0.044**	0.125	0.099	**0.092**	0.060	0.051	**0.045**
	R1	0.094	0.079	**0.069**	0.144	**0.076**	0.102	0.096	0.078	**0.072**
	R2	0.129	0.128	**0.106**	0.141	0.126	**0.097**	0.096	0.107	**0.082**
	T1	0.180	**0.160**	0.168	0.187	0.162	**0.139**	**0.190**	0.195	0.201
	T2	0.185	**0.170**	**0.170**	0.220	0.222	**0.187**	0.115	0.136	**0.099**
	V1	0.089	0.081	**0.070**	0.131	0.135	**0.122**	No data		
	V2	0.114	0.101	**0.100**	0.196	0.182	**0.160**			
	total mean	0.123	0.110	**0.104**	0.164	0.143	**0.128**	0.111	0.113	**0.100**
SSIM→	G2	0.883	0.953	**0.961**	0.774	0.861	**0.889**	0.878	0.948	**0.959**
	R1	0.843	0.913	**0.927**	0.599	0.858	**0.860**	0.741	0.902	**0.931**
	R2	0.847	0.900	**0.927**	0.702	0.828	**0.871**	0.834	0.921	**0.950**
	T1	0.641	0.731	**0.736**	0.665	0.746	**0.802**	0.744	0.870	**0.870**
	T2	0.614	0.726	**0.727**	0.656	0.672	**0.775**	0.757	0.856	**0.885**
	V1	0.868	0.917	**0.932**	0.605	0.740	**0.765**	No data		
	V2	0.821	0.877	**0.882**	0.537	0.639	**0.683**			
	total mean	0.788	0.860	**0.870**	0.648	0.763	**0.807**	0.791	0.899	**0.919**

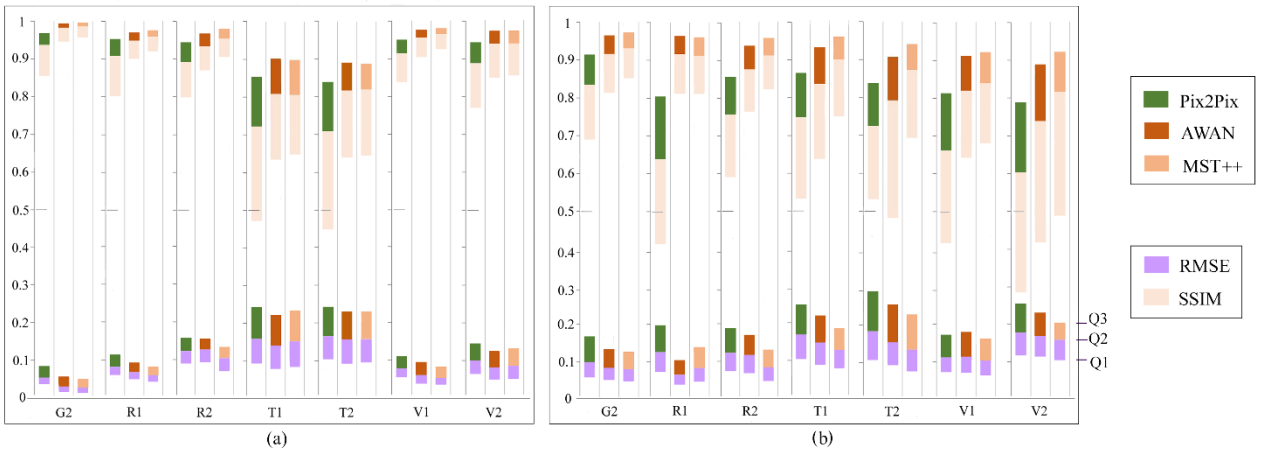

Fig. 3. Boxplots of the evaluation scores in paired ITIT (first stage). (a) Impervious, (b) Vegetation

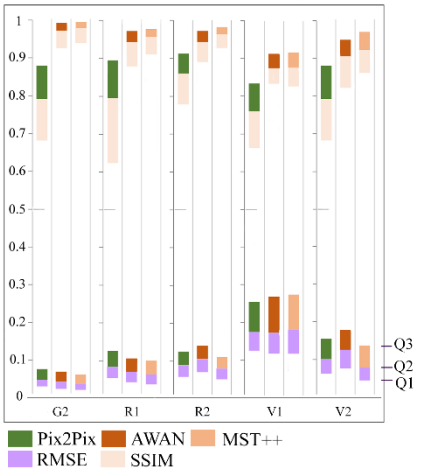

Fig. 4. Boxplots of the evaluation scores in paired ITIT (first stage) for the ground category.

is the average of the mean values for all out-of-domain images.

As explained in section II.B, all models were trained on the G1 image alone and were evaluated on out-of-domain sets (G2, R1, R2, T1, T2, V1, V2). As already mentioned, the out-of-domain evaluation permits the performance estimation when the unavailable NIR image (e.g. airborne photo before 2010s) corresponds to a spectrally similar area with available NIR.

By observing TABLE IV, it can be observed that in all three categories MST++ showed the best overall performance in RMSE and SSIM followed by AWAN and Pix2Pix. The vegetation category appeared to be the most difficult task (total mean RMSE: MST++:0.128, AWAN:0.143, Pix2Pix: 0.164). It is noted that in the impervious category all models showed the worst scores in T1, T2 because the urban fabric spectral characteristics were highly different compared to G1 (different structures/materials). Similarly, regarding vegetation, the least satisfactory performance for all models was also shown in Tønsberg. However, V2 also presented high RMSE and low SSIM. Finally, on the ground category, T1 showed the worst scores.

2) Quantitative evaluation - Boxplots:

Except for quantitatively assessing the paired ITIT by the mean score values, boxplots were also created (Figs. 3, 4). In the boxplots, the first (Q1), second (Q2), and third (Q3) quartiles are depicted. The conclusions produced by observing TABLE IV are consistent with the conclusions drawn by the

boxplots. It is observed that MST++ showed overall the best performance followed by AWAN and Pix2Pix.

3) Quantitative evaluation – Spectral Plots:

To reinforce the visual perception of the results, spectral plots are depicted in Fig. 5. It is noted that since the NIR values correspond to only one band, the spectral curves are diminished to points. Thus, spectral plots were created as follows: At first, the mean values in the ranges: [0.2-0.21], [0.3-0.31], [0.4-0.41], [0.5-0.51], [0.6-0.61], [0.7-0.71] and [0.8-0.81] were calculated for the ground-truth NIR values in each of the three land cover categories. Then, the mean values for the same ranges were calculated for the predictions. As a final step, since predicted and ground-truth values were too close to be discernible in the plot, it was decided to subtract the ground-truth NIR values. Thus, in Fig. 5, y=0 corresponds to the ground-truth NIR values. The closest values to y=0 correspond to mean values closer to the ground-truth.

It was observed that in the impervious category, MST++ showed the closest predictions to the ground-truth for most of the images followed by AWAN and Pix2Pix. The same conclusion can be drawn for the vegetation category. Concerning the ground class, there is less agreement with TABLE IV compared to the impervious and vegetation categories.

The fact that the spectral plots are not always in agreement with the scores in TABLE IV illustrates how different metrics can provide different perspectives on image similarity which can lead to variable conclusions, a known issue in the research community [86]. In this study, RMSE and SSIM employ a

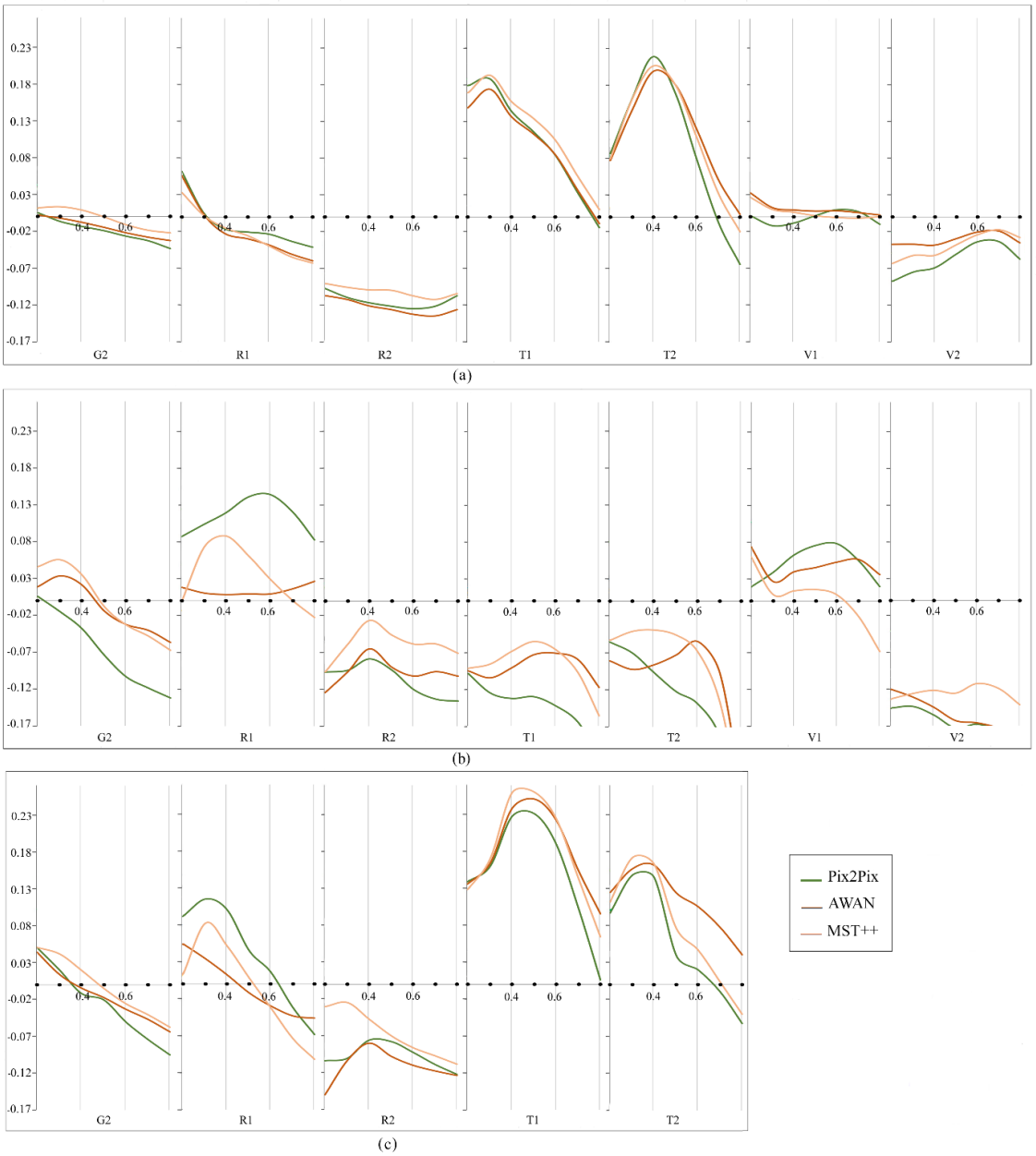

Fig. 5. NIR spectral plots in paired ITIT (first stage). (a) Impervious, (b) Vegetation, (c) Ground

Fig. 6. Samples of pseudo-color composites of the NIR predictions in the paired ITIT (first stage). Red color is assigned to the NIR band, green color to the RED band, and blue color to the GREEN band. (a1-a7) Natural RGB composite, (b1-b7) Ground-truth, (c1-c7) Pix2Pix, (d1-d7) AWAN, (e1-e7) MST++.

detailed spectral comparison through a 7x7 window, thus are more trustworthy metrics than the spectral plots.

4). Qualitative evaluation:

The qualitative evaluation was conducted by visual interpretation and it is complementary to the quantitative (TABLE IV). Samples of pseudo-color composites of the NIR predictions are displayed in Fig. 6. It is noted that it was decided to include pseudo-color composites in the paper instead of the greyscale NIR band to facilitate the visual interpretation. Red color is assigned to the NIR band, green color to the RED band, and blue color to the GREEN band. Fig. 6 shows output samples for each of the seven images (G2, R1, R2, T1, T2, V1, V2) for the three trained models (Pix2Pix, AWAN, MST++). A sample of G1 (training image) is shown in Fig. 7.

(a) (b)

Fig. 7 Sample of G1 training image. (a) Natural RGB composite, (b) Pseudo-color composite containing NIR (Red color: NIR band, green color: RED band, blue color: GREEN band)

Concerning the depiction of the impervious category in Fig. 6, the performance difference among the models can be visually perceived in: a) where Pix2Pix is outperformed in the spectral representation of buildings by MST++ and AWAN; b) R1, with MST++ and AWAN predictions showing better spectral similarity to the ground-truth compared to Pix2Pix; and c) V1 where MST++ outperforms the other two models. It is noted that predictions with RMSE ≲ 0.08 do not visually differ from the ground-truth.

Regarding the depiction of the vegetation category in Fig. 6, the performance differences among the models are visually perceivable in several images: a) G2, where MST++ and AWAN show more satisfactory predicted values compared to Pix2Pix (e.g. top left); a) R1, with AWAN showing the best prediction, followed by MST++; b) R2, where MST++ and AWAN outperform Pix2Pix; c) T2, illustrating MST++ with higher NIR values, which are closer to the original; and d) V1, displaying MST++ as more spectrally similar to the ground-truth. It's noted that predictions with RMSE ≲1 are visually very close to the ground-truth.

Finally, in the ground category, the performance difference among the models can be visually perceived in Fig. 6 in: a) R1 with Pix2Pix showing lower color resemblance compared to the other two models; b) R2 with MST++ depicting higher spectral similarity to the ground-truth; and b) T2 where MST++ (low left edge) predicts more accurately the original values

B. Paired ITIT – after UDA

As explained in section II.B, in this study, UDA (unpaired ITIT) was also implemented to improve the NIR prediction of G2 and V2 (produced by Pix2Pix, AWAN and MST++) regarding the paired ITIT of the first stage. The UDA aimed at adapting G2 data to G1, and V2 data to V1, and thus increasing radiometric similarity based: a) on the logical assumption that the G1-trained models (Fig. 1/Paired ITIT (1)) should perform better on G1 compared to G2, since G1 consisted the training set; and b) based on the fact that the G1-trained models performed better on V1 compared to V2 (TABLE IV).

In more detail, the three G2 and V2 domain-adapted RGB outputs produced by CycleGAN (IN1, IN14, BN14), acted as input to the G1-trained Pix2Pix, AWAN, and MST++ models (Fig 1. /Paired ITIT (2)) and inference was implemented. The results of the inference (NIR predictions with domain-adapted RGB inputs) were compared with the initial predictions (before

TABLE V

NIR EVALUATION SCORES IN PAIRED ITIT AFTER UDA

	RMSE↓						SSIM↑					
	G2			V2			G2			V2		
	impervious	vegetation	ground	impervious	vegetation	ground	impervious	vegetation	ground	impervious	vegetation	ground
Pix2Pix												
Before UDA	**0.069**	**0.125**	**0.060**	0.114	0.196	No data	**0.883**	**0.774**	0.878	0.821	0.537	No data
IN1	0.147	0.222	0.107	0.164	0.194		0.472	0.336	0.361	0.513	0.357	
IN14	0.122	0.207	0.098	0.139	0.169		0.598	0.442	0.502	0.662	0.460	
BN14	0.089	0.142	0.063	**0.101**	**0.130**		0.744	0.607	0.715	0.793	0.586	
HM	0.070	0.152	0.063	0.114	0.156		0.881	0.749	**0.879**	**0.838**	**0.615**	
AWAN												
Before UDA	0.049	**0.099**	**0.051**	0.101	0.182	No data	0.953	**0.861**	**0.948**	**0.877**	**0.639**	No data
IN1	0.146	0.226	0.106	0.182	0.200		0.475	0.334	0.373	0.502	0.366	
IN14	0.117	0.204	0.096	0.151	0.173		0.617	0.462	0.529	0.664	0.491	
BN14	0.081	0.176	0.059	0.115	**0.145**		0.785	0.632	0.770	0.814	0.628	
HM	**0.048**	0.123	0.052	0.134	0.179		**0.955**	0.846	0.947	0.848	0.636	
MST++												
Before UDA	**0.044**	**0.092**	**0.045**	0.100	0.160	No data	**0.961**	**0.889**	**0.959**	**0.882**	**0.683**	No data
IN1	0.144	0.209	0.101	0.167	0.188		0.484	0.357	0.379	0.521	0.386	
IN14	0.115	0.186	0.094	0.137	0.157		0.625	0.492	0.539	0.683	0.517	
BN14	0.078	0.142	0.060	**0.098**	**0.123**		0.795	0.684	0.776	0.833	0.671	
HM	0.045	0.106	0.051	0.124	0.151		**0.961**	0.880	0.955	0.874	0.709	

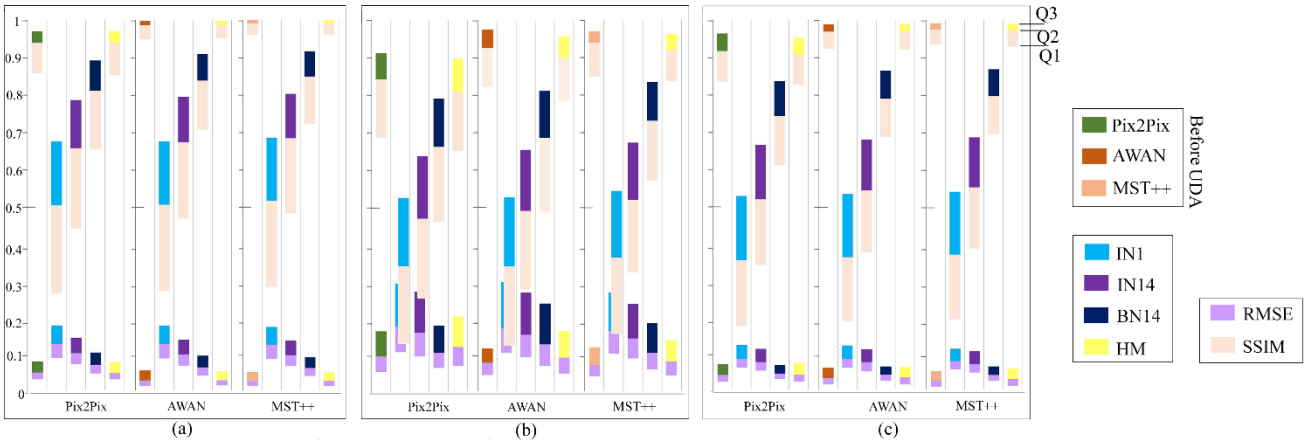

Fig. 8. Boxplots of the evaluation scores in paired ITIT after UDA for the G2 region. (a) Impervious, (b) Vegetation, c) Ground

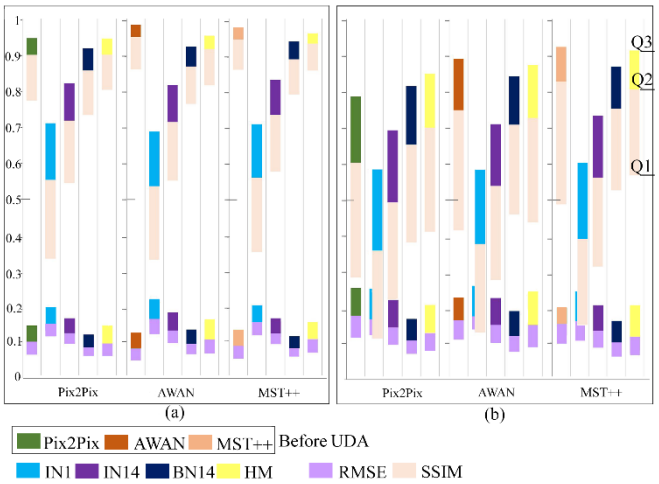

Fig. 9 Boxplots of the evaluation scores in paired ITIT after UDA for the V2 region. (a) Impervious, (b) Vegetation

UDA) through RMSE and SSIM scores (kernel size:7x7), as well as qualitatively. The effect of different normalization techniques (BN, IN) and batch sizes (1, 14) was evaluated. It is noted that besides the CycleGAN-based UDA, HM was also performed for comparison.

As already mentioned in section II.B, the Granada and Venice pairs were chosen to minimize seasonal variation, as they were gathered during the same month. G2 and V2 are deemed out-of-domain data in comparison to G1. G2 was gathered using a different sensor and at a different time, while V2 represents a different region.

1) Quantitative evaluation – Evaluation scores:

The RMSE and SSIM scores are presented in TABLE V. It is observed that the NIR predictions of all three models (Pix2Pix, AWAN, MST++) on the outputs of the BN14 CycleGAN version, were significantly enhanced in terms of RMSE compared to the ones before UDA for the vegetation category in the V2 region. In more detail, the RMSE improved in a) Pix2Pix from 0.196 to 0.130; b) AWAN from 0.182 to 0.145; and c) MST++ from 0.160 to 0.123. The respective V1 vegetation RMSE values (TABLE IV) were 0.131, 0.135, and 0.122. Thus, in Pix2Pix and MST++ the maximum possible NIR prediction enhancement in terms of RMSE was achieved. A smaller RMSE improvement also occurred on the outputs of the IN14 CycleGAN version and HM. Consequently, BN versions of CycleGAN appear the most promising for the enhancement of high RMSE scores. It is noted that significant SSIM change was not observed.

It is also remarked that notable RMSE improvements were not observed in G2 and the impervious V2 category. We believe that the reason is the fact that the initial RMSE scores (before UDA) in G2 were significantly lower than the one of V2 vegetation due to higher radiometric difference in the respective RGB domains, i.e. the G2 RGB radiometry was similar to G1 RGB, while the V2 RGB vegetation radiometry was highly dissimilar to V1 RGB.

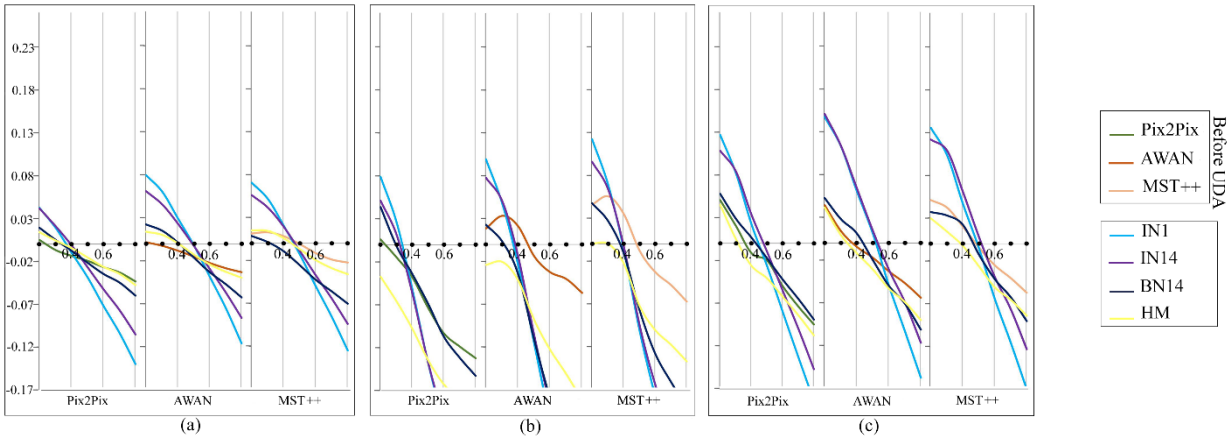

Fig. 10. NIR spectral plots in paired ITIT after UDA for the G2 region. (a) Impervious, (b) Vegetation, (c) Ground

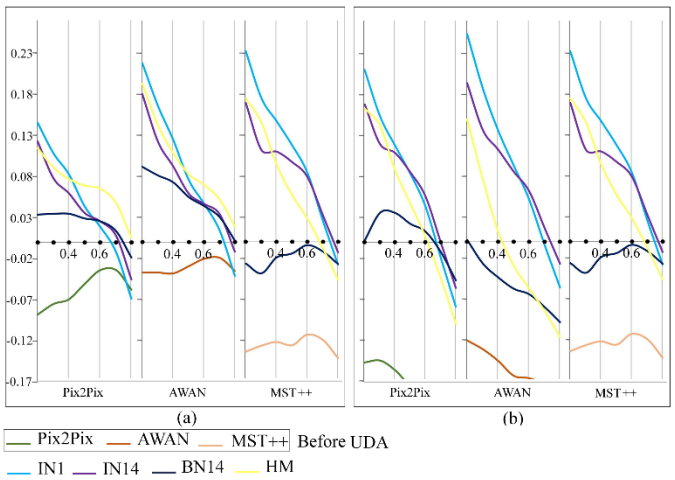

Fig. 11. NIR spectral plots in paired ITIT after UDA for the V2 region. (a) Impervious, (b) Vegetation

2) Quantitative evaluation - Boxplots:

Boxplots were also created for the quantitative assessment (Figs. 8, 9). The observation of the boxplots leads to the same conclusions as the observation of TABLE V. In more detail, the BN14 version of CycleGAN greatly improved the RMSE in all NIR prediction models (Pix2Pix, AWAN, MST++) compared to the ones before UDA for the vegetation category in V2. The outputs of the IN14 version and HM showed a smaller improvement. Significant advancements are not seen in G2 and the impervious V2 category.

3) Quantitative evaluation – Spectral plots:

To enhance the visual interpretation of the results, spectral plots are depicted in Figs. 9, 10. The approach used to create the curves is already explained in section III.A.3.

The spectral plots referring to G2 (Fig. 10) are in agreement with TABLE V, i.e. substantial improvements did not occur for any of the three land cover classes. As already mentioned, the reason for this could be the much lower initial RMSE scores (before UDA) compared to those of the V2 vegetation due to higher radiometric similarity in the respective RGB domains. Regarding the spectral plots referring to V2 (Fig. 11), for the vegetation category, the superiority of the BN14 CycleGAN version is once again highlighted for the three NIR prediction models (Pix2Pix, AWAN, MST++) as it showed the closest predictions to the ground-truth.

Fig. 12. Samples of pseudo-color composites of the NIR predictions in paired ITIT after UDA. (a1-a6) Before UDA, (b1-b6) IN1, (c1-c6) IN14, (d1-d6) BN14, (e1-e6) HM.

It is noted that in the impervious category of V2, the spectral plots corresponding to the BN14 CycleGAN version of Pix2Pix and MST++, also show the closest predictions to the ground-truth. However, since this conclusion is not rigidly supported in TABLE V, it is not considered trustworthy. As already mentioned in section III.A.3, the RMSE score implements a more detailed spectral comparison, thus, in this paper it carries more weight.

4). Qualitative evaluation:

The qualitative evaluation was conducted by visual interpretation additionally to the quantitative (TABLE V). Samples of pseudo-color composites of the NIR predictions are displayed in Fig. 12. This figure shows the outputs of the three NIR prediction models (Pix2Pix, AWAN, MST++) when feeding the domain adapted RGB CycleGAN outputs (IN1, IN14, BN14) as well as the HM adapted images. The ground-truth and the NIR predictions before UDA are also shown.

In Fig. 12 it can be visually perceived that the BN14 version of CycleGAN managed to produce the best enhancement in the spectral similarity of the vehetattion NIR prediction to the ground-truth in Pix2Pix, AWAN and MST++. The IN14 version of CycleGAN and HM produced smaller enhancements. Significant improvement is not seen in G2 and the impervious V2 category.

IV. Conclusion

In this study, a conditional GAN (Pix2Pix) and two attention networks (AWAN, MST++) were initially assessed on predicting NIR on out-of-domain RGB bi-temporal data. The properties of the out-of-domain RGB data were those typically required in RS NIR prediction tasks: different regions/sensors/dates than the training set. The three NIR prediction models were trained in a single Granada image (G1) and evaluated on seven out-of-domain heterogenous images referring to Granada, Rhodes, Tønsberg, and Venice (G2, R1, R2, T1, T2, V1, V2). It is remarked that former research regarding NIR prediction on out-of-domain data is extremely scarce. However, such research is significant for the NIR enrichment of a) airborne RGB images collected before 2010s when NIR imagery was limited; and b) a large number of public annotated RGB datasets often used as benchmarks.

It is also remarked that attention models have not been tested before in the RGB-to-NIR translation.

In a subsequent step, the study's attention was directed toward the increase of the radiometric similarity between the source and target RGB data by employing CycleGAN-based unsupervised domain adaptation (UDA) on unpaired data. The ultimate objective of UDA was the improvement of the initial NIR prediction. The unpaired process followed a more general scenario where the source and target RGB patches used in the CycleGAN training did not geographically correspond. To implement the UDA, three CycleGAN versions (IN1, IN14, BN14) were explored where the effect of different normalization techniques (batch normalization (BN), instance normalization (IN)) and batch sizes (1, 14) was evaluated. To implement the UDA, bi-temporal pairs collected on the same month to avoid seasonal changes were selected. Thus, the aim was to adapt G2 (target) to G1 (source) and V2 (target) to V1 (source) based: a) on the logical assumption that the G1-trained models should perform better on G1 compared to G2, since G1 consisted the training set; and b) based on the fact that the G1-trained models predicted NIR more accurately on V1 compared to V2. It is highlighted that through this study, UDA is employed for the first time in the enhancement of cross-domain band generation.

In all NIR prediction experiments, the assessment was conducted quantitatively and qualitatively on three main land cover categories (impervious surfaces/urban fabric, vegetation, ground), thus contrary to prior work, the impervious and ground classes were not overlooked.

It was shown that MST++ (vision transformer) produced the most satisfactory out-of-domain NIR predictions in all three land cover classes and that UDA through the BN14 CycleGAN version managed to significantly enhance the NIR prediction when there was a substantial RGB radiometric difference (radiometric domain gap).

For future work, the prediction of shortwave-infrared (SWIR) information will be explored. In addition, the feasibility of UDA between different geographic areas with similar spectral characteristics and the spectral enrichment of annotated public datasets will be investigated.

Acknowledgment

This research was conducted in the framework of the THETIDA project (Grant Agreement 101095253) funded by the European Union (HORIZON-CL2-2022-HERITAGE-01).

References

[1] 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. Photogramm. Remote Sens.*, vol. 179, no. July, pp. 14–34, Sep. 2021.

[2] Z. Guo, H. Guo, X. Liu, W. Zhou, Y. Wang, and Y. Fan, "Sar2color: Learning imaging characteristics of SAR images for SAR-to-Optical transformation," *Remote Sens.*, vol. 14, no. 15, p. 3740, Aug. 2022.

[3] J. Wei *et al.*, "CFRWD-GAN for SAR-to-Optical Image Translation," *Remote Sens.*, vol. 15, no. 10, p. 2547, May 2023.

[4] Y. Pan, I. Ahmed Khan, and H. Meng, "SAR-to-optical image translation using multi-stream deep ResCNN of information reconstruction," *Expert Syst. Appl.*, vol. 224, no. November 2022, p. 120040, Aug. 2023.

[5] Q. Luo, H. Li, Z. Chen, and J. Li, "ADD-UNet: An Adjacent Dual-Decoder UNet for SAR-to-Optical Translation," *Remote Sens.*, vol. 15, no. 12, p. 3125, Jun. 2023.

[6] C. Zhang *et al.*, "A domain adaptation neural network for change detection with heterogeneous optical and SAR remote sensing images," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 109, no. April, p. 102769, May 2022.

[7] A. Manocha and Y. Afaq, "Optical and SAR images-based image translation for change detection using generative adversarial network (GAN)," *Multimed. Tools Appl.*, vol. 82, no. 17, pp. 26289–26315, Jul. 2023.

[8] Y. Qing, J. Zhu, H. Feng, W. Liu, and B. Wen, "Two-way generation of high-resolution EO and SAR images via dual distortion-adaptive GANs," *Remote Sens.*, vol. 15, no. 7, 2023.

[9] H. Wang, Z. Zhang, Z. Hu, and Q. Dong, "SAR-to-Optical image translation with hierarchical latent features," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.

[10] Y. Liu *et al.*, "CscGAN: Conditional scale-consistent generation network for multi-level Remote Sensing image to map translation," *Remote Sens.*, vol. 13, no. 10, p. 1936, May 2021.

[11] X. Chen *et al.*, "SMAPGAN: Generative adversarial network-based semisupervised styled map tile generation method," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4388–4406, May 2021.

[12] J. Song, J. Li, H. Chen, and J. Wu, "RSMT: A Remote Sensing image-to-map translation model via adversarial deep transfer learning," *Remote Sens.*, vol. 14, no. 4, p. 919, Feb. 2022.

[13] Y. Fu, Z. Fang, L. Chen, T. Song, and D. Lin, "Level-aware consistent multilevel map translation from satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. c, pp. 1–14, 2023.

[14] J. Song, H. Chen, C. Du, and J. Li, "Semi-MapGen: Translation of Remote Sensing image into map via semisupervised adversarial learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–19, 2023.

[15] P. Ghamisi and N. Yokoya, "IMG2DSM: Height simulation from single imagery using conditional generative adversarial net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 794–798, May 2018.

[16] E. Panagiotou, G. Chochlakis, L. Grammatikopoulos, and E. Charou, "Generating elevation surface from a single RGB remotely sensed image using deep learning," *Remote Sens.*, vol. 12, no. 12, p. 2002, Jun. 2020.

[17] H. A. Amirkolaee and H. Arefi, "Height estimation from single aerial images using a deep convolutional encoder-decoder network," *ISPRS J. Photogramm. Remote Sens.*, vol. 149, no. July 2018, pp. 50–66, Mar. 2019.

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

[19] K. Kim *et al.*, "Nighttime reflectance generation in the visible band of satellites," *Remote Sens.*, vol. 11, no. 18, p. 2087, Sep. 2019.

[20] J. Kim, S. Ryu, J. Jeong, D. So, H. Ban, and S. Hong, "Impact of satellite sounding data on virtual visible imagery generation using conditional generative adversarial network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 4532–4541, 2020.

[21] K.-H. Han, J.-C. Jang, S. Ryu, E.-H. Sohn, and S. Hong, "Hypothetical visible bands of Advanced Meteorological Imager Onboard the Geostationary Korea Multi-Purpose Satellite -2A using data-to-data translation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, no. L, pp. 8378–8388, 2022.

[22] J. Xiao, D. Tortei, E. Roura, and G. Loianno, "Long-range UAV thermal geo-localization with satellite imagery," Jun. 2023.

[23] J.-E. Park, G. Kim, and S. Hong, "Green band generation for Advanced Baseline Imager Sensor using Pix2Pix with Advanced Baseline Imager and Advanced Himawari Imager observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6415–6423, Aug. 2021.

[24] H.-S. Ryu, J.-E. Park, J. Jeong, and S. Hong, "Generation of hypothetical radiances for missing green and red bands in Geostationary Environment Monitoring Spectrometer," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 9025–9037, 2023.

[25] M. Wu *et al.*, "Remote sensing image colorization using symmetrical multi-scale DCGAN in YUV color space," *Vis. Comput.*, vol. 37, no. 7, pp. 1707–1729, Jul. 2021.

[26] M. Wu *et al.*, "Remote Sensing image colorization based on multiscale SEnet GAN," in *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2019, pp. 1–6.

[27] S. Galliani, C. Lanaras, D. Marmanis, E. Baltsavias, and K. Schindler, "Learned spectral super-resolution," Mar. 2017.

[28] S. Paul and D. Nagesh Kumar, "Transformation of multispectral data to quasi-hyperspectral data using convolutional neural network regression," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3352–3368, Apr. 2021.

[29] L. Liu, S. Lei, Z. Shi, N. Zhang, and X. Zhu, "Hyperspectral Remote Sensing imagery generation from RGB images based on Joint Discrimination," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 7624–7636, 2021.

[30] L. Liu, Z. Shi, Y. Zao, and H. Chen, "Hyperspectral image generation from RGB images with semantic and spatial distribution consistency," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, vol. 2022-July, pp. 1804–1807.

[31] S. Liu, H. Li, G. Zhang, B. Hu, and J. Chen, "Using hyperspectral reconstruction for multispectral images change detection," in *2022 7th International Conference on Image, Vision and Computing (ICIVC)*, 2022, pp. 183–188.

[32] L. Deng *et al.*, "M2H-Net: A reconstruction method for hyperspectral remotely sensed imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, no. February, pp. 323–348, Mar. 2021.

[33] T. Li and Y. Gu, "Progressive spatial–spectral joint network for hyperspectral image reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[34] X. Zheng, W. Chen, and X. Lu, "Spectral super-resolution of multispectral images using spatial–spectral residual attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[35] Z. Liu, H. Zhu, and Z. Chen, "Adversarial spectral super-resolution for multispectral imagery using spatial spectral feature attention module," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 1550–1562, 2023.

[36] G. Zhao, Y. He, Z. Wang, and H. Wu, "Hybrid Transformer Architecture for Spectral Super-Resolution Reconstruction of Multispectral Images," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2024, pp. 9468–9471.

[37] N. Wang, S. Mei, Y. Zhang, M. Ma, and X. Zhang, "Hyperspectral Image Reconstruction From RGB Input Through Highlighting Intrinsic Properties," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–13, 2024.

[38] H. Zhou, Z. Liu, Z. Huang, X. Wang, W. Su, and Y. Zhang, "ICTH: Local-to-Global Spectral Reconstruction Network for Heterosource Hyperspectral Images," *Remote Sens.*, vol. 16, no. 18, 2024.

[39] M. Aslahishahri *et al.*, "From RGB to NIR: Predicting of near infrared reflectance from visible spectrum aerial images of crops," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, vol. 2021-Octob, pp. 1312–1322.

[40] S. Illarionova, D. Shadrin, A. Trekin, V. Ignatiev, and I. Oseledets, "Generation of the NIR spectral band for satellite images with convolutional neural networks," *Sensors*, vol. 21, no. 16, p. 5646, Aug. 2021.

[41] D. C. de Lima, D. Saqui, S. A. T. Mpinda, and J. H. Saito, "Pix2Pix network to estimate agricultural near infrared images from RGB Data," *Can. J. Remote Sens.*, vol. 48, no. 2, pp. 299–315, Mar. 2022.

[42] A. Picon *et al.*, "Deep convolutional neural network for damaged vegetation segmentation from RGB images based on virtual NIR-channel estimation," *Artif. Intell. Agric.*, vol. 6, pp. 199–210, 2022.

[43] X. Yuan, J. Tian, and P. Reinartz, "Learning-based near-infrared band simulation with applications on large-scale landcover classification," *Sensors*, vol. 23, no. 9, p. 4179, Apr. 2023.

[44] Y. Cai *et al.*, "BiFDANet: Unsupervised bidirectional domain adaptation for semantic segmentation of Remote Sensing images," *Remote Sens.*, vol. 14, no. 1, p. 190, Jan. 2022.

[45] Y. Zhao, P. Guo, Z. Sun, X. Chen, and H. Gao, "ResiDualGAN: Resize-residual DualGAN for cross-domain Remote Sensing images semantic segmentation," *Remote Sens.*, vol. 15, no. 5, p. 1428, Mar. 2023.

[46] S. Luo, L. Ma, X. Yang, D. Luo, and Q. Du, "Self-Training based Unsupervised Domain Adaptation for Object Detection in Remote Sensing Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024.

[47] J. Zeng *et al.*, "Unsupervised domain adaptation for remote sensing semantic segmentation with the 2D discrete wavelet transform," *Sci. Rep.*, vol. 14, no. 1, p. 23552, 2024.

[48] K. Sasaki, T. Sekine, and W. Emery, "Enhancing the Detection of Coastal Marine Debris in Very High-Resolution Satellite Imagery

via Unsupervised Domain Adaptation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 6014–6028, 2024.

[49] B. Zhang, T. Chen, and B. Wang, "Curriculum-style local-to-global adaptation for cross-domain Remote Sensing image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.

[50] L. Wang, P. Xiao, X. Zhang, and X. Chen, "A fine-grained unsupervised domain adaptation framework for semantic segmentation of Remote Sensing images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 4109–4121, 2023.

[51] L. Wu, M. Lu, and L. Fang, "Deep covariance alignment for domain adaptive Remote Sensing image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.

[52] S. F. Ismael, K. Kayabol, and E. Aptoula, "Unsupervised Domain Adaptation for the Semantic Segmentation of Remote Sensing Images via One-Shot Image-to-Image Translation," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[53] R. Chen, W. Huang, B. Huang, F. Sun, and B. Fang, "Reusing discriminators for encoding: Towards unsupervised image-to-image translation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 8165–8174, Feb. 2020.

[54] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 2242–2251, Mar. 2017.

[55] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[56] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, pp. 234–241.

[57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, vol. 2016-Decem, pp. 770–778.

[58] V. Tiwari, V. Kumar, K. Pandey, R. Ranade, and S. Agrawal, "Simulation of the hyperspectral data using multispectral data," *Int. Geosci. Remote Sens. Symp.*, vol. 2016-Novem, pp. 6157–6160, 2016.

[59] N. T. Hoang and K. Koike, "Transformation of Landsat imagery into pseudo-hyperspectral imagery by a multiple regression-based model with application to metal deposit-related minerals mapping," *ISPRS J. Photogramm. Remote Sens.*, vol. 133, pp. 157–173, Nov. 2017.

[60] X. Han, J. Yu, J. Luo, and W. Sun, "Reconstruction from multispectral to hyperspectral image using spectral library-based dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1325–1335, Mar. 2019.

[61] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Spectral super resolution of hyperspectral images via coupled dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2777–2797, May 2019.

[62] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers Tiramisu: Fully convolutional denseNets for semantic segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, vol. 2017-July, pp. 1175–1183.

[63] K. Arai, K. Gondoh, O. Shigetomi, and - Yuko, "Method for NIR reflectance estimation with visible camera data based on regression for NDVI estimation and its application for insect damage detection of rice paddy fields," *Int. J. Adv. Res. Artif. Intell.*, vol. 5, no. 11, pp. 17–22, 2016.

[64] K. Arai, M. Sakashita, O. Shigetomi, and Y. Miura, "Estimation of protein content in rice crop and nitrogen content in rice leaves through regression analysis with NDVI derived from camera mounted radio-control helicopter," *Int. J. Adv. Res. Artif. Intell*, vol. 3, 2014.

[65] P. J. Navarro, F. Pérez, J. Weiss, and M. Egea-Cortines, "Machine learning and computer vision system for phenotype data acquisition and analysis in plants," *Sensors*, vol. 16, no. 5, p. 641, 2016.

[66] W. Li, R. Dong, H. Fu, and L. Yu, "Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks," *Remote Sens.*, vol. 11, no. 1, p. 11, 2018.

[67] R. Szeliski, *Computer vision: algorithms and applications*. Springer Nature, 2022.

[68] D. C. de Lima, D. Saqui, S. Ataky, L. A. de C. Jorge, E. J. Ferreira, and J. H. Saito, "Estimating agriculture NIR images from Aerial RGB data," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11536 LNCS, 2019, pp. 562–574.

[69] MAXAR, "Satellite imagery for natural disasters." [Online]. Available: https://www.maxar.com/open-data. [Accessed: 17-Jul-2023].

[70] R. Caye Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Underst.*, vol. 187, p. 102783, Oct. 2019.

[71] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for Remote Sensing image change detection," *Remote Sens.*, vol. 12, no. 10, p. 1662, May 2020.

[72] J. Fan, T. Chen, and S. Lu, "Unsupervised feature learning for land-use scene recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2250–2261, Apr. 2017.

[73] R. Jain and R. U. Sharma, "Airborne hyperspectral data for mineral mapping in Southeastern Rajasthan, India," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 81, no. January, pp. 137–145, Sep. 2019.

[74] M. Kremezi *et al.*, "Pansharpening PRISMA data for marine plastic litter detection using plastic litter indexes," *IEEE Access*, vol. 9, pp. 61955–61971, 2021.

[75] M. Kremezi *et al.*, "Increasing the Sentinel-2 potential for marine plastic litter monitoring through image fusion techniques," *Mar. Pollut. Bull.*, vol. 182, p. 113974, Sep. 2022.

[76] X. Huang, D. Xu, Z. Li, and C. Wang, "Translating multispectral imagery to nighttime imagery via conditional generative adversarial networks," in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 6758–6761.

[77] V. Kristollari and V. Karathanassi, "Change detection in VHR imagery with severe co-registration errors using deep learning: A comparative study," *IEEE Access*, vol. 10, pp. 33723–33741, 2022.

[78] J. Li, C. Wu, R. Song, Y. Li, and F. Liu, "Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2020-June, pp. 1894–1903, 2020.

[79] Y. Cai *et al.*, "MST++: Multi-stage Spectral-wise Transformer for Efficient Spectral Reconstruction," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2022-June, pp. 744–754, 2022.

[80] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, vol. 32.

[81] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2014.

[82] Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image Generation," *arXiv Prepr. arXiv1611.02200*, Nov. 2016.

[83] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[84] Adobe Inc., "Adobe Photoshop." .

[85] D. W. Rouse, J.W. , Haas, R.H., SChell, J.A., Deering, "Monitoring vegetation systems in the great plains with ERTS," in *Third Earth Resources Technology Satellite-1 Symposium*, 1973, vol. 1, pp. 309–317.

[86] H. B. Mitchell, "Image Similarity Measures," in *Image Fusion: Theories, Techniques and Applications*, Berlin, Heidelberg:

Springer Berlin Heidelberg, 2010, pp. 167–185.

[87] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. USA: Prentice-Hall, Inc., 2006.

Viktoria Kristollari received the Integrated M.S. degree in rural, surveying, and geoinformatics engineering from the National Technical University of Athens (NTUA), Greece, in 2016. She also received her Ph.D. degree in artificial intelligence and remote sensing from NTUA in 2024. Since 2017, she has been a Researcher with the Laboratory of Remote Sensing, NTUA (RSLab-NTUA), and has participated in several EU-funded projects. She has authored multiple publications for peer-reviewed international scientific journals and conferences. Her research interests focus on artificial neural networks for multispectral/hyperspectral earth observation data applications.

She was a recipient of the Excellence Award of the Limmat Stiftung Foundation, in 2016, and was granted a three-year scholarship by the NTUA Research Committee for her Ph.D. research, in 2017.

Vassilia Karathanassi received the B.S. degree in rural and surveying engineering from the National Technical University of Athens (NTUA), Greece, in 1984, the M.S. degree in urban planning-geography from Paris V, France, in 1985, and the Ph.D. degree in remote sensing from NTUA, in 1990.

Since 2000, she has been a Professor with the School of Rural, Surveying, and Geoinformatics Engineering, NTUA, specialized in hyperspectral/multispectral remote sensing and InSAR/DInSAR processing and applications. She teaches multiple undergraduate and postgraduate courses and she has supervised more than 40 undergraduate, eight master's theses, eight Ph.D. theses (four of them completed), and one postdoctoral research. She has published research work includes more than 100 articles and one chapter in the book Hyperspectral Remote Sensing. Furthermore, she is involved in EU and national excellence/competitive research projects as a Coordinator, a Principal Investigator, and a Researcher toward the design, development, and validation of state-of-the-art methodologies, and cutting-edge technology in remote sensing and earth observation.