*remote sensing*

MDPI

*Article*

# Multispectral Band-Aware Generation of Satellite Images across Domains Using Generative Adversarial Networks and Contrastive Learning

**Arpan Mahara * and Naphtali Rishe**

Knight Foundation School of Computing and Information Sciences (KFSCIS), Florida International University, 11200 SW 8th St CASE 352, Miami, FL 33199, USA; rishen@cs.fiu.edu

* Correspondence: amaha038@fiu.edu; Tel.: +1-1334-492-0242

**Abstract:** Generative models have recently gained popularity in remote sensing, offering substantial benefits for interpreting and utilizing satellite imagery across diverse applications such as climate monitoring, urban planning, and wildfire detection. These models are particularly adept at addressing the challenges posed by satellite images, which often exhibit domain variability due to seasonal changes, sensor characteristics, and, especially, variations in spectral bands. Such variability can significantly impact model performance across various tasks. In response to these challenges, our work introduces an adaptive approach that harnesses the capabilities of generative adversarial networks (GANs), augmented with contrastive learning, to generate target domain images that account for multispectral band variations effectively. By maximizing mutual information between corresponding patches and leveraging the power of GANs, our model aims to generate realistic-looking images across different multispectral domains. We present a comparative analysis of our model against other well-established generative models, demonstrating its efficacy in generating high-quality satellite images while effectively managing domain variations inherent to multispectral diversity.

**Keywords:** contrastive learning; domain variation; generative adversarial networks (GANs); generation; multispectral bands; remote sensing; satellite image



**Citation:** Mahara, A.; Rishe, N. Multispectral Band-Aware Generation of Satellite Images across Domains Using Generative Adversarial Networks and Contrastive Learning. *Remote Sens.* **2024**, *16*, 1154. https://doi.org/10.3390/rs16071154

Academic Editor: Lefei Zhang

Received: 17 January 2024
Revised: 12 March 2024
Accepted: 24 March 2024
Published: 26 March 2024



## 1. Introduction

Remote sensing is a focal area of research that profoundly impacts various positive aspects of human life, including environmental monitoring, urban planning, and disaster management. Advancements in remote sensing technologies, particularly satellite imagery, have transformed our ability to observe and understand the Earth's geographical features. However, the effectiveness of remote sensing applications often depends on the quality and consistency of satellite images, which can vary significantly across different spectral bands and sensors. Satellite images captured under different environmental conditions or with various sensors exhibit inherent feature variations, leading to domain differences [1]. These variations are particularly noticeable across different multispectral bands, where each band captures specific wavelength ranges on the electromagnetic spectrum. Such domain variability constitutes significant challenges for machine learning and deep learning models in classification and semantic segmentation tasks. These models are sensitive to variations in the input data distribution, which can lead to decreased performance and generalizability [2]. Deep learning methods utilized for the detailed analysis and interpretation of satellite images are particularly susceptible to performance degradation due to domain variability. For instance, a semantic segmentation model trained on satellite images from one spectral band or sensor configuration may underperform when applied to images from a different spectral band or sensor, even if the underlying geographical features and structures are similar [3]. Similarly, models trained on satellite images from one spectral configuration may perform poorly for classification tasks when applied to

data from a different spectral range, even if the essential geographical attributes remain the same [4]. This issue manifests in the domain adaptation problem, which aims to adapt models to perform well across varying domains [5–8].

The recent literature has explored various approaches to tackle the domain adaptation challenge with generative adversarial networks (GANs), highlighting their growing significance as a tool in this area [9–11]. In the field of remote sensing, the recent literature, including Benjdira et al. [12] and Zhao et al. [13], has reported the implementation of GANs to generate target domain images from source domain images, effectively bridging the gap between different domains and enhancing model performance on tasks like semantic segmentation. However, while many studies have concentrated on strengthening downstream tasks such as segmentation or classification, the essential initial step of generating high-quality satellite images across multispectral bands or color channels has yet to be sufficiently addressed.

Considering this, our present work takes a step back to emphasize the generation aspect, which has implications for domain adaptation, a concept illustrated in Figure 1. The diagram is divided into two segments, separated by dotted lines. The upper segment conceptually explores the rationale behind generating satellite images across spectral bands. This exploration envisions future initiatives to harness this capability, potentially enhancing the adaptability and performance of deep learning models across various domains. This segment does not delve into specific deep learning tasks but rather serves as a conceptual reasoning of addressing domain variability challenges.

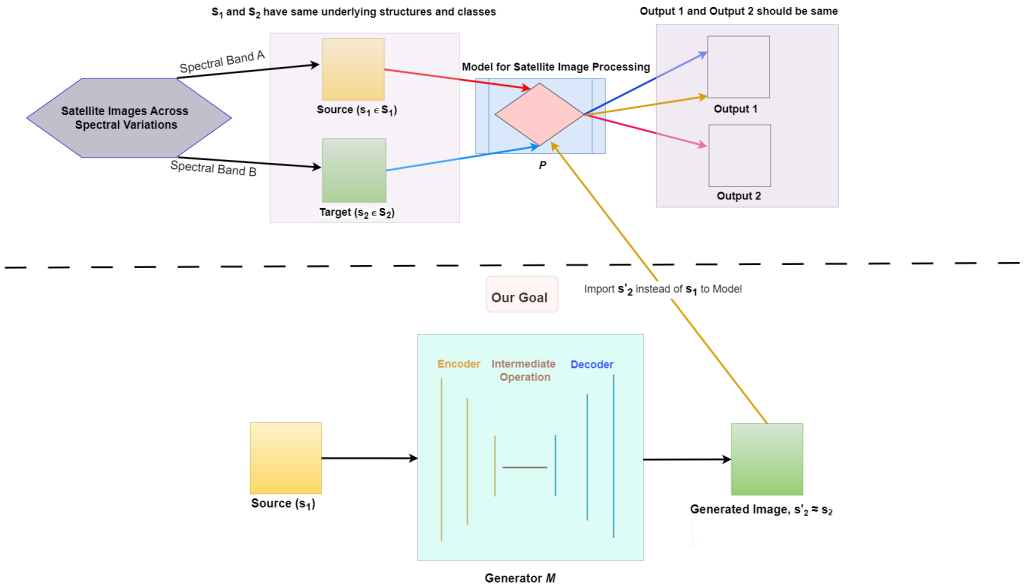


**Figure 1.** Framework overview for multispectral satellite image generation. This figure illustrates our generative approach for transforming satellite images from a source spectral band $A$ to a target spectral band $B$, addressing domain variability. The upper segment conceptually explores potential applications of this transformation in future work, such as improving deep learning model performance across domains. The lower segment specifically details our model's objective and mechanism, highlighting the generation of satellite images in the target domain to bridge spectral discrepancies.

The lower segment of the figure directly addresses our primary objective, i.e., to generate a satellite image $s_2'$ in a target spectral band $B$ from an original image $s_1$ in a source spectral band $A$. For illustrative clarity, in Figure 1, $P$ denotes a hypothetical deep learning model. Suppose $P$ is trained with a dataset from $S_2$ to convert satellite images into expected images (Output 1). Considering that $s_1$ and $s_2$ share identical underlying structures and classes, ideally, feeding $s_1$ into model $P$ should yield the same result as $s_2$. However, discrepancies in color or spectral composition may lead to divergent results (Output 2), as depicted by the red arrow in the figure. To prevent this divergence, the first

step would be generating image $s_2'$ from $s_1$ with a generative model. As shown in the figure, a generator $M$ integrated with an encoder–decoder architecture as shown in the figure can be trained on the source domain $S_1$ to generate images corresponding to the target domain $S_2$. Upon successful training, selecting an image $s_1$ to perform a deep learning task, for instance, allows the generation of $s_2'$, mirroring the spectral band of $s_2$. The resultant image $s_2'$ can then be processed with model $P$ to achieve the desired outcome, as illustrated by yellow-colored arrows in the diagram.

Continuing, we propose a GAN architecture integrated with contrastive learning, specifically designed to generate realistic-looking satellite images across multispectral bands, motivated by the work of Han et al. [14]. By focusing on generating high-quality, cross-domain satellite images, our approach addresses the inherent channel variability and lays a foundation for subsequent domain adaptation applications.

## 2. Related Work

Domain shift or variation has been an enduring problem in the remote sensing domain. Various models have identified and addressed several related issues, yet some aspects of domain variation still require further exploration and solutions. The domain variability problem in satellite images can be traced back to the work of Sharma et al. [15] in 2014. Sharma et al. tackled the challenge of land cover classification in multitemporal remotely sensed images, mainly focusing on scenarios where labeled data are available only for the source domain. This situation is complicated by variability arising from atmospheric and ground reflectance differences. To address this, they employed an innovative approach using ant colony optimization [16] for cross-domain cluster mapping. The target domain data is overclustered in their method and then strategically matched to source domain classes using algorithms inspired by ant movement behavior. In the same year, Yilun Liu and Xia Li developed a method to address a similar challenge of insufficient labeled data in land use classification due to domain variability in satellite images [17]. Using the TrCbrBoost model, their approach harnessed old domain data and fuzzy case-based reasoning for effective classifier training in the target domain. This technique demonstrated significant improvement in classification accuracy, highlighting its effectiveness in overcoming the constraints of domain variability. Similarly, Banerjee and Chaudhuri addressed the problem of unsupervised domain adaptation in remote sensing [18], focusing on classifying multitemporal remote sensing images with inherent data overlapping and variability in semantic class properties. They introduced a hierarchical subspace learning approach, organizing source domain samples in a binary tree and adapting target domain samples at different tree levels. The method proposed by Banerjee and Chaudhuri demonstrated enhanced cross-domain classification performance for remote sensing datasets, effectively managing the challenges of data overlapping and semantic variability [18].

Building on previous advancements in domain adaptation to address domain variability for remote sensing, Postadjian et al.'s work addressed large-scale classification challenges in very high resolution (VHR) satellite images, considering issues such as intra-class variability, diachrony between surveys, and the emergence of new classes not included in predefined labels [19]. Postadjian et al. [19] utilized deep convolutional neural networks (DCNNs) and fine-tuning techniques to adapt to these complexities, effectively handling geographic, temporal, and semantic variations. Following the innovative approaches in domain adaptation to address domain variability, Hofman et al. uniquely applied the CycleGAN [20] network technique to generate target domain images to bridge domain gaps [21]. This approach, leveraging cycle-consistent adversarial networks, enhances the adaptability of deep convolutional neural networks across varied environmental conditions and sensor bands, facilitating effective domain adaptation in unsupervised adaptation tasks, such as classification and semantic segmentation of roads, effectively overcoming pixel-level and high-level domain shifts.

Building on the momentum in addressing domain variability with generative adversarial networks, Zhang et al.'s work addressed the challenge of adapting neural networks

to classify multiband SAR images [22]. This work by Zhang et al. [22] integrated adversarial learning in their proposed MLADA method to align the features of images from different frequency bands in a shared latent space. This approach effectively bridged the gap between bands, demonstrating how adversarial learning can be strategically used to enhance the adaptability and accuracy of neural networks in multiband SAR image classification [22]. Similarly, a methodology proposed by Benjdira et al. focused on improving the semantic segmentation of aerial images through domain adaptation [12]. This work utilized a CycleGAN-inspired adversarial approach, similar to the method employed by Hofman et al. [21]. However, the approach by Benjdira et al. is distinguished by integrating a U-Net model [23] within the generator. This adaptation enables the generation of target domain images that more closely resemble those of the source domain, effectively reducing domain shift related to sensor variation and image quality. Their approach demonstrated substantial improvement in segmentation accuracy across different domains, underscoring the potential of GANs to address domain adaptation challenges in aerial imagery segmentation. Along with this, to address a similar kind of domain variability, Tasar et al. introduced an innovative data augmentation approach, SemI2I, that employed generative adversarial networks to transfer the style of test data to training data, utilizing adaptive instance normalization and adversarial losses for style transfer [24]. The approach, highlighted by its ability to generate semantically consistent target domain images, has outperformed existing domain adaptation methods, paving the way for more accurate and robust segmentation models with the generative adversarial mechanism in varied remote sensing environments.

Expanding on the work to address domain variability in remote sensing, another work by Tasar et al. [25] effectively harnessed the power of GANs to mitigate the multispectral band shifts between satellite images from different geographic locations. Through ColorMapGANs, this work adeptly generated training images that were semantically consistent with original images yet spectrally adapted to resemble the test domain, substantially enhancing the segmentation accuracy. This intelligent use of GANs demonstrates their growing significance in addressing complex domain adaptation challenges in the remote sensing field. Consequently, Zhao et al. introduced an advanced method to minimize the pixel-level domain gap in remote sensing [13]. The ResiDualGAN framework incorporates a resizing module and residual connections into DualGAN [26] to address scale discrepancies and stabilize the training process effectively. Demonstrating its efficacy, the authors showcased significant improvements in segmentation accuracy on datasets collected from the cities of Potsdam and Vaihingen, open-source remote sensing semantic segmentation datasets [27], proving that their approach robustly handles the domain variability and improves cross-domain semantic segmentation with a generative adversarial model. In the current literature on remote sensing, GANs and image-to-image translation mechanisms [20,26,28,29] have been promising in solving the domain variation problem. Since the translation of images from the source domain to the target domain is one of the fundamental building blocks in solving the domain variability problem, we present a GAN model inspired by the work of Han et al. [14]. Han et al. presented an interesting approach for mapping two domains using generators within a GAN model, uniquely employing an unpaired fashion without a cyclic procedure. This approach demonstrates that reverse mapping between domains can be learned without relying on the generated images, leading to nonrestrictive mapping, in contrast to the restrictive mapping approach presented by Zhu et al. [20]. Expanding on nonrestrictive mapping, our model integrates contrastive learning in a GAN with two generators and aims to perform better than well-established GAN models by generating realistic satellite images from one multispectral band to another. This capability is applicable for image generation and potentially beneficial for domain adaptation tasks.

## 3. Materials and Methods

This study aims to generate satellite images from one multispectral band mode to another, a process that can potentially help in domain adaptation within remote sensing.

For this purpose, we conducted a thorough review of available datasets and selected an open-source dataset available from the ISPRS 2D Open-Source benchmark [27], which has been frequently utilized in recent domain adaptation research [12,13]. This collected dataset includes different multispectral bands. For our analysis, we focused on two subsets: one from Potsdam City, with a pixel resolution of 5 cm and comprising the RGB (red, green, blue) spectral bands, and another from Vaihingen City, characterized by a 9 cm pixel resolution and including the IRRG (infrared, red, green) bands.

We consider two domains of satellite images, $S_1$ and $S_2$, where the underlying features of an image $s_1 \in S_1$ and an image $s_2 \in S_2$ vary based on different multispectral bands. This scenario is formalized as the input domain $S_1 \subset \mathbb{R}^{H \times W \times C_1}$ and the target domain $S_2 \subset \mathbb{R}^{H \times W \times C_2}$, where $C_1$ and $C_2$ represent the number of channels in each domain, respectively. Given sets of unpaired satellite images from these domains, we aim to generate a new image $s_2'$ from an input image $s_1$ in domain $S_1$, such that $s_2'$ closely follows the true data distribution of domain $S_2$. We aim to accomplish this by constructing two mapping functions, $M_1$ and $M_2$, such that $M_1(S)$: $S_1 \rightarrow S_2$ and $M_2(S)$: $S_2 \rightarrow S_1$, where $S$ denotes a set of satellite images from the respective input domains. In this study, these two mapping functions, $M_1$ and $M_2$, are demonstrated as two generators, to which there are two corresponding discriminators, $D_1$ and $D_2$, as shown in Figure 2.

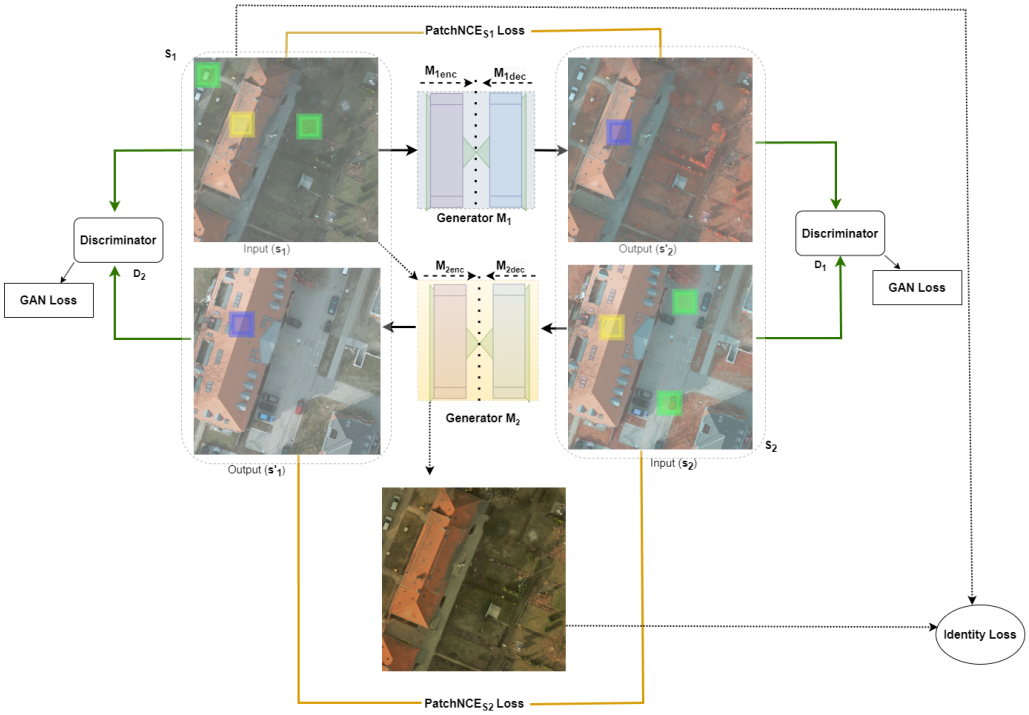


**Figure 2.** Overview of the proposed model's architecture. This figure illustrates the process designed to generate images in domain $S_2$ from domain $S_1$. The generation relies on two generators: $M_1$ and $M_2$. $M_1$ is trained to translate an input $s_1$ to output $s_2'$, while $M_2$ is trained to translate an input $s_2$ to output $s_1'$. Encoders $M_{1_{enc}}$ and $M_{2_{enc}}$ are involved in computing PatchNCE$_{s1}$ and PatchNCE$_{s2}$ losses, depicted with yellow lines. These loss functions ensure that corresponding areas in the source and target images (indicated by yellow and blue square patches, respectively) retain mutual information in contrast to dissimilar areas (depicted with green square patches). Discriminators $D_1$ and $D_2$ validate the authenticity of generated images, facilitating the generation of realistic-looking images via GAN losses represented by rectangles. Identity loss (shown in an ellipse) is calculated by passing $s_1$ to $M_2$ instead of $s_2$, as depicted with dotted arrows.

### *3.1. Efficient Mapping with Generators*

In the present work, each generator comprises an encoder-decoder architecture (as shown in Figure 3), drawing inspiration from CUT [28] and DCLGAN [14] to generate

designated satellite images. For each mapping and to better utilize the features in satellite images, we extract features from $L = 4$ encoder layers and propagate them to a two-layer MLP network ($H_1$ and $H_2$), as performed in SimCLR [30].

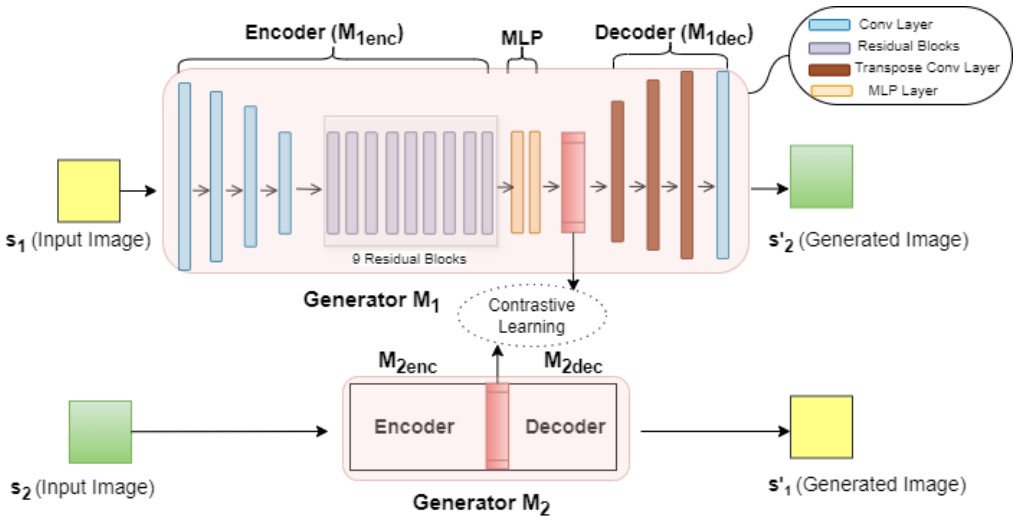


**Figure 3.** Overview of the proposed model's generators. The figure illustrates the architecture of generators $M_1$ and $M_2$, integral to transforming input images across spectral domains. $M_1$, designated for translating images from $S_1$ to $S_2$, features an encoder ($M_{1enc}$), a multilayer perceptron (MLP) for processing the encoded features, and a decoder ($M_{1dec}$) for generating the output image $s'_2$. Highlighted in purple, the core of $M_1$ includes 9 residual blocks designed for deep feature refinement. $M_2$ operates in reverse, translating $S_2$ to $S_1$, and similarly incorporates an encoder ($M_{2enc}$) and decoder ($M_{2dec}$). The dotted lines represent the process of contrastive learning, utilized to enhance the fidelity of generated images by aligning features across the two domains. Key components such as convolution layers, residual blocks, transposed convolution layers, and MLP layers are denoted by distinct colors.

#### 3.1.1. ResNet-Based Generator

The generators employed in this work are based on a ResNet architecture, as depicted in Figure 3, which has been proven successful in various generative models [31]. This choice is integral for synthesizing satellite images within our generative adversarial network (GAN) framework. The generator is designed to capture and translate the complex spatial and textural information in satellite images into corresponding images of different multispectral band representations. Motivated by [20], each generator integrates nine residual blocks to help the encoding and decoding processes. These blocks enable the model to handle complex features essential for high-quality satellite images.

#### 3.1.2. Encoder and Decoder Architecture

Building upon the initial framework, where the two mapping functions are represented as generators $M_1$ and $M_2$, to which two corresponding discriminators, $D_1$ and $D_2$, are assigned, we delve deeper into the architecture of these generators. Each generator consists of an encoder and a decoder component. The encoders, $M_{1_{enc}}$ and $M_{2_{enc}}$, are used to capture and compress the spectral features of the satellite images from their respective domains. This is achieved by extracting features from $L = 4$ layers of the encoder, as previously mentioned, which are then propagated to a two-layer MLP network to enhance feature utilization and facilitate effective domain translation.

The decoders, $M_{1_{dec}}$ and $M_{2_{dec}}$ (illustrated in Figure 3), are responsible for reconstructing the image in the new domain while preserving spatial coherence and relevant features. They take the encoded features and, through a series of transformations, generate the output image that corresponds to the target domain. This process ensures that the translated

images maintain the target domain's essential characteristics while reflecting the source domain's content.

With this procedure, the encoder and decoder collaborate within each generator to facilitate a robust and efficient translation between the multispectral bands of satellite images.

### *3.2. Discriminator Architecture*

Discriminators are crucial for the adversarial training mechanism within the GAN framework. Our model incorporates two discriminators, $D_1$ and $D_2$, each corresponding to generators, $M_1$ and $M_2$, respectively. These discriminators distinguish between authentic satellite images and those synthesized by their respective generators. Their primary role is to provide critical feedback to the generators, driving them to produce more accurate and realistic translations of satellite images.

Our architecture employs a PatchGAN [32] discriminator, chosen for its effectiveness in generative tasks. The architecture of the discriminator is depicted in Figure 4. Unlike traditional discriminators that assess the authenticity of an entire image, PatchGAN divides the image into smaller $p \times p$ patches and evaluates the realism of each patch individually. The size of each patch, $p$, is selected to be less than or equal to the height $H$ of the image, ensuring that each patch is reasonably sized to maintain a balance between training efficiency and feature learning. All the patches are processed through various convolution layers (depicted with blue-colored rectangles in Figure 4, with deeper layers having an increasing number of filters. These convolution layers build up to a final convolution layer that assigns scores, indicating whether each patch is real or fake. The scores assigned to all patches by the discriminator are then aggregated to determine the overall authenticity of the image. This approach guides the generator in refining its output and facilitates the production of high-quality, realistic textures and patterns at the patch level.

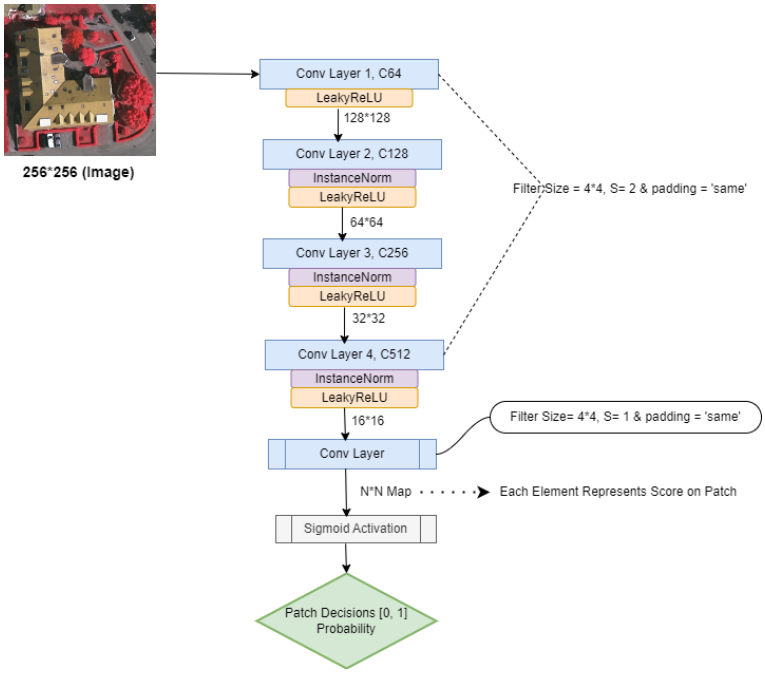


**Figure 4.** Discriminator architecture in our model. The depicted architecture outlines a PatchGAN discriminator, which operates on a patch-wise basis to evaluate the image's authenticity.

Furthermore, we utilize instance normalization over batch normalization, as illustrated with purple-colored rounded rectangles in the diagram, a decision aligned with the best practices in image-to-image translation tasks [20,26,28]. The integration of instance normalization contributes to the stability and performance of the model, particularly in the context of diverse and variable satellite imagery.

### *3.3. Contrastive Learning in the Present Work*

Recently, contrastive learning has significantly advanced unsupervised learning in image processing, often outperforming conventional methods in scenarios with limited data [30,33]. It identifies negative and positive pairs and maximizes the mutual information relevant to the target objective.

#### 3.3.1. Mutual Information Maximization

In our satellite image translation task, we devise an efficient method to maximize the mutual correlation between corresponding segments of the input and generated images across different multispectral bands. Specifically, a patch representing a particular land feature in the input image from domain $S_1$ should be strongly associated with the corresponding element in the generated image in domain $S_2$, as shown in Figure 2, where images $s_1$ and $s_2'$ display the same building roof structure, highlighted with yellow and blue square patches, respectively. This approach ensures that relevant geographical and textural information is preserved and accurately represented in the translated images.

In the general procedure of the contrastive learning mechanism, we strategically select an anchor feature $a \in \mathbb{R}^K$, extracted from $s_2$ (as shown in Figure 2, highlighted with a blue-colored square), and a corresponding positive feature $p \in \mathbb{R}^K$, extracted from $s_1$ (as shown in Figure 2, depicted with a yellow-colored square), where $K$ represents the feature space dimensionality. Additionally, we sample a set of negative features $\{n_i\}_{i=1}^N$, with each $n_i \in \mathbb{R}^K$, from $s_1$. These negative features are drawn from different locations within the same image from which the corresponding $p$ was specified (depicted with green-colored squares in Figure 2). All these features are represented as vectors and normalized with $L_2$-normalization [34]. Subsequently, we construct an $(N+1)$-way classification task. In this task, the model evaluates the similarity between an anchor and various positive or negative elements. We direct the model to adjust the scale of these similarities using a temperature parameter, $\tau = 0.07$, before converting them into probabilities [28]. This temperature scaling fine-tunes the model's confidence in its predictions, facilitating a more effective learning process through the contrastive mechanism. After establishing this setup for normalization and predictions, we compute the probability that the positive element is more aligned to the anchor than any negative element. This probability is obtained using a cross-entropy loss, expressed mathematically as

$$L(a, p, \{n_i\}) = -\log\left(\frac{\exp(\text{sim}(a, p)/\tau)}{\exp(\text{sim}(a, p)/\tau) + \sum_{i=1}^{N} \exp(\text{sim}(a, n_i)/\tau)}\right) \quad (1)$$

where $\text{sim}(a, p)$ denotes the cosine similarity between $a$ and $p$, calculated as $\frac{a \cdot p}{\|a\|\|p\|}$ [30]. Through this loss equation, we effectively maximize the mutual correlation between matching segments, ensuring that the translation between multispectral bands of satellite images retains high fidelity and relevance.

#### 3.3.2. PatchNCE Loss Mapping of Two Domains of Satellite Images

Utilizing the encoders $M_{1_{enc}}$ and $M_{2_{enc}}$ from the generators $M_1$ and $M_2$, we extract features from the satellite images in domains $S_1$ and $S_2$, respectively. For each domain, features are extracted from selected layers of the respective encoder and then propagated through a two-layer MLP network. This process results in a stack of feature layers $\{z_l\}_L$, where each $z_l$ represents the output of the $l$-th selected layer after processing through the MLP network. Specifically, for an input image $s_1$ from domain $S_1$, the feature stack can be represented as $\{z_l\}_L = \{H_1^l(M_{1_{enc}}^{(l)}(s_1))\}_{l=1}^L$, where $M_{1_{enc}}^{(l)}$ denotes the output of the $l$-th layer of encoder $M_{1_{enc}}$, and $H_1^l$ represents the corresponding MLP processing for that layer. Similarly, for domain $S_2$, the feature stack is obtained using encoder $M_{2_{enc}}$ and its corresponding MLP layers, which is given by $\{\hat{z}_l\}_L = \{H_2^l(M_{2_{enc}}^{(l)}(s_2'))\}_{l=1}^L$, where $s_2'$ is a generated image.

Our framework defines spatial locations within a layer to refer to distinct areas in the generated feature maps. Each spatial location is associated with a specific region of the input image, as determined by the network's convolutional operations. We denote the layers by $l \in \{1, 2, \ldots, L\}$ and the spatial locations within these layers by $y \in \{1, \ldots, Y_l\}$, where $Y_l$ represents the total number of spatial locations in layer $l$. We then define an anchor patch and its corresponding positive feature as $z_l^{(y)} \in \mathbb{R}^{C_l}$ and all other features (the "negatives") as $z_l^{(Y \setminus y)} \in \mathbb{R}^{(Y_l - 1) \times C_l}$, where $C_l$ denotes the number of feature channels in each layer, and Y represents the conceptual set containing all possible indices of these spatial locations within a layer.

By obtaining insights from CUT [28] and DCLGAN [14], the present work incorporates patch-based PatchNCE loss that aims to align analogous patches of input and output satellite images across multiple layers. For the mapping $M_1 : S_1 \to S_2$, the PatchNCE loss is expressed as

$$L_{\text{PatchNCE}_{S_1}}(M_1, H_1, H_2, S_1) = \mathbb{E}_{s_1 \sim S_1} \sum_{l=1}^{L} \sum_{y=1}^{Y_l} \ell(\hat{z}_l^{(y)}, z_l^{(y)}, z_l^{(Y \setminus y)}) \tag{2}$$

Similarly, for the reverse mapping $M_2 : S_2 \to S_1$, we utilized similar PatchNCE loss:

$$L_{\text{PatchNCE}_{S_2}}(M_2, H_1, H_2, S_2) = \mathbb{E}_{s_2 \sim S_2} \sum_{l=1}^{L} \sum_{y=1}^{Y_l} \ell(\hat{z}_l^{(y)}, z_l^{(y)}, z_l^{(Y \setminus y)}) \tag{3}$$

Through these formulations, the PatchNCE losses effectively encourage the model to learn translations that maintain the essential characteristics and patterns of the geographic features in satellite images, ensuring that the translated images retain the contextual and spectral integrity necessary for accurate interpretation and analysis.

### *3.4. Adversarial Loss*

The present work utilizes an adversarial loss function to ensure the generation of realistic-looking satellite images from one domain to another based on different multispectral bands [35]. The objective is to maintain balanced training of the generator and the discriminator to produce satellite images in the target domain that are indistinguishable from ground-truth satellite images. The discriminator learns to differentiate between the original and synthetic satellite images. The training is guided by the adversarial loss function with backpropagation and iterative updates of the layers' weights in the model. Each generator, $M_1$ and $M_2$, has a corresponding discriminator, $D_1$ and $D_2$, respectively, ensuring a targeted adversarial relationship. The GAN loss for each generator–discriminator pair can be formulated as

$$\begin{aligned} L_{\text{GAN}}(M_1, D_1, S_1, S_2) = &\mathbb{E}_{s_2 \sim S_2}[\log D_1(s_2)] + \\ &\mathbb{E}_{s_1 \sim S_1}[\log(1 - D_1(M_1(s_1)))] \end{aligned} \tag{4}$$

$$\begin{aligned} L_{\text{GAN}}(M_2, D_2, S_2, S_1) = &\mathbb{E}_{s_1 \sim S_1}[\log D_2(s_1)] + \\ &\mathbb{E}_{s_2 \sim S_2}[\log(1 - D_2(M_2(s_2)))] \end{aligned} \tag{5}$$

In these equations, $D_1(s_2)$ and $D_2(s_1)$ represent the discriminator's decision for ground-truth satellite images, $s_2$ and $s_1$, respectively. $M_1(s_1)$ and $M_2(s_2)$ are the images generated from the input satellite images $s_1$ and $s_2$ that should hypothetically correspond to $s_2$ and $s_1$, respectively. Real satellite images form $S_1$ and $S_2$ distributions from each domain. The generators $M_1$ and $M_2$ aim to minimize these losses, while the discriminators $D_1$ and $D_2$ aim to maximize them [35]. Hence, the given loss function is termed adversarial, as the given generator and discriminator compete to get better and produce visually promising satellite images.

### *3.5. Identity Loss*

Identity loss with mean squared error (MSE) is implemented to preserve the essential characteristics of the input image when it already belongs to the target domain. This approach ensures the generator minimizes alterations when the input image exhibits the target domain's characteristics. Specifically, in our satellite image translation task, when generator $M_1$ is trained to convert an image $s_1$ from domain $S_1$ to an equivalent image in domain $S_2$, it should ideally introduce minimal changes if an image from $S_2$ is provided as input. This strategy encourages the generator to maintain the identity of the input when it aligns with the target domain.

Mathematically, the identity loss for generator $M_1$ when an image from $S_2$ is fed as input using MSE can be expressed as

$$L^{S_2}_{\text{identity-MSE}}(M_1, S_2) = \mathbb{E}_{s_2 \sim S_2}[\|M_1(s_2) - s_2\|_2^2],$$

where $\|.\|_2^2$ denotes the squared L2 norm, representing the sum of the squared differences between the generated and input images. A similar identity loss is applied for generator $M_2$ when an image from $S_1$ is fed:

$$L^{S_1}_{\text{identity-MSE}}(M_2, S_1) = \mathbb{E}_{s_1 \sim S_1}[\|M_2(s_1) - s_1\|_2^2].$$

Total identity loss, combining the contributions from both generators, is given as

$$\begin{aligned} L_{\text{identity-MSE}}(M_1, M_2) = {} & L^{S_1}_{\text{identity-MSE}}(M_2, S_1) \\ & + L^{S_2}_{\text{identity-MSE}}(M_1, S_2), \end{aligned} \tag{6}$$

Although we experimented with identity loss using the L1 norm, i.e., mean absolute error (MAE), we found that the L2 norm provided better results in our model. Therefore, we chose to utilize the L2 norm for identity loss. In practice, identity loss aids in stabilizing the training of the generators by ensuring they do not introduce unnecessary changes to images that already possess the desired characteristics of the target domain. This concept, inspired by the work of Zhu et al. [20], is particularly beneficial in maintaining the structural and spectral integrity of satellite images during the translation process.

### *3.6. Final Objective*

Our satellite image generation framework is depicted in Figure 2. In our framework, the objective is to generate images that are not only realistic but also maintain a correspondence between patches in the input and output images. To achieve our goal, we integrate the combination of the GAN loss, PatchNCE loss, and identity loss to have our final loss function, which is given as

$$\begin{aligned} L(M_1, M_2, D_1, D_2, H_1, H_2) = {} & \lambda_{\text{GAN}}(L_{\text{GAN}}(M_2, D_2, S_2, S_1) \\ & + L_{\text{GAN}}(M_2, D_1, S_2, S_1)) \\ & + \lambda_{\text{NCE}}(L_{\text{PatchNCE}_{S_1}}(M_1, H_1, H_2, S_1) \\ & + L_{\text{PatchNCE}_{S_2}}(M_2, H_1, H_2, S_2)) \\ & + \lambda_{\text{idt}} L_{\text{identity-MSE}}(M_1, M_2), \end{aligned} \tag{7}$$

## 4. Results

This section is divided into two main parts. The first part details the experimental setup, including metrics for evaluation and the experimental environment; the second part discusses the results obtained from comparing our approach with baseline models.

### *4.1. Evaluation Metrics*

The present work utilizes a set of evaluation metrics to measure the quality and accuracy of the generated satellite images. These metrics include the root mean square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM).

#### 4.1.1. Root Mean Square Error (RMSE)

The RMSE is a widely used metric that quantifies the differences between values predicted by a model and the observed actual values [36]. It is beneficial in evaluating the accuracy of generated or reconstructed images compared to the original images. Mathematically, the RMSE is defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - \hat{p}_i)^2}$$

where $N$ is the number of pixels in the image, $p_i$ is the actual value of the pixel, and $\hat{p}_i$ is the predicted or generated value for the same pixel.

#### 4.1.2. Peak Signal-to-Noise Ratio (PSNR)

The PSNR is a metric that measures the quality of reconstruction of lossy compression [37] and is widely utilized to assess the visual quality of generated or reconstructed images. It compares the maximum potential power of a signal (in images, this corresponds to the maximum possible pixel value) to the power of the noise that affects the fidelity of its representation. Mathematically, the PSNR is defined as

$$PSNR = 10 \cdot \log_{10}\left(\frac{Max_s^2}{MSE}\right)$$

where $Max_s$ represents the maximum possible pixel value of the image $s$ and $MSE$ is the mean squared error between the input image $s$ and the reconstructed image. A higher PSNR value indicates better quality of the generated image.

#### 4.1.3. Structural Similarity Index (SSIM)

The SSIM assesses the perceived quality of digital images and videos by evaluating changes in luminance, contrast, and structure between the original image $s$ and the predicted image $s'$. Mathematically, the SSIM is defined as

$$SSIM(s, s') = \frac{(2\mu_s\mu_{s'} + C)(2\sigma_{ss'} + C)}{(\mu_s^2 + \mu_{s'}^2 + C)(\sigma_s^2 + \sigma_{s'}^2 + C)}$$

In this equation, $\mu_s$ and $\mu_{s'}$ are the average luminance values of images $s$ and $s'$, respectively; $\sigma_s^2$ and $\sigma_{s'}^2$ are their variances, and $\sigma_{ss'}$ is the covariance between the images. $C$ is a constant introduced to prevent the division of the SSIM's numerator by zero, thus ensuring a well-defined value for the SSIM. A higher SSIM value indicates better image quality, with a value of 1 signifying perfectly identical images.

### *4.2. Experimental Environment and Baselines*

#### 4.2.1. Experiment Setting

We conducted our experiments in a Python 3.6.8 environment with the PyTorch framework for all the training and testing tasks. The computational workload was executed on a system equipped with dual NVIDIA TITAN RTX GPUs. Each of these GPUs features 24 GB of GDDR6 memory with the power of the TU102 architecture. The system operated with CUDA version 12.3 and NVIDIA driver version 545.23.08, ensuring high performance and efficiency for our computational tasks. To maintain uniform training for all the models, we selected 800 images for both the Potsdam and Vaihingen training datasets and 500 images

for the testing dataset. We resized all the images to dimensions of 256 × 256. We trained all the models for 200 epochs with a learning rate of 0.0001. During the first half of these epochs, we maintained a steady learning rate, whereas in the latter half, the learning rate decayed linearly to have better convergence. We set the temperature parameter $\tau = 0.07$. Along with this, we set the loss functions' parameters: $\lambda_{\text{GAN}} = 1$, $\lambda_{\text{NCE}} = 1$, and $\lambda_{\text{idt}} = 1$.

#### 4.2.2. Baselines

We selected four well-established unsupervised generative adversarial network models and compared them to our model for qualitative and quantitative analysis. These selected four methods are DualGAN [26], CUT [28], CycleGAN [20], and GcGAN [29]. These models were chosen due to their success in image generation tasks and effective utilization to address domain variation problems [12,13,38], which is relevant to the objectives of our study.

### *4.3. Comparison and Results*

Our model's comparative performance against other generative adversarial network models is detailed in Table 1 and visually illustrated in Figure 5. These results demonstrate the better performance of our model in generating satellite images across multispectral bands. Specifically, our model achieved an RMSE of 23.9320, a PSNR of 20.5512, and an SSIM of 0.7888, as highlighted in bold in Table 1. In comparison, the CUT model [28] recorded slightly lower metrics with an RMSE of 28.2995, PSNR of 19.0952, and SSIM of 0.7082. It is worth mentioning that a lower RMSE value indicates better quality of the generated image [39]. Furthermore, compared to our model and CUT, CycleGAN [26] achieved marginally lower values in PSNR, at 18.6886, and in RMSE, at 29.6556. However, it achieved a higher SSIM value of 0.7461 compared to CUT, yet this value remained lower than that of our model. The DualGAN model scored slightly lower than our model, CUT, and CycleGAN on all metrics, except that it performed slightly better than CUT in SSIM with 0.7145. Subsequently, the GcGAN model [29] recorded the lowest values among all models, including ours, with a PSNR of 17.2172, SSIM of 0.5786, and RMSE of 35.1302.

**Table 1.** Performance Comparison: Our Model vs. Other GANs.

| Models | RMSE | PSNR | SSIM | Training Time (h:min) |
|---|---|---|---|---|
| Our model | **23.9320** | **20.5512** | **0.7888** | 10 h:52 min |
| CUT | 28.2995 | 19.0952 | 0.7082 | 10 h:13 min |
| CycleGAN | 29.6556 | 18.6886 | 0.7461 | 10 h:43 min |
| DualGAN | 31.7752 | 18.0890 | 0.7145 | 10 h:48 min |
| GcGAN | 35.1302 | 17.2172 | 0.5786 | 9 h:01 min |

Similarly, the total training time for each model was recorded and is presented in Table 1. Under the experimental setup mentioned, our model required approximately 11 h for 200 epochs of training, which is comparable to the time taken by DualGAN and CycleGAN but longer than that required for CUT and GcGAN. The relatively shorter training time for CUT and GcGAN can be attributed to their simplified architectures, utilizing only one generator and discriminator for image generation. However, despite GcGAN being faster to train by approximately 2 h, the substantial difference in SSIM values (0.5786 compared to our 0.7888) establishes our model as the preferable choice for generating higher-quality images.

Further insights into our model's training behavior are illustrated in Figure 6, which presents the average losses obtained by our model's generators and discriminators over 200 epochs. The figure reveals a balanced learning trajectory between generator $M_2$ (illustrated in red) and discriminator $D_2$ (illustrated in green). Consequently, discriminator $D_1$ (illustrated in blue) demonstrates effective learning across epochs. Although generator $M_1$ appears to exhibit a slight upward trend in the curve towards the latter epochs, its loss stabilizes around a value of 0.6, demonstrating continuous learning over time. This

coherent convergence of loss curves illustrates a well-regulated learning process between the discriminators and generators, reinforcing the efficacy of our GAN framework.

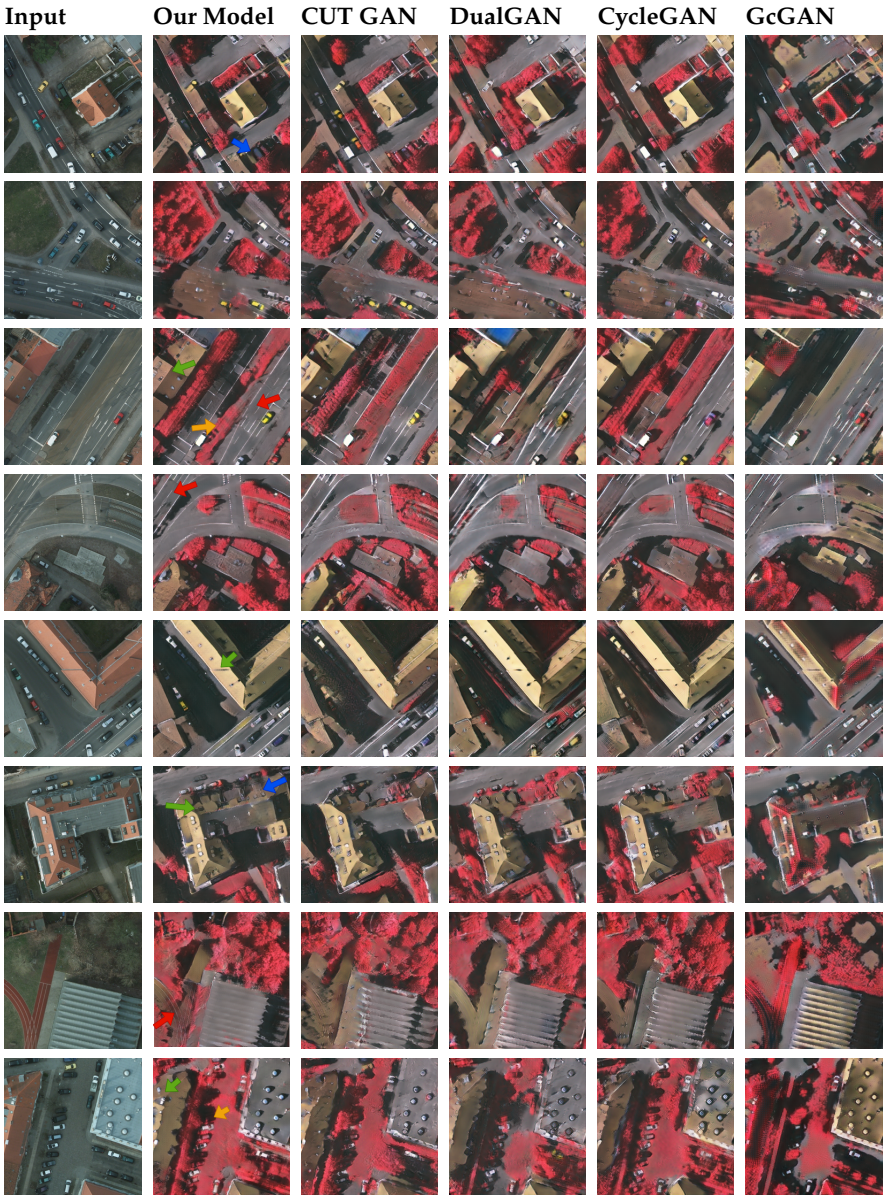


**Figure 5.** Comparative results of satellite image generation via visualization. Green-colored arrows (rows 3, 5, 6, 8) illustrate areas where our model preserves structural details on roofs more accurately than other models. Red-colored arrows (rows 3, 4, 7) denote the effective retention of white road markings, showcasing our model's strengths. Blue-colored arrows (rows 1, 6) highlight the successful prevention of inappropriate color generation on vehicles and roofs. However, yellow-colored arrows (rows 3, 8) reveal our model's deficiencies, such as incorrect color generation and merging issues, which are also observed in comparative models. These results highlight the intricate balance of successes and challenges in satellite image synthesis across multispectral bands.

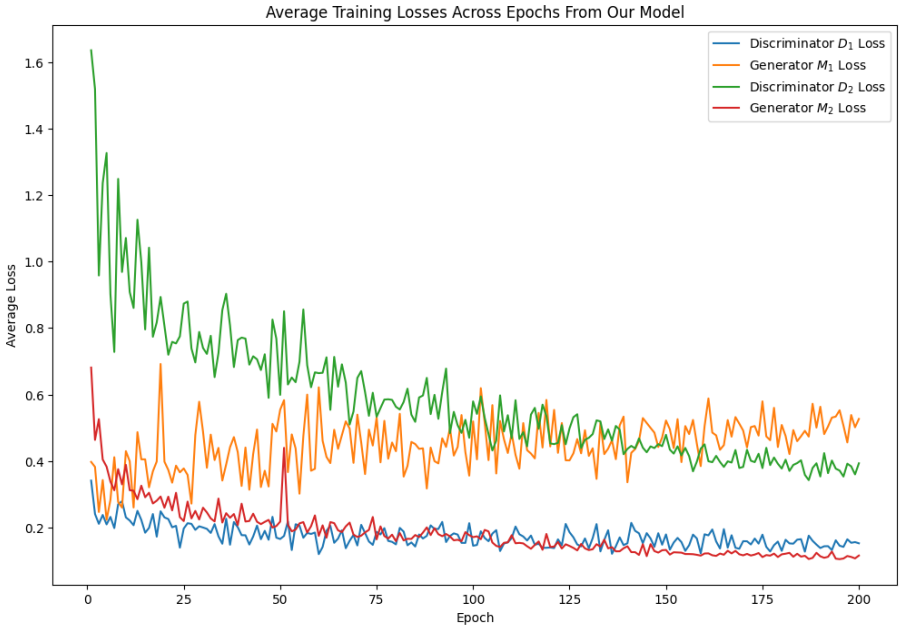


**Figure 6.** Training loss curves across epochs for our model. The figure illustrates the training dynamics over 200 epochs observed in our model. The blue and green curves represent the losses for discriminators $D_1$ and $D_2$, respectively. The generator $M_1$ and $M_2$ losses, depicted in orange and red, respectively, reflect the generators' performance in creating images that are increasingly indistinguishable from real images.

## 5. Discussion

In selecting datasets available from [27], our goal was to generate satellite images with IRRG color composition similar to those collected from Vaihingen City (example images are shown in Figure 7) by training the model on satellite images with RGB color composition collected from Potsdam City (images are shown in Figure 5 in the first column, i.e., Input). These two datasets generally have similar underlying features, such as buildings and roads, but the images in the input and target domains do not have one-to-one correspondence as they are taken from two different locations. During this, our training approach yielded promising results in image generation, as demonstrated in Figure 5.

**Figure 7.** Ground-truth IRRG images from Vaihingen City's dataset.

For the evaluation metrics SSIM, PSNR, and RMSE, it is generally necessary to have the exact generated image corresponding to the input image and the exact ground-truth image for the input image. Since the Potsdam and Vaihingen datasets are collected from two different cities, if the input image $A$ is from Potsdam and the target image $B$ is from

Vaihingen, the ground truth image should maintain the underlying structure of *A* but adopt the color composition of *B* to align with our work of transferring between two different multispectral bands. The Potsdam dataset provides corresponding IRRG images for each RGB image; hence, during the calculation of SSIM, PSNR, and RMSE, we used related Potsdam IRRG color composition images (as shown in Figure 8) as ground truth for each corresponding generated image.

**Figure 8.** Ground truth IRRG images from Potsdam City's dataset.

Similarly, as observed in Figure 5, the generated images exhibit a color composition similar to the target dataset, i.e., IRRG color composition satellite images from Vaihingen City, as shown in Figure 7, aligning with our primary objective of replicating the target spectral characteristics. The presence of diverse objects such as buildings, cars, houses, and roads complicates the clear differentiation between the successes and failures of our model, especially in focusing on specific details. Upon detailed examination, we identified certain areas where our model outperforms or falls short compared to other models, as indicated by arrows in various colors. Specifically, as seen in Figure 5 in rows 3, 5, 6, and 8, our model exhibits marginally better performance in preserving the structural details of certain parts of the roofs on houses and buildings, depicted with green-colored arrows. Furthermore, our model more effectively retains white-colored road markings, another underlying structure, as indicated by red-colored arrows in rows 3, 4, and 7. Additionally, in rows 1 and 6, our model effectively prevents inappropriate color generation, such as on cars and roofs, as highlighted by blue-colored arrows. Although these improvements are subtle, they underline the model's enhanced capability to prevent inaccurate color generation.

There are, however, some deficiencies in our model, such as the incorrect generation of red color on roads, particularly in row 3, as indicated by a yellow-colored arrow. It is worth mentioning that all comparative models exhibit this same misrepresentation, suggesting that the flaw may be attributed to an imbalance in the dataset. Similarly, our model incorrectly merges black-colored cars with shadows, as depicted by yellow-colored arrows in row 8, a problem also common among the other models. This issue may be attributed to either dataset imbalance or the presence of shadows in the input images. Again, the same observation can be seen with all other models, showing the same performance of blending those cars into the shadow. This failure may also be attributable to the imbalanced dataset or shadows in the input images. An interesting observation was made on GcGAN [29] regarding the retainment of specific structures. Despite inaccurately generating red color on roofs of buildings and houses, it successfully retained white-colored road markings similar to our results, where other models failed.

## 6. Conclusions

Satellite imagery, characterized by its diverse spectral bands, inherently exhibits domain variability that significantly impacts the performance of various analytical models.

In response to this challenge, our work introduced an approach that leverages the capabilities of generative adversarial networks (GANs) combined with contrastive learning. Specifically, the present work incorporates a dual translating strategy, ensuring that each translation direction uses original images from the respective domains ($S_1$ and $S_2$) rather than relying on previously generated images, maintaining the integrity and quality of the translation process. Our model effectively translates images across multispectral bands by integrating adversarial mechanisms between generators and discriminators, maximizing mutual information among important patches and utilizing identity loss with mean squared error (MSE). The quantitative results, as evidenced by metrics such as SSIM, PSNR, and RMSE, alongside qualitative visualization, demonstrate that our model performs better than well-established methods, including CycleGAN, CUT, DualGAN, and GcGAN. Furthermore, for training over 200 epochs, our model requires a comparable amount of time, around eleven hours, similar to that of the other models except for GcGAN, which took only nine hours to train but obtained a significantly lower SSIM value of 0.5786 compared to our model's 0.7888. These findings highlight the practicability of our model in generating high-quality satellite images. These images accurately reflect the desired spectral characteristics while retaining important underlying structures, highlighting advancements in remote sensing applications.

It is important to note that the present work integrates a GAN model to generate satellite images across different multispectral bands. While our current work focuses on image generation, the natural progression is to extend these capabilities to domain adaptation applications in several subdomains of remote sensing. We aim to explore various fields within and beyond remote sensing, applying our model to generate images that can then be used with developed models like U-Net for semantic segmentation and other tasks where domain variability is a significant challenge. By continuing to refine and apply our model, we hope to contribute to advancing remote sensing techniques and the broader field of image processing.

**Author Contributions:** Conceptualization, A.M.; Methodology, A.M.; Validation, N.R.; Investigation, A.M.; Writing—original draft, A.M.; Writing—review & editing, N.R.; Visualization, A.M.; Supervision, N.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This material is based in part upon work supported by the National Science Foundation under Grant No. CNS-2018611.

**Data Availability Statement:** The dataset used in this article is a publicly available dataset that can be obtained from the ISPRS 2D Open-Source benchmark dataset [27]. A few examples of images generated in this study are presented in Figure 5. The complete dataset generated in this research is available upon request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zhao, W.; Lu, H.; Wang, D. Multisensor image fusion and enhancement in spectral total variation domain. *IEEE Trans. Multimed.* **2017**, *20*, 866–879.
2. Laborte, A.G.; Maunahan, A.A.; Hijmans, R.J. Spectral signature generalization and expansion can improve the accuracy of satellite image classification. *PLoS ONE* **2010**, *5*, e10516.
3. Toldo, M.; Maracani, A.; Michieli, U.; Zanuttigh, P. Unsupervised domain adaptation in semantic segmentation: A review. *Technologies* **2020**, *8*, 35.
4. Lunga, D.; Yang, H.L.; Reith, A.; Weaver, J.; Yuan, J.; Bhaduri, B. Domain-adapted convolutional networks for satellite image classification: A large-scale interactive learning workflow. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 962–977.
5. Farahani, A.; Voghoei, S.; Rasheed, K.; Arabnia, H.R. A brief review of domain adaptation. In *Advances in Data Science and Information Engineering*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 877–894.
6. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **2010**, *22*, 199–210.
7. Ganin, Y.; Lempitsky, V. Unsupervised domain adaptation by backpropagation. In Proceedings of the International Conference on Machine Learning (PMLR), Lille, France, 6–11 July 2015; pp. 1180–1189.

8. You, K.; Long, M.; Cao, Z.; Wang, J.; Jordan, M.I. Universal domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2720–2729.
9. Bousmalis, K.; Silberman, N.; Dohan, D.; Erhan, D.; Krishnan, D. Unsupervised pixel-level domain adaptation with generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3722–3731.
10. Hong, W.; Wang, Z.; Yang, M.; Yuan, J. Conditional generative adversarial network for structured domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 July 2018; pp. 1335–1344.
11. Rau, A.; Edwards, P.E.; Ahmad, O.F.; Riordan, P.; Janatka, M.; Lovat, L.B.; Stoyanov, D. Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy. *Int. J. Comput. Assist. Radiol. Surg.* **2019**, *14*, 1167–1176.
12. Benjdira, B.; Bazi, Y.; Koubaa, A.; Ouni, K. Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images. *Remote Sens.* **2019**, *11*, 1369.
13. Zhao, Y.; Guo, P.; Sun, Z.; Chen, X.; Gao, H. ResiDualGAN: Resize-residual DualGAN for cross-domain remote sensing images semantic segmentation. *Remote Sens.* **2023**, *15*, 1428.
14. Han, J.; Shoeiby, M.; Petersson, L.; Armin, M.A. Dual contrastive learning for unsupervised image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 746–755.
15. Sharma, S.; Buddhiraju, K.M.; Banerjee, B. An ant colony optimization based inter-domain cluster mapping for domain adaptation in remote sensing. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 2158–2161. https://doi.org/10.1109/IGARSS.2014.6946894.
16. Dorigo, M.; Birattari, M.; Stutzle, T. Ant colony optimization. *IEEE Comput. Intell. Mag.* **2006**, *1*, 28–39.
17. Liu, Y.; Li, X. Domain adaptation for land use classification: A spatio-temporal knowledge reusing method. *ISPRS J. Photogramm. Remote Sens.* **2014**, *98*, 133–144.
18. Banerjee, B.; Chaudhuri, S. Hierarchical Subspace Learning Based Unsupervised Domain Adaptation for Cross-Domain Classification of Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5099–5109. https://doi.org/10.1109/JSTARS.2017.2732682.
19. Postadjian, T.; Bris, A.L.; Sahbi, H.; Malle, C. Domain Adaptation for Large Scale Classification of Very High Resolution Satellite Images with Deep Convolutional Neural Networks. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018), Valencia, Spain, 22–27 July 2018; pp. 3623–3626. https://doi.org/10.1109/IGARSS.2018.8518799.
20. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
21. Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.Y.; Isola, P.; Saenko, K.; Efros, A.; Darrell, T. Cycada: Cycle-consistent adversarial domain adaptation. In Proceedings of the International Conference on Machine Learning (PMLR), Stockholm, Sweden, 10–15 July 2018; pp. 1989–1998.
22. Zhang, W.; Zhu, Y.; Fu, Q. Adversarial Deep Domain Adaptation for Multi-Band SAR Images Classification. *IEEE Access* **2019**, *7*, 78571–78583. https://doi.org/10.1109/ACCESS.2019.2922844.
23. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
24. Tasar, O.; Happy, S.; Tarabalka, Y.; Alliez, P. SemI2I: Semantically consistent image-to-image translation for domain adaptation of remote sensing data. In Proceedings of the 2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2020), Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: New York, NY, USA, 2020; pp. 1837–1840.
25. Tasar, O.; Happy, S.; Tarabalka, Y.; Alliez, P. ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7178–7193.
26. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2849–2857.
27. Rottensteiner, F.; Sohn, G.; Jung, J.; Gerke, M.; Baillard, C.; Bnitez, S.; Breitkopf, U.; International Society for Photogrammetry and Remote Sensing. 2d Semantic Labeling Contest. 2020. Volume 29. Available online: https://www.isprs.org/education/benchmarks/UrbanSemLab/semantic-labeling.aspx (accessed on 31 October 2023).
28. Park, T.; Efros, A.A.; Zhang, R.; Zhu, J.Y. Contrastive learning for unpaired image-to-image translation. In Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part IX 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 319–345.
29. Fu, H.; Gong, M.; Wang, C.; Batmanghelich, K.; Zhang, K.; Tao, D. Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2427–2436.
30. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual Event, 13–18 July 2020; pp. 1597–1607.
31. Chen, Y.; Zhao, Y.; Jia, W.; Cao, L.; Liu, X. Adversarial-learning-based image-to-image transformation: A survey. *Neurocomputing* **2020**, *411*, 468–486.

32. Alqahtani, H.; Kavakli-Thorne, M.; Kumar, G. Applications of generative adversarial networks (gans): An updated review. *Arch. Comput. Methods Eng.* **2021**, *28*, 525–552.
33. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 9729–9738.
34. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv* **2018**, arXiv:1802.05957.
35. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*.
36. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE). *Geosci. Model Dev. Discuss.* **2014**, *7*, 1525–1534.
37. Yuanji, W.; Jianhua, L.; Yi, L.; Yao, F.; Qinzhong, J. Image quality evaluation based on image weighted separating block peak signal to noise ratio. In Proceedings of the International Conference on Neural Networks and Signal Processing, Nanjing, China, 14–17 December 2003; IEEE: New York, NY, USA, 2003; Volume 2, pp. 994–997.
38. Bai, L.; Du, S.; Zhang, X.; Wang, H.; Liu, B.; Ouyang, S. Domain adaptation for remote sensing image semantic segmentation: An integrated approach of contrastive learning and adversarial learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13.
39. Piñeiro, G.; Perelman, S.; Guerschman, J.P.; Paruelo, J.M. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol. Model.* **2008**, *216*, 316–322.