Generative Adversarial Model Equipped with Contrastive Learning in Map Synthesis

Arpan Mahara
amaha038@fiu.edu
Florida International University
Miami, Florida, USA

Naphtali D. Rishe
rishen@cs.fiu.edu
Florida International University
Miami, Florida, USA

Abstract

In the dynamic field of urban planning and the context of unprecedented natural events, such as hurricanes, the fast generation of accurate maps from satellite imagery is paramount. While several studies have utilized Generative Adversarial Networks (GANs) for map generation from satellite images, the present work introduces a new approach by integrating contrastive learning into the GAN framework for enhanced map synthesis. Our methodology distinctively employs positive sampling by aligning similar features (e.g., roads) in both satellite images and their corresponding map outputs, and contrasts this with negative samples for disparate elements. This approach effectively replaces the conventional cyclic process in GANs with a more streamlined, unidirectional procedure, leading to improvements in both the quality of the synthesized maps and computational efficiency. We show the effectiveness of our proposed model, offering an advancement in map generation for remote sensing applications.

CCS Concepts: • **Image Processing**; • **Synthesis and Visualization**; • **Computational Imaging**; • **Deep Learning for Images**;

Keywords: Anchor, Contrastive Learning, Generative Adversarial Networks (GANs), CycleGAN, Generation, Map, PatchNCE, Satellite Imagery, Synthesis

ACM Reference Format:
Arpan Mahara and Naphtali D. Rishe. 2024. Generative Adversarial Model Equipped with Contrastive Learning in Map Synthesis. In *Proceedings of 2024 6th International Conference on Image Processing and Machine Vision (IPMV 2024) (IPMV '24).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

IPMV '24, Jan 12–14, 2024, Macau, China

ACM ISBN 979-8-4007-0847-3/24/01
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The emergence of satellite technology has revolutionized the way we capture and utilize images of Earth's surface. Access to these images, whether in real-time or offline, has become important in various applications. Among these, the generation of readable maps from satellite imagery stands out as one of the major applications, especially for navigation and geographic information systems. This need has motivated researchers to develop computational models capable of transforming satellite images into maps that are more intuitive for human interpretation.

Initially, the focus was on paired image translation methods, where models were trained on datasets consisting of corresponding pairs of satellite images and maps. Innovative works like that of Isola et al. [19] demonstrated the feasibility of using Generative Adversarial Networks (GANs) for map generation tasks with remarkable success. However, a significant limitation of these approaches is their reliance on paired datasets, which are time-consuming and costly to compile. This challenge led to the emergence of unsupervised methods, such as CycleGAN [32], which can learn to translate between domains from unpaired datasets. Despite not being specifically designed for map generation, CycleGAN and other models integrating and adapting a cyclic procedure in synthesis have shown promising results in this domain, as evidenced by the works of Ganguli et al. [11], Song et al. [27], and Chen et al. [7]. The work of Seo et al. [26] demonstrates the adaptability of GANs to a wide range of image generation tasks, adapting the DCGAN architecture for the unique challenges of colorizing grayscale images using a one-to-one training approach.

Map synthesis holds enormous potential, particularly for emergency response during natural disasters, such as earthquakes, wildfires, and floods, and for urban planning [28]. In these scenarios, the ability to synthesize up-to-date maps from satellite images at a fast pace is invaluable. However, the cyclic nature of models like CycleGAN introduces significant computational overhead, making them less practical for time-sensitive applications [18]. Responsive to this challenge, the integration of the emergent unsupervised learning paradigm of contrastive learning within the GAN framework shows promise for map synthesis. The utilization of contrastive learning techniques is particularly appealing given their demonstrated success in the remote sensing domain, as

shown in studies by Bai et al. [4] and Abbasnejad et al. [1], along with their broader impact evidenced by revolutionary methods such as SimCLR [6] and MoCo [16] across various fields.

Gutmann and Hyvärinen's work is foundational in contrastive learning, focusing on distinguishing actual data from noise, laying the groundwork for advanced machine learning applications [14]. Although primarily focused on statistical models rather than direct applications in GANs for image translation, the principles in their paper highlight key aspects of contrastive learning. Subsequently, Oord et al. introduce a method for learning high-dimensional data representations by capturing shared information across different parts of a signal, enhancing data synthesis quality [24]. Building on the insights from the work of Oord et al. [24], Park et al. demonstrate a practical application of contrastive learning concepts in a GAN framework [25], particularly through representation learning at the patch level of images. The work of Park et al. suggests the potential for applying these advanced contrastive learning techniques to map synthesis from satellite imagery.

2 Related Works

In the current literature, map synthesis is primarily conducted using conditional GANs, which can be broadly categorized into two approaches: paired and unpaired translation. In the paired translation context, the utilization of the Pix2Pix model by Isola et al. has shown promising results in map synthesis [19]. However, the real-world scarcity of paired satellite and map images necessitated exploration into GANs capable of practical map synthesis with unpaired datasets. Zhu et al.'s work stands out as a foundational contribution in this domain [32].

Following the core idea of CycleGAN [32], several literature studies explored the generation of maps from satellite images. Expanding on CycleGAN's concept, the GeoGAN model proposed by Ganguli, Garzon, and Glaser significantly advances map generation from satellite images [11]. It uniquely combines reconstruction and style transfer losses with a conditional GAN to enhance the quality of map synthesis. This innovative approach, particularly its third model architecture, yields more accurate map features, showcasing the evolving techniques in map synthesis from satellite imagery. Following the development of GeoGAN, the researchers introduce the Semantic-regulated Geographic GAN (SG-GAN), which integrates crowdsourced vehicle GPS coordinates into the map synthesis process [31]. This model adopts the Pix2Pix framework as its backbone, enhancing it with additional layers of GPS data and semantic estimations to reduce noise and improve accuracy in areas with sparse geographic information. The SG-GAN [31] approach not only enriches the map generation process with external geographic data but also maintains the standard adversarial training, demonstrating an advanced application of GANs in satellite-to-map image conversion. In the subsequent year, Song et al. introduced MapGen-GAN [28], an enhancement of the CycleGAN approach in map generation. This model incorporates circularity and geometrical consistency constraints to refine the translation of remote-sensing images into maps. These innovations enable MapGen-GAN to achieve more accurate and reliable map generation, particularly in emergency response scenarios.

Consequently, it is important to note that the above generative models focus on either the pix2pix model (paired mechanism) or the cyclic approach (unpaired translation). These approaches are often not realistically applicable due to the lack of paired data or the computational overhead associated with the cyclic process, as suggested in the work by Kazemi et al. [20]. Not specifically designed and tested on map synthesis, contrastive learning in GAN proposed by Park et al. [25] successfully replaced the cyclic approach and yielded good results on image quality and complexity in image-to-image translation. Extending upon the work of Park et al., Han et al. [15] presents a novel method for unsupervised image-to-image translation that utilizes a dual learning setting with two encoders. This approach smartly integrates contrastive learning within a cyclic procedure, enabling the model to learn mappings between two unpaired domains effectively. The given work achieved better performance in image-to-image translation tasks compared to previous cyclic and contrastive models, as evidenced by the Fréchet Inception Distance (FID) score [17], a metric used to measure the similarity between distributions of real images and generated images, reflecting improvements in both variety and quality. Similarly, SCONE-GAN, as presented by Abbasnejad et al. [1], advances image translation by incorporating contrastive learning with graph convolutional networks to generate more realistic and diverse scenery images. This approach effectively maintains image structure, maximizes mutual information between style and output, and demonstrates enhanced performance on four datasets. It is worth mentioning that contrastive learning with mutual information maximization at the patch level has not yet been explored dominantly on map synthesis given satellite images. The present work revisits the contrastive learning approach proposed by Park et al. [25] in map synthesis to achieve better results based on visualization and time complexity.

3 The Proposed Approach

This section presents a detailed procedure on how Contrastive Learning was utilized by selecting positive and negative samples on patch level adapting for our specific domain in the present work for Map Synthesis. The architecture of our model is depicted in Figure 1, which illustrates the integration of patch-based contrastive learning within the generative adversarial framework.

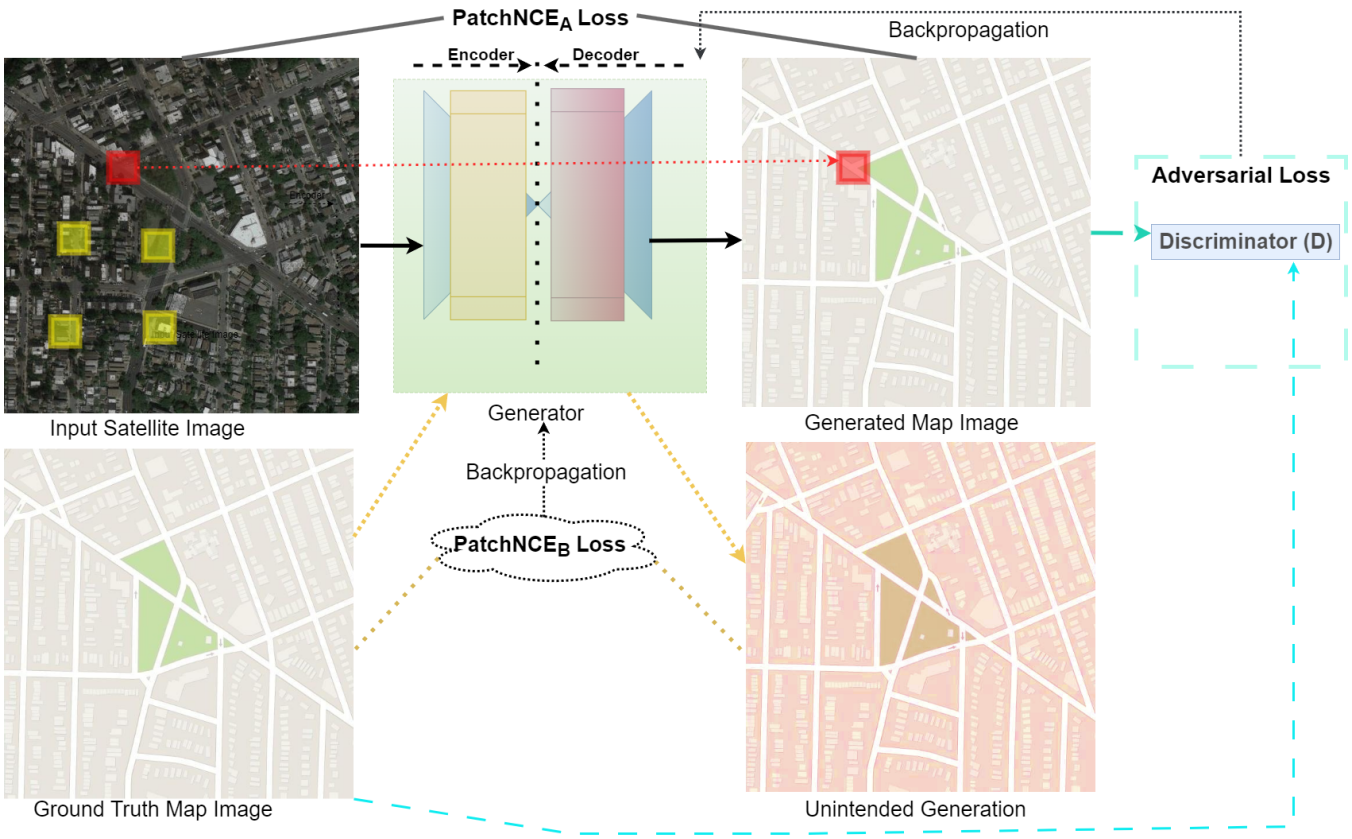

Figure 1. Overview of the Proposed Model's Architecture. The model processes an input satellite image using a generator implemented with patch-based contrastive learning. The PatchNCE$_A$ loss function is utilized, which maximizes the mutual information between corresponding objects within the red highlighted areas, while minimizing the same against negative samples depicted by the yellow rectangles. To prevent unintended translations during synthesis, the model incorporates PatchNCE$_B$ loss together with adversarial loss, ensuring the generation of realistic-looking map images.

3.1 Overall Architecture

3.1.1 Goal. The goal of this work is to generate maps in domain B from the input satellite images from domain A, by learning the reasonable mapping between two images, a and b, given unpaired instances $A = \{a \in A\}$, $B = \{b \in B\}$. The problem can also be understood as translating images from domain $A \subset \mathbb{R}^{H \times W \times C}$, to target domain $B \subset \mathbb{R}^{H \times W \times C}$, where H is height, W is width, and C is the number of channels of given satellite images respectively. In our translation challenge, the focus is on rendering roads in map images with distinct coloration, differentiating them from other features, such as greenery subjects (vegetation), which is represented in green and can be seen in Figure 1. We aim to simplify complex urban areas present in satellite images, retaining only essential details to aid navigation. This selective detailing ensures that important navigational elements are emphasized, while extraneous information is minimized, resulting in a clear and user-friendly map output. CycleGANs implement a cyclic approach to achieve the final goal by first translating $a \in A$ to $b \in B$ and then back to a and vice versa by using two different generators and two corresponding discriminators to discriminate map images generated by the given generators. This methodology and its variations have been explored and demonstrated in the works of Hsieh et al. [18], Ganguli et al. [11], and Song et al. [28]. This cyclic procedure of generating images, in general, is very time consuming and restrictive because the process considers the relation between the input domain and the target domain is a bijection [13]. To address this, we adopt an alternative method, i.e., Contrastive Learning in conditional GAN, as presented by Park et al. [25], by learning translation in one direction using only one generator and one discriminator.

3.1.2 Generator Architecture in present work. The present work employs a ResNet-based generator, proven successful in generative models [8], integral for synthesizing map images from satellite images. This generator is an essential component of our Generative Adversarial Network (GAN) framework, designed to capture and translate the

complex spatial and textural information present in satellite images into the corresponding map representations.

ResNet-based Generator. Motivated by the work of Zhu et al. [32], the present work's generator features a series of 9 residual blocks, which form the backbone of both the encoding and decoding processes. These blocks enable the model to handle deep feature extraction efficiently, preserving crucial details and structures essential for accurate map generation.

Encoder-Decoder Structure. Structured into an encoder and a decoder, the generator first encodes the satellite image into a latent space, extracting key features and patterns. The decoder then reconstructs these features into a map image, ensuring that the generated map retains fidelity to the input satellite image while introducing the stylistic elements of map visuals.

Support for Contrastive Learning. The architecture of the generator is specifically designed to support the contrastive learning approach and the PatchNCE loss, as presented in the study by Park et al. [25], which are detailed in the following subsections. The effectiveness of these methods in our model support the robust feature extraction and synthesis capabilities of the generator.

3.1.3 Discriminator Architecture. Following the architecture of the ResNet-based generator adapted in the present work, our model adapts a discriminator essential for the adversarial training process in our GAN framework. In the adversarial training setup, the discriminator's task is to accurately classify real and synthesized map images, providing a learning signal to the generator to improve the quality of its output. This process forms a crucial part of the GAN framework, enabling the generation of high-fidelity map images from satellite data. For the discriminator architecture, we select PatchGAN with a 70x70 patch size due to its successful application in generative models as noted by Alqahtani et al. [2] to determine if the given image is synthesized or real. This particular discriminator, instead of directly determining if the entire image is real or fake, first divides the given generated map image or ground truth map image into 70x70 patches. Then, it evaluates and scores each patch individually on its authenticity. The final decision about the image being real or generated is based on the average of these scores. This design choice helps the generator generate realistic-looking map images by focusing detail at the patch level [9]. The discriminator uses LeakyReLU [22] for non-linearity and has been initialized with an emphasis on stability in the training process. Also, we stick to instance normalization rather than batch normalization to have high-quality map images since instance normalization was successful in generating image-to-image translation on similar but different domains [32].

Having established the foundation of our model's generator and discriminator architecture, we now delve into the specifics of our contrastive learning approach and the effective use of the PatchNCE loss function in map synthesis, adapted from CUT model as detailed by Park et al. [25].

3.1.4 Contrastive Learning Approach. In this work, the essence of our contrastive learning approach lies in transforming satellite imagery into simplified map representations, as illustrated in Figure 2. Starting with a satellite image a from domain A and a corresponding map image b from domain B, we define an anchor feature $f \in \mathbb{R}^K$ extracted from b, and a corresponding positive feature $p \in \mathbb{R}^K$ extracted from a, 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 domain A. These negative features are sourced either from different locations within the same image or from different images, as depicted in Figure 2, and distinctly highlighted with yellow colored squares.

The contrastive loss function in this work aims to decrease the embedding space between f and p, without collapsing them into a single point. Simultaneously, it tends to widen the separation in the feature space between f and the set of negative features $\{n_i\}$. These actions allow inherent variability and help to avoid overly restrictive mappings [3]. For example, we do not want to make vegetation and forest (two different components) strictly as negative or positive, and we do not want the distance sampled to be zero. So, considering this, our loss is defined as:

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

Here, the similarity measure $\text{sim}(f, p)$ is defined as the dot product between ℓ_2 normalized vectors f and p, which is essentially the cosine similarity (i.e., $\text{sim}(f, p) = \frac{f \cdot p}{\|f\| \|p\|}$), as mentioned by Chen et al. [6]. τ is a temperature parameter that scales the similarity scores. By minimizing this loss, our framework facilitates a feature space where positive pairs are closer compared to anchor-negative pairs, yet not identical, mirroring the complex relationship between satellite imagery and map representations.

3.2 Formulation

3.2.1 Modified PatchNCE Loss for Suitable Contrastive Learning in Map Synthesis. In our refined approach to the PatchNCE loss for translating satellite images to map images, we enhance the negative sampling strategy. This strategy involves selecting negative samples from various locations within the same image, as well as from entirely different satellite images.

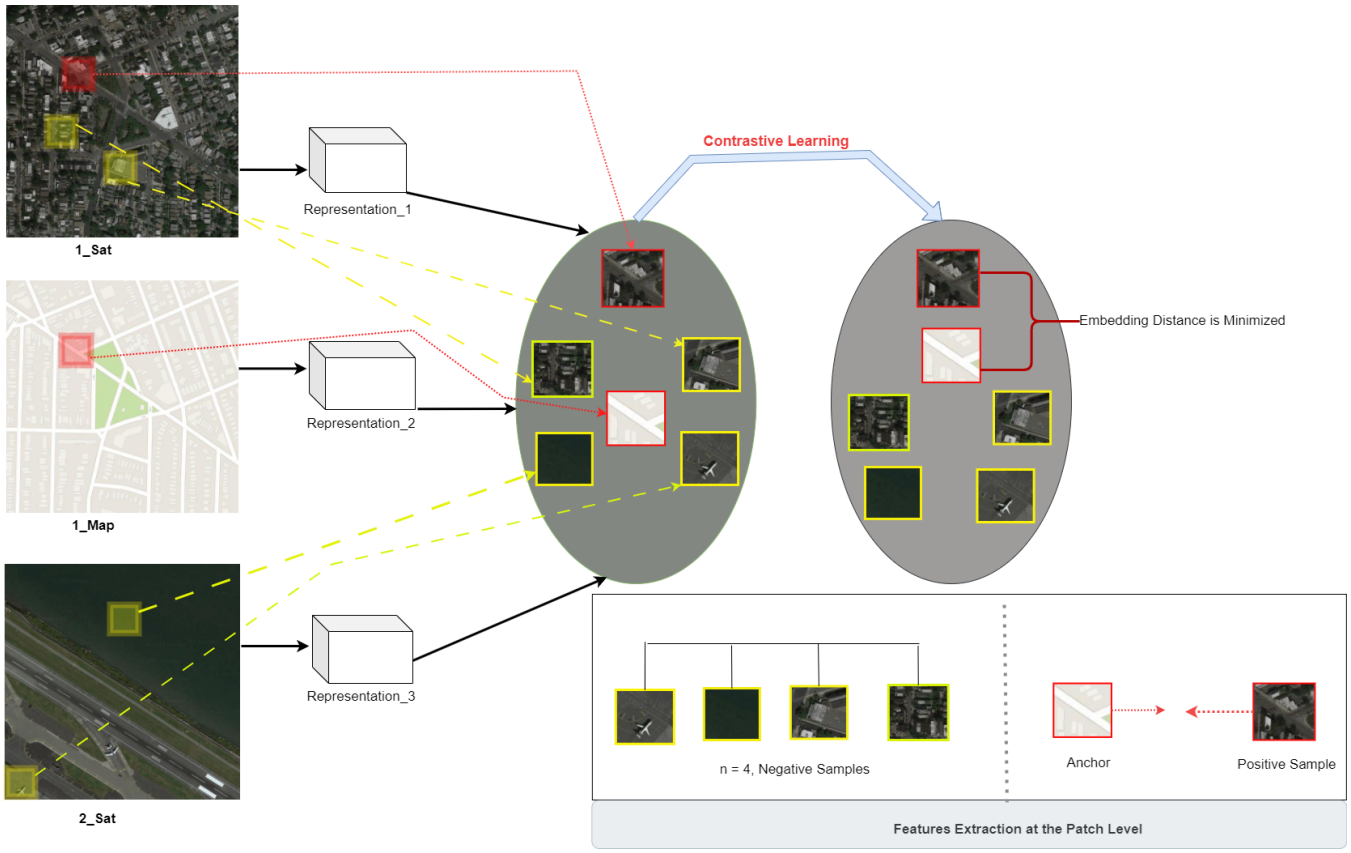

Figure 2. Contrastive Learning and Feature Selection at the Patch Level. Our model samples an anchor patch from the output b (1_Map) and compares it with the corresponding input patch at the same location in a (1_sat), both highlighted with red colored squares. 'n' negative patches are drawn from the same input image (1_Sat) and from a randomly selected different image within domain A (2_Sat), these are highlighted with yellow colored squares. This contrastive learning approach is designed to efficiently leverage both input and output patches within a shared embedding space, aiming to enhance the precision of our map synthesis.

Multilayer, Patchwise Contrastive Learning. Building upon the methodology outlined by Park et al. [25], our approach integrates a two-layer Multilayer Perceptron (MLP) network represented as H_x, within specific encoder layers. This MLP network is used to transform the feature maps from each layer into an enhanced feature stack. In this architecture, each layer, along with its respective spatial location in the encoder's feature hierarchy, corresponds to a distinct patch of the initial image. As we delve into deeper layers, these patches increase in size. We focus on X key layers and pass the feature maps through the MLP network, producing a set of enhanced features ${z_x}^X = H_x(G_x^{enc}(a))^X$, where G_x^{enc} denotes the output of the x-th selected layer of the encoder.

PatchNCE Loss with Enhanced Features. The adapted PatchNCE loss, termed PatchNCE-SAT, involves a generator G, a set of layers H within the network, and an input data distribution A is formulated as:

$$L_{\text{PatchNCE-SAT}}(G,H,A) = \mathbb{E}_{a\sim A}\Bigg[\sum_{x=1}^{X}\sum_{s=1}^{S_x} X\Big(\hat{z}_x^s, z_x^s, \{z_x^{S\setminus s}\}, \{z_x^{\text{diff,diffloc}}\}\Big)\Bigg] \tag{2}$$

In equation (2), $\hat{z}_x^s$ represents the feature representation from the generated image at layer x, and patch s. z_x^s is the corresponding feature from the real image. $\{z_x^{S\setminus s}\}$ are features from other patches within the same image at the same layer. $\{z_x^{\text{diff,diffloc}}\}$ are the negative samples drawn from different locations in other satellite images at the same layer. X denotes the number of layers, and S_x is the number of sampled patches at layer x.

This PatchNCE-SAT loss, by utilizing the refined features from the MLP network H_x, ensures a robust and nuanced contrastive learning mechanism, vital for the generation of

accurate and contextually coherent map images from satellite data.

3.2.2 GAN Loss for Generation of Realistic-Looking Maps. To ensure the generation of realistic-looking maps from satellite images, the present work utilizes an adversarial loss function, also known as GAN loss, initially proposed by Goodfellow et al. [12]. The GAN loss comprises two main components: the loss for the generator and the loss for the discriminator. The objective is to train the generator to produce map images that are indistinguishable from real map images while the discriminator learns to differentiate between the real and generated maps. The GAN loss can be formulated as:

$$\begin{aligned} L_{\text{GAN}} =& \mathbb{E}_{b\sim B}[\log D(b)] + \\ & \mathbb{E}_{a\sim A}[\log(1 - D(G(a)))] \end{aligned} \tag{3}$$

In equation (3), $D(b)$ represents the discriminator's decision for a ground truth map image b. $G(a)$ is the generated map image from the input satellite image a that should hypothetically correspond to b. A is the distribution formed by real satellite images, and B is the distribution formed by real map images. The generator G aims to minimize this loss, while the discriminator D aims to maximize it, leading to a minimax game between the two [12].

3.2.3 Final Loss Functions. The objective of our model is twofold: first, to generate realistic-looking map images from satellite images (domain A), and second, to ensure that the corresponding patches between the input satellite images and the output map images are closely aligned. Simultaneously, our model aims to differentiate between the anchor patch and other non-corresponding patches within the same satellite image and across different satellite images, employing these as negatives in the contrastive learning process. To accomplish these goals, we integrate multiple loss functions. This includes the GAN loss for realism, the PatchNCE-SAT loss for ensuring patch-level correspondence in domain A, and a similar PatchNCE loss for domain B to prevent inappropriate translation by the generator, similar to the identity loss presented by Zhu et al. [32]. Our work's combined final loss function is formulated as follows:

$$\begin{aligned} L_{\text{final}} =& \lambda_{\text{GAN}} L_{\text{GAN}}(G, D, A, B) + \\ & \lambda_{\text{PatchNCE-A}} L_{\text{PatchNCE-SAT}}(G, H, A) + \\ & \lambda_{\text{PatchNCE-B}} L_{\text{PatchNCE-SAT}}(G, H, B) \end{aligned} \tag{4}$$

4 Experiments

For the experiments and comparison, we used a public Google map dataset collected by Zhu et al. [32]. We selected 1096 images for training and 500 for testing our model and compared it to state-of-the-art methods. All the images have dimensions of 256x256. For comparison, we utilized different evaluation metrics, as presented below.

4.1 Evaluation Metrics

In this study, we utilize a set of evaluation metrics to assess the performance of our model, focusing on the accuracy and quality of the image generation results. These metrics include the Root Mean Square Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), and the Structural Similarity Index Measure (SSIM).

4.1.1 Root Mean Square Error (RMSE). The RMSE is a widely used measure of the differences between values predicted by a model and the values observed [5]. For an original image O and its estimated counterpart E, each of size $m \times n$, the RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(O_{ij} - E_{ij})^2} \tag{5}$$

In the context of GANs for image generation, a lower RMSE value typically indicates better quality of the generated images, as it signifies a smaller average difference between the generated and the original images [21].

4.1.2 Peak Signal-to-Noise Ratio (PSNR). PSNR is a popular metric for measuring the quality of reconstruction of lossy compression codecs [30]. Essentially, it measures the ratio of the maximum potential power of a signal to the power of the noise that distorts its representation. As a function of RMSE, the PSNR is computed as follows:

$$\text{PSNR} = 20 \cdot \log_{10}\left(\frac{\text{MAX}_O}{\text{RMSE}}\right) \tag{6}$$

where MAX_O represents the maximum possible pixel value of the image.

4.1.3 Structural Similarity Index (SSIM). The SSIM index is a metric used to assess the perceived quality of digital images and videos. By analyzing different segments of an image, SSIM evaluates the visual impact caused by variations in luminance, contrast, and structure. This comparison is done between a predicted image and the original one [29]. The SSIM index is represented by the following formula:

$$SSIM(x, y) = \frac{2xy + C}{x^2 + y^2 + C}. \tag{13}$$

In this equation, x and y have similarity values within the range of [0,1]. The term C is a constant, introduced to avoid division by zero, which prevents the output of an undetermined value. The SSIM value lies between 0 and 1, where a value of 1 signifies ideal identical images.

4.2 Experimental Environment and Baselines

Experiment Setting. We conducted our experiments in a Python 3.10.12 environment, utilizing the PyTorch framework for all deep learning tasks. The computational workload was handled by an NVIDIA V100 GPU, which comes

with 15GB of GPU RAM, ideal for handling demanding image generation tasks. Our experimental procedure spanned 40 epochs. During the first half of these epochs, we maintained a steady learning rate, whereas, in the latter half, we allowed the learning rate to decrease linearly. This approach facilitated more precise adjustments as the model neared its convergence point. We set the temperature parameter $\tau = 0.07$. Along with this, we set the loss functions' parameters: $\lambda_{\text{GAN}} = 1$, $\lambda_{\text{PatchNCE-A}} = 1$, and $\lambda_{\text{PatchNCE-B}} = 2$. We didn't just focus on standard performance metrics; we also paid close attention to the model's time complexity. This involved careful monitoring and recording of the time taken for each epoch to complete, from which we calculated an average epoch duration. These observations helped us estimate the training time required for the model, adding a valuable perspective for comparing models based on time efficiency.

Baselines. For our comparative analysis, we selected three well-established generative adversarial network models as our baselines, DCLGAN [15], CycleGAN [32], and GcGAN [10]. These models were chosen due to their relevance and proven effectiveness in image generation and translation tasks, which aligns with the objectives of our study. Our proposed method was evaluated against these models using the chosen dataset and the specified evaluation metrics (RMSE, PSNR, and SSIM), providing a comprehensive view of its effectiveness in generating map images.

4.3 Comparison and Results

The comparative performance of our model against other GANs, as mentioned above, is presented in Table 1, complemented by visualizations in Figure 3. These results demonstrate the better performance of our model. The detailed performances of the proposed model and the comparative models are summarized as follows:

- **Our Model:** Achieved an RMSE value of 43.8872, attained a PSNR value of 28.2572, and reached an SSIM value of 0.6255. Our model was trained in approximately 1.76 hours, equivalent to about 1 hour and 45 minutes.
- **CycleGAN:** Recorded an RMSE of 48.1944 and a PSNR of 27.7143, both slightly lower in performance than our model. The SSIM was 0.5915, indicating less structural similarity to the target images compared to our model. The training duration for CycleGAN was marginally longer at approximately 1.97 hours, close to 2 hours.
- **DCLGAN:** Showed an RMSE of 47.4574 and a PSNR of 28.2961, slightly better than CycleGAN and our model. It achieved a better SSIM value of 0.6336 compared to our model, yet the training time was the longest at approximately 2.67 hours, or about 2 hours and 40 minutes.
- **GcGAN:** Recorded the highest RMSE of 58.1702 and the lowest PSNR of 26.3392, indicating lower performance in map synthesis. It achieved an SSIM value of 0.4680, the lowest among the models. Although it was the fastest to train, taking only about 0.89 hours, or approximately 53 minutes, it failed to generate visually appealing results.

The quantitative results in Table 1 and the qualitative results, as visualized in Figure 3, collectively highlight the better image quality and training efficiency of our model in comparison to existing methods. While DCLGAN [15] demonstrates slightly better performance metrics, it does so at the cost of longer training times. In scenarios where a balance between computational efficiency and accuracy is important, our model emerges as a preferable choice. Moreover, despite GcGAN's [10] faster training capability, it falls short in producing visually appealing results, emphasizing the trade-offs inherent in model selection.

5 Conclusion and Future Work

In this study, we have demonstrated the practical application of contrastive learning within a GAN framework for map synthesis from satellite imagery. Our generative approach involves drawing an anchor from the output map image, aligning it with the corresponding positive sample from the input image, and contrasting it with negative samples from different locations within the same and different satellite images. These selections are made at the patch level of the images, employing PatchNCE-SAT, as detailed in the previous sections, to maximize mutual information between two corresponding elements while minimizing information between unrelated elements. This method has successfully enabled the GAN model to establish a more accurate mapping between input and output, thus enhancing the quality of the generated map images. The experimental results, assessed using RMSE, PSNR, and SSIM metrics, have shown that incorporating contrastive learning into GANs yields better outcomes in map synthesis compared to existing GAN models. Notably, while DCLGAN demonstrates slightly better performance in some aspects, it requires longer training times, making our model a more efficient choice in terms of computational resources and time. Similarly, despite its faster training, GcGAN does not match the visual quality achieved by our model. These comparisons emphasize our model's balance of efficiency and quality, making it a strong candidate for practical applications like emergency rescue operations. This study not only highlights the effectiveness of our approach in map synthesis but also lays possibilities for future advancements in image processing in remote sensing with generative models.

This study primarily focused on image data for map synthesis without incorporating additional parameters. Moving forward, we aim to integrate geographical features, such

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

Models	RMSE	PSNR	SSIM	Time for Training (in Hr)
Our Model	43.8872	28.2572	0.6255	1.76
CycleGAN	48.1944	27.7143	0.5915	1.97
DCLGAN	47.4574	28.2961	0.6336	2.67
GcGAN	58.1702	26.3392	0.4680	0.89

Figure 3. Comparative Results of Map Synthesis. The figure presents a side-by-side comparison of map images generated by various models, including Our Model, against the Ground Truth, underscoring the effectiveness of each approach in synthesizing accurate map details.

as latitude and longitude, to the potential enhancement of the map generation process. Additionally, we plan to collect and utilize our dataset, to be downloaded using the TerraFly Mapping System developed and managed by the High Performance Database Research Center at Florida International University, as previously detailed by Mahara and Rishe [23]. This approach is predicted to provide us with satellite images with additional parameters, varying resolutions, and imagery from different seasons. These variations may potentially contribute to broadening the generalizability of our model in generating maps from satellite imagery. Given the significant role of data augmentation in contrastive learning, we plan to explore and design new augmentation techniques specifically designed for our domain.

6 Acknowledgment

This material is based in part upon work supported by the National Science Foundation under Grant MRI20 CNS-2018611

References

[1] Iman Abbasnejad, Fabio Zambetta, Flora Salim, Timothy Wiley, Jeffrey Chan, Russell Gallagher, and Ehsan Abbasnejad. 2023. SCONE-GAN: Semantic Contrastive Learning-Based Generative Adversarial Network for an End-to-End Image Translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1111–1120.

[2] Hamed Alqahtani, Manolya Kavakli-Thorne, and Gulshan Kumar. 2021. Applications of generative adversarial networks (gans): An updated review. *Archives of Computational Methods in Engineering* 28 (2021), 525–552.

[3] Asha Anoosheh, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. 2018. Combogan: Unrestrained scalability for image domain translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 783–790.

[4] Lubin Bai, Shihong Du, Xiuyuan Zhang, Haoyu Wang, Bo Liu, and Song Ouyang. 2022. Domain Adaptation for Remote Sensing Image Semantic Segmentation: An Integrated Approach of Contrastive Learning and Adversarial Learning. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–13. https://doi.org/10.1109/TGRS.2022.3198972

[5] Tianfeng Chai and Roland R Draxler. 2014. Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific model development discussions* 7, 1 (2014), 1525–1534.

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*. PMLR, 1597–1607.

[7] Xu Chen, Songqiang Chen, Tian Xu, Bangguo Yin, Jian Peng, Xiaoming Mei, and Haifeng Li. 2020. SMAPGAN: Generative adversarial network-based semisupervised styled map tile generation method. *IEEE Transactions on Geoscience and Remote Sensing* 59, 5 (2020), 4388–4406.

[8] Yuan Chen, Yang Zhao, Wei Jia, Li Cao, and Xiaoping Liu. 2020. Adversarial-learning-based image-to-image transformation: A survey. *Neurocomputing* 411 (2020), 468–486.

[9] Ugur Demir and Gozde Unal. 2018. Patch-based image inpainting with generative adversarial networks. *arXiv preprint arXiv:1803.07422* (2018).

[10] Huan Fu, Mingming Gong, Chaohui Wang, Kayhan Batmanghelich, Kun Zhang, and Dacheng Tao. 2019. Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2427–2436.

[11] Swetava Ganguli, Pedro Garzon, and Noa Glaser. 2019. GeoGAN: A conditional GAN with reconstruction and style loss to generate standard layer of maps from satellite images. *arXiv preprint arXiv:1902.05611* (2019).

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[13] Yao Gou, Min Li, Yu Song, Yujie He, and Litao Wang. 2023. Multi-feature contrastive learning for unpaired image-to-image translation. *Complex & Intelligent Systems* 9, 4 (2023), 4111–4122.

[14] Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. 9)*, Yee Whye Teh and Mike Titterington (Eds.). PMLR, Chia Laguna Resort, Sardinia, Italy, 297–304. https://proceedings.mlr.press/v9/gutmann10a.html

[15] Junlin Han, Mehrdad Shoeiby, Lars Petersson, and Mohammad Ali Armin. 2021. Dual contrastive learning for unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 746–755.

[16] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9729–9738.

[17] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* 30 (2017).

[18] Yi-Yen Hsieh, Yu-Chi Lee, and Chia-Hsiang Yang. 2020. A CycleGAN Accelerator for Unsupervised Learning on Mobile Devices. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. 1–5. https://doi.org/10.1109/ISCAS45731.2020.9180845

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. Image-To-Image Translation With Conditional Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[20] Hadi Kazemi, Sobhan Soleymani, Fariborz Taherkhani, Seyed Iranmanesh, and Nasser Nasrabadi. 2018. Unsupervised image-to-image translation using domain-specific variational information bound. *Advances in neural information processing systems* 31 (2018).

[21] Ashish Kumar, Abeer Alsadoon, PWC Prasad, Salma Abdullah, Tarik A Rashid, Duong Thu Hang Pham, and Tran Quoc Vinh Nguyen. 2022. Generative adversarial network (GAN) and enhanced root mean square error (ERMSE): deep learning for stock price movement prediction. *Multimedia Tools and Applications* (2022), 1–19.

[22] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, Vol. 30. Atlanta, GA, 3.

[23] Arpan Mahara and Naphtali Rishe. [n. d.]. Integrating Location Information as Geohash Codes in Convolutional Neural Network-Based Satellite Image Classification. ([n. d.]).

[24] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).

[25] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. 2020. Contrastive learning for unpaired image-to-image translation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*. Springer, 319–345.

[26] Junghoon Seo, Taewon Yoon, Jinwoo Kim, and Kin Choong Yow. 2017. One-to-one example-based automatic image coloring using deep convolutional generative adversarial network. *Journal of Advances in Information Technology Vol* 8, 2 (2017).

[27] Jieqiong Song, Hao Chen, Chun Du, and Jun Li. 2023. Semi-MapGen: Translation of Remote Sensing Image Into Map via Semisupervised Adversarial Learning. *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023), 1–19. https://doi.org/10.1109/TGRS.2023.3263897

[28] Jieqiong Song, Jun Li, Hao Chen, and Jiangjiang Wu. 2021. MapGen-GAN: A fast translator for remote sensing image to map via unsupervised adversarial learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), 2341–2357.

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[30] Wang Yuanji, Li Jianhua, Lu Yi, Fu Yao, and Jiang Qinzhong. 2003. Image quality evaluation based on image weighted separating block peak signal to noise ratio. In *International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003*, Vol. 2. IEEE, 994–997.

[31] Ying Zhang, Yifang Yin, Roger Zimmermann, Guanfeng Wang, Jagannadan Varadarajan, and See-Kiong Ng. 2020. An Enhanced GAN Model for Automatic Satellite-to-Map Image Conversion. *IEEE Access* 8 (2020), 176704–176716. https://doi.org/10.1109/ACCESS.2020.3025008

[32] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2223–2232.