Cloud Removal from Satellite Imagery Using Generative Adversarial Networks

Manas Khanger
Department of Computing Technologies SRM Institute of Science and Technology Chennai, India
mk0011@srmist.edu.in

Nikhil Sharad Sharma
Department of Computing Technologies SRM Institute of Science and Technology Chennai, India
ns3411@srmist.edu.in

Dr R Subash
Associate Professor
Department of Computing Technologies SRM Institute of Science and Technology Chennai, India
subashr@srmist.edu.in

***Abstract*—Cloud cover remains one of the major challenges in the effective utilization of optical satellite imagery for Earth observation applications such as agriculture monitoring, land-use analysis, and climate studies. A significant portion of satellite data becomes unusable due to cloud obstruction, resulting in spatial and temporal discontinuities in analysis. Traditional cloud removal techniques based on interpolation, temporal compositing, or physical modeling often fail to accurately reconstruct occluded regions or preserve spectral integrity. This paper presents a supervised cloud removal framework based on a Spatial Attention Generative Adversarial Network (SpA GAN), which leverages spatial attention mechanisms to learn a direct pixel-to-pixel mapping between cloudy satellite images and their corresponding cloud-free counterparts. The model is trained using temporally aligned Copernicus Sentinel-2 image pairs separated by a two-day interval to minimize land-cover variation. Experimental results demonstrate that the proposed approach effectively removes cloud cover while preserving spatial structure and spectral consistency, making it suitable for real-world remote sensing applications.**

***Index Terms*—Cloud Removal, Satellite Imagery, Generative Adversarial Networks, Spatial Attention, SpA GAN, Sentinel-2, Image-to-Image Translation**

I. Introduction

Optical satellite imagery plays a critical role in Earth observation by enabling continuous monitoring of land surfaces, vegetation, water bodies, and urban regions. However, cloud contamination significantly reduces the usability of optical imagery, with a large percentage of satellite acquisitions being partially or completely obscured. This limitation disrupts downstream applications such as precision agriculture, disaster assessment, and environmental monitoring.

Conventional cloud removal methods rely on threshold-based cloud detection followed by interpolation or multi-temporal compositing [2]. Although computationally efficient, these approaches often introduce visual artifacts and fail to preserve fine spatial and spectral details. Recent advances in deep learning, particularly Generative Adversarial Networks (GANs), have shown strong potential in image restoration tasks by learning data-driven reconstruction patterns.

In this work, we propose a supervised cloud removal framework using a Spatial Attention Generative Adversarial Network (SpA GAN) trained on paired cloudy and cloud-free Sentinel-2 images. By incorporating spatial attention mechanisms and learning a direct pixel-to-pixel mapping, the model reconstructs cloud-free imagery while preserving spatial continuity and spectral fidelity.

II. Literature Survey

Early cloud mitigation techniques primarily focused on radiometric correction and rule-based cloud masking, followed by interpolation of missing regions [3]. These methods attempt to detect cloud pixels using spectral thresholds and reconstruct the hidden areas using spatial interpolation. However, such approaches lack robustness under dense cloud cover and complex terrain conditions, often resulting in artifacts or inaccurate reconstruction of the underlying surface. Multi-temporal compositing approaches improved performance by combining observations from multiple time instances, but they depend heavily on the availability of frequent cloud-free observations.

Recent research has explored deep learning-based approaches for cloud removal in satellite imagery. Convolutional Neural Networks (CNNs) have been applied to learn spatial patterns in remote sensing data and reconstruct missing image regions. While these models improve reconstruction quality compared to traditional methods, they often produce overly smooth outputs and struggle to preserve fine textures.

More recently, Generative Adversarial Networks (GANs) have been widely used for image restoration and image-to-image translation tasks. GAN-based models such as Cloud-GAN introduced adversarial learning to enhance visual realism in reconstructed images [4]. Multispectral GANs further improved reconstruction by exploiting spectral correlations between satellite bands, while SAR–optical fusion approaches addressed cloud opacity limitations by integrating radar imagery. However, many existing methods rely on unpaired data or emphasize perceptual quality rather than pixel-level accuracy.

The proposed approach addresses these limitations by employing a Spatial Attention Generative Adversarial Network (SpA GAN) trained in a supervised manner on temporally aligned cloudy and cloud-free Sentinel-2 image pairs. The spatial attention mechanism enables the network to focus on cloud-affected regions while preserving unaffected areas, resulting in accurate pixel-to-pixel reconstruction with preserved spatial structures and spectral characteristics.

III. Dataset Description

The proposed model is trained using Copernicus Sentinel-2 imagery obtained from the CloudSEN12 benchmark dataset. Sentinel-2 provides multispectral images with spatial resolutions ranging from 10 to 60 meters and a revisit cycle of approximately five days. Cloudy and cloud-free image pairs are selected with a temporal gap of two days to minimize land-cover variation. All images are resized to 256×256 pixels and normalized to the range $[-1, 1]$ prior to training.

IV. System Overview

The proposed cloud removal system is designed as an **end-to-end supervised image-to-image translation framework** that transforms cloud-contaminated optical satellite imagery into corresponding cloud-free representations. The system operates on multispectral satellite images and is specifically optimized for cloud removal in Copernicus Sentinel-2 data.

Initially, **cloudy satellite images** are ingested into the system and subjected to a preprocessing stage. This stage includes spatial resizing, pixel normalization, and band alignment to ensure compatibility with the SpA GAN model architecture. Preprocessing also ensures consistent spatial correspondence between cloudy and cloud-free image pairs, which is critical for supervised pixel-to-pixel learning.

The preprocessed images are then fed into the **Spatial Attention Generative Adversarial Network (SpA GAN)**. The generator network, implemented using a spatial attention architecture with encoder–decoder layers and skip connections, learns to reconstruct cloud-free imagery by synthesizing missing surface information at the pixel level. The spatial attention mechanism enables the network to selectively focus on cloud-contaminated regions while preserving unaffected areas. Skip connections enable the preservation of low-level spatial features, thereby maintaining structural continuity in the reconstructed output. Simultaneously, a **PatchGAN discriminator** is employed to enforce realism in the generated outputs. Rather than evaluating the image as a whole, the discriminator operates on local image patches and distinguishes between real cloud-free images and generated reconstructions conditioned on the input cloudy image. This localized adversarial supervision encourages the generator to produce high-frequency details and reduces smoothing artifacts commonly observed in image restoration tasks.

During the **training phase**, both the generator and discriminator are optimized jointly using adversarial learning, guided by a combination of adversarial loss, pixel-level reconstruction loss, and attention loss. This optimization ensures that the generated outputs are not only visually realistic but also closely aligned with the ground-truth cloud-free images.

During the **inference phase**, the discriminator is discarded, and only the trained generator is utilized. Given an unseen cloudy satellite image, the generator directly produces a corresponding cloud-free image in a single forward pass. This design enables efficient deployment and fast inference, making the system suitable for practical remote sensing applications requiring large-scale image processing.

architecture.png

Fig. 1. Proposed SpA GAN-based cloud removal architecture.

V. Methodology

A. Problem Formulation

Cloud removal from optical satellite imagery is formulated as a **supervised image-to-image translation problem**, where the goal is to reconstruct a cloud-free satellite image from a cloud-contaminated input. Let $x \in \mathbb{R}^{H \times W \times C}$ represent a cloudy satellite image with height H, width W, and C spectral channels, and let $y \in \mathbb{R}^{H \times W \times C}$ denote the corresponding cloud-free image of the same scene.

The objective of the proposed approach is to learn a deterministic mapping function G, parameterized by a deep neural network, such that:

$$G(x) \rightarrow y \tag{1}$$

This mapping is learned using **paired cloudy and cloud-free satellite images**, enabling pixel-to-pixel correspondence during training. The cloudy image x serves as the input, while the cloud-free image y acts as the ground truth target. By leveraging supervised learning, the model aims to minimize the discrepancy between the generated output $G(x)$ and the true cloud-free image y.

The learning process focuses on reconstructing cloud-obscured regions while preserving spatial continuity and spectral consistency in non-cloudy areas. To achieve this, the generator is optimized to produce visually realistic and spectrally accurate cloud-free images that are indistinguishable from real observations. The formulation ensures that for every cloud-affected pixel in the input image, a corresponding cloud-free pixel is synthesized, enabling precise pixel-level restoration.

This problem formulation is particularly suitable for cloud removal tasks in remote sensing, where accurate reconstruction of surface information is critical for downstream applications such as land-use analysis, vegetation monitoring, and environmental assessment.

B. SpA GAN Architecture

The Spatial Attention Generative Adversarial Network (SpA GAN) consists of a generator G and a discriminator D. The generator leverages spatial attention mechanisms to translate cloudy images into cloud-free images, while the discriminator evaluates the authenticity of generated outputs.

1) Generator Architecture: The generator employs a spatial attention network (SPANet) architecture. The spatial attention mechanism generates an attention map that identifies cloud-affected regions, allowing the network to focus reconstruction efforts on occluded areas while preserving cloud-free regions. The architecture comprises encoder–decoder layers with skip connections that preserve spatial details by directly transferring low-level features from the encoder to the decoder, enabling accurate pixel-level reconstruction.

2) Discriminator Architecture: The discriminator follows a PatchGAN architecture, classifying local image patches as real or fake. This design encourages the preservation of high-frequency details and reduces blurring artifacts in the generated outputs. The discriminator consists of convolution layers, batch normalization, and Leaky ReLU activations.

3) Loss Function: The total loss of SpA GAN combines three components: adversarial loss, L_1 reconstruction loss, and attention loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{GAN}}(G, D) + \lambda_c \mathcal{L}_{L1}(G) + \mathcal{L}_{\text{att}}(A, M) \quad (2)$$

The first component is the adversarial loss:

$$\mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}[\log D(y)] + \mathbb{E}[\log(1 - D(G(x)))] \quad (3)$$

The second component is the standard L_1 loss, where λ_c is a hyper-parameter controlling the weight of each channel:

$$\mathcal{L}_{L1}(G) = \mathbb{E}\left[\|y - G(x)\|_1\right] \quad (4)$$

The third component is the attention loss, where A is the attention map and M is the cloud mask computed from $M = |I_{in} - I_{gt}|_1$:

$$\mathcal{L}_{\text{att}}(A, M) = \mathbb{E}\left[\|A - M\|_1\right] \quad (5)$$

VI. Training and Inference

During training, paired cloudy and cloud-free images are provided to the model. The generator produces cloud-free predictions, and the discriminator provides adversarial feedback. Both networks are optimized iteratively. During inference, only the trained generator is used to reconstruct cloud-free images from unseen cloudy inputs.

VII. Experimental Setup

The proposed cloud removal model is trained using the **Adam optimizer**, which is well-suited for deep neural network optimization due to its adaptive learning rate and momentum-based updates. The initial learning rate is set to 2×10^{-4}, following standard practices for training Generative Adversarial Networks, ensuring stable convergence during adversarial training.

Training is conducted over multiple epochs on a **GPU-enabled computing environment**, enabling efficient processing of high-resolution satellite imagery and accelerating the iterative optimization process. The model is trained using **mini-batch gradient descent**, where each batch consists of paired cloudy and cloud-free satellite images. This batch-wise training strategy improves generalization and reduces memory overhead.

To ensure stable adversarial learning, the generator and discriminator are trained alternately. During each iteration, the generator produces a cloud-free image from a cloudy input, while the discriminator evaluates the authenticity of the generated image against the ground-truth cloud-free image. The training objective combines adversarial loss with pixel-level reconstruction loss and attention loss, enabling the generator to balance visual realism with structural accuracy.

Model performance is evaluated using widely adopted quantitative metrics for image quality assessment. **Peak Signal-to-Noise Ratio (PSNR)** is used to measure the overall reconstruction fidelity, and **Structural Similarity Index (SSIM)** assesses the preservation of spatial structures and textures. These complementary metrics provide a comprehensive assessment of the model's effectiveness in reconstructing cloud-free satellite imagery.

All experiments are conducted under consistent training and evaluation settings to ensure reproducibility and fair comparison. The evaluation is performed on a held-out test set consisting of unseen cloudy satellite images, allowing objective assessment of the model's generalization capability.

VIII. Results and Discussion

Experimental results demonstrate that the proposed SpA GAN-based framework is capable of effectively removing cloud cover from satellite imagery while preserving important spatial structures and surface details. The generator network successfully reconstructs cloud-free regions by learning the relationship between cloudy inputs and their corresponding cloud-free counterparts. Visual inspection of the generated outputs shows that the model is able to recover land features such as vegetation patterns, water bodies, and urban regions that were previously obscured by cloud cover.

The use of **temporally aligned training data** plays a crucial role in improving reconstruction accuracy. By training the model on cloudy and cloud-free image pairs captured within a short time interval, the system minimizes inconsistencies caused by land-cover changes. This temporal alignment allows the network to learn meaningful pixel-to-pixel correspondences, resulting in more accurate and realistic reconstructions.

The **spatial attention mechanism** contributes significantly to the quality of the generated images by enabling the network to focus on cloud-affected regions. The attention map generated by the model identifies areas requiring reconstruction, allowing the generator to allocate more capacity to heavily occluded regions while preserving already cloud-free areas. The **PatchGAN discriminator** further enforces local image realism by examining smaller image patches and encouraging the generator to produce sharper textures and fine spatial details. At the same time, the L_1 **reconstruction loss** ensures that the generated output remains close to the ground-truth cloud-free image, preserving pixel-level accuracy and reducing structural distortions.

A. Quantitative Comparison

Table I presents a quantitative comparison of the proposed SpA GAN with conditional GAN (cGAN) and cycle GAN on the RICE1 dataset using PSNR and SSIM metrics. The results demonstrate that SpA GAN significantly outperforms both baseline methods across both metrics.

TABLE I
QUANTITATIVE COMPARISON ON RICE1 DATASET

Method	**PSNR**	**SSIM**
cGAN	26.547	0.903
cycle GAN	25.880	0.893
SpA GAN	**30.232**	**0.954**

SpA GAN achieves a PSNR of 30.232 dB and an SSIM of 0.954, outperforming cGAN by 3.685 dB in PSNR and 0.051 in SSIM, and cycle GAN by 4.352 dB in PSNR and 0.061 in SSIM. These results confirm that the spatial attention mechanism enables more accurate reconstruction of cloud-free imagery suitable for remote sensing analysis.

Despite these promising results, certain limitations remain. The model requires **paired cloudy and cloud-free datasets**, which may not always be available for all geographic regions. Additionally, training GAN-based models can be computationally intensive and requires access to GPU resources for efficient processing. Future work can focus on improving training efficiency and exploring approaches that reduce dependency on paired datasets.

IX. CONCLUSION AND FUTURE WORK

This paper presented a Spatial Attention Generative Adversarial Network (SpA GAN) based framework for cloud removal from satellite imagery. By leveraging supervised learning, spatial attention mechanisms, and temporally aligned Sentinel-2 data, the model achieves accurate pixel-to-pixel cloud reconstruction. The proposed SpA GAN significantly outperforms conventional cGAN and cycle GAN approaches in terms of both PSNR and SSIM metrics. Future work will focus on integrating SAR data, improving model generalization across diverse regions, and deploying the framework as a real-time cloud removal service.

REFERENCES

[1] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1125–1134, 2017.

[2] P. Singh and N. Komodakis, "Cloud-GAN: Cloud removal for Sentinel-2 imagery using cyclic consistent GAN," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1772–1775, 2018.

[3] X. Liu, Z. Pan, and J. Lei, "Thin cloud removal from remote sensing images using conditional generative adversarial networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3573–3583, 2019.

[4] C. Hasan, R. Horne, S. Mauw, and A. Mizera, "Cloud removal from satellite imagery using multispectral edge-filtered conditional generative adversarial networks," *International Journal of Remote Sensing*, vol. 43, no. 5, pp. 1881–1893, 2022.

[5] P. Ebel, M. Schmitt, and X. X. Zhu, "Cloud removal in unpaired Sentinel-2 imagery using cycle-consistent adversarial networks," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 929–932, 2020.

[6] Y. Zhang, L. Zhang, X. Bai, and L. Zhang, "Cloud and cloud shadow removal from optical satellite imagery using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 12, pp. 1795–1799, 2018.

[7] C. Gröhnfeldt, X. X. Zhu, and M. Schmitt, "Learning to remove clouds from Sentinel-2 imagery using multitemporal data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 6203–6218, 2018.

[8] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2672–2680, 2014.

[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.

[10] European Space Agency, "Copernicus Sentinel-2 mission," ESA, 2020. [Online]. Available: https://sentinel.esa.int

[11] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2223–2232, 2017.

[12] T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8798–8807, 2018.

[13] Z. Zheng, Y. Yang, and Y. Liu, "Multitemporal cloud removal from remote sensing images using deep learning," *Remote Sensing of Environment*, vol. 237, pp. 111–123, 2020.

[14] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS: A curated dataset of georeferenced multispectral Sentinel-1/2 imagery for deep learning," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W7, pp. 153–160, 2019.

[15] Q. Zhang, Y. Yuan, and Q. Wang, "Learning-based cloud removal for remote sensing images using generative adversarial networks," *IEEE Access*, vol. 7, pp. 171123–171134, 2019.

[16] H. Li, Y. Zhang, and L. Zhang, "Deep learning-based cloud detection and removal for satellite imagery," *Remote Sensing*, vol. 11, no. 1, pp. 44–56, 2019.