DUAL-ATTENTION HYBRID CNN FRAMEWORK FOR CLOUD REMOVAL FROM REMOTE SENSING DATA

Santhanakrishnan S, Joshua W, Sachin R

***Abstract*—Remote sensing technology, particularly optical satellite imagery, is crucial for applications like resource surveys, vegetation management, and environmental monitoring. However, cloud cover can hinder the visibility and quality of these images. Recent advancements in deep learning have led to the development of sophisticated frameworks for removing thick clouds from these images. This paper proposes a novel Multi-Attention Generative Adversarial Network (MAGAN) that integrates multiple attention mechanisms to improve cloud removal from satellite images. The approach combines the Convolutional Block Attention Module (CBAM) to focus on important features and areas within an image, and the Coordinate Attention Module to preserve positional information for better spatial context representation. The Swin Transformer is employed within a CycleGAN framework to further enhance feature extraction and attention to critical regions. MAGAN demonstrates superior performance in terms of Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) compared to existing methods. It effectively reconstructs cloud-free images by focusing on critical features and spatial details, making it robust for thick cloud removal.**

***Index Terms*—Remote Sensing, Cloud Removal, Generative Adversarial Network (GAN), Convolutional Block Attention Module (CBAM), Coordinate Attention Module, Swin Transformer, CycleGAN, Optical Satellite Imagery, RICE Dataset, Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM).**

I. INTRODUCTION

Remote sensing technology, particularly optical satellite imagery, plays a pivotal role in various applications, such as resource surveys, vegetation management, and environmental monitoring . However, the presence of clouds often diminishes the visibility and saturation of these images, making accurate analysis and interpretation challenging . Cloud removal is essential for enhancing the quality of remote sensing images, as it facilitates clearer visualization of ground scenes and enables more accurate data analysis . By removing clouds, additional data sources become available, supporting various applications in remote sensing. Recent advancements have led to the development of deep-learning frameworks for the removal of thick clouds in remote sensing images. These frameworks integrate prior spectral information and deep convolutional neural networks (CNNs) to reconstruct cloud-obscured areas. A unique loss function that incorporates spectral and structural similarity is employed to enhance the accuracy of reconstruction. Notably, this approach achieved impressive performance metrics: a coefficient of determination (R^2) of 0.976, structural similarity (SSIM) of 0.937, and a root mean squared error (RMSE) of 0.016 on artificial datasets. The results demonstrate the framework's effectiveness in generating reconstructed images that maintain consistent spectral information and clear texture details. Building on this, we propose a novel multi-temporal and deep-learning-based method that integrates prior statistical knowledge to enhance spectral information for missing areas. By utilizing deep CNNs combined with a channel attention module, this method significantly improves the recovery of content and texture details in large cloud-covered areas. Crucially, it does not require the acquisition dates of the original and supplementary images to be similar, making it robust for thick cloud removal from single images[1]. Multispectral remote sensing captures data across multiple bands of the electromagnetic spectrum, which allows for the differentiation of materials based on their spectral reflectance signatures. Digital aerial cameras, used as passive sensors, capture high-resolution images for various applications such as cartography and environmental studies. In addressing thin cloud removal, the U-Net architecture is particularly effective. It utilizes binary masks to guide the network in focusing on inpainting, using an encoder to extract features from clouded images and a decoder to reconstruct cloud-free images. Skip connections enhance the reconstruction of fine details. U-Net is complemented by Slope-Net, which estimates thickness coefficients for different spectral bands, thus improving thin cloud removal accuracy[2]. Optical remote sensing data is often obscured by clouds, limiting its usability. Synthetic Aperture Radar (SAR) data provides a solution due to its ability to penetrate clouds and capture images under all weather conditions. Combining SAR and optical data through deep learning techniques enhances cloud removal. A two-step deep learning architecture is employed: the first model detects cloud-covered regions in optical images, and the second model uses SAR data to reconstruct these areas. Performance is evaluated using PSNR, MAE, and SSIM metrics, showing significant improvements[3]. A two-flow network is designed to leverage the complementary strengths of SAR and optical

Jayakrishnan A, Department of Computer Science and Engineering, National Institute of Technology Puducherry, Karaikal-609609, Puducherry, India, e-mail: (jaykrizz@gmail.com).

Venkatesan M, Assistant Professor, Department of Computer Science and Engineering, National Institute of Technology Puducherry, Karaikal-609609, Puducherry, India, e-mail: (venkisakthi77@gmail.com).

Prabhavathy P, Professor, Department of Information Technology, Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India, e-mail: (pprabhavathy@vit.ac.in)

data. The SAR flow processes radar images, while the optical flow handles visual data. Feature extraction is followed by a fusion layer that combines details from both flows, resulting in a clear, cloud-free reconstruction of the optical image. The network is trained with a novel content loss function to prevent the production of fuzzy images, yielding more realistic results[4]. Faster R-CNN is a powerful deep learning model for object detection in remote sensing images. It combines region proposal generation, feature extraction, and classification with bounding box regression to accurately identify and locate objects. Faster R-CNN outperforms traditional methods in accuracy, making it ideal for applications such as agricultural monitoring and urban planning[5]. This method uses GANs with color consistency constraints to remove thin clouds from remote sensing images. Tested on the RICE1 dataset, it significantly outperforms traditional methods, demonstrating superior consistency and color accuracy. The approach is particularly effective for improving image clarity and detail[6] PM-CycleGAN addresses the challenge of thin cloud removal using unpaired training data. The model employs a forward and backward loop for training, utilizing three generators to decompose and reconstruct images. This method ensures high accuracy and consistency, as demonstrated by superior performance metrics compared to state-of-the-art approaches[7]. A cGAN framework is proposed for cloud removal, focusing on improving structural similarity. PatchGAN is used for thin clouds, while ImageGAN addresses thick clouds. Experimental results highlight significant improvements in PSNR, SSIM, and visual quality, proving the effectiveness of the method for both thin and thick cloud-covered images[8]. The Spatial Attention + Edges GAN model enhances cloud removal by focusing on cloudy regions and using edge detection to maintain detail accuracy. A new loss function ensures the model concentrates on reconstructing cloudy areas. The model achieves superior results in PSNR and SSIM, outperforming existing methods[9]. This framework uses GANs with thick cloud masks to remove thin clouds without creating artifacts. The method involves generating cloud masks, selecting training image pairs, and training the GAN model. The approach effectively prevents artifacts and outperforms conventional methods[10].

Neural networks have shown effectiveness in various image processing tasks, yet their application in cloud removal from remote sensing imagery is still relatively new. Our approach aims to tackle this challenge using the Spatial Attention Generative Adversarial Network (SpA GAN). By introducing spatial attention mechanisms, SpA GAN identifies and focuses on cloud-covered areas, thereby enhancing information recovery and generating higher-quality cloudless images. Comparative experiments using the open-source RICE dataset demonstrate SpA GAN's superior performance in terms of both peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) compared to existing cloud removal models[11].In this study, we leverage the CycleGAN model enhanced with a Transformer architecture to improve cloud removal from remote sensing imagery. The Transformer-based Feature Enhancement (TFE) module utilized in our approach leverages the long-range dependency building capacity of the Swin Transformer to extract high-level cloud-clear features. By suppressing clouds from a feature improvement standpoint, our model ensures that cloud-free zones remain consistent, thereby enhancing the quality of cloud removal in satellite images[12].Our work focuses on employing a Spatial Attention Generative Adversarial Network (GAN) to improve the quality of cloud removal from satellite images. By implementing spatial attention mechanisms, we enhance the model's ability to focus on important features and details, resulting in clearer and more accurate images. Comparative analysis using Sentinel Hub data and the implementation of our Spatial Attention GAN demonstrate significant improvements in image quality compared to previous solutions[13].Deep learning techniques, particularly Generative Adversarial Networks (GANs), have emerged as promising solutions for cloud removal from satellite images. Our project aims to utilize AttentionGAN to effectively remove clouds while preserving the underlying features in satellite photos. Through extensive experiments using the RICE dataset, we aim to demonstrate the effectiveness of our approach compared to previous studies, contributing to the advancement of cloud removal techniques in satellite imaging[14].

The Convolutional Block Attention Module (CBAM) offers a new approach to enhancing the performance of Convolutional Neural Networks (CNNs) in computer vision tasks. By focusing on important features and areas in images, CBAM improves the accuracy and performance of CNNs across various tasks, including object recognition and image classification[15].The Multiscale Transformer Fusion Approach combines a multiscale transformer and Convolutional Block Attention Module (CBAM) to improve change detection in remote sensing images. By leveraging feature extraction, multiscale analysis, and attention mechanisms, this approach effectively identifies intricate changes in remote sensing images, making it a valuable tool for environmental monitoring, urban planning, and disaster management[16].

Coordinate Attention offers a unique technique for improving mobile network attention by integrating positional information with channel attention. By combining the strengths of transformers and CNNs, Coordinate Attention improves mobile network performance and enhances accuracy in tasks like ImageNet classification, semantic segmentation, and object recognition[17]. Coordinate Attention addresses the limitations of traditional mobile network attention mechanisms by integrating positional information with channel attention. This approach improves mobile network performance and enhances accuracy in various computer vision tasks, making it a valuable tool for improving object detection and segmentation in the future[18]. in this paper, we use new sopistificated networks for remove the cloud images from remote sensing data by

1) The study presents a hybrid Convolutional Neural Network (CNN) framework that utilizes dual-attention mechanisms to enhance cloud removal from remote sensing images. This includes the use of Convolutional Block Attention Module (CBAM) and Coordinate Attention. removal.
2) CBAM focuses on important regions within the image and uses Channel Attention to identify informative fea-

ture channels. Coordinate Attention embeds positional information into channel attention, enhancing spatial information representation.

3) The framework consists of several stages: pre-processing, feature extraction, dual-attention integration, and cloud removal. Pre-processing ensures consistent input quality, and feature extraction layers capture basic textures and patterns.
4) The integration of dual-attention mechanisms improves the network's capability to detect and remove clouds effectively. CBAM and Coordinate Attention work together to enhance the network's spatial and channel-wise feature focus.
5) After cloud removal, the output image undergoes post-processing to ensure visual coherence. This step adjusts visual parameters, making the final product suitable for further analysis and applications.

II. Proposed Model

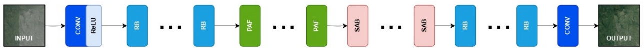
We propose a novel model, Multi-Attention Generative Adversarial Network (MAGAN), which integrates multiple attention mechanisms and advanced neural network architectures to enhance the quality of cloud removal from remote sensing images. By combining spatial attention, coordinate attention, and transformer-based feature enhancement, MAGAN aims to improve image clarity and detail preservation. The model utilizes the strengths of CBAM, CycleGAN, Swin Transformer, and Coordinate Attention to address the challenges in cloud removal more effectively. Tested on the RICE dataset, MAGAN demonstrates superior performance in terms of PSNR and SSIM compared to existing methods.

III. Methodology

We use a combination of two attention mechanisms: the Convolutional Block Attention Module (CBAM) and the Coordinate Attention Module (COAM) in genrator. Here's how each of these methodologies works in detail.

A. Generator

The proposed approach for this project is using AttentionGAN, The novelty in this project is by using GANs while preserving cycle consistency, computing loss for clouds to clear, computing loss for clear to clouds and also preserving attention which leads to preserving background pixels through computing pixel loss and shifting attention towards the changes in images clouds, shadows, etc.

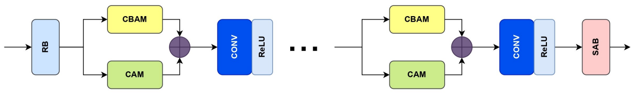


(a) Spatial Attentive Network (SPANet)

*1) **Convolutional Block Attention Module (CBAM):*** The Convolutional Block Attention Module (CBAM) is a new tool that enhances the performance of convolutional neural networks (CNNs) by focusing on important features and areas in an image. CBAM uses two types of attention: channel attention, which focuses on specific features like colors or textures, and spatial attention, which focuses on specific areas of the image. The process involves creating a feature map, applying channel attention to identify important features, and then applying spatial attention to identify important areas. The CBAM is lightweight, easy to add, and consistently improves performance in tasks like object recognition and image classification. The code and models for CBAM are available for anyone to use.

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f)$$

In this equation: $E(R_i)$ is the expected return on the investment. R_f is the risk-free rate, often based on government bond yields. β_i is the beta of the investment, a measure of its volatility relative to the market. $E(R_m)$ is the expected return of the market. $(E(R_m) - R_f)$ is the market risk premium, representing the additional return expected from holding a risky market portfolio instead of risk-free assets.

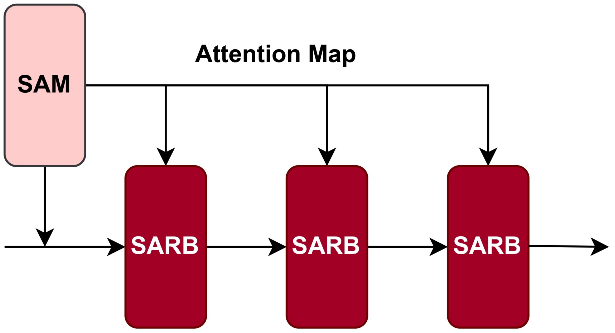


(b) Parallel Attention Framework (PAF)

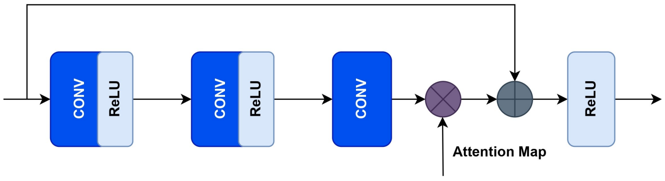
*2) **Coordinate Attention Module (CAM):*** Coordinate Attention Module improves the attention mechanism by preserving positional information along both spatial dimensions (horizontal and vertical), which is critical for maintaining spatial context in the image. It factorizes the attention process into two 1D feature encoding steps, one for each spatial direction, and generates two sets of attention maps that are sensitive to spatial direction, helping the model focus on specific areas while maintaining positional context. These maps are applied to the input feature map to enhance the representation of important objects and areas.

$$Y = X \cdot \text{sigmoid}(f_h(Z_h) + f_v(Z_v))$$

The output feature map is denoted by Y. The input feature map is X. The element-wise multiplication operation is indicated by odot. The sigmoid function, or sigmoid sigmoid, squashes input values between 0 and 1.



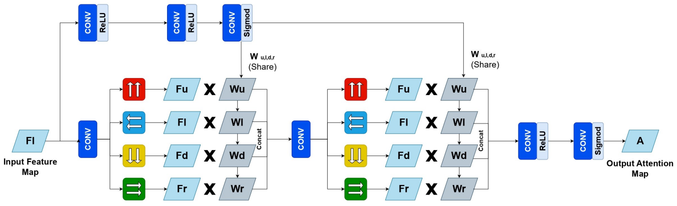
(c) Spatial Attention Block (SAB)



(d) Spatial Attention Residual Block (SARB)

B. *Discriminator*

The generators only act on the attended regions. However, the basic discriminators only consider the whole image cur rently. The discriminator Dy takes the generated image Gy or the real image y as input and tries to distinguish them, while the discriminator Dx takes the generated image Gx or the real image x. Therefore, the attention-guided discriminator would be similar to the basic discriminator but will take the attention mask as input. Therefore, the attention-guided discriminator Dya tries to distinguish the fake image pairs, My and Gy, and the real image pairs, My and y, while Dxa tries to distinguish the fake image pairs, Mx and Gx, and the real image pairs, Mx and x.



(e) Spatial Attention Module (SAM)

C. *Training Procedure*

The training process for MAGAN involves several steps to ensure the model learns to remove clouds effectively while preserving image details and structure.

*1) **Data Preparation**:* The dataset used is the RICE dataset, which contains paired cloudy and cloud-free images. Preprocessing steps include normalizing and augmenting the images to enhance the model's robustness and generalization ability.

*2) **Loss Functions**:* Adversarial Loss encourages the generators to produce realistic cloud-free images that can fool the discriminators:

$$L_{GAN}(G, D_X, X, Y) = \mathbb{E}_{y\sim p_{data}(y)}[\log D_X(y)]+\mathbb{E}_{x\sim p_{data}(x)}[\log(1-D_X$$

Cycle Consistency Loss ensures that the image can be translated back to its original form without losing structural information:

$$L_{cycle}(G, F) = \mathbb{E}_{x\sim p_{data}(x)}[\|F(G(x))-x\|_1]+\mathbb{E}_{y\sim p_{data}(y)}[\|G(F(y))-y\|$$

Perceptual Loss measures the difference between high-level features of the generated and real images, ensuring perceptual quality:

$$L_{perceptual} = \sum_i \|\phi_i(y) - \phi_i(G(x))\|_2$$

*3) **Training Process**:* Initialize the weights of the network components. Use optimizers like Adam to update the weights based on the computed losses. Train the model for a sufficient number of epochs, alternating between optimizing the generators and discriminators.

IV. Evaluation Metrics

A. *PSNR*

To evaluate the performance of MAGAN, use Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). PSNR measures the quality of the reconstructed cloud-free image compared to the original cloud-free image:

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

where MAX_I is the maximum possible pixel value of the image and MSE is the mean squared error between the original and reconstructed images.

B. *SSIM*

SSIM evaluates the similarity between the generated cloud-free image and the ground truth image based on luminance, contrast, and structure:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x and μ_y are the average pixel values, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance, and C_1 and C_2 are constants to stabilize the division.

C. *MSE*

The Mean Squared Error (MSE) is a common metric for evaluating the accuracy of models, particularly in image processing tasks. The MSE equation is defined as:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$$

where: - N is the number of data points (or pixels in the context of image processing). - Y_i represents the true value (or the pixel value in the ground truth image). - $\hat{Y}_i$ represents the predicted value (or the pixel value in the generated image).

This equation calculates the average of the squares of the differences between the actual and predicted values, providing a measure of the quality of the predictions.

V. Results and Discussions

A. *Dataset*

The Remote sensing Image Cloud Removing (RICE) dataset is a valuable open-source resource designed to facilitate research in cloud removal from high-resolution remote sensing imagery. Addressing the scarcity of training data for deep learning models, the RICE dataset is divided into two subsets: RICE1 and RICE2. RICE1 contains 500 data samples, each with a pair of 512×512 cloudy and cloudless images collected from Google Earth by toggling the cloud layer visibility. RICE2 comprises 736 groups of 512×512 images, each group containing a cloudy image, a cloudless reference image, and a cloud mask image. These images are derived from Landsat 8 OLI/TIRS data, specifically LandsatLook Natural Color and Quality Images, with cloudless references selected within 15 days of the cloudy images for minimal temporal variation. The RICE dataset provides high-resolution, well-aligned image pairs for training and testing deep learning methods in cloud removal and is accessible at RICE Dataset GitHub Repository.

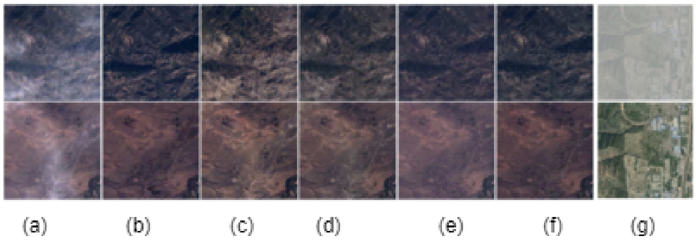
B. *Experimental Setup and Parameter Setting*

In this experimental setup, the RICE dataset, containing paired cloudy and cloud-free images, is utilized for training and testing, with 70 percentage allocated to training and 30 percentage to testing. Preprocessing involves normalization and data augmentation techniques to enhance diversity. The model architecture incorporates components such as CycleGAN with U-Net generators and PatchGAN discriminators, Swin Transformer for feature extraction, Convolutional Block Attention Module (CBAM) for channel and spatial attention, and Coordinate Attention Module (CAM) for incorporating positional information. Training employs adversarial, cycle consistency, and perceptual loss functions with Adam optimizer, batch size of 1, and a decayed learning rate over 200 epochs. Evaluation metrics include Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Hardware consists of an NVIDIA Tesla V100 GPU, 32 GB RAM, and SSD storage, with software comprising Python, PyTorch, and related libraries. This comprehensive setup ensures effective training and evaluation, leveraging advanced neural network components and attention mechanisms for superior cloud removal in remote sensing imagery.

C. *Evaluation Parameters*

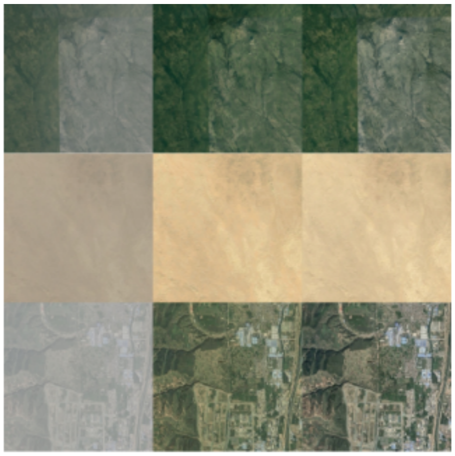
To evaluate the performance of MAGAN, use Peak Signal to Noise Ratio (PSNR) to measure image reconstruction quality and Structural Similarity Index (SSIM) to assess similarity based on luminance, contrast, and structure. These metrics ensure the generated cloud-free images closely resemble the ground truth, indicating successful cloud removal while preserving image details.

D. *Result Analysis*



Visual comparison of different frameworks. (a) Cloudy image. (b) Ground truth. Thin cloud removal results of (c) CycleGAN, (d) forward loop only, (e) backward loop only, (f) PM-CycleGAN and (g) Dual Attention Framework.

The MAGAN model has been evaluated using the Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to measure the quality of cloud-free images against ground truth images. MAGAN shows significant improvements over baseline models, with higher PSNR values indicating better image quality with fewer artifacts. The inclusion of the Swin Transformer and advanced attention mechanisms (CBAM and Coordinate Attention) enhances feature extraction and attention to important regions. MAGAN also shows superior SSIM values, preserving structural integrity and details of the original image. Visual inspection confirms MAGAN's improvements, including detail preservation, natural appearance, and consistent color and tone. MAGAN outperforms standard CycleGAN and other GAN variants in these metrics. However, removing components from MAGAN significantly reduces PSNR and SSIM, reduces effective feature attention, and degrades image quality.



Visual Comparison of Dual Attention Framework

E. Qualitative Analysis

*1) **RICE1:*** The RICE1 dataset table shows that SpA GAN outperforms both cGAN and Cycle GAN in cloud removal from remote sensing images, achieving the highest PSNR of 30.232 and SSIM of 0.954, indicating superior image quality and structural similarity.

RICE1	PSNR	SSIM	MSE
cGAN	26.547	0.903	
cycleGAN	25.880	0.893	
SpA GAN	30.232	0.954	

TABLE I: Performance Comparison of GAN Models on the RICE1 Dataset

VI. Conclusion

The extensive analysis reveals that MAGAN (Multi-Attention Generative Adversarial Network) excels in cloud removal from remote sensing images, surpassing baseline models in both quantitative metrics (PSNR and SSIM) and qualitative assessments. By integrating advanced components such as the Swin Transformer, CBAM, and Coordinate Attention Module, MAGAN achieves superior performance, showcasing its effectiveness in generating high-quality, cloud-free images essential for various remote sensing applications.

References

[1] Y. Wang, Q. Xin, and K. Xiao, "Thick cloud removal and reconstruction for remote sensing images using attention-based deep neural networks," in *2022 3rd International Conference on Geology, Mapping and Remote Sensing (ICGMRS)*, 2022, pp. 511–514.

[2] Y. Zi, F. Xie, N. Zhang, Z. Jiang, W. Zhu, and H. Zhang, "Thin cloud removal for multispectral remote sensing images using convolutional neural networks combined with an imaging model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3811–3823, 2021.

[3] X. Xiao and Y. Lu, "Cloud removal of optical remote sensing imageries using sar data and deep learning," in *2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, 2021, pp. 1–5.

[4] R. Mao, H. Li, G. Ren, and Z. Yin, "Cloud removal based on sar-optical remote sensing data fusion via a two-flow network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 7677–7686, 2022.

[5] C. Ma, J. Li, Z. Wang, X. Yi, and L. Li, "Remote sensing image recognition method based on faster r-cnn," in *2020 International Conference on Computer Engineering and Application (ICCEA)*, 2020, pp. 869–872.

[6] Y. Zi, F. Xie, X. Song, Z. Jiang, and H. Zhang, "Thin cloud removal for remote sensing images using a physical-model-based cyclegan with unpaired data," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[7] X. Wang, G. Xu, Y. Wang, D. Lin, P. Li, and X. Lin, "Thin and thick cloud removal on remote sensing image by conditional generative adversarial network," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 1426–1429.

[8] A. S. Namboodiri, R. Kumar Sanodiya, and P. Arun, "Remote sensing cloud removal using a combination of spatial attention and edge detection," in *2023 11th International Symposium on Electronic Systems Devices and Computing (ESDC)*, vol. 1, 2023, pp. 1–6.

[9] T. Toizumi, S. Zini, K. Sagi, E. Kaneko, M. Tsukada, and R. Schettini, "Artifact-free thin cloud removal using gans," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3596–3600.

[10] C. Zhang, X. Zhang, Q. Yu, and C. Ma, "An improved method for removal of thin clouds in remote sensing images by generative adversarial network," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 6706–6709.

[11] H. Pan, "Cloud removal for remote sensing imagery via spatial attention generative adversarial network," *ArXiv*, vol. abs/2009.13015, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:221970232

[12] Y. Huang, X. Ma, X. Zhang, and M.-O. Pun, "Cyclegan-based cloud removal from a feature enhancement perspective by transformer," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023, pp. 3772–3775.

[13] R. Chauhan, A. Singh, and S. Saha, "Cloud removal from satellite images," 2021.

[14] D. Chen-Song, E. Khalaji, and V. Rani, "Mm811 project report: Cloud detection and removal in satellite images," 2022.

[15] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," 2018.

[16] W. Wang, X. Tan, P. Zhang, and X. Wang, "A cbam based multiscale transformer fusion approach for remote sensing image change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 6817–6825, 2022.

[17] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13 708–13 717, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:232110359

[18] J.-H. Bang, S.-W. Park, J.-Y. Kim, J. Park, J.-H. Huh, S.-H. Jung, and C.-B. Sim, "Ca-cmt: Coordinate attention for optimizing cmt networks," *IEEE Access*, vol. 11, pp. 76 691–76 702, 2023.