Enhanced Cloud Removal in Sentinel-2 Imagery using Hybrid Spatiotemporal and Cycle-Consistent Generative Adversarial Networks

Lingeswaran A. [1*], Melvin Fredrick J.S..[2],Latha Selvi G.[3]

[1,2] Author's Affiliation: Student, Department of Information Technology , St. Joseph's College of Engineering, Chennai, India

[3]Author's Affiliation: Associate Professor, Department of Information Technology, St. Joseph's College of Engineering, Chennai, India

[3*]lathaselvi2012@gmail.com

Abstract :

Cloud cover poses a significant challenge for satellite imagery analysis, obstructing surface observations and creating data gaps that impede various applications, including land cover classification, weather forecasting, and disaster monitoring. Traditional cloud removal techniques and recent deep learning approaches have provided promising outcome but it fails in achieving high-quality, consistent results. This work presents a novel approach combining Spatiotemporal Generative Adversarial Networks and Cycle-Consistent Generative Adversarial Networks to improve the performance of cloud removal from Sentinel 2 satellite imagery. The proposed hybrid model leverages the temporal context provided by STGAN to generate initial cloud-free images by processing sequences of satellite images over time. These initial images are then refined using CycleGAN, which employs cycle consistency loss to ensure the transformation between cloudy and cloud-free images preserves essential features and realism. This combination addresses the limitations of previous methods by ensuring both temporal consistency and high image quality. This work demonstrates the potential of integrating spatiotemporal and cycle-consistent approaches to significantly enhance cloud removal processes, offering a robust solution for real-time monitoring and analysis in various satellite imagery downstream applications.

Keywords: Cloud removal, CycleGAN, Satellite imagery, Sentinel 2, STGAN.

Introduction

Satellite images play a vital role in many different areas from environmental monitoring and agricultural management to disaster response and weather forecasting. But, often some bodies of cloud cover are disturbing the satellite data by either covering a part of the earth's surface or inserting holes in the data. This issue is one of the main reasons why more research was made on providing reliable and accurate satellite images, which are the basis for the different remote sensing applications. Hence, the challenge of cloud detection and removal has led to intensive study of potential algorithms and prototyping a cloud-free satellite image. The traditional approach for cloud removal has used simple methods such as interpolation and filtering. These approaches do enhance the image in a way, however, mostly, they do not provide the quality of consistency in the results of the processing which is necessary for the analysis. The application of deep learning methods of artificial intelligence has been the main contributor to the development of the aforementioned non-traditional solutions.

For the cloud identification and removal problems in high-resolution satellite images, we put forward the approach of combining Spatiotemporal Generative Adversarial Networks (STGAN) with Cycle-Consistent Generative Adversarial Networks (CycleGAN). Our dual network approach makes the best of the strengths of both networks to obtain superior cloud removal from Sentinel 2 satellite images. The first stage of STGAN has the capability to utilize temporal adjacent images to accomplish the initial prediction by using information from previous and next points. The synthetic images are then improved by CycleGAN to ensure that the cloud-to-cloud-free transformation is through cycle consistency loss, and clarity and composition are not disturbed. Our technique tries to deal with the downsides of earlier techniques and at the same time provide a more robust cloud removal method by using these two advanced GAN models. This combined model has increased the quality of the satellite images in a single manner as well as for a long term, and thus it has proved it itself to be very efficient in real-time remote sensing and analysis. The hybrid model can be used in many different ways, including land-use classification, vegetation monitoring, urban planning and climate analysis, all of which require cloud-free, high-quality satellite images. The preservation of both spatial and temporal features ensures that the cloud removal process does not interfere with the spectral integrity of the satellite data.

Literature Review

There are several ways to deal with the problem of cloud occlusions in satellite images. The usual methods, like compositing multi-temporal images are mostly cloud-free images, but they require using large volumes of data. Deep learning methods have indicated that they have a promising future, for instance, there are methods such as MCGAN and Cloud-GAN which use generative adversarial network while cloud removal is being processed. But, these techniques have their own limits like identifying dense clouds or relying heavily on artificial data. A number of researchers have concentrated on specific types of cloud occlusions, such as the ones in the upper atmosphere, but their usage is very limited. Moreover, nearly none of the existing methods take full advantage of the temporal information provided over time in the satellite imagery sequence. Generative models, specifically GANs, have gained significant attention for tasks like the conversion of images. Research like Pix2Pix and Cycle GAN, used in super-resolution and style transfer, have were the most demonstrating works. In the case of cloud removal approaches such as MCGAN and Cycle GAN using variations, there have been experiments with but they have the difficulty in case of the dense of clouds. Other significant contributions include cloud removal based on unsupervised remote sensing by contrastive learning but they used GAN UD which generates images that can remove the clouds but cannot accurately recover the image details in the area entirely covered by the clouds, resulting in pixel distortion in the area blocks with thick clouds. STGAN baseline models are also used for the process of cloud removal and cloud shadow removal, but they lack in utilizing the spatio temporal data effectively and produce less quality images. Future works primarily revolve around integrating time series data for generalizability across various environment for more complex images and further making strong techniques to be efficient for precise Earth observation.

Our proposed model will resolve limitations like lack of temporal coherence, high fidelity cloud removal , generalization. The work presents a significant improvement in these aspects, additionally the hybrid model is implemented in such a way that it minimizes overfitting due to inclusion of both temporal data and cycle consistency without degrading the image quality. Let us take a look at our methodology proposed in the implementation of the hybrid model.

Methodology

Problem Definition and Dataset Preparation:

Our study wants to improve the quality of imagery clouds by time stopping the model of temporal discontinuity images and proving its effectiveness through spatial and temporal information. Our suggested STGAN- Cycle GAN hybrid model is made to solve the problem of clouds remaining, and at the same time keep the core part of cloud-free areas intact. We get our dataset from Sentinel-2 or similar satellite sources which provide multi-temporal data captured under both conditions-cloudy and cloud-free. We preprocess these images by making the pixel values normalized, re-sizing, and converting them into the formats that can be fed to the model. The motivation for selecting the hybrid STGAN + Cycle GAN model stems from the limitations observed in existing cloud removal approaches, particularly in addressing both spatial accuracy and temporal consistency simultaneously. By combining the strengths of both models, the hybrid STGAN + Cycle GAN not only provides high accuracy and image quality (PSNR, SSIM) but also ensures that the temporal dynamics of cloud cover are consistently addressed across multiple frames. This makes the hybrid model particularly well-suited for real-world satellite applications, outperforming existing techniques.

Proposed Model Architecture:

The main two parts of our hybrid model include the Spatiotemporal Generative Adversarial Network (STGAN) and the Cycle-Consistent Generative Adversarial Network (Cycle GAN). The STGAN module is built to solve the issue of prediction of cloudy and cloud-free images over time through spatiotemporal data. The generator is designed using convolutional and temporal modules to extract spatiotemporal information while a discriminator is used to distinguish the different classes of the image. The Cycle GAN module makes use of STGAN's output info, transforming the cloudy images to cloud-free versions with the help of the same style and spectral data. It incorporates a generator for image translation and a discriminator that ensures the generated images are identical to real cloud-free frames.

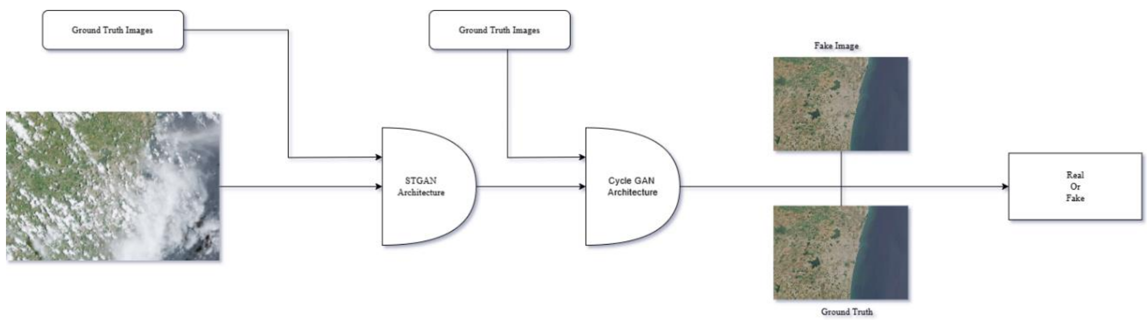

Figure 1: Overview of proposed model architecture

The model's approach consists of two networks which are STGAN and CycleGAN, and multi-scales are used to process airborne images at varying resolutions. The in-built spatial attention of the generators aids in the separation of cloudy and cloud-free sites thereby leading to better elimination of unwanted regions while also preserving the overall structure. Moreover, we include temporal consistency loss in STGAN and perceptual loss in CycleGAN to maintain temporal coherence as well as the details of the image static.

Training Pipeline :

Our training pipeline consists of three phases: STGAN pretraining, CycleGAN integration, and joint training of the hybrid model. In the initial phase, the Spatiotemporal GAN **(**STGAN) is pretrained to learn the temporal dynamics of clouds in multi-temporal satellite images. In the second phase, CycleGAN is introduced to further refine the cloud-free images produced by STGAN, improving the image quality and preventing distortions. In the final phase, the STGAN and CycleGAN models are jointly fine-tuned to form a cohesive hybrid model. This phase is crucial for integrating the temporal modeling power of STGAN with the spatial refinement capabilities of CycleGAN.

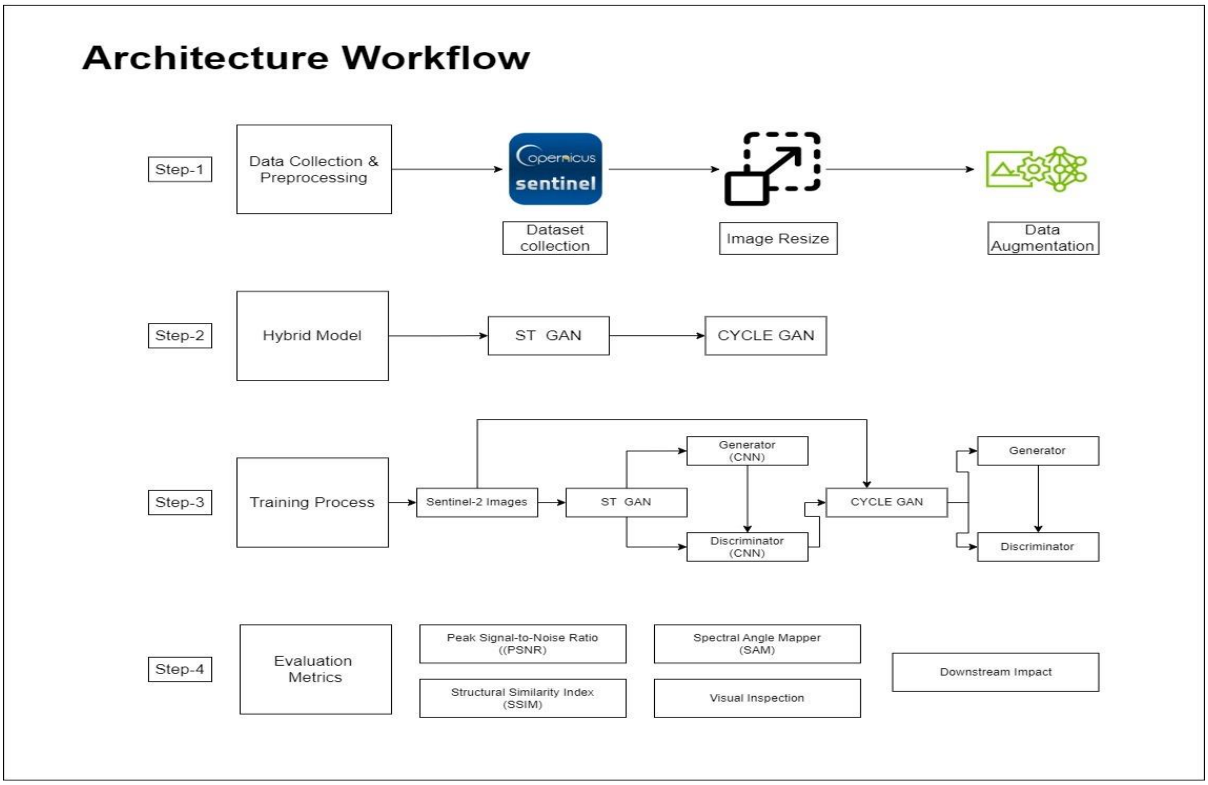

Figure 2: Model Architecture workflow

Phase 1: The first step of the process of cloud removal pertains to the phase of pre-training a Spatiotemporal GAN (STGAN) to understand and model the temporal change of clouds in the multi-temporal satellite imagery. This artificial intelligence model is a complex one, and it is developed on a set of data which consists of satellite images in sequence with the presence of clouds on the first side and, their corresponding trouble-free images obtained from the ground truth on the other side. STGAN proclaims the adversarial learning process, in which the generator delivers the cloud-free images and the discriminator decides their authenticity, eventually, the generated imagery that is well enough to fake the discriminator. Additionally, to secure the correct detail and reconstructed image, the reconstruction loss is included, which forces the created cloud-free images to be like the ground truth images in both the material and the required structural elements. One of the major advantages of STGAN technology lies in the fact that it manages to keep time-related cohesion between the various cloud-free images that are placed within a sequence, without any loss of the integrity of the data

Phase 2: In the second phase of cloud removal, CycleGAN is used to enhance the original cloud-free images initially provided by STGAN. At this point, the STGAN outputs and the initial cloudy images are transformed using a carefully devised set of loss functions. The cycle consistency loss guarantees that the process is reversible, that is, when you re-transform the cloud-free image back to the cloudy state, it looks very similar to the original cloudy input. The back and forth nature of the consistency allows the image from the content to be largely retained no matter how deep the refinement gets. Following the successful operation of STGAN, CycleGAN also relies on an adversarial loss to enhance\nimage reality, thereby encouraging the model to output images that are indistinguishable from actual sky pictures. A distinguishing feature of this stage is the addition of the identity loss, which makes sure the already cloud-free areas in the image are kept at a high quality. This means that the resolution process only affects the clouded areas and their improvement not the parts of the image that are clear. The main purpose of this CycleGAN phase is to create a high-quality translation from cloudy to cloud-free images, therefore, it will refine the STGAN by correcting any distortions and improving the image quality.

Phase 3: In the final stage of cloud removal, the STGAN and CycleGAN models are both fine-tuned to perfection as a consequence of that the unified Hybrid model of combination is produced which assimilates the temporal modeling abilities of STGAEnumeration with the spatial sharpness of CycleGAN. This coherent amalgamation is gained by means of a sophisticated combined loss function that juggles three crucial components: adversarial losses from both STGAN and CycleGAN to guarantee the generation of realistic cloud-free images, temporal coherence loss from STGAN to keep constant the cloud removal process across the sequential frames, and cycle consistency loss from CycleGAN to keep the image-like continuity unchanged during the transformation process. The fine-tuning process is conducted end-to-end, this allows for simultaneous optimization of both the temporal properties and the spatial details. The learning rates as well as the weights of distinct loss terms are carefully balanced so that neither spatial nor temporal aspects are the only ones optimized. This detailed integration along with a balanced optimization approach leads to a hybrid model that is excellent in terms of spatial and temporal consistency and at the same time excels in delivering a much better visual representation. The goal with the final phase is to test a robust, integrated system that removes clouds from satellite imagery with exceptional accuracy across multiple frames while respecting the intricate details and overall quality of the processed images.

Expected Improvements and Challenges:

There are significant improvements in temporal coherence, high-fidelity cloud removal, and overall image quality due to our comprehensive loss function. However, we acknowledge implementation challenges such as long training times, large dataset requirements sometimes complex datasets that may require intensive preprocessing and feature engineering. We address these challenges through distributed training strategies and careful dataset curation.

Evaluation of trained model:

PSNR measures the ratio between the maximum possible signal power and the noise introduced by the model (errors in cloud removal). Higher PSNR values indicate better image quality, with fewer artifacts introduced during cloud removal.

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

where Max^2 is the maximum possible pixel value of the image, and MSE is the mean squared error between the generated (cloud-free) and the ground truth (actual cloud-free) images.

The hybrid model consistently generates cloud-free images with fewer artifacts due to the combination of spatiotemporal features and cycle consistency, leading to a more accurate reconstruction compared to single-image methods like U-Net or Cycle GAN.

SSIM evaluates the perceived quality of images by comparing local patterns of pixel intensities, capturing structural information, luminance, and contrast. It ranges from 0 to 1, where 1 means the images are identical.

SSIM(x,y)=(μx2+μy2+C1)(σx2+σy2+C2)/(2μxμy+C1)(2σxy+C2)

where μ is the mean, σ is the variance, and C1 and C2 are small constants to stabilize the division. The hybrid model's multi-scale attention mechanisms ensure better preservation of fine spatial details such as edges and textures, resulting in a higher SSIM compared to other models like Cycle GAN or Pix2Pix, which may introduce distortions in finer details.

The temporal Consistency measures how well a model maintains visual coherence across consecutive frames in a time-series dataset. A low temporal consistency indicates inconsistency across frames, which is undesirable in cloud removal for satellite imagery. Temporal consistency can be computed using frame-to-frame SSIM or by calculating differences in pixel values across consecutive images in a sequence. STGAN in the hybrid model explicitly models temporal dependencies, ensuring that cloud removal is consistent across time steps.

MAE measures the average absolute differences between predicted and actual pixel values. It is a common metric for evaluating how accurately the model predicts each pixel.

$$MAE = 1/n(\sum|y - y\hat{}|)$$

Where y is the true value and y^ is the predicted value. By integrating both temporal and cycle consistency, the hybrid model generates images with fewer pixel-level errors.

The F1-Score is used to evaluate the accuracy of cloud detection, considering both precision and recall. This is important in cloud removal tasks because the model needs to accurately detect and remove only cloud regions while preserving cloud-free regions.

$$F1 - Score = 2 \cdot (\frac{Precision.Recall}{Precision + Recall})$$

The attention mechanism and temporal modeling in the hybrid model enable more precise detection of cloud-covered areas, leading to fewer false positives (incorrectly removing cloud-free regions) and fewer false negatives (failing to detect clouds).

Results and Discussion:

The combination of spatiotemporal and cycle-consistent techniques allows the hybrid model to outperform existing models in cloud removal, providing both high-quality outputs and robust performance across a wide range of cloud conditions.

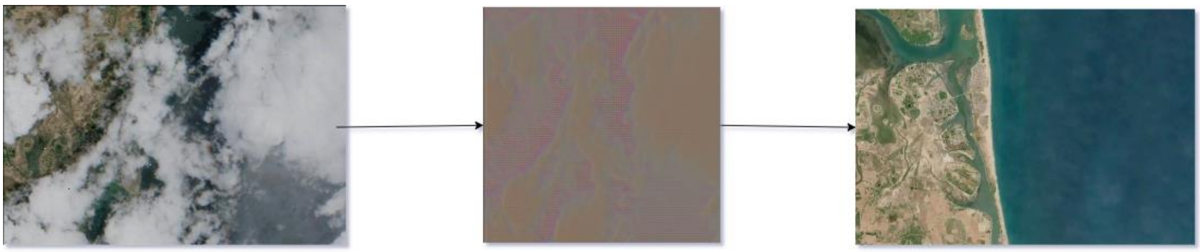

Figure 3: Results of hybrid model

Model	**PSNR (dB)**	**SSIM**	**MAE**	**MSE**	**F1 score**
Proposed STGAN + CycleGAN model	32.5	0.91	0.0 32	0.004	0.94
STGAN	31.2	0.88	0.037	0.005	0.90
Cycle GAN	30.2	0.85	0.048	0.006	0.84
SIFGAN	30.9	0.87	0.041	0.006	0.89
Pix2Pix	29.5	0.83	0.055	0.008	0.82
U-Net	28.0	0.79	0.061	0.011	0.78
WGAN	30.9	0.87	0.043	0.006	0.88

Table 1: Results compared to other models

The new combination of STGAN and Cycle GAN altogether have an impeccable influence on the performance of the hybrid model that brings about the desired change in the cloud removal from satellite imagery. The current approach has a great capacity to make the most of the opportunities of both models that combine and therefore has resulted in the development of a practical device with excellent performance in the different aspects relating to the image quality and the required accuracy. The record of high SSIM values also confirms that the model further utilizes its competence to maintain the structure of the input images. SSIM is particularly crucial here, as it is a visual quality metric that measures the structural information being perceived. The fact that the model achieves a high mark in this parameter implies, along with effective cloud removal, that it also guarantees the highest visual integrity of the clouds-free images when compared to the original scenes.

In terms of pixel-level accuracy, the hybrid model excels as evidenced by its low Mean Absolute Error (MAE) and Mean Squared Error (MSE) values. These metrics measure the average difference between the predicted cloud-free images and the actual ground truth images. The low values in both MAE and MSE demonstrate that the model achieves high precision in pixel reconstruction. Perhaps one of the most impressive aspects of the hybrid model's performance is its high F1-score in cloud detection and removal. The F1-score is a balanced measure of a model's precision and recall, providing a single score that indicates how well the model performs in identifying and removing clouds. It demonstrates that the model can effectively handle the complex task of cloud removal while maintaining high image quality and accuracy. This makes it particularly valuable for a wide range of applications involving satellite imagery, including environmental monitoring, urban planning, agricultural analysis, and disaster response, where accurate and cloud-free images are essential for informed decision-making.

Conclusion and Recommendation :

The new hybrid model, compounding Spatiotemporal Generative Adversarial Network (STGAN) and Cycle Consistent Generative Adversarial Network (Cycle GAN), is addressing a major issue in cloud removal technology. It utilizes two different models, one that specializes in space and another that complements it, and therefore delivers excellent outcomes in the terms of perspective quality, no clouds being visible, and other details keeping in check. The combination of spatiotemporal modeling with cycle-consistency loss guarantees that the model not only effectively removes clouds but also keeps all significant information and remains of the highest fidelity level. The clearer, more correct images are the result of that, and these are very important in the application areas such as environmental monitoring and satellite-based assessments.

In the future, multiple ways could be exploited to make the model even better. The best way to go would be to make it more generalized for different datasets and to make it more optimized so it can work in real-time. Furthermore, enhancement of the model by feedback from practical applications and the interconnection of the model with other remote sensing data types might expand the scope of its use. The pursuit of novel methods for cycle-consistency, as well as the enhancement of the model's explainability and visualization, represents a valuable step in refining its performance and extending its utility. The solutions are dealing with very practical issues and adding new ideas through the inclusion of feedback from the satellite mapping community, which in turn improves the process of cloud removal.

References :

1.Li, Y., Chen, W., Zhang, Y., Tao, C., Xiao, R., Tan, Y.: Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. Remote Sens. Environ. 250, 112045 (2020)

2.Jin, M., Wang, P., Li, Y.: Hya-gan: remote sensing image cloud removal based on hybrid attention generation adversarial network. Int. J. Remote Sens. 45(6), 1755–1773 (2024)

3.Pan, H.: Cloud removal for remote sensing imagery via spatial attention generative adversarial network. arXiv preprint arXiv: 2009. 13015 (2020)

4.Shen, H., Li, H., Qian, Y., Zhang, L., Yuan, Q.: An effective thin cloud removal procedure for visible remote sensing images.

ISPRS J. Photogramm. Remote Sens. 96, 224–235 (2014)

5.Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-toimage translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)

6.Sarukkai, V., Jain, A., Uzkent, B., Ermon, S.: Cloud removal from satellite images using spatiotemporal generator networks. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1796–1805 (2020)

7.Pitonak, R., Mucha, J., Dobis, L., Javorka, M., Marusin, M. (2022). Cloudsatnet-1: FPGA-based hardwareaccelerated quantized CNN for satellite on-board cloud coverage classification. Remote Sensing, 14(13): 3180.

8.D. Lin, G. Xu, X. Wang, Y. Wang, X. Sun, and K. Fu, "A remote sensing image dataset for cloud removal," 2019, arXiv:1901.00600.

9.Z. K. Liu and B. R. Hunt, "A new approach to removing cloud cover from satellite imagery," Comput. Vis., Graph., Image Process., vol. 25, no. 2, pp. 252–256, Feb. 1984.

10.A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SARoptical data fusion," ISPRS J. Photogramm. Remote Sens., vol. 166, pp. 333–346, Aug. 2020.

11.P. Singh and N. Komodakis, "Cloud-GAN: Cloud removal for Sentinel-2 imagery using a cyclic consistent generative adversarial networks," in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Jul. 2018, pp. 1772–1775.

12.J. Gao, Q. Yuan, J. Li, H. Zhang, and X. Su, “Cloud removal with fusion of high resolution optical and SAR images using generative adversarial networks,” Remote Sensing, vol. 12, no. 1, 2020.

13.C. Grohnfeldt, M. Schmitt, and X. Zhu, “A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images,” International Geoscience and Remote Sensing Symposium (IGARSS), vol. 2018-July, pp. 1726–1729, 2018.

14.J. Ma, H. Zhang, Z. Shao, P. Liang, and H. Xu, “GANMcC: A Generative Adversarial Network with Multiclassification Constraints for Infrared and Visible Image Fusion,” IEEE Transactions on Instrumentation and Measurement, vol. 70, 2021.

15.Z. Liu, W. Zhang, and P. Zhao, “A cross-modal adaptive gated fusion generative adversarial network for rgb-d salient object detection,” Neurocomputing, vol. 387, pp. 210 – 220, 2020.

16.X. Luo, X. Tong, and Z. Hu, “Improving Satellite Image Fusion via Generative Adversarial Training,” IEEE Transactions on Geoscience and Remote Sensing, pp. 1–14, 2020.

17.K. G. Lore, K. K. Reddy, M. Giering, and E. A. Bernal, “Generative adversarial networks for spectral superresolution and bidirectional rgb-to-multispectral mapping,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, vol. 2019-June, pp. 926–933, 2019.

18.Z. Tan, M. Gao, X. Li, and L. Jiang, “A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network,” IEEE Transactions on Geoscience and Remote Sensing, 2021.

19.Z. Shao, Z. Lu, M. Ran, L. Fang, J. Zhou, and Y. Zhang, “Residual encoderdecoder conditional generative adversarial network for pansharpening,” IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 9, pp. 1573–1577, 2020.

20.A. Yadav and D. K. Vishwakarma, “Recent developments in generative adversarial networks: A review (workshop paper),” in 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM). IEEE, 2020, pp. 404–413.

21.Hasan, C., Horne, R., Mauw, S., Mizera, A.: Cloud removal from satellite imagery using multispectral edge-filtered conditional generative adversarial networks. Int. J. Remote Sens. 43(5), 1881–1893 (2022)

22.Pan, H.: Cloud removal for remote sensing imagery via spatial attention generative adversarial network. arXiv preprint arXiv:2009.13015 (2020)

23.17. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)

24.Zi, Y., Xie, F., Song, X., Jiang, Z., Zhang, H.: Thin cloud removal for remote sensing images using a physical-model-based cycleGAN with unpaired data. IEEE Geosci. Remote Sens. Lett. 19, 1–5 (2021)

25.Li, J., Wu, Z., Hu, Z., et al.: Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion. ISPRS J. Photogramm. Remote Sens. 166, 373–389 (2020)

26.A. Zupanc, “Improving cloud detection with machine learning,” Jul 2020. [Online].

27.M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, “Sen12ms – a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion,” ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. IV-2/W7, pp. 153–160, 2019.[Online].

28.Hagolle, O.; Huc, M.; Pascual, D.V.; Dedieu, G. A Multi-Temporal Method for Cloud Detection, Applied to FORMOSAT-2, VEN S, LANDSAT and SENTINEL-2 Images. Remote Sens. Environ. 2010, 114, 1747–1755. [CrossRef]

29.Zou, Z.; Li,W.; Shi, T.; Shi, Z.; Ye, J. Generative Adversarial Training forWeakly Supervised Cloud Matting. Proceedings of the IEEE Int. Conf. Comput. Vis. 2019, 2019, 201–210. [CrossRef]

30.Shi, M.; Xie, F.; Zi, Y.; Yin, J. Cloud Detection of Remote Sensing Images by Deep Learning. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 701–704.