Clear-GAN: Cloud Removal in Satellite Imagery via Cyclic Multispectral Generative Adversarial Networks

Adam Hazimeh
Dept. of Electrical & Computer Eng.
American University of Beirut
Beirut, Lebanon
amh116@mail.aub.edu

Rana Hazimeh
Dept. of Electrical & Computer Eng.
American University of Beirut
Beirut, Lebanon
rah115@mail.aub.edu

Yasser Nasser
Dept. of Electrical & Computer Eng.
American University of Beirut
Beirut, Lebanon
yhn08@mail.aub.edu

***Abstract*—Remote sensing technologies that rely on satellite imagery are vulnerable to the impediment of cloud covers, leading to partially unusable data. Several cloud removal techniques have been explored in the literature, ranging from statistical methods to advanced deep learning models. In this work, we propose Clear-GAN, a cyclic, multispectral Generative Adversarial Network. Clear-GAN's cyclic architecture allows it to avoid the constraint of collecting paired image datasets. Our model's reliance on the Near Infra-red wavelength, which has partial cloud penetration properties, in addition to wavelengths from the visible spectrum, helps the network capture low-level details in cloud-covered landscapes. To allow for multispectral support, the input channels of the Generative Adversarial Networks in our model were be modified. The model was trained on images from the Sentinel-2 dataset and its peak signal-to-noise ratio was compared to a state-of-the-art cyclic network architecture implementing the Least Squares Generative Adversarial Network. Our model's performance was also evaluated on high-opacity cloud covers.**

***Index Terms*—remote sensing, cloud removal, deep learning, generative adversarial networks**

I. Introduction

Remote sensing technology is crucial for many applications, such as development, meteorology, and agriculture. Satellites acquire this spatial information and aid in creating road maps, observing climate change, and identifying crops [1]. The collection of image data is performed by capturing multispectral radiations emitted from objects, and is often hindered by cloud covers intruding the troposphere. This disruption can be especially troublesome in monitored areas characterized by murky skies, as it stagnates progressive findings and projects. The utilization of synthetic aperture radar (SAR) technology mitigates these drawbacks and allows for clear images in cloudy areas by employing signals that acquire data through attributes such as structure and moisture instead of visible light. SAR is however considerably limiting in terms of its spatial resolution, and is known for poor visibility as a result of its increased signal wavelength [2]. Researchers have been making efforts towards designing algorithmic and statistical image processing methods, some of which can be applied for the cloud removal task. One such method is dark channel prior (DCP), an a priori statistical algorithm that has been shown to be effective at thin cloud removals, but suffers from high noise sensitivity [3].

Eliminating the haze effect caused by clouds in satellite imagery encompasses several challenges, some of which have been handled by the literature, and others which remain open. The main challenge lies in designing a model that is capable of simulating landscapes found beneath cloud covers, while still preserving as much of the real landscape contents as possible. This issue has been addressed using different Generative Adversarial Network (GAN) variations, including models such as CycleGANs [4], Spatial Attention GANs (SpA-GANs) [5], and conditional GANs (cGANs) [6]. These implementations have shown promising results in successfully removing thin cloud covers, yet they fail at yielding realistic images in the case of high-opacity cloud removal as shown in figure (1) (this remains an open challenge).

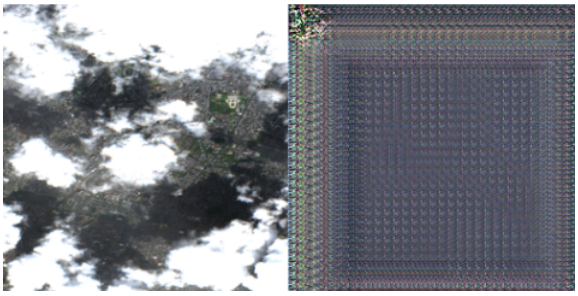


Fig. 1. Complete failure on thick cloud covers in Cloud-GAN [4]

A variation of the classical GAN model, called Wasserstein GAN + Gradient Penalty (WGAN-GP) [7], has proven to attain the highest quality generated images among other models in certain image inpainting tasks [8] and is yet to be used for cloud removal tasks. However, experimentation with WGAN-GP in this work led to sub-optimal results and so a Least Square GAN (LSGAN) approach was implemented instead.

In addition, CycleGANs [4] have demonstrated a remarkable ability at avoiding the constraint of collecting paired image datasets (cloudy and cloud-free images of the same area). Paired image datasets, especially in remote sensing applications, are notably difficult to collect due to highly variable climatic conditions. Since Near Infra-red (NIR) light has partial cloud penetration capabilities, as displayed in McGAN [6], expanding LSGAN input channels to accommodate NIR and incorporating the resultant model into a CycleGAN architecture could lead to more realistic cloud removal. In this work, we propose Clear-GAN: a cyclic, Multispectral GAN for cloud removal in satellite imagery. We also explore the proposed model's performance on thick (high-opacity) cloud covers.

II. Related Work

Developments in the field of Deep Learning (DL) have proven that some DL techniques, such as Convolutional Neural Networks (CNNs), perform exceptionally in computer vision tasks. In particular, Generative Adversarial Networks (GANs) [9], have demonstrated a great capability in handling image generation, image inpainting [8], and image up-scaling tasks. The GAN architecture consists of two networks: a Generator (G) and a Discriminator (D). In computer vision applications, G generates fake images in an attempt to fool D, which is a binary classifier that categorizes input images as "real" (ground-truth) or "fake" (generated by G). The two networks engage in a loss minimax game until Nash equilibrium is reached, at which point G is able to fool D in 50% of attempts.

Recently, several variations of GANs have been used to tackle de-hazing and cloud removal tasks in the field of remote sensing. One paper [5] proposes a model named SpA-GAN to handle the removal of thin cloud covers in satellite images. This model introduces a spatial awareness mechanism into its architecture by using a type of spatial attention network (SPANet) as a generator. While SpA-GAN outperforms several other GAN architectures by achieving a higher peak signal-to-noise ratio (PSNR), it requires a paired dataset of cloudy and cloud-free images of the same geographical area during the training process, which can be difficult to obtain.

Another paper [6] proposes a different GAN architecture called Multispectral conditional GAN (McGAN), which is a variation of cGAN with a modification to support a NIR wavelength input channel. Owing to its partial cloud penetration properties, NIR reveals more low-level features to the network and results in higher generated image quality [6]. However, even though McGAN obtains a paired image dataset by synthesizing clouds in cloud-free images, the resulting cloudy images do not accurately represent real cloud covers, which impacts the model's generalization performance [4].

Cloud-GAN [4], another GAN solution for thin cloud removal in RGB satellite images, relies on the CycleGAN architecture. Instead of consisting of one generator and one discriminator as proposed in the original GAN paper [9], a CycleGAN includes two of each. This change frees the model of the paired dataset constraint, allowing a dataset containing cloudy and cloud-free images from different geographical locations to be used in training. In its CycleGAN implementation, Cloud-GAN uses Least Squares GANs (LSGANs), which are capable of remedying the issue of vanishing gradients. Accordingly, Cloud-GAN avoids obstacles such as paired dataset constraints and obtaining non-RGB band images (e.g., NIR/SAR), though it runs the risk of over-smoothing certain inputs or completely failing on others (e.g., images with thick cloud covers).

TABLE I
Qualitative comparison of the literature contributions

Model	Contributions	
SpA-GAN	Spatial Awareness	High PSNR
McGAN	Synthesized clouds	Uses NIR for cloud penetration
Cloud-GAN	No paired dataset	No vanishing gradients
Clear-GAN	No paired dataset	Uses NIR for cloud penetration

III. Proposed Framework

A. Baseline Model

As this article proposes the use of a CycleGAN architecture relying on two LSGAN networks to tackle the task of cloud removal, setting Cloud-GAN (a model which relies on the CycleGAN architecture) as a baseline for comparison would be the optimal course of action. This would help establish the significance of implementing a CycleGAN architecture and adding an NIR channel to the model's inputs.

B. Model

Similar to the Cloud-GAN model, the proposed network (Clear-GAN) models a mapping that translates cloudy images (inputs) into cloud-free images (outputs). Clear-GAN's architecture are decomposed into two parts:

- **CycleGAN:** CycleGAN [11] was chosen because it overcomes the necessity of training on a paired (cloudy/cloud-free) image dataset, which can be especially difficult to obtain for this application. The difficulty of obtaining paired image datasets in this context arises from the variability of climatic conditions and landscape features. CycleGAN's architecture consists of two discriminators (D_A and D_B) and two generators ($G_{A\to B}$ and $G_{B\to A}$). In this context, domain A represents cloudy images, while domain B represents cloudless images.

  The discriminators, in the context of a CycleGAN architecture, are deep convolutional neural networks that perform binary image classification, where D_A classifies images from domain A and D_B classifies images from domain B (real/fake). Each discriminator consists of repeated blocks of convolutional layers,

instance normalization layers, and LeakyReLU activation functions.

On the other hand, the generators follow an encoder-decoder architecture which down-samples (encodes) the input images using convolutional layers, passes the encoding into a bottleneck layer consisting of a series of residual convolutional blocks [12], and finally up-samples (decodes) the output of the bottleneck using convolutional layers back to the original input dimensions. The first generator, $G_{A\to B}$, maps an input from domain A (cloudy) to domain B (cloudless), while the second generator, $G_{B\to A}$, does the opposite. Including $G_{B\to A}$ in the model allows for creating a cycle consistency (i.e., mapping the input back to the source domain by feeding it through both generators), which creates a regularizing effect for the model. In addition, identity mappings (i.e., mapping an input back to its domain by feeding it through one generator) are also optimized, as they have been shown to lead to better color-profile matching in image generation tasks.

- **LSGAN:** For certain image inpainting tasks [8], WGAN-GP loss, which relies on the Wasserstein metric [7], was observed to reach faster and more stable convergence than standard adversarial loss functions. However, implementing Wasserstein loss in the proposed model led to highly unstable gradient descent and was deemed unsatisfactory. Consequently, the adversarial loss function implemented in this model was the least squares loss (as seen in Cloud-GAN). As for the cycle consistency and identity mapping losses, the mean absolute error (MAE) was considered. In addition, to allow for training the network on images containing the NIR wavelength, it is necessary to expand the number of input channels in the model architecture by one. This therefore yields a multispectral LSGAN model, which we name Clear-GAN.

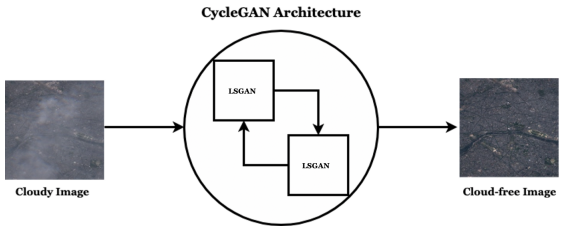


Fig. 2. Model Architecture

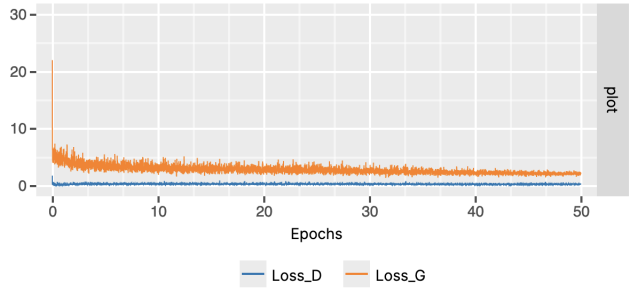
C. Dataset

The dataset used in this work consists of openly-accessible satellite images collected by Sentinel-2, a land monitoring satellite mission launched as part of the Copernicus Programme [10]. The dataset contains unpaired cloudy and cloud-free images of different geographical areas with similar landscapes. In addition to RGB light bands, the NIR wavelength was also included in the dataset. This is because, as previously mentioned, NIR is more powerful at penetrating cloud covers than RGB bands and may thus help the model capture more landscape details.

The images were collected from Sentinel-2 using Sentinel Hub's Process API. Each spectral band (R, G, B, NIR) was stored as a separate, grayscale image to allow for easier data processing. Cloudless and cloudy images were selected to have 0% and $10-100\%$ cloud cover respectively. Six large regions were manually selected, and 512 x 512 image tiles were retrieved from them, leading to a total of 1735 unpaired images of each wavelength and category (cloudy/cloudless) for the training dataset. As for the evaluation dataset, the same approach was implemented. However, the evaluation dataset contains paired images (required for PSNR evaluation), which led to a total of 608 evaluation images. Images in this dataset were selected to be paired so that the PSNR metric could be calculated to evaluate the proposed model's performance. All images were scaled to 256 x 256 for computational efficiency.

D. Training

For the sake of comparison, all hyperparameter choices for Clear-GAN were chosen to be identical to those of Cloud-GAN [4]. The key difference in training Clear-GAN is that it was trained on a dataset of satellite images with the image dimensions (256x256x4), with the 4th channel added for the NIR band. All loss functions were optimized using ADAM [13] with a batch size of 1 and a learning rate of 0.0002. Stochastic Gradient Descent (SGD) optimization was also tested, but led to immediate loss divergence. Even though it was proposed that the model be trained for 200 epochs, we were only able to train it for 70 epochs due to computational constraints. Hyperparameter optimization was also one of the goals of this work, but time and computational constraints formed unanticipated obstacles (10 epochs require more than 1 hour to train on a GPU).



E. Experiments and Results

After training, Clear-GAN's performance was first evaluated qualitatively on the evaluation dataset. Cloudy images were passed through the model (namely, $G_{A\to B}$), and results were compared visually to the ground-truth cloudless images. Selected images are displayed (in RGB) below:

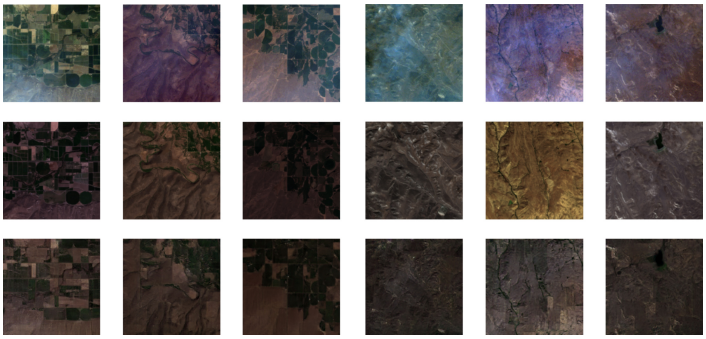


Fig. 4. Top Row: Cloudy Images; Middle Row: Ground Truth Cloudless Images; Bottom Row: Cloudless Images Generated by Clear-GAN

For many images with thin cloud covers, Clear-GAN was successfully able to generate cloudless images that are visually similar to ground-truth images as seen in the examples above. However, due to a lack of diversity in the landscapes of the images in the training dataset relative to Cloud-GAN, and our inability to train the model for more than 70 epochs, our model led to two failure cases: The first case was that Clear-GAN generated some images which displayed overfitting patterns to the training images (e.g., it learned to generate random agricultural crop tiles when the input image was relatively different from the training dataset distribution) (fig. 5). The second failure case was that the model generated random (but relatively realistic) images when thick cloud covers where present in the input (fig. 5). However, unlike Cloud-GAN, our model did not seem to generate random noise when encountering thick cloud covers (as seen in fig. 1), which may bring us a step closer towards solving the open challenge of thick cloud removal.

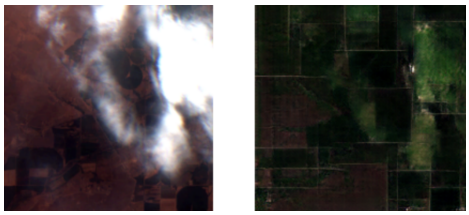


Fig. 5. Failure: Cloudy Image (Left), Generated Cloudless Image (Right)

Our model was also quantitatively evaluated using the Peak Signal-to-Noise Ratio (PSNR), which is commonly used to quantify image reconstruction quality when paired datasets (e.g., our evaluation dataset) are present. Clear-GAN achieved 18.1 dB average PSNR for cloud removal (compared to 24 dB for Cloud-GAN), and 16.47 dB average PSNR for cloud reconstruction (compared to 27.5 dB for Cloud-GAN). Even though our PSNR values were lower than those of Cloud-GAN (which may be largely due to incomplete training), our PSNR values were measured on a paired dataset. On the other hand, Cloud-GAN's PSNR evaluation entailed generating synthetic clouds and applying cloud removal, and synthetic clouds are likely not to realistically model real clouds.

IV. Conclusion

With this, we have developed Clear-GAN by fitting the model on unpaired Sentinel-2 images as previously done, and augmenting the inputs with near infra red wavelengths to obtain 256x256x4 training images. This in turn allowed for better cloud penetrative capabilities. Clear-GAN was then evaluated through PSNR tests on manually collected paired image datasets for strengthened validation, and was able to produce photo-realistic images through light to medium cloud covers. Terrain was generated underneath heavy cloud sheets, but did not bare accurate resemblance. However, given unrestricted time and resource, Clear-GAN can be further improved through elaborate data collection on a larger variety of terrain, and higher number of training hours.

References

[1] Saibi, Hakim & Bersi, Mohand & Mia, Md & Saadi, Nureddin & Al Bloushi, Khalid & Avakian, Robert. (2018). Applications of Remote Sensing in Geoscience. 10.5772/intechopen.75995.

[2] Flores, Africa & Herndon, K. & Thapa, Rajesh & Cherrington, Emil. (2019). The SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation. 10.25966/nr2c-s697.

[3] Li, Jiayuan & Hu, Qingwu & Ai, Mingyao. (2018). Haze and Thin Cloud Removal via Sphere Model Improved Dark Channel Prior. IEEE Geoscience and Remote Sensing Letters. PP. 1-5. 10.1109/LGRS.2018.2874084.

[4] Singh, Praveer & Komodakis, Nikos. (2018). Cloud-Gan: Cloud Removal for Sentinel-2 Imagery Using a Cyclic Consistent Generative Adversarial Networks. 1772-1775. 10.1109/IGARSS.2018.8519033.

[5] Pan, Heng. (2020). Cloud Removal for Remote Sensing Imagery via Spatial Attention Generative Adversarial Network.

[6] Enomoto, Kenji & Sakurada, Ken & Wang, Weimin & Fukui, Hiroshi & Matsuoka, Masashi & Nakamura, Ryosuke & Kawaguchi, Nobuo. (2017). Filmy Cloud Removal on Satellite Imagery with Multispectral Conditional Generative Adversarial Nets. 1533-1541. 10.1109/CVPRW.2017.197.

[7] Gulrajani, Ishaan & Ahmed, Faruk & Arjovsky, Martin & Dumoulin, Vincent & Courville, Aaron. (2017). Improved Training of Wasserstein GANs.

[8] Yu, Jiahui & Lin, Zhe & Yang, Jimei & Shen, Xiaohui & Lu, Xin. (2018). Generative Image Inpainting with Contextual Attention.

[9] Goodfellow, Ian & Pouget-Abadie, Jean & Mirza, Mehdi & Xu, Bing & Warde-Farley, David & Ozair, Sherjil & Courville, Aaron & Bengio, Y.. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems. 3. 10.1145/3422622.

[10] Phiri, Darius & Simwanda, Matamyo & Salekin, Serajis & Nyirenda, Vincent & Murayama, Yuji & Ranagalage, Manjula. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. Remote Sensing. 12. 2291. 10.3390/rs12142291.

[11] Zhu, Jun-Yan & Park, Taesung & Isola, Phillip & Efros, Alexei. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2242-2251. 10.1109/ICCV.2017.244.

[12] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2015). Deep Residual Learning for Image Recognition. 7.

[13] Kingma, Diederik & Ba, Jimmy. (2014). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations.