



HAL
open science

CLOUD AND FOG REMOVAL FROM SATELLITE IMAGES USING GENERATIVE ADVERSARIAL NETWORKS (GANs)

Akash Ramjyothi, Santonu Goswami

► **To cite this version:**

Akash Ramjyothi, Santonu Goswami. CLOUD AND FOG REMOVAL FROM SATELLITE IMAGES USING GENERATIVE ADVERSARIAL NETWORKS (GANs). 2021. hal-03462652

HAL Id: hal-03462652

<https://hal.science/hal-03462652>

Preprint submitted on 2 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CLOUD AND FOG REMOVAL FROM SATELLITE IMAGES USING GENERATIVE ADVERSARIAL NETWORKS (GANs)

Akash Ramjyothi
B.Tech Student,
Computer Science Engineering
SRM University
Chennai, Tamil Nadu, India
akash.ramjyothi@gmail.com

Santonu Goswami
Scientist - SF
Earth and Climate Science Area
National Remote Sensing Centre
Indian Space Research Organization
Hyderabad, Telangana, India
santonu@gmail.com

ABSTRACT

Advanced Earth observation satellites produce a constant stream of data. These datasets have wide applications in ecosystem research such as vegetation health monitoring, water quality, disaster monitoring, and a range of other areas. Among the earth observation satellites, multispectral remote sensing data have been widely used for the last several decades and continue to be at the center of earth observation datasets for the scientific communities. Multispectral remote sensing data suffer from environmental effects due to cloud and fog and hence a major chunk of it becomes unusable. Therefore, being able to use advanced AI/ML techniques to remove these environmental impurities from these data to produce uninterrupted datastream can be quite valuable. Here we demonstrate the use of Generative Adversarial Network (GAN) to remove fog and cloud from Landsat multispectral data to regenerate and fill in the gaps for uninterrupted satellite data.

1 Introduction

Remote Sensing technologies have a wide scope of applications in earth and environmental sciences. Whether urban planning, agriculture, environmental management, or climate change, remote sensing products provide indispensable information in our daily life activities as well as in research work. This limitation makes it difficult to analyze data and derive the required information about [4]. But recently, the picture has changed. Nowadays, the number of operating satellites increases periodically, each of which has a specific mission to achieve with different types of attached sensors. Different remote sensing datasets become more complicated in terms of structure and size. Therefore, the bigger the volume of data received from satellites, the more storage and computing power is required to analyze these datasets [8]. Volume, Variety and Velocity act as a challenge for remote sensing scientists and experts nowadays. In order to process and analyze this large amount of data, it is not possible to use traditional algorithms, storage servers and computing power. Previously, traditional servers were being used to process earth observation data. But these servers were limited to the hardware capabilities and it was very expensive to upgrade the server's hardware [14]. Also, Cloud Computing can reduce the costs, since it is possible to pay as use and on demand, instead of purchasing advanced hardware and software licenses to process these big data.

2 Image processing techniques for dehazing

A number of image dehazing techniques was introduced both for RGB images only and for RGB and NIR images[3].For the former, some prior knowledge or assumptions on the colours of the hazed image is required. For the latter, NIR images are exploited to guide the dehazing of the RGB images.Generative Adversarial Networks is a framework that recently gained notable interest in the area of remote sensing due to its remarkable capabilities in image generation[11]. GANs consist of two kinds of artificial neural networks: a generator and a discriminator. The two networks are trained against each other[2]. Generator is trained to generate images that cannot be distinguished by the discriminator from images belonging to a target domain, while discriminator is trained to better distinguish artificial images generated by the generator. By using an adversarial loss during learning, the GANs can be trained to generate images which

are virtually indistinguishable from the real images of the target domain. A number of variants of GANs has been introduced in the literature: conditional Generative Adversarial Networks (cGANs) [10] and Deep Convolutional Generative Adversarial Networks (DcGANs) [5]. They have been widely used for image restoration tasks [14]. McGANs were introduced for filmy cloud removal from RGB images with additional information provided by accompanying NIR images. This approach was modified to fuse SAR and optical multi-spectral image input data to generate cloud- and haze-free multispectral optical images. Further developments on fusion of SAR and multi-spectral optical data with cGANs for cloud removal were presented in [15], where the input consisted of three co-registered images, i.e. two SAR images at time points t_1 and t_2 and an optical image at t_2 and the cGAN was used to generate a plausible optical image at t_1 . Another GANs-based approach was proposed in [16], where the authors introduced a Cloud-GAN to learn the mapping from cloudy images to cloud-free images. The advantage of this method is twofold. First, it does not require a paired cloudy/cloud-free dataset. Second, it does not need any kind of cloud-penetration sources of imagery such as SAR or NIR. However, this approach is limited by the capabilities of performing in-painting over not so thin cloud regions.

2.1 Image Matting

Image matting refers to a group of methods that aim to extract the foreground from an image, which is important in image and video editing. The matting task usually produces an “alpha matte” that can be used to separate foreground from the background in a given image, which naturally corresponds to the cloud detection and removal process. Traditional image matting methods can be divided into two groups: 1) sampling-based methods and 2) propagation based methods, where the former produces the alpha matte by a predefined metric given a set of the foreground and background sampling regions, while the latter formulates the prediction as the propagation of the foreground and background regions. As the matting is an ill-posed problem, some methods also take in the user interactions (e.g. trimap or scribbles as additional inputs which specify the predefined foreground, background, and unknown regions to produce more accurate predictions. In recent years, deep learning techniques have greatly promoted the image matting.

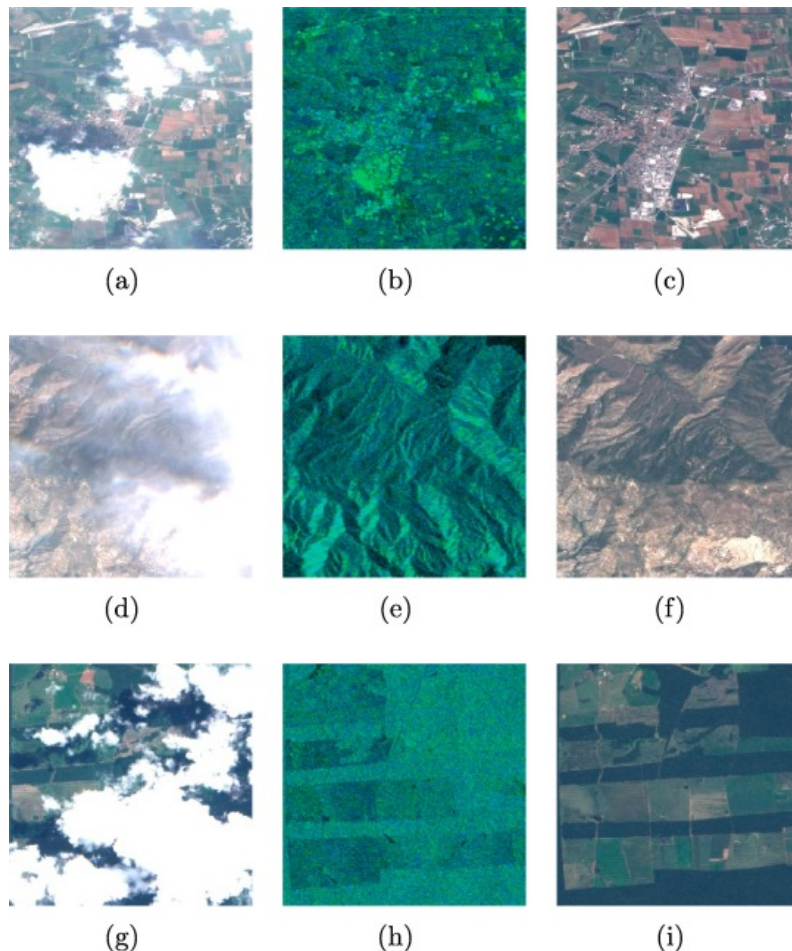


Figure 1: Comparisons of thin cloud removal on synthetic images.

2.2 Generative Adversarial Networks

The Generative Adversarial Network (GAN) has received great attention in recent years, and has achieved impressive results in various tasks such as image generation, image style transfer and image superresolution. A typical GAN consists of two neural networks: a generator network and a discriminator network, where the former learns to map from a latent space to a particular data distribution of interest, while the latter aims to discriminate between the different instances from the true data distribution and those to the generated ones. The key to GAN's success is the idea of an adversarial training framework under which the two networks will contest with each other in a minimax two player game and forces the generated data to be, in principle, indistinguishable from real ones.

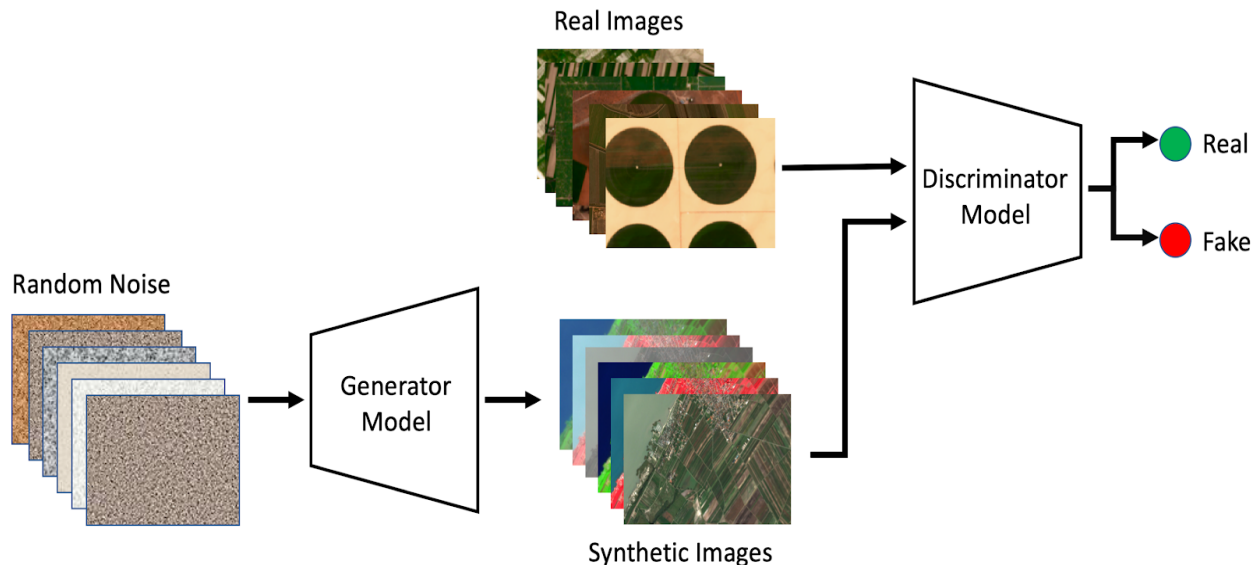


Figure 2: Network Architecture.

2.3 HF method

The HF method proposed by Mitchell et al. [14] is a popular thin cloud-imaging model. In this work, the authors argue that cloud noise is mainly distributed in the low-frequency domain of one image. Thus, the thin cloud or fog removal work can be conducted using a suitable filter to reduce or eliminate the low-frequency information. However, the cut-off frequency is empirically determined, and the clear regions are easily affected. To keep the quality of the clear regions, Shen et al. [15] proposed an adaptive homomorphism filter to eliminate thin clouds in visible remote sensing images. The authors further suggest using the clear-sky regions in the image to determine the cut-off frequency. This method cannot properly solve the confusion between the spatial frequency of clouds and a low-frequency land surface, which affects the spectral characteristics of pixels in a cloudless region.

2.4 Haze optimization transformation (HOT) method

To address the images with uneven clouds, Zhang et al. [16] proposed the HOT method, in which the pixels in the clear scene can construct the clear skyline. When affected by clouds, the distance of every pixel to the clear skyline is used to assess the relative cloud thickness. However, some sensitive land-surface types, such as snow, bare soil, man-made buildings, and water bodies could induce spurious HOT responses. To solve the above problems, He et al. Proposed an atmospheric correction technique, that is, a virtual cloud point (VCP) method, based on advanced haze optimization transformation (AHOT). However, it still depends on the selection of cloudless samples and requires more manual interventions. Jiang et al. Presented a high-fidelity haze removal method based on semi-automatic HOT transform, but the method only has a good effect on vegetation cover scenes. Chen et al. Proposed a remote sensing cloud correction method based on the cloud detection of iterative haze-optimized transformation (IHOT) and the cloud removal of cloud trajectory. The visual effect of remote sensing images is improved by the relationship between surface reflectance and IHOT. However, the estimation of cloud trajectories is affected by shadow detection error, which leads to inaccurate cloud trajectories estimation.

2.5 Data-driven method

Compared to the above model-driven methods, the effective non-linear expression ability of deep learning has been developed for removing thin clouds. MCGAN [10] is a neural network model trained by near-infrared (NIR) band images and RGB cloud images synthesized by Perlin noise [13]. However, the texture and spectrum of the Perlin noise are different from that of the actual cloud, and the output image of the MCGAN model generates artefact. This method depends on the NIR data, which can only partly penetrate clouds, whereas synthetic aperture radar (SAR) data are almost completely independent of atmospheric conditions and solar irradiation. Bermudez et al. [11] suggest the use of SAR data instead of NIR data. They used SAR data and the conditional GAN method to reconstruct cloudless optical images. The conditional GAN model requires a pair of cloud/cloudless training data. The Cloud-GAN [12] model used unpaired data for network training and realized the style transformation from cloud images to cloudless images. However, in the network training, the output result was very sensitive to the initialization of variables. Thus, to achieve the best effect of background colour without distortion, the training process must be repeatedly conducted.

3 Techniques Used in Land Use Land Cover

The inputs of the SRARC method are a target image which is contaminated by cloud, an auxiliary image which is an adjacent to the specific temporal image that covers the same open space as the target image, and cloud masks of the target image and auxiliary image, which are used as the guidance for the subsequent cloud removal. The masks can be received by the existing cloud detection methods or manual labeling, in which cloud shadow can also be induced and then finally be removed as cloud. In this paper, the acquired target and auxiliary images have already been geometrically registered, and we assume that the regions which are contaminated by clouds in the target image are cloud-free in the auxiliary image. Please note that cloud regions in the target image will not be removed if there is no available cloud-free complementary information in the auxiliary images. The proposed SRARC method consists of three main steps, as shown . Firstly, the boundaries of the target mask are optimized based on the results of superpixel segmentation, to ensure that they go through homogeneous areas in the target image and avoid spatial discontinuity in the boundaries of recovered areas. The complementary areas from the auxiliary image are then normalized, pixel by pixel, and used to fill the cloud-contaminated areas in the target image, which is achieved by stepwise local radiometric adjustment based on the same cloud-free areas in local windows of the target and auxiliary images. Finally, residual correction is conducted on the filled areas to further eliminate any radiometric differences between the filled areas and the cloud-free areas. The final cloud removal result for the target image can then be generated. For the convenience of the method description in the following subsections, we clipped a pair of experimental images to illustrate the detailed process of SRARC, in which the auxiliary image patch is cloud-free, and we further explain how to cope with the case of the auxiliary image patch also being cloudy.

3.1 Mask Optimization Based on Superpixel Segmentation

Since clouds in the target image are randomly distributed across the given input, the boundaries of the labeled clouds in the target mask will also be arbitrarily determined by the system, which can lead to spatial discontinuity in the boundaries of the reconstructed areas. The optimal seam line is determined by ensuring that it goes through continuous homogeneous areas. Before the complementary areas are transferred to fill out the cloud-contaminated areas in the target image, the boundaries of the target cloud mask should be improved to ensure spatial consistency in the reconstruction results, especially for high-resolution images which will have complex land structures. Unlike the seam lines in image mosaicing, the optimized cloud boundaries must form closed areas in the improved cloud mask.

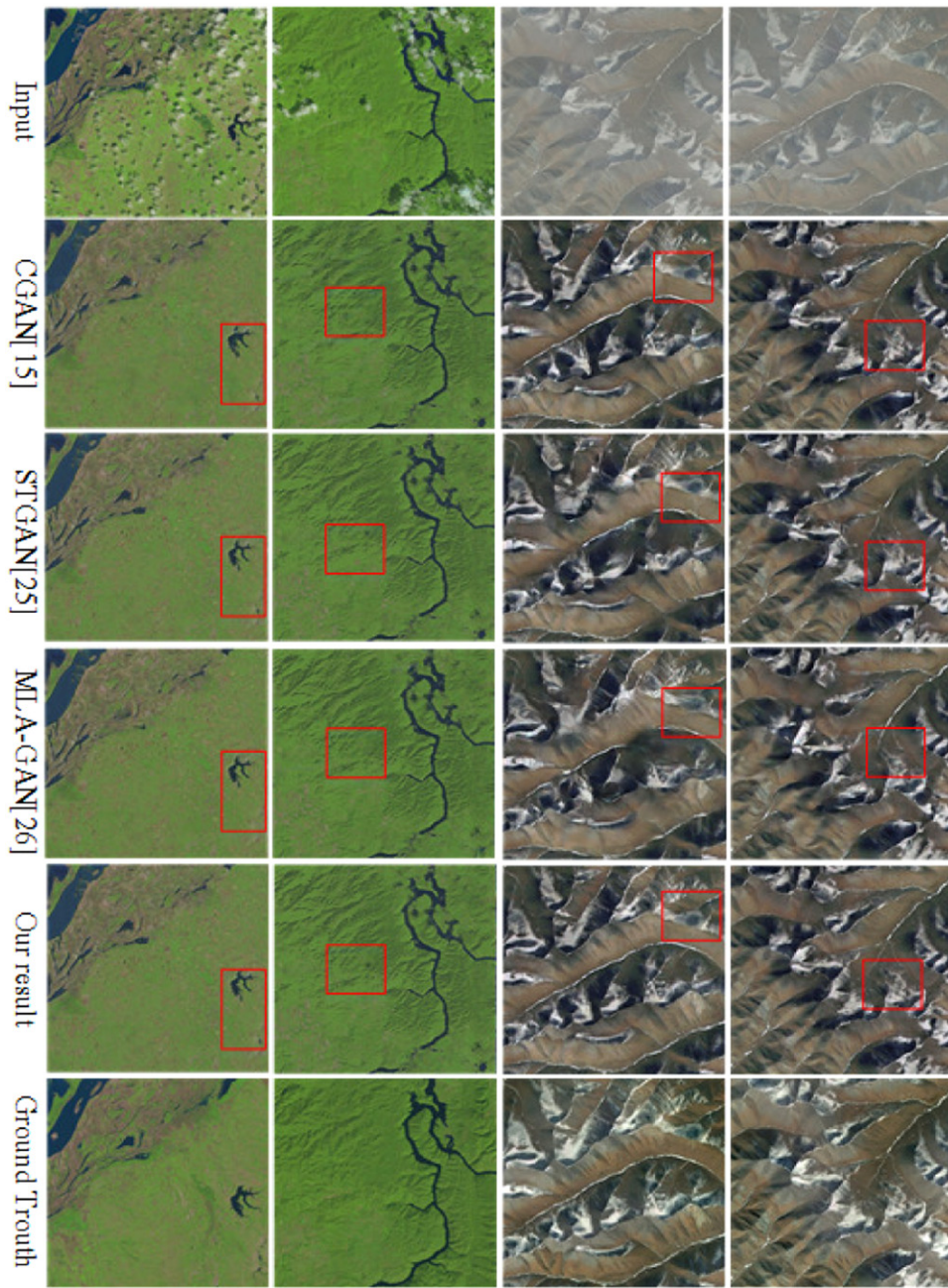


Figure 4: Iterative recovery scheme for large-scale real satellite images.

4 Implementation

4.1 Architecture of the networks

Our cloud generator G consists of an eight-layer encoder and an eight-layer decoder. We add skip connections between all channels at layer i and layer n_i , following the general configuration of the “U-Net” [32] for building both of the high-level semantic features and low-level details. Our cloud discriminator D is a standard convolutional network with 10 convolutional layers followed by 2 fully-connected layers. Our cloud matting network F takes similar configurations with G , but only different in terms of the number of layers and the number of filters.

4.2 Saturation penalty

As the clouds are mostly in white color, the r-g-b channels of the generated cloud reflectance map should be close to each other. To improve the training stability, we add an additional saturation penalty term to the objective of G, i.e. to compute the saturation value of each pixel and penalize those pixels with large saturation values.

Cloud Removal	SAR2OPT	PSNR	SSIM	PSNR (Cloudy Areas)	PSNR (Non-Cloudy Areas)
L1	L1	20.4185	0.5304	18.6577	21.7619
SSIM+L1	L1	21.3423	0.6063	19.0759	23.0808
SSIM	SSIM	21.8548	0.6694	18.5796	24.7527
L1	SSIM+L1	22.0853	0.6309	19.4336	24.3254
SSIM+L1	SSIM+L1	23.0941	0.6909	19.9044	25.9061

Table 1: Performance of method when using different loss functions.

4.3 Training details

The Batch-normalization and ReLU activation function are embedded in G, D and F after all convolution layers, except for their outputs. For the last layer of G and D, we use the sigmoid function to convert the output logits to probabilities. To increase the diversity of the generated images, the background images X and cloud images Y are randomly rotated, flipped and cropped during training. We set the weight for saturation penalty as $\lambda = 1.0$. We use the Adam optimizer for training, with batch size = 4. We use Xavier initialization for all networks. For the first 10 training epochs, we set $\lambda = 0$. For the next 80 epochs, we set $\lambda = 1.0$. All images are resized to 512×512 for training and evaluation. To verify the stability of the training framework, there is a need to take the recent two improvements of GAN into consideration, i.e. WGAN and LSGAN.

Models	Validation Set		Test Set	
	PSNR	SSIM	PSNR	SSIM
Pix2Pix (RGB)	23.13	0.442	22.894	0.437
Pix2Pix (RGB + IR)	21.352	0.485	21.146	0.481
McGAN (RGB + IR)	20.871	0.424	21.013	0.381
Raw Cloudy Images	8.742	0.396	8.778	0.398

Table 2: Running Time Comparisons.

4.4 Cloud removal results

Once we have obtained the cloud reflectance and attenuation maps, the background can be easily recovered by using Eq. 3. In this experiment, we compare our method with a classical cloud removal method: Homomorphic Filter, and two recent proposed methods: DeformedHaze and Adaptive Removal. Then we further estimate our method with a sota image-to-image translation method, CycleGAN. It shows some example results. It can be seen that the thin cloud



Figure 7: Multiple recovery modes of the missing regions.

has nicely been removed by our method and the ground object has been recovered. As the CycleGAN is essentially performing “style transfer” rather than cloud removal, it may introduce unexpected “color shift”. Another limitation of the CycleGAN is that it cannot generate the cloud reflectance or attenuation as it ignores the physical nature or aspects of it. As there is no ground truth for the cloud removal task, to make a quantitative comparison of different methods, we run cloud generator G on our testing set to randomly synthesize 1,390 cloud images and use the original background image as their “ground truth” for evaluation. Table 2 shows the quantitative comparison results of the proposed method and other three cloud removal methods, in which we can see our method performs better than the other three methods under all metrics. We do not further compare with other recent image matting methods because the training of these methods requires the ground truth of alpha matte or the user interactions.

5 Conclusion And Future Scope

In this paper, with the aim of improving the cloud removal results in high-resolution satellite images, which often suffer from the problems of radiometric distortion, noise and artifacts. Here is a proposed thick cloud removal method based on stepwise radiometric adjustment and residual correction (SRARC). The experimental results reveal that the proposed SRARC method is effective in removing the thick clouds in high-resolution satellite images. As a result of the radiometric adjustment and spatial detail preservation ability, SRARC outperforms the other compared cloud removal methods. From this it is known that SRARC has the potential for routine use to support applications based on high-resolution satellite images.

In our future study, multi-source data will be incorporated with the target image to allow the proposed method to better cope with the reconstruction of cloud-contaminated areas suffering significant land-cover changes. Furthermore, the proposed cloud removal method will be applied to generate clear views of desired areas and dates, and to support urban land-use mapping with time-series and high-resolution satellite images.

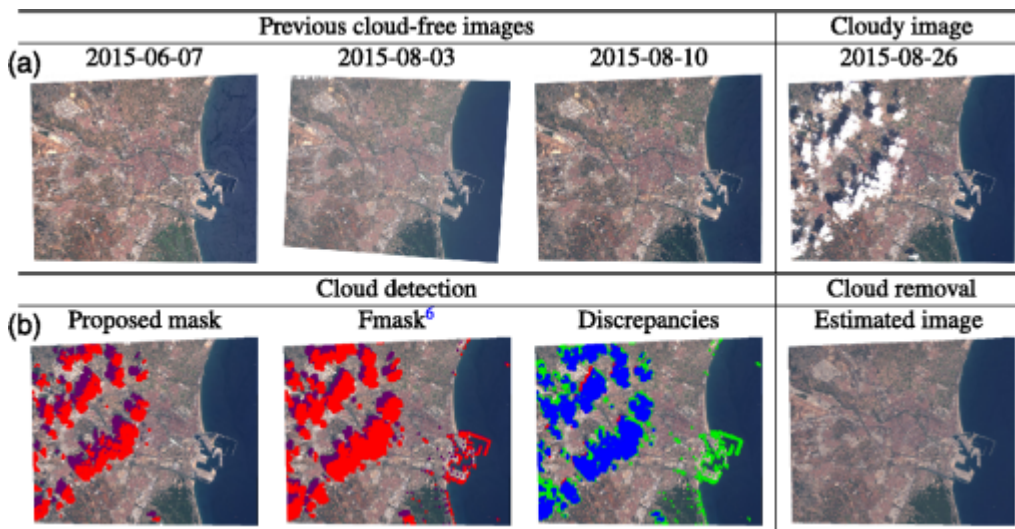


Figure 8: Comparisons of thick cloud removal on real images.

We propose a weakly supervised method for the detection and removal of clouds in remote sensing images based on adversarial training. The proposed method inherently incorporates the cloud imaging mechanism and considers our task as a cloud-background energy separation problem. Our experimental results demonstrate that without ever using any pixel-wise ground truth references during training, our method achieves comparable or even better performance over other methods, which are trained in a fully supervised manner. In addition, the proposed framework can be used for generating cloud images of various styles on any given background. This can be viewed as a new way of performing data augmentation and has great potential for improving occluded object detection and recognition.

References

- [1] Zhenyu An and Zhenwei Shi. Scene learning for cloud detection on remote-sensing images. In *IEEE Journal of selected topics in applied earth observations and remote sensing*, 8(8):4206–4222, 2015.
- [2] Martin Arjovsky, Soumith Chintala, and Leon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223, 2017.
- [3] Guanying Chen, Kai Han, and Kwan-Yee K Wong. Tomnet: Learning transparent object matting from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9233–9241, 2018.
- [4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. In *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018.
- [5] Quan Chen, Tiezheng Ge, Yanyu Xu, Zhiqiang Zhang, Xinxin Yang, and Kun Gai. Semantic human matting. In *In 2018 ACM Multimedia Conference on Multimedia Conference*, pages 618–626. ACM, 2018.
- [6] Qifeng Chen, Dingzeyu Li, and Chi-Keung Tang. Knn matting. In *IEEE transactions on pattern analysis and machine intelligence*, 35(9):2175–2188, 2013.
- [7] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015.
- [8] Kenji Enomoto, Ken Sakurada, Weimin Wang, Hiroshi Fukui, Masashi Matsuoka, Ryosuke Nakamura, and Nobuo Kawaguchi. In Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets. arXiv preprint arXiv:1710.04835, 2017.
- [9] Eduardo SL Gastal and Manuel M Oliveira. Shared sampling for real-time alpha matting. In *In Computer Graphics Forum, volume 29, pages 575–584. Wiley Online Library*, 2010.
- [10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [11] Kaiming He, Christoph Rhemann, Carsten Rother, Xiaoou Tang, and Jian Sun. A global sampling method for alpha matting. In *CVPR 2011, pages 2049–2056. IEEE, 2011*.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [13] Richard R Irish, John L Barker, Samuel N Goward, and Terry Arvidson. Characterization of the landsat-7 etm+ automated cloud-cover assessment (acca) algorithm. In *Photogrammetric engineering remote sensing*, 72(10):1179–1188, 2006.
- [14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [15] Gary J Jedlovec, Stephanie L Haines, and Frank J LaFontaine. Spatial and temporal varying thresholds for cloud detection in goes imagery. In *IEEE Transactions on Geoscience and Remote Sensing*, 46(6):1705–1717, 2008.
- [16] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. In *IEEE transactions on pattern analysis and machine intelligence*, 30(2):228–242, 2008.