Detecting and Removing Clouds Affected Regions from Satellite Images Using Deep Learning

Egharevba, Lawrence; Kumar, Sanjoy; Amini, M. Hadi; Adjouadi, Malek; and Rishe, Naphtali

Abstract: ***Deep Learning is becoming a very popular tool for generating and reconstructing images. Research has shown that deep learning algorithms can perform cutting-edge restoration tasks for various types of images. The performance of these algorithms can be achieved by training Deep Convolutional Neural Networks (DCNNs) with data from a large sample size. The processing of high-resolution satellite imagery becomes difficult when there are only a few images in a dataset. Approaches based on the intrinsic properties of Deep Convolutional Neural Networks (DCNNs) are discussed in this paper for the detection and removal of clouds from remote sensing images without any prior training. Specifically, we focus on reviewing the 2022 study by Czerkawski et al. [10] that proposed deep internal learning for the inpainting of cloud-affected regions in satellite imagery. The technique analyzed performed well when compared to trained algorithms. We also provide an overview of some future research directions.***

Index Terms: ***Artificial Intelligence, Cloud Detection and Removal, Deep Learning, Image Reconstruction, Remote Sensing***

1. *Introduction*

The use of optical remote sensing imagery is undeniably invaluable in several Earth observation applications. Data from satellites are collected consistently and are available on a global scale.
This data can be used for a wide range of purposes, such as monitoring cropland, assessing climate change, assessing land use, and assessing disaster damage. However, clouds pose a major obstacle to surface observations, both in terms of time and space.

Manuscript received March 19, 2023.

Egharevba, Lawrence is with the Knight School of Computing and Information Sciences (KFSCIS), Florida International University (FIU), Miami, FL, USA (e-mail: legha001@fiu.edu)

Kumar, Sanjoy is at the Electrical and Electronic Engineering Department, Shahjalal University of Science and Technology, Dhaka, Bangladesh (e-mail: sanjoy.eee32@gmail.com)

Amini, M. Hadi is with the KFSCIS at FIU (amini@cs.fiu.edu)

Adjouadi, Malek is with the Department of Electrical and Computer Engineering at FIU (e-mail: adjouadi@fiu.edu)

Rishe, Naphtali is with the KFSCIS at FIU (e-mail: rishen@cs.fiu.edu)

Correspondence e-mail is legha001@fiu.edu

Earth observation has increasingly relied on optical remote sensing images for a very long time [2]. Since the launch of the first Earth observation satellite [3], satellite imagery has become an increasingly critical tool for an array of useful applications [4]. These applications are used for Earth observation, natural and anthropogenic disturbances [5], land cover classification, environmental monitoring, disaster monitoring, detection, and the allocation of resources, among others. The usefulness of these applications is largely due to the advancement of remote sensing technology and technological advances in hardware equipment [6]. A satellite sensor's ability to capture many Earth-related processes remotely is impressive for its consistency and repeatability on a large spatial scale [6]. However, this is insufficient for many applications of Earth Observation [7].

Due to weather conditions and changes in the atmosphere, satellite images can be obscured by clouds. Thus, remote-sensing images cannot be interpreted correctly [8]. Approximately 35% of the surface of the Earth is covered by clouds. The transmission of solar radiation through cloud layers results in reflection and absorption.

In remote sensing satellite sensors, electromagnetic waves are lost to a certain extent. This loss makes the information or imagery difficult to process or interpret [9]. Due to the fact clouds obstruct the images, the information the satellites gather cannot be processed or interpreted afterward. This severely limits the use of the images for research. The amount of radiation absorbed, reflected, and emitted by clouds determines their thickness. Analyzing, processing, and interpreting remote-sensing images requires detecting and removing cloud cover [4].

Historically, researchers have been interested in removing clouds and recovering land information. Cloud obscuration is a ubiquitous and inevitable problem. It has hindered the usability of satellite imagery and interfered with accurate geographic mapping, which relies on image interpretation. Long-term Acquisition Plan (LTAP) is utilized by the research community since the launch of the Landsat Enhanced Thematic Mapper Plus (ETM+) to analyze the annual refreshed, cloud-free archive of ETM+ data [7]. A long-term record of land remote sensing data has been provided by the Landsat project/NASA (the U.S. National Aeronautics and Space Administration) [11], [12].

The ease of interpretation and implementation of image enhancement has made it one of the most popular techniques in satellite imagery processing [13]. Since cloud films obscure remote sensing images frequently, satellite images suffer from spatial-temporal discontinuity. This results in reduced quality and utility. The task of predicting what lies under clouds is unrestricted since it allows for a wide range of in-paintings. Satellite images taken at longer wavelengths may enable cloud penetration, which allows for further restrictions. There are many sources of such type of information. Near-infrared (NIR) and Synthetic Aperture Radar (SAR) are the most common types of images [14].

Several methods have been published for detecting and removing clouds by several scholars. In general, these methods can be categorized as traditional and deep learning methods [15]. The detection and removal of clouds from satellite images is a critical preprocessing step in the amelioration of remotely sensed data. This is because clouds are treated as noise on the input image [16].

1.1 Traditional Method

Repairing or reconstructing cloud-contaminated images traditionally involves taking advantage of the spectral or spatial characteristics of clouds. The physical characteristics of clouds are used to determine thresholds in traditional cloud detection techniques. The cloud detection and removal approach studies how clouds reflect light in various bands and how they relate to one another [17]. Setting thresholds for specific physical characteristics can result in a better detection effect by utilizing the difference between the cloud and non-cloud areas [18]. Based on how cloud-free source images are used, traditional methods can be further subdivided into multispectral and multi-temporal techniques [19], [20].

Cloud extraction uses the spatial information in the clouded contaminated regions of an image to its full capacity [15]. Remote sensing images contain spatial structures [21], which tend to be more similar between adjacent pixels than pixels separated by a wider distance. Interpolating missing information requires understanding the magnitude and pattern of spatial variance [8].

Thin cloud removal can be accomplished with the multispectral method when such clouds do not cover the ground object entirely. The multi-temporal method, on the other hand, is a method of restoring images that are obscured by a thick layer of clouds. For instance, Xu et al. [22] proposed a thin cloud removal technique that uses multispectral analysis of the available image. Zhang et al. [23], on the other hand, proposed a technique that uses a haze-optimized transformation (HOT) to eliminate the areas of haze that appear in Landsat images. This HOT method analyzes visible-band space with a high correlation between the spectral response and surface cover classes. However, the haze's spectral response is highly dependent on the wavelength and depth of the haze, which limits its generalization.

A multispectral technique for removing thin clouds was developed by Huanfeng et al. [24]. They are either too complicated or not effective enough to remove locally aggregated thin clouds. The noise-adjusted principal component transform mode (CR-NAPCT) was proposed as a method for removing thin clouds in [19]. A method based on images is developed for cirrus cloud contamination correction [25]. Thick clouds, however, can block most, if not all, the land signals in the optical band if their thickness increases. A lack of auxiliary spectral information makes the above methods ineffective for the removal of thick clouds.

As the Earth's surface reflects light back to space, clouds dynamically impact the absorption and transmission of light. In thick and dense clouds, all signal energy can be blocked, while in thin, transparent clouds, all signal energy can be contaminated and attenuated. Accordingly, different algorithms are implemented for different cloud conditions. To remove thick clouds, multi-temporal image-based techniques have been widely applied. As an auxiliary data source, multi-temporal satellite images are used to create cloud-free images [26]. To create cloud-free images, data are obtained from temporal images that have not been contaminated by clouds [17].

In the paper [27], a dictionary learning technique was developed based on sparse representation. A method for removing clouds using reference information was proposed in [28]. Through the alignment of multi-temporal images, missing data from a satellite image is recovered after cloud-contaminated portions are removed [28]. A method is proposed in [1] for the reconstruction of missing information in remotely sensed data. An approach to reconstructing thick clouds based on multi-temporal dictionary learning was proposed by Li et al. [29]. These methods, however, require the availability of reference images. Cloud-free images can sometimes be difficult to obtain, especially in tropical regions [30]. Multi-temporal methods generally ignore temporal variations in ground cover between time series images [1].

Multi-temporal techniques are equipped to deal with both thin and thick clouds. For example, the multi-temporal dictionary learning technique [22], [31] combines dictionary learning from the target image and the reference image to remove an image. With this approach, contaminated data can be recovered regardless of how thin or thick the clouds are or how prominent the cloud shadow is. Due to the periodic revisiting of the same geographical location by satellites, there is the possibility of acquiring multi-temporal cloud-free images multiple times for the same location. The method, however, requires cloud-free images as auxiliary input data. Thus, these methods are dependent on cloud-free images to perform well [26]. To overcome these limitations, synthetic aperture radar (SAR) images provide an alternative approach for cloud removal tasks [14].

1.2 Deep Learning Method

Deep learning methods have gained popularity in the past few years. They train their networks using pairs of cloudy and cloud-free images, which are then used to map the cloudy regions to the cloud-free regions [5], [33]. Recently, deep learning convolutional neural networks are becoming increasingly popular as a technique for generating and reconstructing images. An image reconstruction technique, known as the Deep Image Prior, is discussed in this paper to reconstruct missing information in remote sensing images. It is a framework based on a deep convolutional neural network (CNN) without any trained data.

Several deep learning networks have been developed to improve the ability to restore remote sensing images. In a recent paper, a deep learning network that combines a convolution and deconvolution operator was proposed for restoring satellite images [34]. A convolutional mapping to deconvolutional network was devised by Li et al. for cloud removal using optical and SAR data. The convolutional layers are used for encoding, the mapping layers are utilized for translating features, and the deconvolutional layers are used for decoding. Another solution involves incorporating spectral information into missing data reconstructions [10].

images, generative adversarial networks (GANs) have been introduced [36]. GANs are generative models based on Deep Learning, including two models trained simultaneously: one model is trained to generate fake data, and the second model is trained to discern the fake data from the real samples [16]. Researchers create artificial naturalistic images using GANs.

Several GAN variations have been used to remove clouds from SAR images. For example, to eliminate clouds from RGB images, Enomoto et al. devised a Multispectral Conditional Generative Adversarial Network (McGAN) [14]. A Cloud-GAN was proposed to map the relationship between cloudy and cloud-free images [37]. Mirza and Osindero, in their paper, suggested that both the generator and discriminator should consider a class label as a second input [38]. This paper introduces generative adversarial networks in a conditional form. Simply by feeding the CGAN the data, it can create images specific to the defined class. For learning unsupervised representations, a GAN model named Deep Convolutional Generative Adversarial Networks was developed (DCGAN) [39].

Cycle GAN is a variation of a generative model that translates images between different domains. The authors explored it in [40], creating a painting of a picture from an image. SRGAN is a Super-resolution GAN sample of low-resolution images for increased spatial resolution [41]. Another variant of GAN is presented in [42], a spatial-spectral GAN model that classifies hyperspectral images into multi-classes.

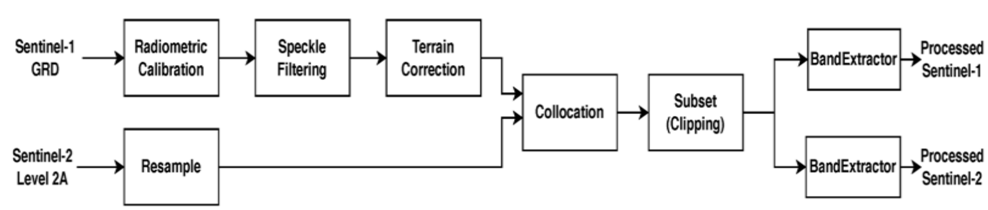

Figure 1: Figure of Sentinel Application Platform (SNAP) iterative process for Sentinel-1&2. Licensed here and infra from [10].

Huang et al. [35] proposed an integrated cloud detection and removal framework. It is based on cascade convolutional neural networks, providing accurate cloud and shadow masks and repairing images. A novel method for removing clouds using image inpainting and denoising is proposed, called Cloud-Aware Generative Network (CAGN) [33]. To produce cloud-free images from SAR

Zhu et al. [43] created a model to increase data for objections in remote sensing images using a multi-branch conditional GAN model (MCGAN). GAN's remarkable feature of generating networks lets Singh and Komodakis build their networks without having to train them on cloudy and cloud-free images [54].

Another effective variance of GAN called DiscoGAN was proposed in [44], in which two

GANs coupled together in DiscoGAN were used to learn the relationships between two domains. Using an image in one domain as a starting point, DiscoGAN can generate an image that corresponds to an image in another domain. As a result of their ability to generate high-quality images and augment data, GANs have received significant attention recently. In [45], YUV-GAN proposes a method for converting RGB images into the YUV color space. Xu et al. [46] processed the cloudy synthesized images at multiple scales. As a result, researchers have developed several GAN variants, including WGAN, CGAN, Progressive GAN, image2image translational GAN (I2I), Cycle GAN (CGAN), text2image GAN (T2I), face inpainting GAN (FI-GAN), and text2speech GAN (T2S) [36].

Deep learning-based techniques, despite their successes over conventional techniques, have introduced several new obstacles in remote sensing. For example, high computational time and dataset requirements. For a deep learning model to be successful, a large number of samples must be used. There is, however, a problem of limited samples within the remote sensing community. As a result, training results in over-fitting, i.e., the data perform well during training but don't generalize. To generalize the decision-making process of supervised machine learning and Deep Learning, large and labeled datasets are needed. Thus, unsupervised Learning is heavily relied upon by the research community.

This paper analyzes a technique based on [10] for reconstructing missing information in remote sensing images, called the deep image prior. An example of unsupervised Learning is deep image prior, a type of generative modeling that utilizes data to create similar data based on its structure. Learning the structure and extracting useful features from the data is completely independent of Learning. When fed with noisy input data, GANs may fail to perform as expected [30].

1.3 Deep image prior

The deep image prior (DIP) is a convolutional neural network that can reconstruct inverse images. Removing noise from remote sensing images or creating super-resolution single images are examples of a deep image prior [20]. It is an effective deep convolutional neural network (ConvNets) for optimizing images without the need for previous training datasets. Deep image prior utilizes an image denoising technique that takes a cloudy image as input and creates a cloudless image as an output [41]. This technique combines learning-based methods using deep convolutional networks and learning-free methods based on manually constructed image priors [47]. Super-resolution refers to the challenge of estimating a high-resolution image from its low-resolution counterpart.

It has been previously assumed that Learning is essential to building reliable image priors; however, recent works show that a significant amount of image information can be derived from the structure of a convolutional image generator independent of Learning**.** This is of paramount importance when solving various problems related to image reconstruction. Image prior is required to incorporate information that has been lost due to noise or contamination. In contrast to previously learned cloud detecting and removal deep learning-based techniques, the structure of a convolutional image generator captures image statistics.

The loss function between the output image and the degraded image is constructed based on the degradation model and is used as input to the deep image prior. It is possible to obtain the reconstruction results after optimizing the network several times. In contrast to these deep learning methods, deep image priors do not require any training data to be input into the network. Deep image priors may, however, not yield satisfactory results if the inverse process is too difficult [47].

Furthermore, deep image priors can be classified as a traditional inpainting method. In essence, it means that it will not work in the case of large gaps. The deep image prior is a type of convolutional neural network (ConvNets) used to enhance images without any previous training data. To help in solving inverse problems, such as noise reduction, super-resolution, and inpainting, neural networks are randomly initialized.

As an input to the deep image prior, the noise map is used to construct a loss function between the degraded output images. The output of the reconstruction can be obtained by optimizing the network multiple times [48].

1.4 Deep Image Prior (DIP) Framework

Deep Image Prior (DIP) transforms spectral information directly, unlike deep neural networks that are conditioned on training. Figure 1 [10] illustrates how a cloudy image is converted directly into a cloud-free image using global information. As a final step in transforming the complex spectral information, the SAR (Synthetic Aperture Radar) Sentinel-1 data are calibrated radiometrically, filtered off speckles, and corrected for terrain. Sentinel-2 multispectral data are reprocessed to integrate the spatial resolution of the two data sets and to facilitate mask stacking.

A deep network consists of two subnetworks, f1 and f2. In DIP, the same cloud-free image is used as input and output, preserving the latent code spatial information in the f1 and f2 subnetworks. Throughout the process, images are optimized.

Unlike other methods based on deep networks, one does not need a training dataset, only a reference image that has intact spatial information and a reconstructed image.

1.5 Related approaches and motivation

Image denoising methods aim to recover the original image from noisy measurements. There have been several approaches proposed for removing noise and recapturing the true image. In Paper [47], the authors propose a simplified approach to basic image reconstruction, including denoising, in-painting, and super-resolution. It is remarkable because the network used isn't based on learning from data. In [49], an approach to low-level vision is presented that combines two key ideas: Image processing architectures based on convolutional networks and unsupervised learning synthesis of training samples based on specific noise models. In various image processing and computer vision tasks, removing noise from the observed image is essential [50].

The deep learning model [51] shows dynamic fitting abilities and learning abilities. Due to this, it is being used in image processing tasks such as in-painting [20] and style transfer [52]. The ability of deep neural networks to reconstruct missing pixels unsupervised has been demonstrated in recent studies [53]. In [47], the authors apply DIP to processing an image with a random-initialized network utilizing image inpainting. Their technique outperforms traditional methods in such applications.

A new reconstruction idea called Image2StyleGAN is introduced in [54]. A trained generative network receives the most suitable input code by propagating backward from the intact parts of the degraded image. By implementing backward propagation, they get high-accuracy reconstruction results using both the optimal code and the noise map [54]. The results of the study show that a deep neural network is capable of reconstructing a single image even without training data. The reliability of the results, however, cannot be guaranteed [55].

1.6 An overview of the paper

This paper is structured as follows. After this introductory section, the methodology being analyzed, including the neural network architecture and custom loss, is explained in Section 2. The characteristics of the used dataset and experiments are discussed in Section 3. Metrics and limitations are then discussed in Section 4. Finally, conclusions are given in Section 5.

1.7 Motivation

Remote sensing images can be reconstructed using various deep learning methods; however, they are useless without training data. The effectiveness of deep learning algorithms is much higher than that of traditional algorithms. A lack of training data, nonetheless, will prevent deep learning methods from working. Additionally, trained networks are not flawless in their ability to generalize. If the real corrupted images diverge greatly from the training dataset, the reconstruction results might not be satisfactory. There is a need to consider the application of deep learning networks to repair only one image in the absence of a training dataset [32], [56], [57]. Data from multiple sources (spatial, spectral, and temporal information) are used as inputs. The process of denoising an image has become more automated and faster thanks to deep Learning [58].

2. Review of Methods

In this paper, we focus discussion on reviewing an unsupervised method widely known as the Deep Image Prior method proposed in [10]. This technique enables the reconstruction of a single remote-sensing image without the need for training datasets. The technique processes a reference image of the cloudy image using a deep self-regression network; the internal structure is thereby extracted. The metrics show that the method outperforms the previously known algorithms.

In their proposed solution, a Deep Image Prior model for satellite imagery detection and removal is described [10]. An image x is mapped to a random code vector z by the DIP architecture. This technique is used to sample real cloud-free satellite imagery from a randomized set of images. These results can be obtained by using an optimizer such as gradient descent. As an example, satellite image recoveries inverse tasks like in-painting, denoising, or removing cloud and super-resolution may be stated as an optimization problem like this:

$$x^* = \min_{x} E(x,\ x_0) + \mathrm{R}(x) \tag{1}$$

The metric $E(x,\ x_0)$ is a task-related metric, x, and x_0 are uncorrupted and corrupted satellite imagery. In most CNN-based methods, $\mathrm{R}(x)$ is a regularization term (image prior) that is applied either manually or automatically. As shown in Figure 1 [10], the algorithm for finding image priors works directly by optimizing the parameters of a network. This algorithm allows for the search for a solution without regularization terms.

$$x^* = f\theta^*(Z), \text{ where } \theta^* = \arg\min_{x} E(f\theta^*(Z), x_0), \tag{2}$$

In this case, f is a DCNN with attributes, and $\mathcal{Z}$ is a fixed input (cloudy). Using only a cloudy image, a cloud-free image can be recovered. The deep image prior approach employs a network structure to find optimal weights based on its intrinsic prior instead of learning them from data. A minimizer approximates the function as follows:

$$\theta^* = \underset{x}{\arg\min}\, E(f\, \theta^*(\mathcal{Z}),\ x_0) \quad (3)$$

Gradient descent, an optimizer, is one method of obtaining these results.

$$\theta^* = \underset{x}{\arg\min}\, E(f\, \theta^*(\mathcal{Z}),\ x_0) \quad (4)$$

3. Experiments

A real-time data experiment was carried out utilizing Sentinel-1 SAR images and Sentinel-2 optical images. Images were captured on February 9, 2022, from two different regions of the Earth (Europe and Asia). The experiments demonstrate the effectiveness of the Deep Image Prior to satellite imagery reconstruction. Since the internal synthesis method is image content-based, it can be used on other image types. Images from Sentinel-2 Level-1C are among the most common image types. Using Level-1C samples, the performance attained by [10] is superior to other models. A match was made between the resulting Sentinel-2 images and their equivalents from Sentinel-1 based on the closest temporal proximity [32], [23].

Two different regions have been analyzed using two-year temporal coverage measurements. The regions for this study in the dataset are Scotland, referred to as Scene I, and India, referred to as Scene II. Approximately 200 images are included in the dataset for each scene. The approach was demonstrated on images with substantial cloud cover to demonstrate its robustness in [10].

3.1 Data Preparation

A data-driven approach is discussed here, which is broadly applicable and sensor-independent. It was trained and tested on Sentinel satellite imagery collected from two different regions (Europe and Asia). Using remote sensing data from the Sentinel-2 satellite, this work developed an algorithm to remove clouds from satellite imagery [59]. The Sentinel-1 satellite provided the Synthetic Aperture Radar (SAR) data utilized in this study [60]. Sentinel-2 images with a cloud fraction of 10% were selected, with a cloudy image within 20% and 60% cloud fraction. As a result, 17 real mask shapes were developed to create cloud covers [10].

3.2 Application

According to the previous explanation, the dataset captured data from two distinct regions over two years. A total of 200 samples were used that were not obscured by clouds. Output pairs were then handled similarly to alter that modality. Images from Sentinel-1 were cropped (VV range: [25.0, 0], VH range: [32.5, 0]), moved to [0, 2], and converted to decibel scale (DS). For all optical channels, Sentinel-2 images were cropped between [0, 104] and split by 2000. Based on the rescaling and combination of cloud-free images with each cloud mask, Table 1 [10] shows sample counts for these two regions:

Table 1.
Samples of Counts for Scene I and II [10]

Dataset/samples	Scene I	Scene II
Incoming (2019)	18	34
Inference (2020)	20	30
Total	340	510

As a result, optical remote sensing images are now available with higher spatial and spectral resolution. Data information from cloud films frequently obscures remote sensing images because of the cover of clouds on images. This results in spatial-temporal discontinuity, which reduces the quality and usefulness of satellite images. Predicting what will happen beneath a cloud is an under-constrained task since it allows for a wide range of in-paintings. Further restrictions in the form of images taken at longer wavelengths, which allow cloud penetration, may be a solution. Synthetic Aperture Radar (SAR) and Near-Infrared (NIR) images are two of many sources of information available.

3.3 Network Architecture

The internal cloud elimination designs discussed here are based on the Deep Image Prior (DIP) approaches [10]. The approach's central tenet is that signals, particularly genuine images, may be stored in the weight of a Deep Convolution Neural Network. This is done by overfitting the model to produce the desired signal for a fixed input. The application shown in the original DIP work is with a single picture. In this case, the approach extends to cloud removal by utilizing extra information, i.e., using many images. However, the extra images may come from other times in the past (multi-temporal synthesis). The outcomes showed that by stacking all the frames to produce an image representation with additional channels, the DIP technique may be successfully used on both multi-temporal and multi-source satellite images. Additionally, even when there was a discrepancy across domains, as there is with optical and SAR sensor data, good-quality images could still be created.

Additional temporal samples must be employed at the inference time in multi-temporal synthesis. This supplementary source may be static (using the same reference for several predictions) or dynamic (using a reference uniquely selected for the synthesized sample). For example, the last clear sky image was taken before the cloudy sample. All 2019 clear sky pictures were averaged to get a static reference. The major goal of calculating a mean over the full year, which included all seasons and weather conditions, was to identify the region's static structural component. However, using the mean also produces a further denoising effect, especially for SAR images. An informed sample mean is shown in Figure 2.

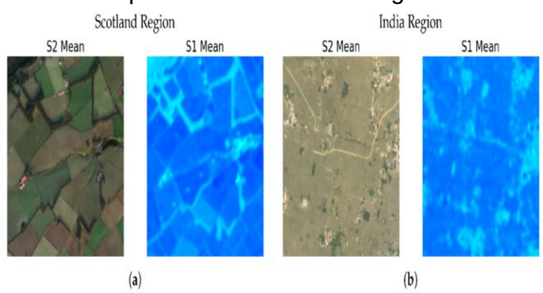

Figure 2: Mean images for 2019 for the region in Scene I (a) and the region in Scene II (b). Sentinel-1 VV (green) and VH (blue) [10].

All the experiments were produced using the Adam optimizer with learning rates of 210^{-2} and exponential decay rates for the first and second-moment coefficients of 0.9 and 0.999, respectively [10].

The metrics used were root-mean-square error (RMSE), structural similarity index measurement (SSIM), and mean absolute error (MAE) -- to conduct a quantitative evaluation of different network configurations.

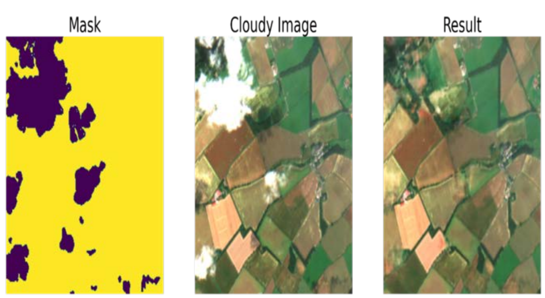

Figure 3: The selection of reconstructed samples across the regions demonstrates the model's adaptability [10].

This approach has a particularly high significance in the domain of Deep Image Prior (DIP). DIP is used for image enhancement, generation, and restoration. Image processing techniques such as Deep Image Prior are widely used. As a result of this technique, realistic satellite images can be learned without training or Learning, as illustrated in Figure 3 [10].

The acquisition of data was streamlined by the simplification of Deep Image implementation processes [10]. As a result, the satellite image detection and reconstruction process were not as complex compared to other deep learning models where the training of data sets is required. It has also made the DIP algorithms increase in demand. The lack of data, however, makes it impossible to use accurate learning-based methods. Even publicly available datasets consist of a relatively small number of images and rarely exceed a hundred. At times, the datasets consist of just one image due to the complexity of the acquisition.

The Formulation used in detecting and removing clouds from satellite imagery is based on the implementation of convolutional encoder-decoders. It is shown above that inverse problems involving noise reduction, super-resolution, and inpainting can be defined as optimization tasks.

4. *Discussion of Existing Work and Potential Future Directions*

A quantitative evaluation was conducted by using performance metrics for the entire image and the in-painted area separately [10]. In the evaluation dataset covering Scotland and India, all synthesis modes have been applied. In both datasets, including the full image and the in-painted regions, stacked MT and MS-MT variants with multi-temporal data performed well. A metric applied to the in-painted region measures the quality of newly generated data and quantifies the distortion caused by the optimization process when applied to the entire image.

Even though the assessment dataset included real cloud masks whose coverage ranged from 10-50%, it did not include a measure of how the system performance fluctuated based on the size of the in-painted zone. A different experiment was conducted in which clear sky samples from each dataset were reconverted to artificial cloud masks with predetermined coverage areas. The in-painted region SSIM for Scotland (Scene I) and India (Scene II) regions was around 0.8 and 0.75, respectively. The in-painting quality remained stable for cloud coverage ratios between 0.1% and 16%.

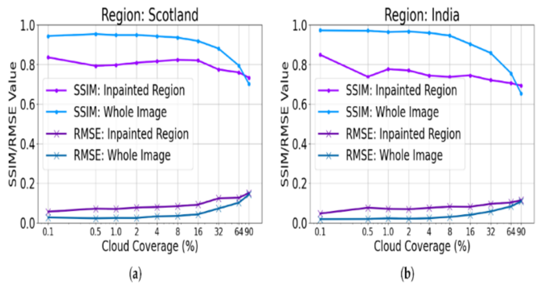

Figure 4: SSIM and RMSE plot for the cloud coverage sweeps for the MS-MT model for the (a) Scotland region and (b) India region [10].

Limitations of Deep Image Prior

There are still limitations to the methods like [10], but they achieve satisfactory results. DIP cannot handle low-resolution reference images in the heterogeneous-based variant. As a result, the framework is time-consuming, demonstrating its inefficiency in comparison with traditional methods and deep learning approaches. It is anticipated that, in the future, multitemporal images will be combined with heterogeneous images to improve reconstruction accuracy. Refinement and enhancement of the method will fill in the gaps of missing information.

In DIP, the remaining parts of cloudy images are assumed to have a large amount of redundant data. The recovery of some unique ground objects is impossible in the presence of clouds. Even though the discussed method [10] yields superior reconstruction results, it is inefficient in terms of time. Instead of optimizing the output of the network by feeding it forward, this method optimizes it by propagating the output backward. Thus, the discussed method takes a long time to implement. Our focus will be on improving its efficiency in the future.

5. Conclusion

This paper included a discussion of a technique for restoring missing information in satellite imagery, called Deep Image Prior, with a focus on reviewing the study in [10]. Their proposed DIP process does not rely on training datasets; instead, it uses a deep image prior network to process the reference image. A feature map is extracted from the hidden layer based on residual information from contaminated data. In this technique, deep convolutional networks are combined with learning-free methods based on self-similarity priors.

In addition to removing clouds from Sentinel-2 data collected in Scotland (Scene I) and India (Scene II), DIP can be transformed into spectral, multitemporal, and heterogeneous methods using different reference images. In the solutions offered, various strategies were used to guide the synthesis process. When dealing with stacked data from the same mode, DIP-based approaches perform better than multi-sourced composites. In the two-band SAR representation, aberrations resulted in higher-quality synthesized samples than historical optical data.

The methods discussed here require reliable cloud detection systems and cloud masks. In addition, the results showed that, depending on the process, the synthesized visuals can differ. There are many possible solutions to in-painting. Further work needs to be done to reduce distortion and unpredictability.

References

[1] Burger, H. C., Schuler, C. J., & Harmeling, S. (2012, June). Image denoising: Can plain neural networks compete with BM3D? In *2012 IEEE conference on computer vision and pattern recognition* (pp. 2392-2399). IEEE.

[2] Li, J., Hu, Q., & Ai, M. (2018). Haze and thin cloud removal via sphere model improved dark channel prior. *IEEE Geoscience and Remote Sensing Letters*, *16*(3), 472-476.

[3] Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., & Zhang, L. (2015). Missing information reconstruction of remote sensing data: A technical review. *IEEE Geoscience and Remote Sensing Magazine*, *3*(3), 61-85.

[4] Zhang, Q., Yuan, Q., Zeng, C., Li, X., & Wei, Y. (2018). Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(8), 4274-4288.

[5] Jeppesen, J. H., Jacobsen, R. H., Inceoglu, F., & Toftegaard, T. S. (2019). A cloud detection algorithm for satellite imagery based on deep Learning. *Remote sensing of environment*, *229*, 247-259.

[6] Enomoto, K., Sakurada, K., Wang, W., Fukui, H., Matsuoka, M., Nakamura, R., & Kawaguchi, N. (2017). Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 48-56).

[7] Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment*, *120*, 25-36.

[8] Meraner, A., Ebel, P., Zhu, X. X., & Schmitt, M. (2020). Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, *166*, 333-346.

[9] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600-612.

[10] Czerkawski, M.; Upadhyay, P.; Davison, C.; Werkmeister, A.; Cardona, J.; Atkinson, R.; Michie, C.; Andonovic, I.; Macdonald, M.; Tachtatzis, C. Deep Internal Learning for Inpainting of Cloud-

Affected Regions in Satellite Imagery. Remote Sens. 2022, 14, 1342. https://doi.org/ 10.3390/rs14061342/. Certain artifacts from this 2022 paper are used in the present paper under the license at https://creativecommons.org/licenses/by/4.0/

[11] Mahajan, S., & Fataniya, B. (2020). Cloud detection methodologies: Variants and development—A review. *Complex & Intelligent Systems*, *6*(2), 251-261.

[12] Chen, J., Zhu, X., Vogelmann, J. E., Gao, F., & Jin, S. (2011). A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. *Remote sensing of environment*, *115*(4), 1053-1064.

[13] Jiayuan Li, Qingwu Hu, and Mingyao Ai. Haze and thin cloud removal via sphere model improved dark channel prior. IEEE Geoscience and Remote Sensing Letters, 16(3):472–476, 2018

[14] Kennedy, R. E., Cohen, W. B., & Schroeder, T. A. (2007). Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sensing of Environment*, *110*(3), 370-386.

[15] Chuvieco, E., Mouillot, F., Van der Werf, G. R., San Miguel, J., Tanase, M., Koutsias, N., ... & Giglio, L. (2019). Historical background and current developments for mapping burned area from satellite Earth observation. *Remote Sensing of Environment*, *225*, 45-64.

[16] Xu, Z., Wu, K., Wang, W., Lyu, X., & Ren, P. (2022). Semi-supervised thin cloud removal with mutually beneficial guides. *ISPRS Journal of Photogrammetry and Remote Sensing*, *192*, 327-343.

[17] Jin, S., & Sader, S. A. (2006). Effects of forest ownership and change on forest harvest rates, types and trends in northern Maine. *Forest Ecology and Management*, *228*(1-3), 177-186.

[18] Coppin, P. R., & Bauer, M. E. (1996). Digital change detection in forest ecosystems with remote sensing imagery. *Remote sensing reviews*, *13*(3-4), 207-234.

[19] Xu, M., Deng, F., Jia, S., Jia, X., & Plaza, A. J. (2022). Attention mechanism-based generative adversarial networks for cloud removal in Landsat images. *Remote Sensing of Environment*, *271*, 112902.

[20] Ju, J., & Roy, D. P. (2008). The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sensing of Environment*, *112*(3), 1196-1211.

[21] Woodcock, C. E., Loveland, T. R., Herold, M., & Bauer, M. E. (2020). Transitioning from change detection to monitoring with remote sensing: A paradigm shift. *Remote Sensing of Environment*, *238*, 111558.

[22] Meng Xu, Mark Pickering, Antonio J Plaza, and Xiuping Jia. Thin cloud removal based on signal transmission principles and spectral mixture analysis. IEEE Transactions on Geoscience and Remote Sensing, 54(3):1659–1669, 2015.

[23] Li, Y., Chen, W., Zhang, Y., Tao, C., Xiao, R., & Tan, Y. (2020). Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep Learning. *Remote Sensing of Environment*, *250*, 112045.

[24] Goward, S. N., Masek, J. G., Williams, D. L., Irons, J. R., & Thompson, R. J. (2001). The Landsat 7 mission: Terrestrial research and applications for the 21st century. *Remote Sensing of Environment*, *78*(1-2), 3-12.

[25] Belward, A. S., & Skøien, J. O. (2015). Who launched what, when, and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 115-128.

[26] Zhang, C., Li, W., & Travis, D. J. (2009). Restoration of clouded pixels in multispectral remotely sensed imagery with cokriging. *International Journal of Remote Sensing*, *30*(9), 2173-2195.

[27] Zhang, J., Wu, J., Wang, H., Wang, Y., & Li, Y. (2021). Cloud detection method using CNN based on cascaded feature attention and channel attention. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-17.

[28] Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote sensing of environment*, *118*, 83-94.

[29] Chen, Y., Weng, Q., Tang, L., Zhang, X., Bilal, M., & Li, Q. (2020). Thick clouds removing from multitemporal Landsat images using spatiotemporal neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-14.

[30] Liu, S., Zhang, Z., Xiao, B., & Cao, X. (2015). Ground-based cloud detection using automatic graph cut. *IEEE Geoscience and remote sensing letters*, *12*(6), 1342-1346.

[31] Zhang, Y., Guindon, B., & Cihlar, J. (2002). An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. *Remote Sensing of Environment*, *82*(2-3), 173-187.

[32] Xu, M., Pickering, M., Plaza, A. J., & Jia, X. (2015). Thin cloud removal based on signal transmission principles and spectral mixture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(3), 1659-1669.

[33] Salim Malek, Farid Melgani, Yakoub Bazi, and Naif Alajlan. Reconstructing cloud-contaminated multispectral images with contextualized autoencoder neural networks. IEEE Transactions on Geoscience and Remote Sensing, 56(4):2270–2282, 2017.

[34] Shen, H., Li, H., Qian, Y., Zhang, L., & Yuan, Q. (2014). An effective thin cloud removal procedure for visible remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *96*, 224-235.

[35] Huang, C., Thomas, N., Goward, S. N., Masek, J. G., Zhu, Z., Townshend, J. R., & Vogelmann, J. E. (2010). Automated masking of cloud and cloud shadow for forest change analysis using Landsat images. *International Journal of Remote Sensing*, *31*(20), 5449-5464.

[36] Xu, M., Jia, X., Pickering, M., & Jia, S. (2019). Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform. *ISPRS Journal of Photogrammetry and Remote Sensing*, *149*, 215-225.

[37] Xu, M., Jia, X., & Pickering, M. (2014, July). Automatic cloud removal for Landsat 8 OLI images using cirrus band. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 2511-2514). IEEE.

[38] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014

[39] Zhu, X., Gao, F., Liu, D., & Chen, J. (2011). A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images. *IEEE Geoscience and Remote Sensing Letters*, *9*(3), 521-525.

[40] Xu, M., Jia, X., Pickering, M., & Plaza, A. J. (2016). Cloud removal based on sparse representation via multitemporal dictionary learning. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(5), 2998-3006.

[41] Lin, C. H., Tsai, P. H., Lai, K. H., & Chen, J. Y. (2012). Cloud removal from multitemporal satellite images using information cloning. *IEEE transactions on geoscience and remote sensing*, *51*(1), 232-241.

[42] Jie Feng, Haipeng Yu, Lin Wang, Xianghai Cao, Xiangrong Zhang, and Licheng Jiao. Classification of hyperspectral images based on multiclass spatial–spectral generative adversarial networks. IEEE Transactions on

[43] Geoscience and Remote Sensing, 57(8):5329–5343, 2019.

[44] Zhang, Q., Yuan, Q., Zeng, C., Li, X., & Wei, Y. (2018). Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(8), 4274-4288.

[45] Li, X., Shen, H., Zhang, L., Zhang, H., Yuan, Q., & Yang, G. (2014). Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(11), 7086-7098.

[46] Martinuzzi, S., Gould, W. A., & González, O. M. R. (2007). Creating cloud-free Landsat ETM+ data sets in tropical landscapes: cloud and cloud-shadow removal. *US Department of Agriculture, Forest Service, International Institute of Tropical Forestry. Gen. Tech. Rep. IITF-32.*, *32*.

[47] Eckardt, R., Berger, C., Thiel, C., & Schmullius, C. (2013). Removal of optically thick clouds from multispectral satellite images using multi-frequency SAR data. *Remote Sensing*, *5*(6), 2973-3006.

[48] Ji, S., Dai, P., Lu, M., & Zhang, Y. (2020). Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(1), 732-748.

[49] Shen, H., Wu, J., Cheng, Q., Aihemaiti, M., Zhang, C., & Li, Z. (2019). A spatiotemporal fusion based cloud removal method for remote sensing images with land cover changes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(3), 862-874.

[50] Viren Jain and Sebastian Seung (2008). Natural image denoising with convo-lutional networks. Advances in neural information processing systems, 21.

[51] Mao, X., Shen, C., & Yang, Y. B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Advances in neural information processing systems*, *29*.

[52] Li, Z., Shen, H., Cheng, Q., Li, W., & Zhang, L. (2019). Thick cloud removal in high-resolution satellite images using stepwise radiometric adjustment and residual correction. *Remote Sensing*, *11*(16), 1925.

[53] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2414–2423, 2016.

[54] Gidaris, S., Singh, P., & Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*.

[55] Singh, P., & Komodakis, N. (2018, July). Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial network. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1772-1775). IEEE.

[56] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

[57] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).

[58] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).

[59] Zhu, L., Chen, Y., Ghamisi, P., & Benediktsson, J. A. (2018). Generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(9), 5046-5063.

[60] Zhu, D., Xia, S., Zhao, J., Zhou, Y., Jian, M., Niu, Q., ... & Chen, Y. (2020). Diverse sample generation with multi-branch conditional generative adversarial network for remote sensing objects detection. *Neurocomputing*, *381*, 40-51.

[61] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139-144.

ABOUT THE AUTORS

Lawrence Egharevba is a PhD candidate at the Knight Foundation School of Computing and Information Sciences at Florida International University, Miami, USA. His research interests include satellite imagery, cloud detection and extraction, image processing, and machine learning.

M. Hadi Amini is an assistant professor at the Knight Foundation School of Computing and Information Sciences at Florida International University, Miami, USA.

Malek Adjouadi is a distinguished professor in the Department of Electrical and Computer Engineering at Florida International University, Miami, USA.

Naphtali Rishe is the inaugural Outstanding University Professor of Florida International University (FIU) and Director of the High Performance Database Research Center at the Knight Foundation School of Computing and Information Sciences at FIU.