IMPLEMENTATION OF DIFFUSION MODEL FOR CLOUD REMOVAL IN SATELLITE IMAGERY

Namdeo Baban Badhe
Research Scholar, Information Technology department FAMT, Mumbai University, Ratnagiri, India

Dr.Vinayak Ashok Bharadi
Professor, Information Technology department FAMT, Mumbai University, Ratnagiri, India

Dr. Nupur Giri
Professor, Computer Department VESIT, Mumbai University, Mumbai, India

Shashank Tolye
Assistant Professor, Information Technology department FAMT, Mumbai University, Ratnagiri, India

ABSTRACT

Remote sensing satellite images provide vital information for a plethora of applications including environmental monitoring, agriculture, urban planning, and resource management.However, thesatellite images often suffer from missing information due to unfavorable weather conditions like dead pixels, deadlines, heavy cloud cover and cloud shadows, rendering the data unusable. Hence, it becomes necessary to fill in the missing information by removing the dense clouds and cloud shadows in the image. The paper suggests utilizing Diffusion Models for satellite image reconstruction as a generative technique. The clouds from satellite images are removed by a mask and the removed regions are generated using unconditional Denoising Diffusion Probabilistic Model (DDPM) based inpainting approach.To compare the performancePeak signal to noise ratio (PSNR) and structural index similarity (SSIM) parameters have been used.Compared with other deep learning models on thin cloud, whereas the proposed model on thick cloud gives better results as PSNR(30.2501) and SSIM (0.9153).

***Keywords**: cloud removal, diffusion model, DDPM, cloud, image reconstruction, satellite imagery, image processing*

I. Introduction

The acquisition of data for monitoring in a range of disciplines, which includes the identification of structures, the mapping of crops, and the detection of changes in land cover, depends heavily on remote sensing. Satellite images are also essential for keeping an eye on the earth's surface. Since these images are crucial for tracking mineral resources, forests, floods, droughts, urban planning, land surface, sea temperature, and other phenomena, they also have been widely employed in a wide range of fields, including environmental research, national defence, security, and weather monitoring. Approximately 67% of the Earth's surface is covered by clouds, as reported by Moderate Resolution Imaging (MODIS). Therefore, cloud elimination methods are being contemplated to address any unwanted issues that may arise in optical remote sensing. Unfortunately, this remote sensing data acquired from the satellite sensors suffers from missing information due to faulty satellite sensors or bad environmental conditions. Problems occur like images with cloud cover, shadows, etc. And solving these problems often requires multitemporal data or other pairs of data.

Removing clouds from satellite images is a pervasive problem because clouds frequently obscure the surface and make it difficult to analyze satellite data. By identifying the cloud pixels in the image and substituting them out for clear pixels from other images that were either shot at a different time or from different sources, the cloud can be removed from the image. Both human and automatic techniques can be used to accomplish this problem. The majority of the time, the manual approaches entail manually modifying the cloud pixels as well as visual image inspection. On the other hand, automated techniques employ algorithms to recognise and eliminate the cloud pixels. The strategies used in automated cloud removal may be based on statistical or machine learning methods and may include thresholding, morphological procedures, or deep learning techniques. Depending on the quality and resolution of the satellite data, the complexity of the cloud structures, and the method used, the accuracy and efficacy of cloud removal can differ.

The aim of this work is to reconstruct the missing data in the satellite images due to cloud covering. The proposes work is not detecting the clouded regions but only reconstructing the missing areas when the clouded pixels are removed using masks. When viewing this issue more broadly, it may also be seen as a form of image inpainting where the clouded parts represent the areas that need to be eliminated. Many people have put out their own solutions for this issue in recent years, including deep learning and conventional image processing techniques. Based on extensive research and previous works, In this paper, Diffusion models approach is explored as this is relatively new and has become very popular in recent times as compared to other deep generative models.

II. Related work

There are two types of techniques for reconstructing data currently available: conventional image processing approaches and deep learning techniques. Traditional image processing methods mainly utilize the basic characteristics of the image, but the models designed for this method often have restricted capability. With the advancement in computing power, deep neural networks have made significant strides in computer vision activities, such as image restoration, image denoising and image super-resolution reconstruction. A few scholars have attempted to apply deep learning methods to resolve the issue.

Researchers in [1] built their model using two GANs. The first individual to recognize the intricate structures in SAR photos was trained using large datasets. The second method is employed to eliminate clouds while preserving the standard of areas that are cloud-free. They recommended dilated residual inception blocks (DRIBs) for generators based on dilated convolutions to improve receptive view and prevent the missing information in satellite images.

Paper [2] proposed a texture complexity-guided sample generation technique that can produce training examples on its own with a balanced distribution of difficulty. The researchers investigated an SPL approach that could train a comprehensive network for cloud elimination across a variety of complexities by intelligently sorting training examples from the simplest to the most challenging. The experimental results indicate that when the size of the cloud is excessively large, the suggested method generates exceedingly impractical content. This is problematic as the cloud-affected region is too extensive to provide enough data for cloud removal.

A novel additive imaging model-based framework for thin cloud removal is proposed for multispectral remote sensing pictures [3]. U-Net has been used to obtain precise data on thin clouds. Slope-Net, a brand-new CNN architecture, is also planned. They proposed a novel thin cloud simulation method for training U-Net and Slope-Net. Compared to other thin cloud removal approaches, the suggested technique will result in cumulative estimation inaccuracy. Because the two kinds of cloudy photos are not totally consistent with one another, the results of eliminating thin clouds from real cloudy photographs are worse than those produced by removing thin clouds from simulated cloudy images.

Paper [4] cloud removal process has two stages. Cloud segmentation, which entails directly extracting the thick clouds using U-Net and segmenting the clouds, is the preliminary stage. The second part of the process is image restoration, which involves eliminating the dense cloud and reconstructing the irregularly missing regions that correlate to it using a generative adversarial network (GAN). When a sizable portion of an image is missing, the suggested method cannot reconstruct it with high accuracy. Another limitation of the two-stage approach is that the generator could provide erroneous images to fill in the gaps when the ratio of thick clouds is large.

Deep learning-based cloud removal technique for a multitemporal ZY-3 satellite image is described in paper [5]. The CNN architectures are made to both identify and get rid of clouds at the same time. The proposed approach cannot be used in scenarios where the land cover may vary dramatically.

In summary, even though there have been few other deep learning solutions for this problem, they have yet to apply a diffusion model approach for this problem, and depending on how it is implemented it can even perform better than previous models.

A. *Diffusion Models*

Diffusion models are based on non-equilibrium thermodynamics and employ a series of random diffusion steps in a Markov chain to introduce noise to data. The objective is to train the model to comprehend how to invert the diffusion procedure and construct preferred data samples from the noise. The DDPM is a generative model that learns the distribution of images in a training set, similar to other generative models. The procedure for generating new images involves selecting a random noise vector and gradually reducing the noise to arrive at a high-quality output image. In summary, diffusion models are neural networks that are trained to predict slightly less noisy images from a noisy input. At inference, they can be used to iteratively transform a random noise to generate an image.

Generative models, such as GANs, VAEs, and flow-based models, have demonstrated significant success in producing high-quality samples, but each model has its own limitations. GANs are infamous for their potential for unstable training and limited diversity during generation due to their adversarial approach. VAEs use a surrogate loss function. Flow models require specialized architectures for reversible transformations. Diffusion models employ set methods during training and have a high-dimensional latent variable that matches the original data. There has been a surge in research into diffusion models in recent years, leading to them producing state-of-the-art image quality. Additionally, diffusion models don't require adversarial training like GANs. Although they may be more computationally demanding compared to other deep learning architectures, they outperform them in specific applications [8].

III. Methodology

This paper explores a new generative approach for removing clouds from satellite images and filling in the missing information through in painting. A growing alternative to generative modelling, Denoising Diffusion Probabilistic Models (DDPM), are used in the process [6]. Recent research has demonstrated that diffusion models can even surpass the most advanced GAN-based image generation technique [8]. By reversing the diffusion process the diffusion model is trained , beginning with a random noise sample and undergoing a specified number of iterations to produce the final image. DDPMs have shown they can produce a variety of excellent photographs despite being based on principled probabilistic modelling [6, 7, 8].

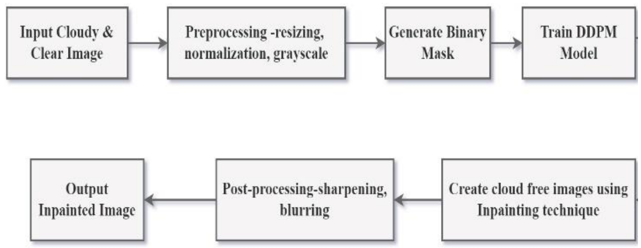

Fig.1.Overview of Proposed System

The Proposed method employs an unconditional image generation DDPM in the same way as paper [9]. Rather than acquiring a mask-based generative model,adaptive sampling procedure is used throughout the reverse diffusion iterations to condition the generation process.As a result, the model is not directly trained for the purpose of eliminating clouds. Nevertheless, with the alterations in denoising procedure is applied to address problem. But, for this still there is a need an unconditional image generation diffusion model which can generate satellite images from random noise. For this a pre-trained model is used to generate random bedroom images and fine-tune it on satellite dataset to generate more images. This has an important advantage that existing powerful DDPM has been trained prior and reduce actual training time as well as data required for training. An enhanced denoising method is implemented which re-samples iterations to improve the conditioning of the image [9]. This technique enables the network to effectively integrate the generated image data of the cloud-covered regions throughout the entire inference process, resulting in better conditioning based on the available satellite image data. Since the original DDPM network remains unaltered, the model produces high-quality output for the generated area.

IV. Implementation

A. Dataset Details

In this paper, the cloud removal dataset from the paper [10] is studied. It consists of two datasets: one that pairs a single cloudy image with a single clear image (singleImage, 97640 images), and the other that pairs three cloudy photos with a single clear image (multipleImage, 3130 images). The singleImage dataset is explored.The pictures have RGB channels with a size of 256x256px.

B. Experimental Setup

All of the trials were carried out using a Google Colab with 16 GB of RAM, GPU Tesla T4. Pytorch, Diffusers numpy Matplotlib, and the skimagetoolkit are used to achieve diverse features andgenerates cloud free images.

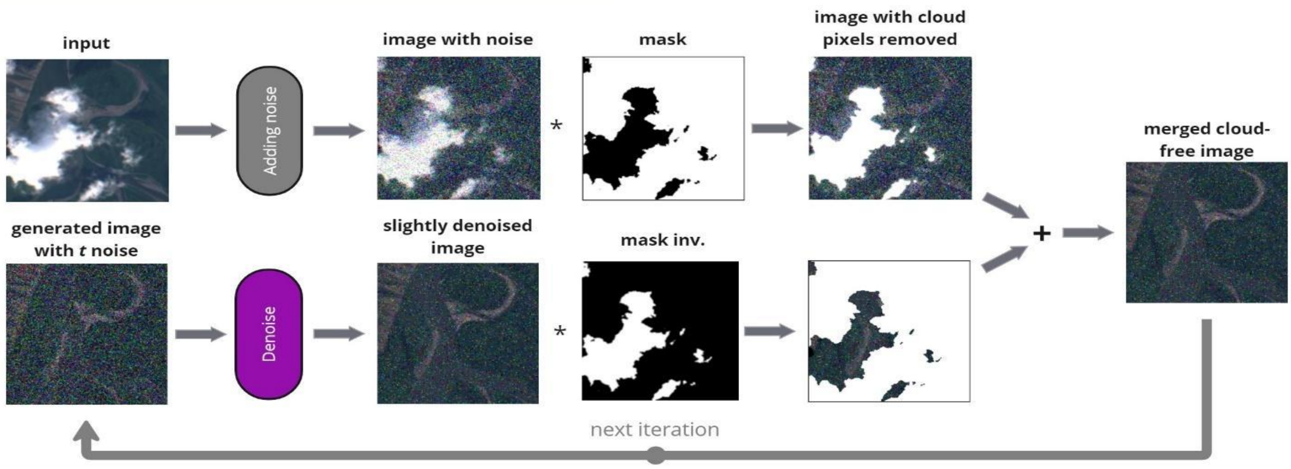

Fig. 2. Overview of Sampling process. Every few step, we merge the cloud-free region of the input and the inpainted part from the DDPM output (bottom) to create a cloud-free image.

C. Model

In this study, the pretrained diffusion model based on UNet architecture from [7] was trained on 256x256px LSUN Bedrooms class to generate unconditional images of bedroom. As only weights are updating during training, the architecture of the model remains the same.

D. Training

For the fine-tuning of the model, only clear satellite images without clouds are needed to generate clear satellite pixels without cloud for the inpainting part. The model takes a timestep *t* (random number during training) which determines the amount of noise in an image and a noisy image x_t (generated by adding some noise in the clear image x_0 based on *t*) then predicts the noise to be removed from the initial noisy image to give clear satellite image. The sampling part which does the image generation from predicted noise does not need to be trained and this part also does the inpainting of the clouded pixels.

E. Sampling

In the sampling part a mask of cloudy regions is used as a condition in the method to forecast missing pixels in the clouded portions of satellite images. As cloud segmentation is not the main aim of this paper,maskarecreated by simply doing the grayscale of the cloudy image and converting it to mask and inverting the mask to remove regions containing cloud pixels. This method does not give proper mask, but it is good enough to use in the sampling process to give a good result. To produce the removed regions, trained unconditional denoising diffusion probabilistic model is utilizedto make use of the context of the known region during the sampling phase.

As the DDPM is trained to generate images resembling satellite images, it inherently strives to create structures that are coherent and consistent. We use this DDPM property to harmonize the input of the model to generate the regions which better harmonize with the satellite image. The sampling process is illustrated in Figure 1. The cloudy regions are removed using the generated mask which is later merged with cloud-free regions generated from model. Then we diffuse the output x_{t-1} (merged cloud-free image) back to x_t (image with *t* noise) by sampling from the model. Although this operation scales back the output and adds noise, some information incorporated in the generated region $x_{t-1}$is still preserved in x_t. It leads to a new x_t which is both more harmonized with *x* (input image) and contains conditional information from it. This resampling also increases the runtime of the reverse diffusion but as a result the generated region matches the neighboring region and semantically correct.

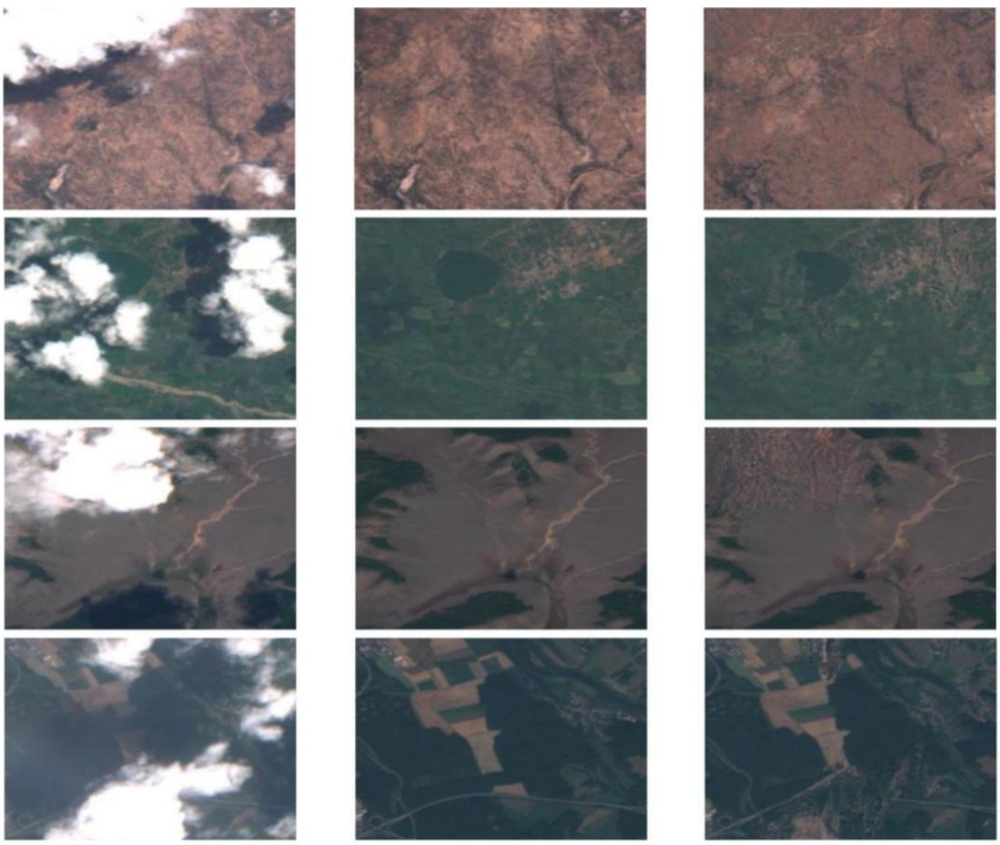

Fig. 2. Results of our cloud removal method. Columns from (left) to (right) show the original cloudy image, original cloud-free image and the cloud removed image from our method.

F. Metrics

The performance of the model quantitatively evaluated using structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) for pixel-level similarity between original clear image and cloud removed image. The PSNR ratio is utilized to evaluate the quality of an image by comparing the original and reconstructed version. A higher PSNR value indicates a better quality of the reconstructed image. On the other hand, the SSIM is a perceptual metric that measures the degradation in image quality by evaluating the perceptual difference between two similar images. Unlike PSNR, SSIM is based on visible structures in the image. The SSIM metric has a range from 0 to 1, where a value of 1 signifies that the reconstructed image matches the original image perfectly. These scores are then compared to some of the previous works using different approaches.

V. Result And Discussion

Figure 2 shows the results of the work.The images obtained can be seen to be almost same as the original cloud-free image. Although some of the generated regions seem to be different from the original clear image this is because no supplementary images were used to produce those regions unlike some of the previous works, so the region is randomly generated. This is not a problem if the cloudy regions are small but if there is cloud covering almost half of the region then the cloud-free image generated will be completely different from the ground truth image. The mask generation can also be improved to better detect the cloud pixels for proper removal of those regions.

Comparison is done with other deep learning-based models [2, 3, 4] in Table I based on PSNR and SSIM score. The PSNR of proposed method is 30.2501 and SSIM is 0.9153. When compared to other methods, score is not the best, but it is still very good and can be further improved. This may be due to not using a supplementary image to get information for generating the missing regions as we are generating thoserandomly. The other reason may also be because the cloudy and clear images have problems such as time difference, light changes, and atmospheric disturbances. This results in lower score for pixelwise metrics as the image generated from cloud removal also have the same problem as original cloudy image. The proposed method can also automatically deal with dead pixels and deadlines in satellite images because of how the diffusion model works. It is currently challenging to utilize the DDPM optimization process for real-time applications because it is much slower than other methods. However, diffusion models have become popular, and new techniques have been developed to improve its efficiency.

TABLE I. Quantitative Evaluation Comparison

Methods	PSNR	SSIM
U-Net and Slope-Net (Thin Cloud)	34.2562	0.9832
U-Net and Generative Adversarial Networks(Thin and Thick Cloud)	20.1	0.807
Texture complexity-guided self paced learning (SPL) framework (Thin Cloud)	30.4864	0.9351
unconditional Denoising Diffusion Probabilistic Model (Thick Cloud)	**30.2501**	**0.9153**

VI. Conclusion

In this paper, a novel approach is presented using diffusion models for reconstruction of missing data in satellite imagery due to cloud covering i.e., to remove clouded regions from satellite images. It is the most recent and popular approach towards image synthesis and never been used before for this problem. We can achieve same or better sample outputs compared to other methods. The proposed method only needs the cloudy image and mask of cloud regions to give cloud-free images. The results show that although the method performs well even for thick cloud removal, the output image is still a bit different from the ground truth image. The proposed method generates completely different region when the cloud-corrupted area is too large. Thus, for large cloud coverage there is a need to use another paired data which can provide sufficient information for the generation of the region.

VII. Future Scope

The entire process can be better optimized with all the new techniques coming out recently for diffusion models and to note that we have used simple DDPM instead of the newer models which produce even better results with less time and processing. The reconstruction model can be utilized for data pre-processing and augmentation for various satellite image applications. The diffusion models can be used in the missing data reconstruction of infrared and other types of satellite images, which is similar to proposed methodology. Data compression and image synthesis and reconstruction may potentially benefit from the diffusion model.

References

1. F. N. Darbaghshahi, M. R. Mohammadi and M. Soryani, "Cloud Removal in Remote Sensing Images Using Generative Adversarial Networks and SAR-to-Optical Image

Translation," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-9, 2022, Art no. 4105309, doi: 10.1109/TGRS.2021.3131035.
2. C. Tao, S. Fu, J. Qi and H. Li, "Thick Cloud Removal in Optical Remote Sensing Images Using a Texture Complexity Guided Self-Paced Learning Method," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-12, 2022, Art no. 5619612, doi: 10.1109/TGRS.2022.3157917.
3. Y. Zi, F. Xie, N. Zhang, Z. Jiang, W. Zhu and H. Zhang, "Thin Cloud Removal for Multispectral Remote Sensing Images Using Convolutional Neural Networks Combined With an Imaging Model," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 3811-3823, 2021, doi: 10.1109/JSTARS.2021.3068166.
4. J. Zheng, X. -Y. Liu and X. Wang, "Single Image Cloud Removal Using U-Net and Generative Adversarial Networks," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 8, pp. 6371-6385, Aug. 2021, doi: 10.1109/TGRS.2020.3027819.
5. Y. Chen, L. Tang, X. Yang, R. Fan, M. Bilal and Q. Li, "Thick Clouds Removal From Multitemporal ZY-3 Satellite Images Using Deep Learning," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 143-153, 2020, doi: 10.1109/JSTARS.2019.2954130.
6. Jonathan Ho, Ajay Jain and Pieter Abbeel, “Denoising Diffusion Probabilistic Models”, 34th Conference on Neural Information Processing Systems (NeurIPS), 2020.
7. Alex Nichol and Prafulla Dhariwa, “Improved Denoising Diffusion Probabilistic Models". arXiv:2102.09672, 2021.
8. Prafulla Dhariwal and Alex Nichol, “Diffusion Models Beat GANs on Image Synthesis”, June 2021.
9. Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte and Luc Van Gool, RePaint: Inpainting using Denoising Diffusion Probabilistic Models, arXiv.2201.09865, 2022.
10. https://dataverse.harvard.edu/dataset.xhtml;jsessionid=6b0716e40f73e6417d6359b5d0bc?persistentId=doi%3A10.7910%2FDVN%2FBSETKZ&version=&q=&fileTypeGroupFacet=&fileAccess=Public&fileTag=&fileSortField=&fileSortOrder=
11. Stefan Oehmcke, Tzu-Hsin Karen Chen, Alexander V. Prishchepov and Fabian Gieseke, Creating Cloud-Free Satellite Imagery from Image Time Series with Deep Learning, Proceedings of the 9th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, BIG SPATIAL 2020, November 2020.
12. Andrea Meranera, Patrick Ebela, Xiao Xiang Zhua and Michael Schmitta, Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion, ISPRS Journal of Photogrammetry and Remote Sensing, 18 May 2020.
13. Heng Pan, Cloud Removal for Remote Sensing Imagery via Spatial Attention Generative Adversarial Network, 28 November 2020.