Cloud Removal from Satellite Imagery

Ali Ashraf Mohammad, Cody Kurpanek, Ryan Chandler, Ivan Lopez, Prisha Bhattacharyya, Kyongsik Yun (Industry Contact) , Ronald Salloum (Faculty Contact)

This work is supported by the DS-PATH Summer Fellowship Program under the National Science Foundation Harnessing Data Revolution Data Science Corps Award #2123444, #2123271, #2123313.

Introduction & Motivation

As part of the NSF DS-PATH fellowship, we have been working on the challenging problem of cloud removal from satellite imagery. Cloud cover significantly hampers the utility of optical remote sensing imagery in Earth observation applications. We conducted a comprehensive survey of cloud removal techniques, sharing our insights during fellowship sessions. Many different approaches have been proposed to solve the cloud-removal problem. One of the state-of-the-art approaches is called DSen2-CR [1], which is a deep-learning-based approach that utilizes the super-resolution Deep Sentinel-2 (DSen2) ResNet architecture presented in [2].

As part of the fellowship, we have been focusing on replicating the results of the DSen2-CR paper. Our efforts entailed reformatting the dataset into the correct structure as well as making code changes in order to setup the environment and train the neural network. We also developed scripts to streamline image visualization. In addition, we have been exploring possible improvements to the proposed DSen2-CR method, and we plan to implement and evaluate these in the near future.

Literature Survey

Earth's surface is cloud-covered 67% of the time (55% over land), emphasizing the need for effective cloud removal algorithms.

The following is a brief summary of the literature survey presented in [1]. Traditional approaches can be grouped into three major categories: multispectral, multitemporal and inpainting techniques. Many methods are a hybrid combination of these categories.

Multispectral methods are applied in the case of haze and thin cirrus clouds, where optical signals are not completely blocked but experience partial wavelength-dependent absorption and reflection. Multispectral methods have the advantage of exploiting information from the original scene without requiring additional data, but are limited to filmy, semi-transparent clouds.

Multitemporal approaches restore cloudy scenes by integrating information from reference images acquired with clear sky conditions. Multitemporal methods are the most popular as they substitute corrupted pixels with real cloud-free observations. However, problems arise when reconstructing scenes with rapidly changing surface conditions because of the time difference between the scene to be reconstructed and the reference acquisition.

Inpainting approaches fill corrupted regions by exploiting surface information from clear parts of the same cloud-affected image. Such direct inpainting methods do not require additional images, but achieve good results only with small clouds.

Data-driven methods using deep learning have recently been receiving a significant amount of attention. One of the state-of-the-art deep-learning-based methods is called DSen2-CR [1]. A key innovation of DSen2-CR involves leveraging SAR-optical data fusion. This approach capitalizes on the complementary attributes of these two imaging systems, enabling the network to remove even optically thick clouds by reconstructing an optical representation of the underlying land surface structure. The model was trained on a diverse and large dataset to ensure its applicability across various scenarios.

DSen2-CR

Dataset

The dataset used in this study is called Sen12MS-CR and it was created specifically for training deep learning models for cloud removal. Each square represents a patch of size 256x256, which is part of a much larger scene captured by the respective satellite.

Quick Facts

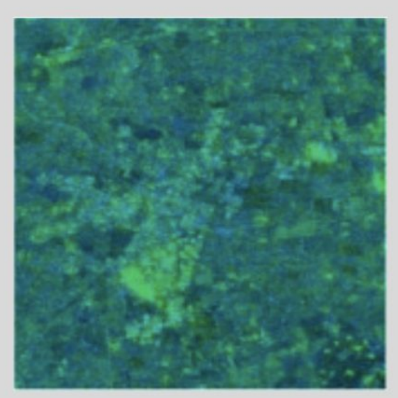
169 Scenes

157K Patch Triplets

620 Gigabytes

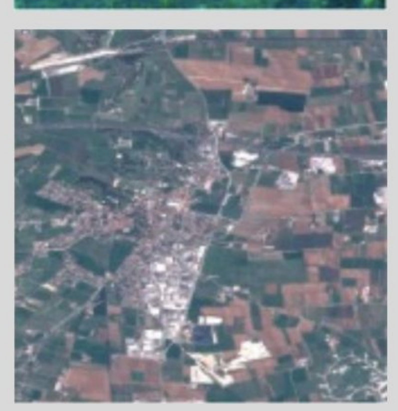
Sentinel-2 Cloudy Image

This represents the input cloudy image, obtained by the Sentinel-2 satellite. Sentinel-2 provides high-resolution optical imagery with a wide coverage of the Earth's surface. This is the input image that we would like to perform cloud removal on.



Sentinel-1 SAR Image

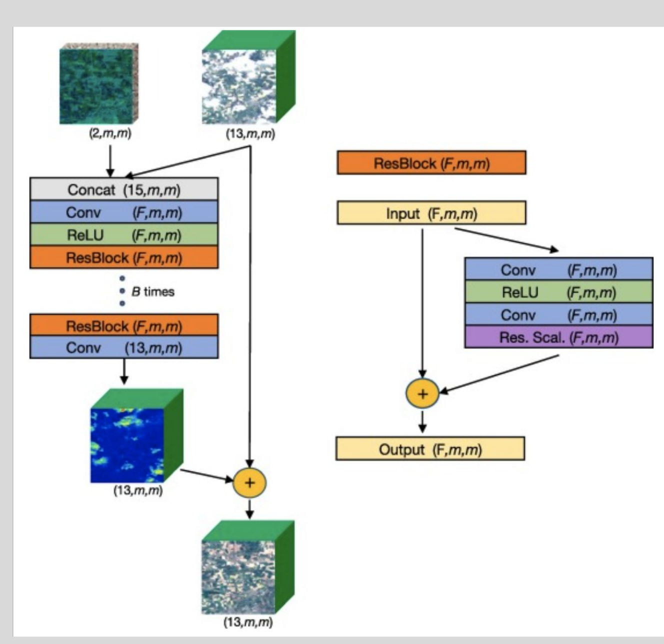
Obtained from the Synthetic Aperture Radar (SAR) satellite system. SAR sensors transmit microwave signals and measure the signals reflected back from the Earth's surface. This technology allows to acquire high-resolution radar images that bypass clouds. Complements Sentinel-2 data by providing further information about the surface characteristics of the target scene.



Sentinel-2 Cloud-Free Image

Used to train DSen2-CR Model by providing a "target" image without clouds. The output image will be compared with the target Sentinel-2 cloud-free image in order to compute the loss.

Proposed Architecture



The DSen2-CR model based on the super-resolution Deep Sentinel-2 (DSen2) ResNet architecture [2].

Experimental Results

Each predicted image can be broken up into two distinct parts: reconstruction areas and reproduction areas. Reconstruction areas are the cloudy parts of the image where the model should have reconstructed the surface below the clouds. Reproduction areas refer to non-cloudy areas where the model should have reproduced the surface. DSen2-CR utilizes 2 loss functions: $\mathcal{L}_T$ (classic target loss) and $\mathcal{L}_{CARL}$ (Cloud-Adaptive Regularized Loss). The two models were compared using MAE, RMSE, and PSNR, popular pixel-by-pixel metrics that measure the differences between predicted and target images.

Test results on pixel-wise metrics

	MAE(p_{TOA})			RMSE (p_{TOA})	PSNR (dB)
Method	Target	Reprod	Recon	Target	Target
DSen2-CR on $\mathcal{L}_{CARL}$	0.0290	0.0204	**0.0266**	0.0366	28.7
DSen2-CR on $\mathcal{L}_T$	**0.0270**	0.0398	**0.0266**	**0.0343**	**29.3**
DSen2-CR on $\mathcal{L}_{CARL}$ w/o SAR	0.0306	**0.0188**	0.0282	0.0387	27.6
DSen2-CR on $\mathcal{L}_T$ w/o SAR	0.0284	0.0389	0.0281	0.0361	28.8
pix2pix	0.0292	0.0210	0.0274	0.0424	28.2

Replicating DSen2-CR

We have been working on replicating the DSen2-CR model and the reported results. The following are some challenges we faced.

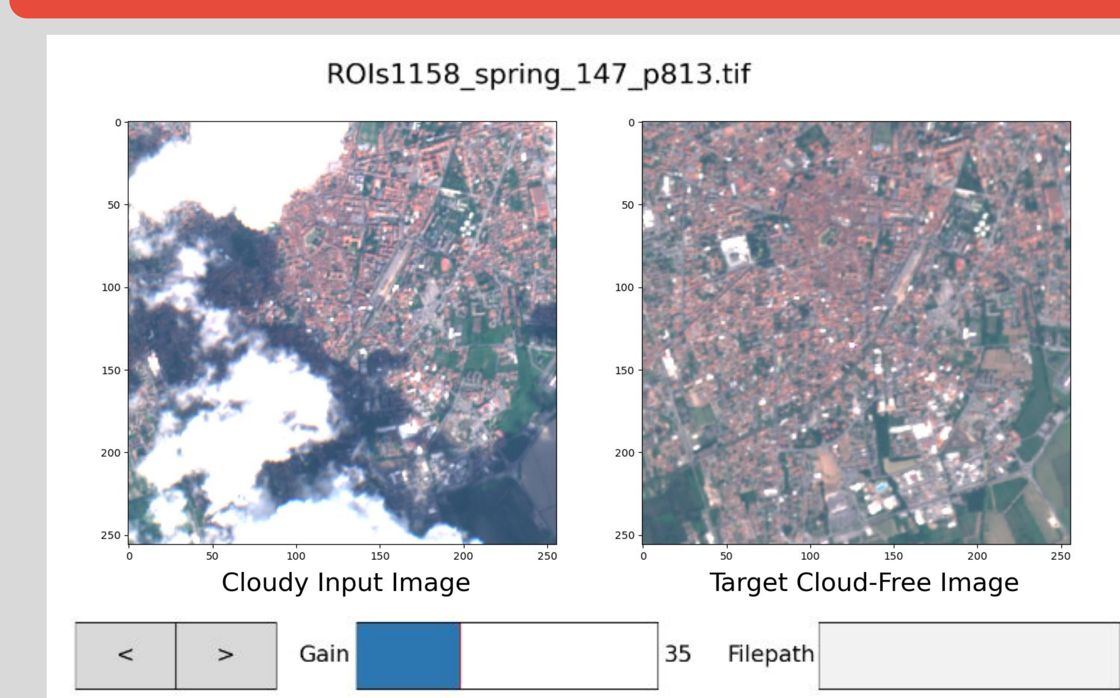
Handling and training a large dataset: The dataset is very large (~620 GB) and has many small files. To achieve large training speedup, we have obtained access to 4 Nvidia GPUs.

Code changes: We made the following code changes in order to successfully train the model. We modified the import statements to correctly import required packages, updated code to fix deprecated functionality, changed the data loader's delimiter to reflect our input data CSV file, and tuned training parameters.

Dataset reformatting: In order to train the DSen2-CR model on the Sen12MS-CR dataset, we reformatted the dataset into the correct directory and file structure.

Reproducibility: It is important for results to be reproducible so that others can continue to make use of and improve on past work. In our GitHub repository [3], we document the changes we have made and provide our exact conda environment for reproducibility. We also include scripts for extracting and reformatting the dataset.

Data Visualization Tool



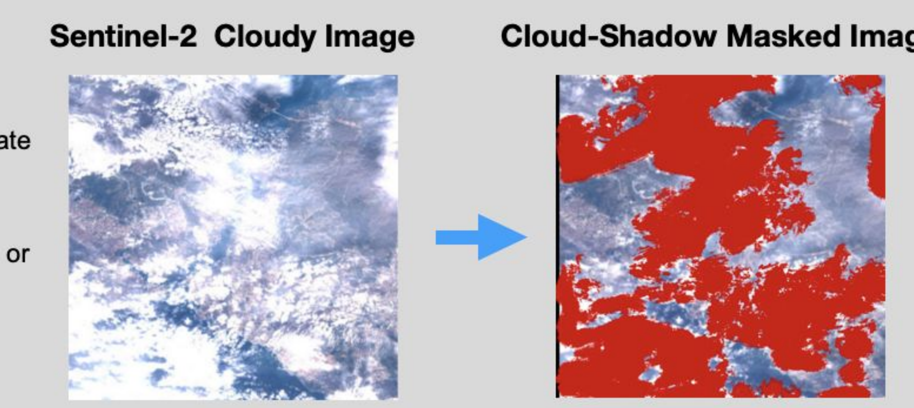
We created an application that visualizes Sentinel-2 cloudy and Sentinel-2 cloud-free images from our dataset by extracting the correct bands and transforming them into RGB images.

Future Work

Past literature shows the benefits of inpainting approaches, which consider global context of the cloudfree areas within an image when reconstructing cloudy areas. We propose a method for cloud removal that takes advantage of both inpainting and SAR data fusion. In particular, we utilize Stable Diffusion [4], a state-of-the-art image model, for inpainting.

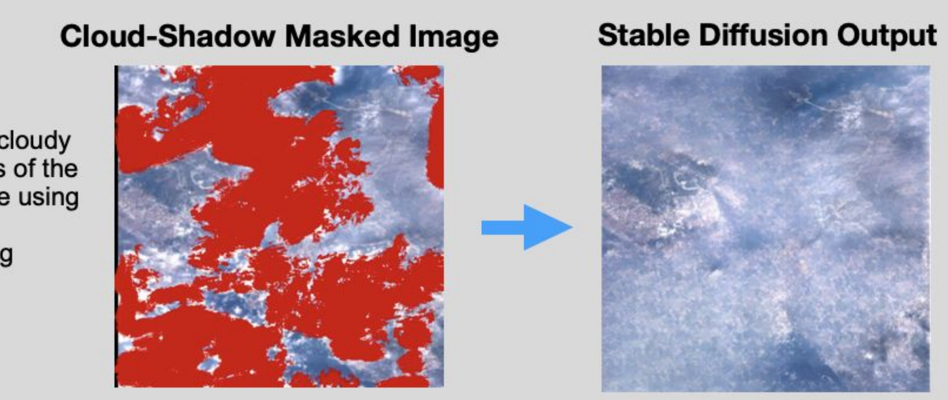
Step #1: Generate Cloud Mask

In this step, we generate the cloud-shadow masked image, which highlights in red the pixels that are clouds, or shadows of clouds.



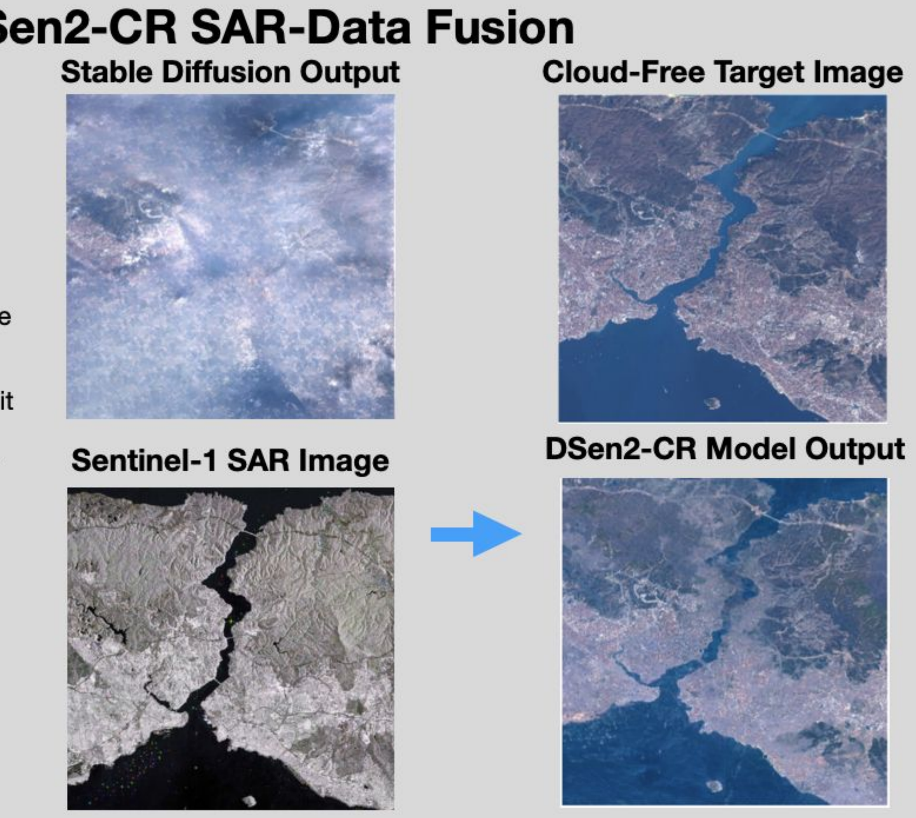
Step #2: Perform Stable Diffusion

Reconstructs the cloudy and shadow areas of the input cloudy image using a Stable Diffusion *Img2Img* inpainting model.



Step #3: DSen2-CR SAR-Data Fusion

We combine the Stable Diffusion output with the patch's respective SAR data and we use it to train the DSen2-CR model with the Cloud-Adaptive Regularized Loss function



References

[1] Meraner, Andrea, et al. "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 (2020): 333-346.

[2] Lanaras, Charis, et al. "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network." ISPRS Journal of Photogrammetry and Remote Sensing 146 (2018): 305-319.

[3] https://github.com/CodyKurpanek/dSen2CR_DSP.

[4] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.