Improving Flood Insights: Diffusion-based SAR to EO Image Translation

Minseok Seo, Youngtack Oh, Doyi Kim, Dongmin Kang, Yeji Choi*
SI Analytics
70, Yuseong-daero 1689beon-gil, Yuseong-gu, Daejeon, Republic of Korea
{minseok.seo, ytoh96, doyikim, dmkang, yejichoi}@si-analytics.ai

Abstract

Driven by the climate crisis, the frequency and intensity of flood events are on the rise. Electro-optical (EO) satellite imagery is commonly used for rapid disaster response. However, its utility in flood situations is limited by cloud cover and during nighttime. An alternative method for flood detection involves using Synthetic Aperture Radar (SAR) data. Despite SAR's advantages over EO in these situations, it has a significant drawback: human analysts often struggle to interpret SAR data. This paper proposes a novel framework, Diffusion-based SAR-to-EO Image Translation (DSE). The DSE framework converts SAR images into EO-like imagery, thereby enhancing their interpretability for human analysis. Experimental results on the Sen1Floods11 and SEN12-FLOOD datasets confirm that the DSE framework provides enhanced visual information and improves performance in all flood segmentation tests.

1 Introduction

Under global warming conditions, the intensity and frequency of heavy precipitation and associated flooding events have increased in most regions [1, 6]. It is important to decide where to deploy the necessary resources to mitigate the damage and quickly recover. Here, allocating needed resources relies on precise information collected manually and remotely.

Herein, Electro-Optical (EO) satellites have provided a broad and comprehensive view of the disaster-stricken region, surpassing the scope of on-site surveys by humans. Satellite-based indexes, such as the Normalized Difference Water Index (NDWI) ([5]), monitor the water bodies and delineate the flood extent. However, in EO observations, cloud cover obstructs the view of the region. It is because the sensors of EO satellites cannot penetrate clouds, but most flood events are due to heavy rains accompanied by thick clouds.

Thus, the EO satellites are not suitable for accurate flood monitoring. As an alternative, approaches that employ Synthetic Aperture Radar (SAR) observation have been proposed ([7]). SAR imaging holds the advantage of being unaffected by cloud cover and nighttime conditions, providing a flexible practice for disaster. However, SAR images often contain impediments to interpretation, such as speckle noise. Hence, although a model appropriately estimates the inundated regions, people cannot easily be reliable without EO imagery. To solve these problems , we propose the Diffusion-based SAR-to-EO Image Translation (DSE) framework, a novel method to generate synthetic EO (SynEO) images from SAR inputs for help flood monitoring.

*corresponding author

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

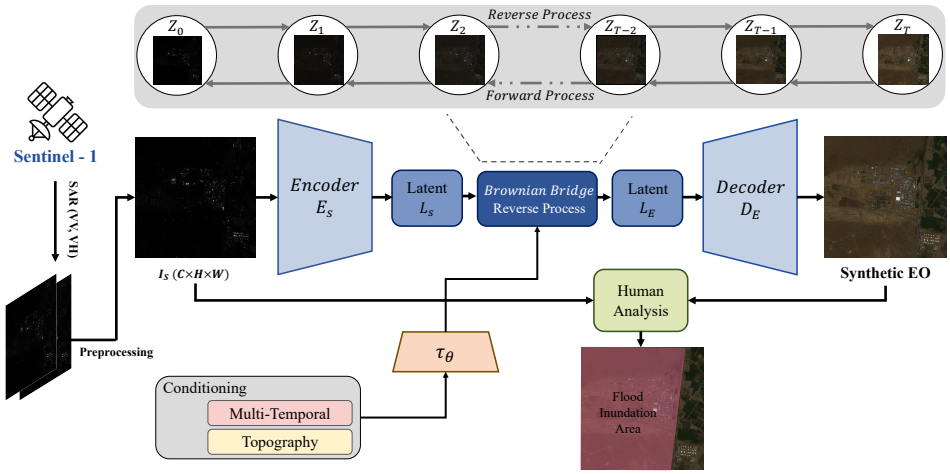


Figure 1: Overview of the DSE framework. The DSE framework takes in the SAR image, applies a self-supervised denoising method, and then carries out diffusion-based SAR2EO image translation. Subsequently, the generated EO and corresponding SAR images are reviewed by the analysts for the purpose of flood mapping.

2 Method

In this section, BBDM [4], which is the basis of the DSE framework, is first briefly described, and then the preprocessing, model, and function are sequentially explained in detail. Please note that the reverse process of DSE aligns strictly with that of BBDM, so we will not delve into the details of the reverse process in this paper.

2.1 Brownian Bridge Diffusion Model (BBDM)

Given two datasets, X_A and X_B, originating from domains A and B respectively, the purpose of image-to-image translation is to ascertain a function that establishes a mapping from domain A to domain B. While numerous image-to-image translation methods based on conditional diffusion models have been proposed, they are not intuitively suited for the task as its translation process seamlessly converts a noise back into an image, not image to image. Moreover they does not have a clear theoretical guarantee because of their complex conditioning algorithm based on attention mechanism. BBDM, however, provides a method for image-to-image translation grounded in the Brownian diffusion process which avoid leveraging complex conditioning algorithm.

Reffering to the original BBDM, we also conduct the process in the latent space of VQGAN[2]. Following the convention, let (x, y) denote the paired training data from X_A and X_B, each. For simplicity, we use x and y to denote the corresponding latent features ($x := L_S(x), y := L_E(y)$). The forward diffusion process of Brownian Bridge is defined as:

$$q_{BB}(x_t|x_0, y) = \mathcal{N}(x_t; (1 - m_t)x_0 + m_t y, \delta_t I), \tag{1}$$

$$x_0 = x, m_t = \frac{t}{T} \tag{2}$$

where T is the total steps of the diffusion process, δ_t is the variance.

The forward diffusion of the Brownian Bridge process provides only the marginal distribution at each time step t, as shown by the transition probability in (1). However, for training and inference, it is essential to deduce the forward transition probability $q_{\text{BB}}(x_t|x_{t-1}, y)$. In the original BBDM, given an initial state x_0 and a destination state y, the intermediate state x_t can be computed in discrete form as follows:

$$x_t = (1 - m_t)x_0 + m_t y + \sqrt{\delta_t}\epsilon_t, \tag{3}$$

$$x_{t-1} = (1 - m_{t-1})x_0 + m_t y + \sqrt{\delta_{t-1}}\epsilon_{t-1} \tag{4}$$

here, ϵ_t, $\epsilon_{t-1} \sim \mathcal{N}(0, I)$.

However, in the SAR2EO task, diversity isn't as crucial as in the original BBDM. Rather, the emphasis is on ***prediction*** that closely aligns with the actual outcome. For instance, in the SAR2EO task, the goal is to generate images that are akin to the actual EO image or resemble the distribution of training EO images, instead of producing a variety of colors and textures like ***generation***. Consequently, we sample ϵ from the target distribution rather than the standard normal distribution $\mathcal{N}(0, I)$. Moreover, in the reverse process, we set the size of ϵ to $\epsilon \times 0.1$. This adjustment brings the SAR2EO task closer to prediction.

2.2 Pre-processing

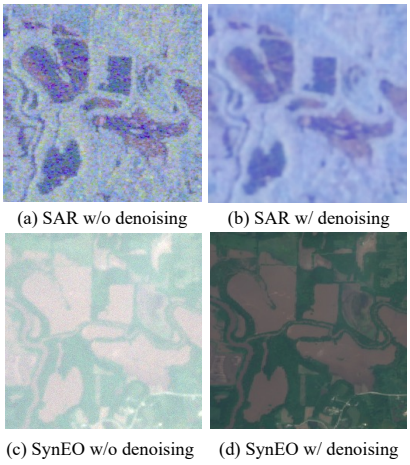
SAR images inherently display speckles due to their generation mechanism. This can be described by the multiplicative speckle noise model:

$$Y = XN, \tag{5}$$

where Y denotes the observed SAR intensity, X represents the clean image, and N is the speckle noise. Typically, N follows a Gamma distribution with mean 1 and variance $1/L$, where L is the number of 'looks' in the multi-look process:

$$p(N) = \frac{1}{\Gamma(N)} L^N N^{L-1} e^{-LN}, \tag{6}$$

The DSE framework uses a diffusion-based translation model. While it can predict SAR images and added noise, distinguishing inherent speckle noise (as in (6)) from added noise is challenging, leading to potential noise residues as seen in Fig. 2-(b).



(a) SAR w/o denoising (b) SAR w/ denoising

(c) SynEO w/o denoising (d) SynEO w/ denoising

Figure 2: Comparison of original and denoised SAR images using a self-supervised method.

To combat this, we pre-denoise SAR images using a blind-spot based self-supervised method before their use in translation. While conventional blind-spot methods assume noise independence from the clean image [3], this is not true for SAR images. Therefore, we adopted the method from [9], a variant leveraging diverse kernels. The denoising results, along with SAR2EO generation using the denoised image within the DSE framework, are presented in Fig. 2-(b,d).

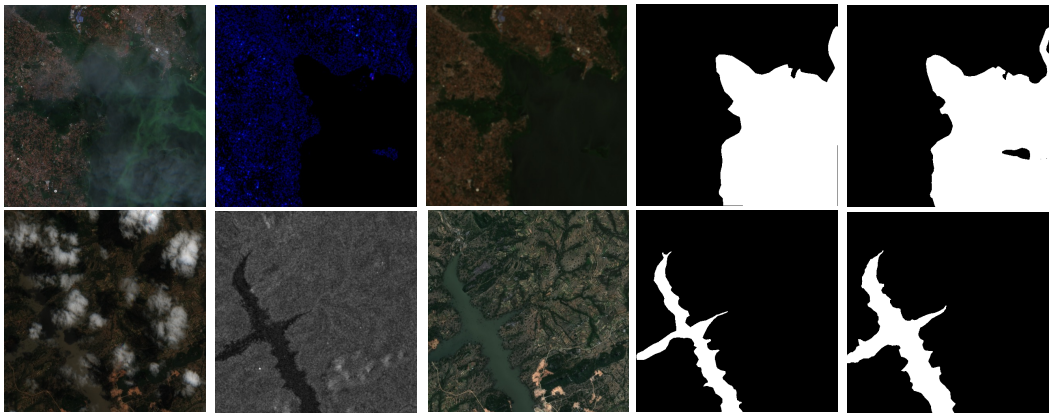
3 Results

3.1 Quantitative results

Table 1: Comparison of the DSE framework results with the commonly employed SAR2EO baselines, pix2pixHD, using a test set derived from the SEN12-FLOOD dataset, where missing or cloud-affected data points have been excluded.

Method	PSNR	SSIM	LPIPS
Pix2PixHD [8]	31.09	0.81	0.116
BBDM [4]	29.20	0.74	0.124
DSE	32.43	0.84	0.109
DSE+multi-temporal	**34.94**	**0.87**	**0.082**

Image-to-Image Translation Table 1 provides the experimental results from the SEN12-FLOOD dataset, with cloud and missing data excluded. The results show the proficiency of the DSE framework in generating SynEO images. Note that the temporal alignment between the multi-temporal SAR and EO datasets is imprecise. Thus, we have matched the EO data from the nearest date to the reference SAR imagery.



(a) EO (b) SAR (c) SynEO (d) Expert (e) Flood Label

Figure 3: omparison of flood detection by SAR experts using pairs of EO and SAR images versus pairs of SynEO and SAR images. Please note that we simultaneously provided SAR experts with three-channel SAR images (VV, VH, (VV+VH)/2) and one-channel images. The SAR images included in the figure have been identified by SAR experts as being more conducive to their analyses.)

Among the compared baselines, the DSE framework demonstrated better performance in terms of PSNR, SSIM, and LPIPS. The most favorable performance was recorded when inputting multi-temporal SAR imageries (it means spatially aligned and temporally random).

Flood mapping and Human Analysis The accuracy of flood area mapping, as interpreted by SAR experts, is represented by scores of **0.5532** for (EO, SAR) image pairs and 0.5464 for (SynEO, SAR) pairs—a marginal decrease of $\downarrow 0.0068$. This indicates that SynEO images can effectively serve in roles similar to EO images, especially when the availability of EO images is constrained. The Intersection over Union (IoU) for the (SynEO, SAR) pairs being only 0.0068 less than that of (EO, SAR) underscores SynEO data's potential as an alternative, particularly when cloud cover and temporal alignment pose challenges for EO imagery. However, potential errors must be considered as our SynEO is derived from SAR imagery. While we don't use SynEO data exclusively in our application, it is recommended as a supplementary resource for SAR interpretation.

As well as quantitative results, Figure3 provides qualitative results of flood area mapping obtained from two data pairs by SAR experts. The use of EO imagery is challenging in some weather conditions. Also, it is complex and imprecise to classify flood regions from raw SAR imagery due to its inherent nature. Under these conditions, SynEO can support SAR experts in mapping flood areas by supplementing the SAR information without EO imagery. Fig.3-(d) shows the result of an expert mapping regions similar to labels, with reference to SynEO.

4 Conclusion

In this paper, we introduced a Diffusion-based SAR-to-EO Image Translation (DSE) framework to improve human analysis of flood area mapping. We designed to address two inevitable issues from satellite observation. First, exploiting EO imagery for flood mapping frequently suffers from impracticability in cloud cover or nighttime. Second, the deep learning-based SAR flood detection model demands a substantial volume of labeled flood datasets. DSE framework exploits the advantages of EO and SAR together by a SAR-to-EO translation scheme and effectively assists analysts. We validated the Sen1Floods11 and SEN12-FLOOD datasets and obtained significant results quantitatively and qualitatively. We hope that our research will be widely used in disaster response tasks.

References

[1] Richard P Allan, Ed Hawkins, Nicolas Bellouin, and Bill Collins. Ipcc, 2021: summary for policymakers. 2021.

[2] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12873–12883, 2021.

[3] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018.

[4] Bo Li, Kaitao Xue, Bin Liu, and Yunyu Lai. Bbdm: Image-to-image translation with brownian bridge diffusion models. 2022.

[5] Stuart K McFeeters. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International journal of remote sensing*, 17(7):1425–1432, 1996.

[6] Melissa M Rohde. Floods and droughts are intensifying globally. *Nature Water*, 1(3):226–227, 2023.

[7] Cheryl WJ Tay, Sang-Ho Yun, Shi Tong Chin, Alok Bhardwaj, Jungkyo Jung, and Emma M Hill. Rapid flood and damage mapping using synthetic aperture radar in response to typhoon hagibis, japan. *Scientific data*, 7(1):100, 2020.

[8] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018.

[9] Dan Zhang, Fangfang Zhou, Yuwen Jiang, and Zhengming Fu. Mm-bsn: Self-supervised image denoising for real-world with multi-mask based on blind-spot network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4188–4197, 2023.