SAR-CONDITIONED FLOW MATCHING FOR SAR-TO-OPTICAL IMAGE TRANSLATION

Jeongwon Ryu[1*] **Youngtack Oh**[1*] **Minseok Seo**[2†]

[1]SI Analytics
[2]Korea Advanced Institute of Science and Technology (KAIST)

{rjw0926,ytoh96}@si-analytics.ai
minseok.seo@kaist.ac.kr

ABSTRACT

Synthetic Aperture Radar (SAR) imagery enables reliable observation regardless of weather conditions or acquisition time; however, its interpretation is challenging due to speckle noise and complex scattering characteristics that obscure spatial structures. To alleviate this limitation, SAR-to-Optical image translation has been proposed to translate SAR imagery into optical-like images. Existing SAR-to-optical image translation models can struggle to preserve structural consistency, such as object layouts and spatial relationships, as they primarily optimize for distribution matching rather than structure-level fidelity. In this paper, we formulate SAR-to-optical image translation from the perspective of structural consistency, where preserving spatial correspondence between input and output is essential. We introduce a flow-based framework that learns SAR-conditioned transformations defined by ordinary differential equations (ODE), modeling translation as a trajectory that continuously deforms SAR-aligned latent representations into their optical counterparts. By explicitly tracking how each SAR structure evolves along this trajectory, the framework naturally preserves geometric layouts and spatial relationships encoded in SAR imagery. Quantitative and qualitative results demonstrate improved image quality over existing image translation baselines, while preserving structural properties inherent in SAR images.

1 INTRODUCTION

Synthetic Aperture Radar (SAR) imagery is widely used in various remote sensing applications due to its ability to provide reliable surface observations regardless of weather conditions and day–night cycles (Zhou et al., 2024; Zhang & Xia, 2014). However, SAR images are inherently difficult for humans to interpret compared to optical imagery, owing to speckle noise and their high sensitivity to terrain geometry (Singh et al., 2021; Molini et al., 2021). To improve the interpretability of SAR imagery, SAR-to-Optical (S2O) image translation has been proposed as an effective means for facilitating the understanding and analysis of SAR data (Yang et al., 2022; Bai et al., 2023).

In particular, S2O translation plays a critical role in scenarios where optical imagery is missing due to cloud cover, as it enables the generation of interpretable optical images from SAR observations. Satellite observations over the same region are conducted at fixed revisit cycles. For optical imagery, valid observations may not be available when cloud cover or adverse weather conditions occur at the acquisition time. In contrast, SAR imagery can provide consistent observations at each revisit. As a result, S2O translation leveraging SAR–optical multimodal data enables higher temporal resolution than relying on optical imagery alone. This is particularly important for time-sensitive tasks such as disaster monitoring and environmental surveillance, where rapid situational awareness is required under cloudy or adverse conditions (Ryu et al.; Wang et al., 2025).

Early S2O translation methods primarily relied on GAN-based frameworks, including paired and unpaired models such as Pix2Pix (Isola et al., 2017) and CycleGAN (Zhu et al., 2017), which bridge

[*]Equal contribution
[†]Corresponding author

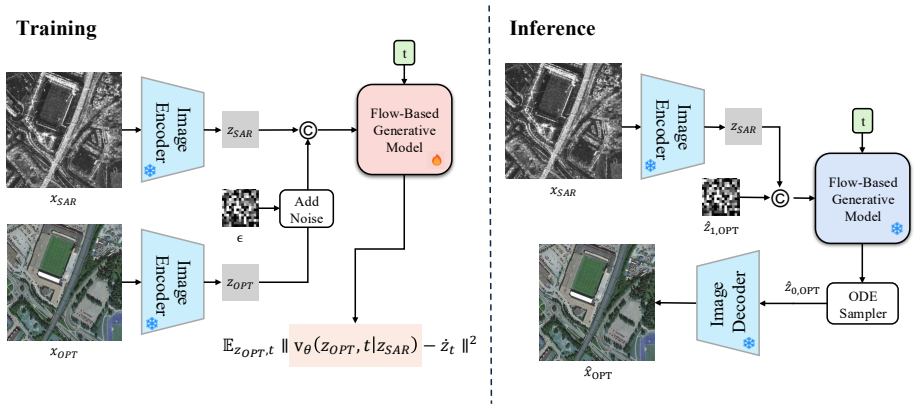

Figure 1: Overview of the proposed SAR-to-optical image translation framework.

SAR and optical domains through adversarial, cycle-consistency, or contrastive objectives (Park et al., 2020). More recently, diffusion-based methods have gained attention by leveraging denoising diffusion probabilistic models (DDPM) (Brune & Ban, 2024; Qin et al., 2024) and latent diffusion models (LDM) (Guo et al., 2024; Seo et al., 2025; Kim & Chung, 2025) to improve translation quality. Building on this trend, CycleGAN-Turbo (Parmar et al., 2024) shows the effectiveness of pretrained text-to-image latent diffusion models for image-to-image translation, a paradigm later extended to S2O translation by Do et al. (2024). In parallel, flow-based generative models have emerged as a competitive class of generative models across a wide range of applications (Lipman et al., 2023; Albergo et al., 2025).

In contrast to diffusion-based models, flow-based generative models learn continuous transformation trajectories between distributions through deterministic flows. This deterministic formulation is particularly well suited for SAR-to-optical image translation, where preserving geometric structure and maintaining consistent spatial correspondence between the input SAR image and the generated optical image are essential. From this perspective, SAR-to-optical image translation can be viewed as a deterministic, SAR-conditioned transformation process that preserves structural consistency as an inherent property of the mapping. Indeed, similar deterministic formulations have been successfully adopted in image and video editing and synthesis tasks, where maintaining semantic consistency across transformations is critical (Song et al., 2020; Grathwohl et al., 2019; Chen et al., 2025; Zhao et al., 2024).

In this paper, we apply flow-based generative modeling to the SAR-to-Optical image translation problem and present a robust framework that learns consistent mappings from the SAR domain to the optical domain while preserving the structural characteristics of the input SAR images. The proposed method learns continuous transformation trajectories conditioned on SAR latent representations, enabling the generation of optical images that maintain structural consistency between inputs and outputs.

2 Methodology

2.1 Preliminaries

In this work, we formulate SAR-to-optical image translation as a flow-based generative modeling problem. As a foundation for the proposed SAR-conditioned model, we briefly review the flow matching framework introduced in Scalable Interpolant Transformer (SiT) (Ma et al., 2024).

Following the latent diffusion paradigm, let $z_0 \sim p(z)$ denote the latent representation of an optical image and $\epsilon \sim \mathcal{N}(0, I)$ a standard Gaussian noise variable. A continuous latent interpolant between the two distributions is defined for $t \in [0, 1]$ as

$$z_t = \alpha(t)\, z_0 + \sigma(t)\, \epsilon, \tag{1}$$

where $\alpha(t)$ and $\sigma(t)$ are monotonic functions satisfying $\alpha(0) = 1$, $\alpha(1) = 0$ and $\sigma(0) = 0$, $\sigma(1) = 1$, respectively.

The latent process defined by the interpolant can be expressed as a probability flow ordinary differential equation (ODE):

$$\dot{z}_t = \frac{dz_t}{dt} = v(z_t, t), \tag{2}$$

where the velocity function $v(z_t, t)$ represents the instantaneous direction of the latent trajectory. Following the SiT formulation, this velocity is defined as the conditional expectation of the time derivative:

$$v(z_t, t) = \mathbb{E}[\dot{z}_t \mid z_t = z]\,. \tag{3}$$

To approximate this velocity field, we introduce a neural network parameterized by θ, denoted as $v_\theta(z_t, t)$, and train it using the flow matching objective:

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{z_t,\,t}\left[\left\|v_\theta(z_t, t) - \dot{z}_t\right\|^2\right]. \tag{4}$$

After training, the learned velocity field defines a deterministic probability flow ODE, which is integrated backward from the terminal noise state to obtain a latent sample $\hat{z}_0$ from the data distribution.

2.2 SAR-Conditioned Flow-based SAR-to-Optical image Translation

In this section, we describe a *SAR-conditioned flow-based SAR-to-optical image translation* method, which generates an optical image conditioned on a SAR observation. Both SAR and optical images are mapped into a common latent space, and the optical latent generation is guided by the SAR latent. An overview of the proposed framework is shown in Figure 1.

Latent Representation Following recent studies (Do et al., 2024; Bellier & Audebert, 2025), the input SAR image x_{SAR} and its paired optical image x_{OPT} are encoded into latent representations using a pretrained KL-VAE encoder $E(\cdot)$ Rombach et al. (2022):

$$z_{\mathrm{SAR}} = E(x_{\mathrm{SAR}}), \quad z_{\mathrm{OPT}} = E(x_{\mathrm{OPT}}). \tag{5}$$

After generation, the reconstructed optical latent is decoded back to the pixel space using the same KL-VAE decoder $D(\cdot)$:

$$\hat{x}_{\mathrm{OPT}} = D(\hat{z}_{\mathrm{OPT}}), \quad \text{where } \hat{z}_{\mathrm{OPT}} \equiv \hat{z}_{0,\mathrm{OPT}}. \tag{6}$$

By employing a single pretrained KL-VAE for both SAR and optical images, the two modalities are mapped into a common latent space, which facilitates conditional generation of optical imagery from SAR observations.

SAR-Conditioned Latent Flow Matching We formulate SAR-conditioned latent generation as a deterministic probability flow that transforms an initial noise latent into an optical latent. Let $\hat{z}_{1,\mathrm{OPT}} \sim \mathcal{N}(0, I)$ denote the terminal (noise) latent, and let $\hat{z}_{0,\mathrm{OPT}}$ denote the recovered optical latent at $t = 0$. We write the intermediate optical latent along the trajectory as $z_{t,\mathrm{OPT}}$ for $t \in [0, 1]$.

Extending the latent flow defined in Section 3.1 to a conditional form, we define the SAR-conditioned probability flow ODE:

$$\dot{z}_{t,\mathrm{OPT}} = v_\theta(z_{t,\mathrm{OPT}}, t \mid z_{\mathrm{SAR}})\,, \tag{7}$$

where $v_\theta(\cdot)$ is a neural velocity field that takes as input the time-dependent optical latent $z_{t,\mathrm{OPT}}$ and the SAR latent z_{SAR}, and predicts the instantaneous velocity along the latent trajectory.

Starting from the noise state $\hat{z}_{1,\mathrm{OPT}}$, we integrate Eq. equation 7 backward from $t = 1$ to $t = 0$ to obtain the recovered optical latent $\hat{z}_{0,\mathrm{OPT}}$ (i.e., $\hat{z}_{\mathrm{OPT}}$), which is then decoded to yield the final optical image $\hat{x}_{\mathrm{OPT}}$. This SAR-conditioned latent flow enables optical image generation that is consistent with the given SAR observation while preserving structural cues encoded in the SAR modality.

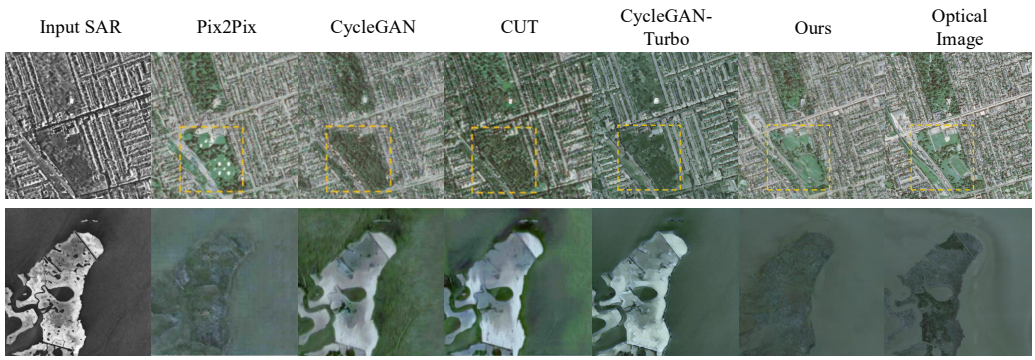

Figure 2: Qualitative comparison of SAR-to-Optical image translation results on the SAR2Opt dataset.

Table 1: Quantitative results on the SAR2OPT dataset.

Method	PSNR↑	SSIM↑	LPIPS↓	FID↓
Pix2Pix (Isola et al., 2017)	<u>15.4220</u>	0.2185	<u>0.5408</u>	196.87
CycleGAN (Zhu et al., 2017)	13.1573	0.1554	0.6172	195.92
CUT (Park et al., 2020)	13.2900	0.1239	0.6023	193.54
CycleGAN-Turbo (Parmar et al., 2024)	13.1577	<u>0.2386</u>	0.6609	<u>171.21</u>
Latent Diffusion (Rombach et al., 2022)	**15.6088**	0.2218	0.5619	153.81
Ours	15.5053	**0.2782**	**0.3955**	**125.95**

3 EXPERIMENTS

We evaluate the proposed SAR-to-Optical image translation model on the SAR2Opt dataset (Zhao et al., 2022). SAR2Opt consists of paired SAR–optical image patches collected from ten cities and includes 1,450 training samples and 627 test samples of size 600×600. The SAR images are acquired from TerraSAR-X with a ground sampling distance (GSD) of 1 m, and the corresponding optical images are obtained from Google Earth.

All models follow the default settings of their original implementations. For a fair comparison, the latent diffusion baseline (fine-tuned from Stable Diffusion v2) and our method use the same inference budget of 50 NFE. Our model uses a SiT-L/2 Transformer velocity network with an Euler ODE sampler.

Quantitative evaluation is conducted using PSNR, SSIM (Wang et al., 2004), LPIPS (Zhang et al., 2018), and FID (Heusel et al., 2017). Table 1 reports the quantitative results on the SAR2Opt test set. The proposed method achieves superior performance across all metrics, with notable improvements in LPIPS and FID, indicating improved perceptual quality and distributional similarity.

Qualitative results are shown in Figure 2. Existing methods often produce unstable results or blurred structures, whereas the proposed approach preserves spatial layout more consistently and generates optical images closer to the reference distribution.

4 CONCLUSION

In this paper, we presented a flow-based framework for SAR-to-Optical image translation that explicitly targets structure preservation. By modeling the translation process as a deterministic and continuous transformation in the latent space, the proposed method enables consistent mapping from SAR imagery to the optical domain while maintaining geometric correspondence. Experimental results on the SAR2Opt dataset demonstrate that our approach outperforms existing methods across both distortion-based and perceptual metrics, with particularly strong improvements in LPIPS and FID. These results highlight the effectiveness of deterministic flow-based modeling for SAR-to-Optical image translation tasks where structural consistency is critical.

REFERENCES

Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Xinyu Bai, Xinyang Pu, and Feng Xu. Conditional diffusion for sar to optical image translation. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2023.

Georges Le Bellier and Nicolas Audebert. Floweo: Generative unsupervised domain adaptation for earth observation. *arXiv preprint arXiv:2512.05140*, 2025.

Eric Brune and Yifang Ban. Sar-to-optical translation using conditional diffusion models for wildfire-burned area segmentation. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 7011–7015. IEEE, 2024.

Shoufa Chen, Chongjian Ge, Yuqi Zhang, Yida Zhang, Fengda Zhu, Hao Yang, Hongxiang Hao, Hui Wu, Zhichao Lai, Yifei Hu, et al. Goku: Flow based video generative foundation models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 23516–23527, 2025.

Jeonghyeok Do, Jaehyup Lee, and Munchurl Kim. C-diffset: Leveraging latent diffusion for sar-to-eo image translation with confidence-guided reliable object generation. *arXiv preprint arXiv:2411.10788*, 2024.

Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. In *International Conference on Learning Representations*, 2019.

Zhe Guo, Jiayi Liu, Qinglin Cai, Zhibo Zhang, and Shaohui Mei. Learning sar-to-optical image translation via diffusion models with color memory. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.

Seon-Hoon Kim and Daewon Chung. Conditional brownian bridge diffusion model for vhr sar to optical image translation. *IEEE Geoscience and Remote Sensing Letters*, 2025.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, pp. 23–40. Springer, 2024.

Andrea Bordone Molini, Diego Valsesia, Giulia Fracastoro, and Enrico Magli. Speckle2void: Deep self-supervised sar despeckling with blind-spot convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2021.

Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European conference on computer vision*, pp. 319–345. Springer, 2020.

Gaurav Parmar, Taesung Park, Srinivasa Narasimhan, and Jun-Yan Zhu. One-step image translation with text-to-image models. *arXiv preprint arXiv:2403.12036*, 2024.

Jiang Qin, Bin Zou, Lamei Zhang, and Yu Qiu. Sar-to-optical image translation using conditional denoising diffusion probabilistic models. In *2024 International Radar Conference (RADAR)*, pp. 1–5. IEEE, 2024.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Jeongwon Ryu, Youngtack Oh, and Minseok Seo. Sar2earth: A sar-to-eo translation dataset for remote sensing applications.

Minseok Seo, Jinwook Jung, and Dong-Geol Choi. Improved flood insights: Diffusion-based sar-to-eo image translation. *Remote Sensing*, 17(13):2260, 2025.

Prabhishek Singh, Manoj Diwakar, Achyut Shankar, Raj Shree, and Manoj Kumar. A review on sar image and its despeckling. *Archives of Computational Methods in Engineering*, 28(7):4633–4653, 2021.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Peng Wang, Yongkang Chen, Bo Huang, Daiyin Zhu, Tongwei Lu, Mauro Dalla Mura, and Jocelyn Chanussot. Mt_gan: A sar-to-optical image translation method for cloud removal. *ISPRS Journal of Photogrammetry and Remote Sensing*, 225:180–195, 2025.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

Xi Yang, Jingyi Zhao, Ziyu Wei, Nannan Wang, and Xinbo Gao. Sar-to-optical image translation based on improved cgan. *Pattern Recognition*, 121:108208, 2022.

Richard Yi Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595. IEEE, 2018.

Tianxian Zhang and Xiang-Gen Xia. Ofdm synthetic aperture radar imaging with sufficient cyclic prefix. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1):394–404, 2014.

Wenliang Zhao, Minglei Shi, Xumin Yu, Jie Zhou, and Jiwen Lu. Flowturbo: Towards real-time flow-based image generation with velocity refiner. *Advances in Neural Information Processing Systems*, 37:4148–4176, 2024.

Yitao Zhao, Turgay Celik, Nanqing Liu, and Heng-Chao Li. A comparative analysis of gan-based methods for sar-to-optical image translation. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5, 2022.

Jie Zhou, Chao Xiao, Bo Peng, Zhen Liu, Li Liu, Yongxiang Liu, and Xiang Li. Diffdet4sar: Diffusion-based aircraft target detection network for sar images. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.