Technische Hochschule Deggendorf
Fakultät Angewandte Informatik
Bachelor Künstliche Intelligenz

Erzeugung optischer Fernerkundungsdaten (Sentinel-2) auf Basis von Radar-Fernerkundungsdaten (Sentinel-1) mittels generativer KI

Generation of optical remote sensing data (Sentinel-2) based on radar remote sensing data (Sentinel-1) using generative AI

Bachelorarbeit zur Erlangung des akademischen Grades:
Bachelor of Science (B.Sc.)
an der Technischen Hochschule Deggendorf

Vorgelegt von:
Ahmed Attia
Matrikelnummer: 00815907

Prüfungsleitung:
Dr. Peter Hofmann

Am: XX. Monat 20XX

Erklärung

Name des Studierenden: Ahmed Attia

Name des Betreuenden: Dr. Peter Hofmann

Thema der Abschlussarbeit:

Erzeugung optischer Fernerkundungsdaten (Sentinel-2) auf Basis von Radar-Fernerkundungsdaten (Sentinel-1) mittels generativer KI . . . .

. . . .

. . . .

. . . .

1. Ich erkläre hiermit, dass ich die Abschlussarbeit gemäß § 35 Abs. 7 RaPO (Rahmenprüfungsordnung für die Fachhochschulen in Bayern, BayRS 2210-4-1-4-1-WFK) selbständig verfasst, noch nicht anderweitig für Prüfungszwecke vorgelegt, keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.

   Deggendorf, ...
   Datum Unterschrift des Studierenden

2. Ich bin damit einverstanden, dass die von mir angefertigte Abschlussarbeit über die Bibliothek der Hochschule einer breiteren Öffentlichkeit zugänglich gemacht wird:

   ◯ Nein

   ◯ Ja, nach Abschluss des Prüfungsverfahrens

   ◯ Ja, nach Ablauf einer Sperrfrist von . . . Jahren.

   Deggendorf, ...
   Datum Unterschrift des Studierenden

Bei Einverständnis des Verfassenden vom Betreuenden auszufüllen:

Eine Aufnahme eines Exemplars der Abschlussarbeit in den Bestand der Bibliothek und die Ausleihe des Exemplars wird:

◯ Befürwortet

◯ Nicht befürwortet

Deggendorf, ...

Datum Unterschrift des Betreuenden

Abstract

Contents

1 Introduction

2 Background

2.1 Introduction to Remote Sensing

2.2 Types of Remote Sensing Data

2.3 Relevant Types and their Applications: Sentinals 1 and 2

2.4 Data Availability and Limitations

2.4.1 Focus: Cloud Removal

2.4.2 Other Challenges

2.5 Generative AI

2.5.1 Pre-GenAI: Classical Approaches

2.5.2 Deep Learning and Computer Vision

2.5.3 Generative Adversarial Networks (GANs)

2.5.4 Diffusion Models

2.5.5 Vision Transformer

2.5.6 Vision Mamba

2.6 Application and Relevance to KIWA

3 Literature Review

Cloud contamination in optical remote sensing imagery hinders continuous Earth observation, limiting applications such as crop monitoring and land cover classification. Synthetic Aperture Radar (SAR) systems can penetrate clouds, enabling data acquisition in all weather conditions. This capability makes SAR data valuable for filling gaps in optical time series, driving research into SAR–optical fusion and SAR-to-optical image translation for cloud removal.

Early methods leveraged traditional signal processing techniques. Huang et al. [1] introduced sparse representation-based cloud removal using SAR data, which Xu et al. [2] extended via multi-temporal dictionary learning. These approaches, however, struggled under heavy cloud cover or highly dynamic surface changes.

Deep learning transformed the field. The foundational GAN framework was introduced by Goodfellow et al. [3], and Mirza & Osindero [4] extended it to conditional GANs (cGANs), ideal for image-to-image tasks like SAR-to-optical translation. Enomoto et al. [5] applied cGANs for cloud removal using NIR input, though dense clouds remained problematic. To address this, Grohnfeldt et al. [6] proposed SAR-Opt-cGAN to fuse Sentinel-1 SAR and Sentinel-2 optical data; Bermudez et al. [7] explored cGAN-based SAR-to-optical synthesis for crop classification. The SEN1-2 [8] and SEN12MS [9] datasets were pivotal for training deep models.

Advancements continued with Fuentes Reyes et al. [10] on cGAN optimization, Wang et al. [11] with supervised CycleGANs for translation, Meraner et al. [12] introducing DSen2-CR networks, Abady et al. [13] leveraging ProGANs, and Pan [14] with spatial-attention models. Gao et al. [15] developed fusion-based GAN approaches for high-resolution images.

Recent models show increasing complexity: Naderi Darbaghshahi et al. [16] proposed a two-GAN model with DRIBs, Ebel et al. [17] introduced UnCRtainTS with uncertainty prediction, Kwak & Park [18] proposed MTcGANs, and Liu et al. [19] developed S2MS-GAN.

Diffusion models have found their way into this field: Bai et al. [20] proposed a conditional diffusion model; Bai et al. [21] extended it with color supervision; Zou et al. [22] introduced the efficient DiffCR framework.

Vision Transformers (ViT) also made inroads: Dosovitskiy et al. [23] introduced the original ViT, while Park et al. [24] integrated multiscale ViT blocks into a cGAN for SAR-optical tasks. However, ViTs are computationally intensive.

Alternatives emerged via SSMs: Gu & Dao [25] introduced Mamba—an efficient, linear-complexity sequence model. U-Mamba [26] adapted this into a U-Net architecture, and Swin-UMamba [27] enhanced it further using ImageNet pretraining. Swin-UNet [28] remains a benchmark transformer-based segmentation model.

4 Methodology

5 Experiments

Table 5.1: Data cropped to 64x64, channels kept raw

Model	**MAE** $\downarrow$	**RMSE** $\downarrow$	**PSNR (dB)** $\uparrow$	**SSIM** $\uparrow$
pix2pix	68.06	291.62	37.54	0.9545
SwinUnet	41.58	181.62	41.11	0.9687
UMamba	35.14	158.47	42.67	0.9757
SiMaVP	**31.41**	**139.95**	**43.75**	**0.9797**

Table 5.2: Data cropped to 64x64, only channel 0 is used

Model	**MAE** $\downarrow$	**RMSE** $\downarrow$	**PSNR (dB)** $\uparrow$	**SSIM** $\uparrow$
pix2pix	70.48	300.88	37.10	0.9490
SwinUnet	47.65	203.04	40.15	0.9648
UMamba	33.60	149.58	42.96	0.9749
SiMaVP	**30.09**	**134.05**	**44.14**	**0.9812**

Table 5.3: Data cropped to 64x64, only channel 1 is used

Model	**MAE** $\downarrow$	**RMSE** $\downarrow$	**PSNR (dB)** $\uparrow$	**SSIM** $\uparrow$
pix2pix	69.78	298.66	37.25	0.9491
SwinUnet	45.00	194.12	40.54	0.9662
UMamba	38.83	175.52	41.92	0.9739
SiMaVP	**32.12**	**143.12**	**43.58**	**0.9787**

6 Discussion

7 Conclusion

Bibliography

[1] B. Huang, Y. Li, X. Han, Y. Cui, W. Li, and R. Li, "Cloud removal from optical satellite imagery with sar imagery using sparse representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 1046–1050, 2015.

[2] M. Xu, X. Jia, M. Pickering, and A. J. Plaza, "Cloud removal based on sparse representation via multitemporal dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2998–3006, 2016.

[3] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[4] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014.

[5] R. Enomoto *et al.*, "Image restoration from satellite cloud contamination using multispectral conditional gan," *Remote Sensing (likely)*, 2017.

[6] C. Grohnfeldt, M. Schmitt, and X. X. Zhu, "A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images," in *ISPRS TC III Mid-term Symposium*, 2018.

[7] J. Bermudez, P. Happ, A. Boulch *et al.*, "Synthesis of multispectral optical images from sar/optical multitemporal data using conditional gans," in *IGARSS*, 2018.

[8] M. Schmitt, L. H. Hughes, and X. X. Zhu, "The sen1-2 dataset for deep learning in sar-optical data fusion," in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-1, 2018, pp. 141–146.

[9] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "Sen12ms–a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W7, 2019, pp. 153–160.

[10] M. Fuentes Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt, "Sar-to-optical image translation based on conditional generative adversarial networks—optimization, opportunities and limits," *Remote Sensing*, vol. 11, no. 17, p. 2067, 2019.

[11] L. Wang, X. Xu, Y. Yu, R. Yang, R. Gui, Z. Xu, and F. Pu, "Sar-to-optical image translation using supervised cycle-consistent adversarial networks," *IEEE Access*, vol. 7, pp. 129 136–129 149, 2019.

[12] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 333–346, 2020.

[13] L. Abady *et al.*, "Gan generation of synthetic multispectral satellite images," 2020, online; ResearchGate.

[14] B. Pan *et al.*, "Cloud removal for remote sensing imagery via spatial attention generative adversarial network," 2020.

[15] F. Gao *et al.*, "Cloud removal with fusion of high-resolution optical and sar images using generative adversarial networks," *Remote Sensing*, vol. 12, no. 1, p. 191, 2020.

[16] F. Naderi Darbaghshahi and M. R. Mohammadi, "Cloud removal in remote sensing images using generative adversarial networks with dilated residual inception blocks," in *IGARSS*, 2021.

[17] P. Ebel *et al.*, "Uncrtaints: Uncertainty quantification for cloud removal in optical satellite time series," *arXiv preprint*, 2022.

[18] H. Kwak and S. Park, "Assessing the potential of multi-temporal conditional gans in sar-to-optical image translation for early-stage crop monitoring," *Remote Sensing*, vol. 16, no. 7, p. 1199, 2024.

[19] J. Liu *et al.*, "High-resolution sar-to-multispectral image translation based on s2ms-gan," *Remote Sensing*, vol. 16, no. 21, p. 4045, 2024.

[20] W. Bai *et al.*, "Conditional diffusion for sar to optical image translation," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[21] ——, "Sar to optical image translation with color supervised diffusion model," 2024.

[22] H. Zou *et al.*, "Diffcr: A fast conditional diffusion framework for cloud removal from optical satellite images," 2023.

[23] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," *arXiv preprint*, 2020.

[24] S. Park *et al.*, "Sar-to-optical image translation using vision transformer-based cgan," in *IGARSS*, 2025.

[25] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint*, 2023.

[26] J. Ma, F. Li, and B. Wang, "U-mamba: Enhancing long-range dependency for biomedical image segmentation," *arXiv preprint*, 2024.

[27] J. Liu, H. Yang, H.-Y. Zhou, Y. Xi, L. Yu, Y. Yu, Y. Liang, G. Shi, S. Zhang, H. Zheng, and S. Wang, "Swin-umamba: Mamba-based unet with imagenet-based pretraining," *arXiv preprint*, 2024.

[28] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-UNet: *Unet*-like pure transformer for medical image segmentation," pp. 205–218, 2023.

[29] J. Liu *et al.*, "Swin-umamba: Mamba-based unet with imagenet-based pretraining," in *MICCAI 2024 Workshop*, 2024.