

**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.

Deggen Dorf, .....  
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.

Deggen Dorf, .....  
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

|  |           |
|--|-----------|
| <b>Abstract</b>  | <b>v</b>  |
| <b>1 Introduction</b>  | <b>1</b>  |
| <b>2 Background</b>  | <b>3</b>  |
| 2.1 Introduction to Remote Sensing . . . . .                           | 3         |
| 2.2 Types of Remote Sensing Data . . . . .                             | 3         |
| 2.3 Relevant Types and their Applications: Sentinals 1 and 2 . . . . . | 3         |
| 2.4 Data Availability and Limitations . . . . .                        | 3         |
| 2.4.1 Focus: Cloud Removal . . . . .                                   | 3         |
| 2.4.2 Other Challenges . . . . .                                       | 3         |
| 2.5 Generative AI . . . . .  | 3         |
| 2.5.1 Pre-GenAI: Classical Approaches . . . . .                        | 3         |
| 2.5.2 Deep Learning and Computer Vision . . . . .                      | 3         |
| 2.5.3 Generative Adversarial Networks (GANs) . . . . .                 | 3         |
| 2.5.4 Diffusion Models . . . . .                                       | 3         |
| 2.5.5 Vision Transformer . . . . .                                     | 3         |
| 2.5.6 Vision Mamba . . . . .   | 3         |
| 2.6 Application and Relevance to KIWA . . . . .                        | 3         |
| <b>3 Literature Review</b>   | <b>5</b>  |
| <b>4 Methodology</b>   | <b>7</b>  |
| <b>5 Experiments</b>   | <b>9</b>  |
| <b>6 Discussion</b>  | <b>11</b> |
| <b>7 Conclusion</b>  | <b>13</b> |





# 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. Consequently, research has focused on SAR-optical data fusion and SAR-to-optical image translation for cloud removal to improve data availability.

Early cloud removal methods utilized traditional image processing techniques like sparse representation and multi-temporal dictionary learning. For instance, Huang et al. (2015) proposed using SAR data for cloud removal via sparse representation. This approach was subsequently extended by Xu et al. (2016) using multi-temporal dictionary learning. However, these methods exhibited limitations when confronted with extensive cloud coverage or dynamic surface conditions.

The advent of deep learning, particularly Generative Adversarial Networks (GANs), transformed remote sensing image processing for tasks like cloud removal and SAR-to-optical translation. Goodfellow et al. (2014) introduced the foundational GAN framework for generating synthetic data. This was extended by Mirza and Osindero (2014) with Conditional GANs (cGANs), which generate outputs conditioned on specific inputs. These models provide a powerful framework for image-to-image translation tasks central to SAR-to-optical conversion and cloud removal.

Enomoto et al. (2017) applied cGANs to generate cloud-free RGB images from cloudy RGB and near-infrared (NIR) data, though this approach was limited by the opacity of dense clouds to both visible and infrared light. To overcome this, Grohnfeldt et al. (2018) developed the **SAR-Opt-cGAN** to fuse Sentinel-2 optical and Sentinel-1 SAR imagery, demonstrating superior reconstruction performance by utilising SAR information. Concurrently, Bermudez et al. (2018) investigated SAR-to-optical synthesis using cGANs for crop classification. The development of large-scale datasets, namely **SEN1-2** (Schmitt et al., 2018) and its successor **SEN12MS** (Schmitt et al., 2019b), was critical for this field, providing large quantities of co-registered SAR and optical image pairs.

Research in GAN-based methods continued to advance. Fuentes Reyes et al. (2019) analyzed the optimization and limitations of cGANs for SAR-to-optical translation. Wang et al. (2019) introduced a supervised **CycleGAN** that incorporated cycle-consistency to improve translation performance. Meraner et al. (2020) proposed **DSen2-CR**, a deep residual network for SAR-optical fusion that demonstrated robustness to extensive cloud cover. Abady et al. (2020) used **ProGAN** to generate synthetic multispectral images, showing it could preserve spectral relationships. Pan (2020) developed the **Spatial Attention Generative Adversarial Network (SpA GAN)**, which incorporates a spatial attention mechanism to focus on cloud-affected areas, enhancing information recovery and generating higher quality cloudless images. Gao et

al. (2020) also contributed a GAN-based method for cloud removal by fusing high-resolution optical and SAR images.

More recent GAN models have incorporated greater complexity. Naderi Darbaghshahi et al. (2021) proposed a two-GAN model for cloud removal, using a first GAN for SAR-to-optical translation and a second for cloud removal, incorporating **dilated residual inception blocks (DRIBs)** to increase receptive fields. Ebel et al. (2022) introduced **UnCRtainTS**, which combines an attention-based architecture with multivariate uncertainty prediction for multi-temporal cloud removal, achieving state-of-the-art results on challenging datasets. Kwak and Park (2024) evaluated **Multi-Temporal Conditional Generative Adversarial Networks (MTcGANs)** for crop monitoring, highlighting their advantages over conventional methods by allowing diverse input data compositions. Liu et al. (2024) developed **S2MS-GAN**, which includes a **TV-BM3D module** for speckle noise reduction and spectral attention to enhance spectral features in high-resolution SAR-to-multispectral translation.

Diffusion Models have also been applied to these tasks. Bai et al. (2023) proposed a conditional diffusion model for SAR-to-optical translation, using the SAR image as a constraint during the sampling process to produce high-quality optical images with clearer borders, aiding interpretation. A subsequent work introduced a **color-supervised diffusion model** to address color shift issues in the generated images. Zou et al. (2023) presented **DiffCR**, a fast conditional diffusion framework for cloud removal that uses a decoupled encoder and an efficient fusion block to achieve state-of-the-art performance with high computational efficiency.

Vision Transformer (ViT) architectures, known for their ability to capture global image relationships, have also been adapted for SAR-to-optical translation. Dosovitskiy et al. (2020) first demonstrated the power of transformers for image recognition, establishing a new paradigm for vision tasks. In the context of remote sensing, Park et al. (2025) incorporated **multiscale Vision Transformer (ViT) blocks** into a cGAN, leveraging a ViT processor to extract multi-level terrain and color features, which improved the detail and colorization of generated optical images. This successful integration highlights the potential of transformer-based architectures for extracting the complex, large-scale features inherent in satellite imagery.

While ViTs have proven effective, their quadratic complexity in relation to image size presents computational challenges, particularly for high-resolution remote sensing data. A more recent development, the Mamba architecture, introduced by Gu & Dao (2023), addresses this limitation. Mamba and its vision-centric variants are based on State Space Models (SSMs) that can model long-range dependencies with linear complexity. This efficiency has led to their rapid adoption in various remote sensing tasks. For instance, recent studies have demonstrated the success of Vision Mamba in demanding applications such as semantic segmentation, change detection, and even in related optical-to-optical translation tasks.

The proven ability of SSMs to effectively model complex spatial data in other remote sensing applications, combined with their significant computational advantages over ViTs, makes them a highly promising candidate for the more challenging cross-modality problem of SAR-to-optical translation. However, the application of Vision Mamba to this specific task, which requires bridging the gap between active microwave and passive optical data, remains an open research question. This thesis aims to address this gap by implementing and evaluating a Vision Mamba-based model for generating optical Sentinel-2 data from Sentinel-1 SAR imagery.

## 4 Methodology





## 5 Experiments

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

| <b>Model</b> | <b>MAE ↓</b> | <b>RMSE ↓</b> | <b>PSNR (dB) ↑</b> | <b>SSIM ↑</b> |
|--------------|--------------|---------------|--------------------|---------------|
| 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       | <b>31.41</b> | <b>139.95</b> | <b>43.75</b>       | <b>0.9797</b> |

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

| <b>Model</b> | <b>MAE ↓</b> | <b>RMSE ↓</b> | <b>PSNR (dB) ↑</b> | <b>SSIM ↑</b> |
|--------------|--------------|---------------|--------------------|---------------|
| 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       | <b>30.09</b> | <b>134.05</b> | <b>44.14</b>       | <b>0.9812</b> |

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

| <b>Model</b> | <b>MAE ↓</b> | <b>RMSE ↓</b> | <b>PSNR (dB) ↑</b> | <b>SSIM ↑</b> |
|--------------|--------------|---------------|--------------------|---------------|
| 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       | <b>32.12</b> | <b>143.12</b> | <b>43.58</b>       | <b>0.9787</b> |

## **6 Discussion**



## **7 Conclusion**



# Bibliography

- [1] N. author really, “Example reference.”
- [2] Stack Exchange, “Reference to a website.” [Online]. Available: <https://tex.stackexchange.com/>
- [3] Y. Liu, Q. Han, H. Yang, and H. Hu, “High-resolution sar-to-multispectral image translation based on s2ms-gan,” *Remote Sensing*, vol. 16, no. 21, 2024. [Online]. Available: <https://www.mdpi.com/2072-4292/16/21/4045>