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:	Prüfungsleitung:
Ahmed Attia	Dr. Peter Hofmann
Matrikelnummer: 00815907	

Am: XX. Monat 20XX

Contents

List of Abbreviations

Abbreviation	Full Form
RS	Remote Sensing
SAR	Synthetic Aperture Radar
GAN	Generative Adversarial Network
cGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
DDPM	Denoising Diffusion Probabilistic Model
ESA	European Space Agency
GRD	Ground Range Detected
EW	Extra-Wide Swath Mode
IW	Interferometric Wide Swath Mode
WV	Wave Mode
LULC	Land Use and Land Cover
MODIS	Moderate Resolution Imaging Spectroradiometer
ROI	Region of Interest
SWIR	Shortwave Infrared
VNIR	Visible and Near Infrared
VV	Vertical–Vertical Polarization
VH	Vertical–Horizontal Polarization
NIR	Near Infrared
IQA	Image Quality Assessment
SSIM	Structural Similarity Index Measurement
FSIM	Feature Similarity Index Measurement
FSIM	Image Quality Assessment
DISTS	Deep Image Structure and Texture Similarity
PSNR	Peak Signal-to-Noise Ratio
SAM	Spectral Angle Mapper
FID	Fréchet Inception Distance
LPIPS	Learned Perceptual Image Patch Similarity

1 Background

1.1 Remote Sensing

Remote sensing (RS) is commonly defined as the acquisition of information about an object through sensors without direct physical contact. This information is obtained by detecting and measuring the modifications the object induces in its surrounding fields, which may include electromagnetic, acoustic, or potential fields [1].

RS is a relatively recent scientific discipline characterized by its strong interdisciplinary nature. It draws upon a wide spectrum of fields, requiring practitioners to develop a broad foundational understanding of both natural and applied sciences. Effective research in remote sensing often involves collaboration with specialists in electromagnetic theory, spectroscopy, applied physics, geology, atmospheric sciences, oceanography, electrical engineering, and optical engineering [1].

Remote observations require an interaction of energy between the target and the sensor. In the case of passive sensors, the detected energy originates from external or natural sources, such as solar radiation reflected by the Earth's surface or thermal radiation emitted by the object itself. A prominent example is the *Landsat* program [1], which represents the longest continuously operating Earth observation mission. Over several decades, Landsat has generated a continuous global record, contributing significantly to environmental monitoring and Earth system science.

By contrast, active sensors generate their own energy pulses to illuminate the target and subsequently measure the portion of the signal that is reflected or backscattered. This capability allows them to operate independently of solar illumination and under a wide range of environmental conditions, including day or night and, in the case of microwave systems, through cloud cover and adverse weather [2]. The most widely used active sensing technologies are Radar (Radio Detection and Ranging), which transmits and receives microwave radiation, and LiDAR (Light Detection and Ranging), which employs laser pulses in the optical domain. Both systems record the properties of the reflected signals to extract information about the target.

The term *Remote Sensing* was introduced in the early 1960s to denote techniques for observ-

[1] https://landsat.gsfc.nasa.gov/

ing the Earth from a distance, with particular reference to aerial photography, which represented the predominant sensing technology at that time [3].

With the advent of satellites, global and synoptic observations of Earth and other planetary environments have become possible. Earth-orbiting sensors provide essential data on atmospheric dynamics, cloud distribution, vegetation cover, and its seasonal variability. Their long-term operation and repetitive coverage enable the monitoring of rapidly changing processes, such as polar ice dynamics and tropical deforestation. Beyond Earth, planetary missions (orbiters, flybys, landers, and rovers) have extended similar observations to all major planets in the solar system. To date, every planet has been visited at least once [1].

The origins of remote sensing date back to the invention of photography in 1839, which soon after was applied to topographic mapping. By the mid-19th century, aerial photographs were obtained from balloons, followed later by kites, pigeons, and eventually airplanes—the latter marking a decisive step with Wilbur Wright's first aerial photographs in 1909. Aerial photography became essential during World War I and advanced further in the 1930s-1940s with the introduction of color and infrared-sensitive films, widely used during World War II for reconnaissance and camouflage detection [1, 3].

The postwar decades brought rapid technological progress with the development of radar and synthetic aperture radar (SAR), enabling high-resolution imaging independent of daylight or weather. Early rocket experiments in the late 1940s foreshadowed the space age, initiated by the launch of Sputnik in 1957. NASA's TIROS-1 satellite (1960) delivered the first global meteorological observations, while the launch of Landsat-1 in 1972 introduced systematic multispectral Earth observation, a program that continues today as the longest-running record of land surface change [1, 3].

A symbolic milestone came with the Apollo 8 mission in 1968, when astronaut William Anders captured the famous Earthrise photograph, showing Earth rising above the lunar horizon (see Figure 1.1). This image not only had profound cultural, philosophical, and scientific impact but also highlighted the scientific value of spaceborne Earth observation.

Since the 1980s, remote sensing has expanded through international efforts such as SPOT (France, 1986), MOS-1 (Japan, 1987), and IRS-1 (India, 1988). The European Space Agency (ESA) [2] launched its first radar satellite, ERS-1, in 1991, and a second with comparable specifications in 1995. The 1990s and 2000s saw the rise of commercial satellites like IKONOS and QuickBird, offering very high-resolution imagery. Today, constellations of small satellites operated by private companies provide near-daily global coverage at meter-scale resolution. These advances—driven by improvements in optics, sensors, data transmission, and digital processing—have transformed remote sensing into a cornerstone of Earth system science, environ-

[2] https://www.esa.int/

Figure 1.1: The iconic “Earthrise” photograph taken by astronaut William Anders during the Apollo 8 mission in 1968. Source: NASA.

mental monitoring, disaster response, and planetary exploration [3].

A summary of major milestones in the historical development of remote sensing platforms, from early balloon photography to modern satellite constellations, is illustrated in Figure 1.2.

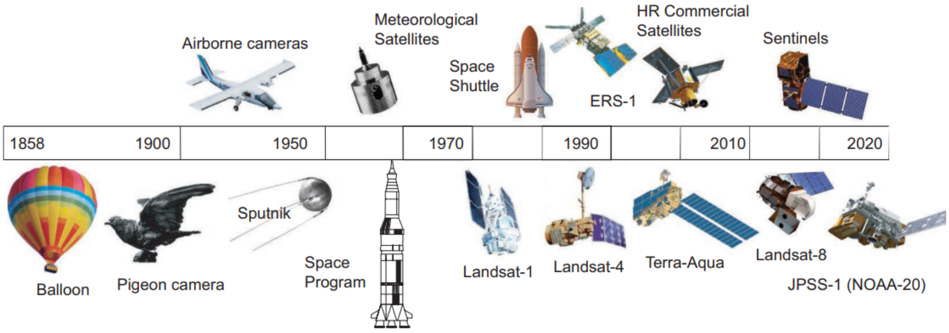

Figure 1.2: Timeline of remote sensing platform development, from early airborne cameras to modern Earth observation satellites. Adapted from [3].

1.2 Copernicus: Europe's eyes on Earth

Copernicus, known as the most ambitious Earth observation programme, is the Earth observation component of the European Union's Space Programme. It is funded, coordinated, and managed by the European Commission in cooperation with partners such as the European Space Agency (ESA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)[3]. The programme was named after the European scientist and observer Nicolaus Copernicus[4]. It integrates satellite and in situ observations (e.g., ground stations, airborne and seaborne instruments) to provide reliable, up-to-date information. Its services cover six domains: land, marine, atmosphere, emergency management, security, and climate change.

The Copernicus Space Component features a new family of dedicated satellites, called Sentinels, depicted in Figure 1.3, specifically designed for the operational needs of the Copernicus programme.

On 3 April 2014, the deployment of the Copernicus Space Component began with the launch of the Sentinel-1 radar satellite, operating in the C-band and providing all-weather, day-and-night radar imagery. It was followed by its radar successors in 2016 and 2024. Its synthetic aperture radar (SAR) instruments are crucial for monitoring land deformation, subsidence, sea-ice dynamics, and emergency situations such as flooding and earthquakes [4, 5].

Sentinel-2, launched in 2015, 2017, and 2024, is designed to deliver high-resolution, multispectral optical images, supporting applications such as agriculture, forestry, land use, disaster management, and climate studies. Its 13 spectral bands enable detailed analysis of vegetation health, water quality, and land cover dynamics [4].

[3] https://www.eumetsat.int/
[4] https://www.biography.com/scientists/nicolaus-copernicus

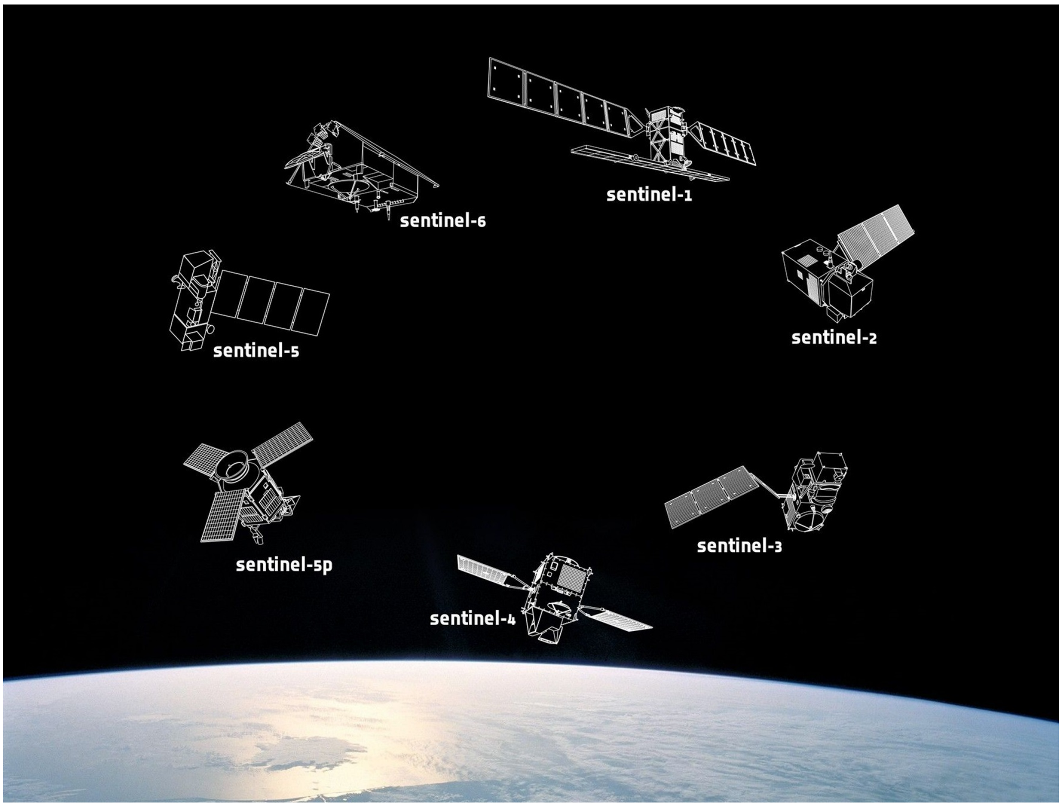

Figure 1.3: An artist's impression of the Copernicus Sentinel Missions. Source: ESA.

The two Sentinel-3 satellites, launched on 16 February 2016 and 25 April 2018, provide data for services relevant to the ocean and land. They carry instruments to measure sea surface topography, sea and land surface temperature, and ocean and land colour, providing essential data for oceanography, marine resource management, and climate monitoring [4, 5].

Sentinel-4 is an ultraviolet, visible, and near-infrared spectrometer carried on the Meteosat Third Generation Sounder satellites. Launched on 1 July 2025, it is dedicated to monitoring atmospheric composition and air quality over Europe and parts of North Africa. It provides hourly measurements of key trace gases and aerosols, enabling near-real-time assessments of air pollution, UV radiation, and climate-relevant processes [4].

Launched on 13 October 2017, the Sentinel-5P mission (Sentinel-5 Precursor) is the first Copernicus mission dedicated to monitoring the atmosphere. It provides high spatio-temporal resolution data for air quality, ozone and UV radiation, as well as climate monitoring and forecasting [5].

Sentinel-6 is dedicated to high-precision ocean monitoring, focusing on sea surface topography. It continues the long-term record of satellite altimetry, measuring global sea level rise and ocean circulation patterns. These data are critical for climate change research, weather forecasting, and operational oceanography. Sentinel-6 was launched on 21 November 2020 [5].

Looking into the future, six Sentinel Expansion missions will join the fleet. These include, among others, the Hyperspectral Imaging Mission, the Polar Ice and Snow Topography Altimeter, and the Anthropogenic Carbon Dioxide Monitoring mission [4].

Since the focus of this work is on SAR and Optical data, only Sentinel-1 and 2 will be discussed in details in the next sections.

1.2.1 Sentinel-1

The following description is based on the official SentiWiki resource provided by the European Space Agency [6].

Sentinel-1, launched on 3 April 2014, constitutes the radar component of the European Copernicus Programme. The mission is designed as a constellation of two sun-synchronous, near-polar orbiting satellites in the same orbital plane, separated by 180° in phase. Equipped with C-band synthetic aperture radar (SAR) operating at 5.4 GHz, Sentinel-1 provides continuous, all-weather, day-and-night imaging capability. Sentinel-1A was followed by Sentinel-1B in 2016, which ceased operations after an anomaly in 2021 and was subsequently replaced by Sentinel-1C in 2024.

The SAR instrument actively transmits microwave signals towards the Earth and records the backscattered response. Both amplitude and phase are preserved, enabling the reconstruction of high-resolution images. Polarisation diversity further enhances information extraction, as different surfaces exhibit characteristic scattering signatures, supporting classification and retrieval of geophysical parameters.

Sentinel-1 operates in four exclusive acquisition modes: Stripmap (SM), Interferometric Wide Swath (IW), Extra-Wide Swath (EW), and Wave (WV). These modes achieve spatial resolutions down to 5 m and swath widths of up to 400 km. The system supports single (HH or VV) and dual (HH+HV or VV+VH) polarisation. While SM, IW, and EW modes allow a duty cycle of up to 30 minutes per orbit, WV mode extends this to 75 minutes. Over land, IW mode with VV+VH polarisation is the primary operational configuration, balancing revisit performance, service requirements, and the creation of a consistent long-term archive. For open-ocean observations, WV mode with VV polarisation is predominantly employed, while EW mode is mainly used for sea-ice monitoring and maritime surveillance in high-latitude regions. SM mode is activated only for small islands or in response to emergencies. Across all modes, products are provided at multiple processing levels, from raw SAR data (Level-0) to geophysical ocean products (Level-2 OCN).

The revisit capabilities of Sentinel-1 are particularly notable. In IW mode, a single satellite can achieve global coverage every 12 days, while the two-satellite constellation reduces the repeat cycle to six days, completing 175 orbits per cycle. These systematic observations, combined with advanced interferometric capabilities, enable the precise detection of land subsidence, structural deformation, and ground movements that are otherwise imperceptible. Such data are invaluable for urban planning, geohazard monitoring, and applications in mining, geology, and risk assessment for infrastructure and natural hazards [6].

1.2.2 Sentinel-2

The following description is based on the official SentiWiki resource provided by the European Space Agency [6].

Sentinel-2 is the optical imaging mission of the Copernicus Programme, designed to provide systematic, high-resolution observations over land and coastal regions. The mission consists of a constellation of two sun-synchronous satellites in the same orbital plane, phased 180° apart, ensuring global coverage with a revisit frequency of five days at the Equator. Sentinel-2A was launched in 2015, followed by Sentinel-2B in 2017 and Sentinel-2C in September 2024, the latter ensuring mission continuity as Sentinel-2A approaches the end of its operational lifetime.

Each satellite carries a single payload: the Multi-Spectral Instrument (MSI). This passive optical sensor collects sunlight reflected from the Earth's surface, splitting the incoming radiation into two focal plane assemblies: one covering the visible and near-infrared (VNIR) and the other the shortwave infrared (SWIR). The instrument has a swath width of 290 km, which is considerably wider than comparable missions such as Landsat 5/7 (185 km) or SPOT-5 (120 km).

The MSI samples 13 spectral bands at three spatial resolutions: four bands at 10 m (Blue, Green, Red, and Near-Infrared), six bands at 20 m (red-edge and SWIR), and three bands at 60 m (aerosol, water vapour, and cirrus). These bands span the VNIR to SWIR regions of the electromagnetic spectrum and are tailored to applications including vegetation and crop monitoring, land cover mapping, water quality assessment, snow and ice monitoring, cloud screening, and atmospheric correction. An overview of the spectral bands is provided in Table 1.1.

Sentinel-2 imagery is systematically and freely available, supporting several Copernicus services. The Copernicus Land Monitoring Service (CLMS) employs Sentinel-2 for land cover and forest mapping, crop monitoring, ecosystem assessment, and climate change adaptation. The Copernicus Marine Environment Monitoring Service (CMEMS) relies on Sentinel-2 to derive products such as turbidity, chlorophyll, suspended particulate matter, bathymetry, and ice analysis. The Copernicus Emergency Management Service (CEMS) uses Sentinel-2 extensively in disaster response, particularly for rapid mapping of floods, fires, and earthquakes. By enabling systematic, frequent, and global observations, Sentinel-2 has become a cornerstone of Copernicus services, supporting environmental monitoring, resource management, and disaster re-

Table 1.1: Sentinel-2 MSI spectral bands with central wavelength and spatial resolution [6].

Band	Central Wavelength [nm]	Resolution [m]
B1	443 (Aerosols)	60
B2	490 (Blue)	10
B3	560 (Green)	10
B4	665 (Red)	10
B5	705 (Red edge)	20
B6	740 (Red edge)	20
B7	783 (Red edge)	20
B8	842 (NIR)	10
B8a	865 (Red edge)	20
B9	945 (Water vapour)	60
B10	1375 (Cirrus)	60
B11	1610 (SWIR)	20
B12	2190 (SWIR)	20

sponse worldwide [6].

Together, Sentinel-1 and Sentinel-2 provide complementary SAR and optical observations, which form the basis of this thesis aiming to translate SAR imagery into its optical counterpart.

1.3 Cloud Removal

As briefly mentioned in the sections above, optical remote sensing imagery, such as Sentinel-2 products, represents a key source of Earth observation data. Compared to SAR observations, multispectral images contain rich spectral information and are readily interpretable by the human eye. Such data play an essential role in a wide range of applications, including environmental monitoring, resource exploration, and disaster assessment. While the quality and quantity of satellite observations have dramatically increased in recent years, one common problem persists for optical remote sensing imagery: **cloud cover**.

Based on findings from the International Satellite Cloud Climatology Project (ISCCP), average global cloud cover surpasses 66% [7–9], with 55% over land surface alone [9], preventing optical satellites from acquiring valuable information about the Earth's surface due to the frequent presence of clouds in the imagery. In contrast to SAR instruments, optical sensors cannot penetrate clouds, resulting in considerable data gaps in both the spatial and temporal domains. For applications requiring consistent time series, e.g., agricultural monitoring, or where a specific scene must be observed at a given time, e.g., disaster monitoring, cloud cover represents a serious limitation [9]. The diversity of clouds —including thin and thick clouds as well as haze— together with the wide range of occlusion scenarios and their uneven distribution, poses an additional challenge for image reconstruction and the generalizability of cloud removal techniques [10].

Consequently, removing clouds and obtaining cloud-free optical data to retrieve surface information is both of theoretical importance and practical necessity. Cloud removal in optical remote sensing imagery aims to mitigate or eliminate the influence of clouds, thereby revealing more accurate and complete surface details [10]. In response to this challenge, a wide range of approaches have been proposed. These methods can broadly be divided into three categories: (i) single-image methods, (ii) multimodal-based methods, and (iii) multitemporal-based methods [10]. The main categories and their characteristics are summarized below

(A) **Single-image methods:** Constrained by the limited acquisition capabilities of early remote sensing data, single-image cloud removal techniques attempt to restore surface information using only the cloudy optical image. Classical approaches employ statistical and physical models such as spatial similarity, frequency filtering, or atmospheric scattering models. For example, Zhang et al. [11] proposed the *Haze Optimized Transformation (HOT)*, which detects and compensates for thin cloud and haze contamination in Landsat images by exploiting the spectral correlation of clear-sky bands and quantifying deviations caused by haze. Similarly, He et al. [12] introduced the *dark channel prior*, a widely used statistical prior that estimates haze thickness from local image patches to recover clear radiance, later adapted for thin cloud removal in optical remote sensing. With the advent of deep learning, CNNs, U-Nets, and GAN-based architectures have been applied to learn the mapping from cloudy to cloud-free domains, sometimes extended with unpaired learning schemes like CycleGANs. Notably, U-Net-based methods have been widely used for their encoder-decoder structures, while CycleGAN approaches exploit cycle consistency loss to preserve colors and textures during cloudy-to-clear translation. While these methods demonstrate effectiveness for thin or semi-transparent clouds, their reliance on information present in a single image limits their applicability to dense cloud cover. In such cases, they cannot reliably reconstruct surface features, which has motivated the integration of external data sources such as SAR imagery [10].

(B) **Multimodal-based methods:** Multimodal strategies explicitly integrate auxiliary data from other sensors to improve optical image restoration. Multispectral-based methods exploit the differential sensitivity of spectral bands, but the most notable progress has been achieved by fusing synthetic aperture radar (SAR) with optical imagery. A representative work is Meraner et al. [9], who proposed the DSen2-CR framework, a deep residual network that combines Sentinel-1 and Sentinel-2 data to improve reconstructions under thick cloud cover and preserve spectral fidelity. Likewise, Grohnfeldt et al. [13] demonstrated the potential of conditional GANs (cGANs) to fuse SAR and multispectral data for cloud removal, highlighting the advantages of adversarial training in capturing nonlinear relationships between modalities. More recently, Xu et al. [14] presented the GLF-CR model, which applies a global–local fusion strategy to better exploit SAR features for cloud removal. SAR-to-optical image translation has thus emerged as a powerful paradigm in this context, as SAR penetrates cloud layers and provides structural information that can guide optical reconstruction. A wide range of approaches have been proposed, including CNN-based fusion, cGANs, and CycleGAN-style frameworks, which either translate SAR features into optical-like imagery or combine them with partially corrupted optical inputs. These methods have proven especially effective in recovering surface information under dense and persistent cloud conditions, although challenges remain in terms of data registration, modality differences, and SAR-induced speckle noise.

(C) **Multitemporal-based methods:** Multitemporal approaches leverage repeated acquisitions of the same location at different times to fill in cloud-covered areas. *Non-blind* methods use cloud masks to guide restoration, whereas *blind* methods directly infer cloud-free information from temporal sequences. A classical example is the work of Xu et al. [15], who proposed a sparse representation framework with multitemporal dictionary learning (MDL) that learns dictionaries from both cloudy and clear images, effectively reconstructing areas obscured by thin and thick clouds without requiring explicit cloud masks. More recently, Ebel et al. [16] introduced UnCRtainTS, an attention-based deep learning model that not only reconstructs cloud-free images from Sentinel-1/2 time series but also quantifies pixel-wise uncertainty, providing reliability measures alongside the reconstructed outputs. Techniques therefore range from traditional model-driven approaches, such as low-rank tensor decomposition and sparse representation, to data-driven deep learning frameworks that learn spatio-temporal mappings. Recent research has also begun to combine multitemporal optical data with SAR, creating hybrid SAR–optical time series methods that enhance robustness under persistent cloud cover and enable more accurate SAR-to-optical translation. Although highly effective for dense cloud removal, these approaches face challenges such as geometric misalignment, temporal variability in land cover [17], and the need for large, paired training datasets [10]. In this context, mono-temporal data offers an advantage, as it requires less data and avoids the need for co-registration compared to multi-temporal approaches [18].

As shown in Table 1.2, research on cloud removal has been uneven across categories. Single-image methods have been the most extensively studied due to their simplicity and minimal data requirements, though their effectiveness is limited under dense clouds. Multimodal approaches, particularly SAR–optical fusion, have gained significant traction in recent years and are currently the most active research direction. By contrast, multitemporal methods, while highly effective in principle, are less frequently explored because of the challenges in acquiring consistent, well-aligned time series data.

Table 1.2: Summary of cloud removal categories, their advantages and limitations [10, 19].

Category	Advantages	Limitations	Representative literature
Single-image	• No auxiliary data required (cost- and time-efficient). • Effective for thin or semi-transparent clouds. • Straightforward implementation with statistical/physical models or deep learning.	• Ineffective for dense or opaque clouds. • Often introduces artifacts or color distortions. • Deep learning requires large paired datasets, which are difficult to obtain.	[11] [12] [20] [21] [22] [23] [24] [25] [26]
Multimodal	• Integrates complementary information from other sensors. • Multispectral bands provide spectral redundancy. • SAR–optical fusion enables SAR-to-optical translation, penetrating cloud layers. • Suitable for both thin and thick clouds.	• Requires accurate registration of heterogeneous data. • SAR data introduces speckle noise. • High computational complexity and preprocessing effort.	[13] [27] [28] [9] [29] [30] [31] [32] [33] [34] [35] [36]
Multitemporal	• Exploits temporal redundancy to reconstruct cloudy regions. • Effective for dense and extensive cloud cover. • Deep learning models can capture spatio-temporal correlations. • Can be extended with SAR–optical time series for improved robustness.	• Sensitive to geometric misalignment and temporal variability. • Requires consistent multitemporal datasets, which may be unavailable. • Landscape or seasonal changes reduce restoration accuracy.	[28] [37] [15] [16] [38] [39]

In summary, cloud removal research spans single-image, multimodal, and multitemporal strategies, each with distinct advantages and limitations. Among these, SAR-to-optical image translation has recently emerged as a particularly promising direction, as it leverages the cloud-penetrating capability of SAR while producing optical-like imagery suitable for interpretation and analysis. This thesis builds on this line of research by systematically investigating and advancing SAR-to-optical translation methods for cloud removal.

1.4 Generative Artificial Intelligence

Generative Artificial Intelligence (GenAI) refers to a class of machine learning models designed to generate new data samples that resemble a given training distribution, such as images, text, or audio. Unlike discriminative models, which focus on classifying or predicting labels, generative models learn (an approximation to) the underlying probability distribution of the data to create novel instances [40]. This capability has revolutionized fields like computer vision, where GenAI is used for tasks including image synthesis, style transfer, and domain adaptation. In the context of remote sensing, GenAI enables, among others, the creation of synthetic imagery, such as translating radar data to optical-like representations, which is particularly useful for overcoming environmental limitations like cloud cover, as well as data fusion for both heterogeneous and homogeneous imagery, enhancing spatial, spectral, and temporal resolution and mitigating the limitations of individual sensors [41].

One of the foundational frameworks in GenAI is the Generative Adversarial Network (GAN), introduced in 2014 by Goodfellow et al. [42]. As depicted in Figure 1.4, a GAN consists of two neural networks: a generator (G) that produces synthetic data from random noise, and a discriminator (D) that evaluates whether the generated data is real or fake. These components are trained adversarially—the generator aims to fool the discriminator, while the discriminator improves its ability to distinguish real from generated samples—leading to increasingly realistic outputs.

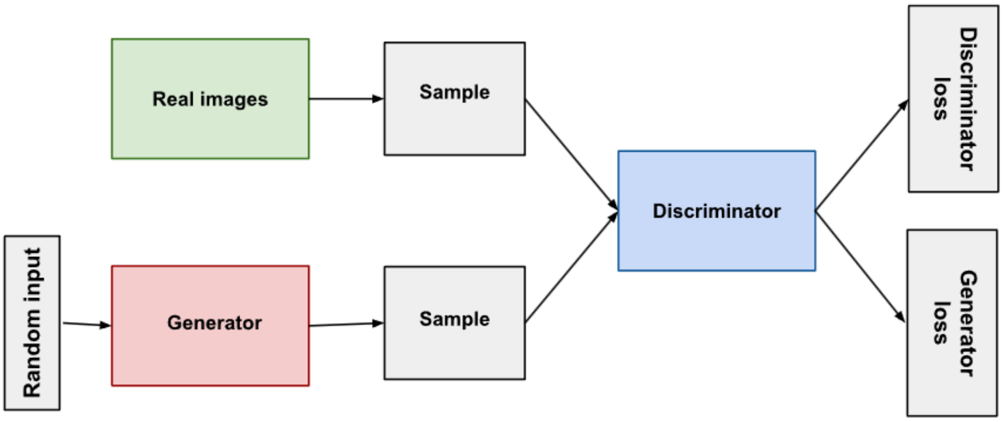

Figure 1.4: GANs Architecture: Source: https://developers.google.com/

This adversarial process minimizes a minimax loss function, allowing GANs to capture complex data distributions without explicit probabilistic modeling. Formally, the optimization problem is defined as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))], \tag{1.1}$$

where $D(x)$ denotes the discriminator's estimate of the probability that x comes from the real data distribution p_{data}, and $G(z)$ generates samples from latent noise $z \sim p_z$.

Building on GANs, conditional GANs (cGANs) [43] incorporate additional input conditions, such as class labels or reference images, to guide the generation process toward specific outputs. In image-to-image translation settings (e.g., pix2pix-style formulations), cGANs typically require paired datasets for supervised training, which is non-trivial given the complexity of acquiring and registering satellite imagery. A key extension for unpaired image translation is the Cycle-Consistent GAN (CycleGAN) [44], proposed in 2017. CycleGAN addresses the challenge of learning mappings between two domains (e.g., SAR to optical) without paired training examples by enforcing cycle consistency: translating an image from domain X to Y and back to X should reconstruct the original. This is achieved through a cycle-consistency loss, combined with adversarial losses, making it suitable for RS applications where perfectly aligned SAR–optical pairs are scarce. Nevertheless, given the weak supervision during CycleGAN training, it is prone to texture/detail distortions [41], often resulting in performance comparable to other GAN-based models [38].

For the problem of RS data fusion, many studies do not directly reuse off-the-shelf GAN architectures but instead adapt them to better accommodate multisource fusion. Among these, cGANs and CycleGANs are the most commonly adopted families in numerous fusion studies [41].

Despite the remarkable performance of GAN-based approaches in cloud removal, several challenges remain. They are inherently difficult to train and can suffer from issues such as mode collapse, leading to distorted or repetitive outputs, especially at high spatial resolutions [29]. Moreover, cloud morphology and distribution exhibit substantial diversity and complexity, imposing demanding requirements on training and application of GAN-based methods [10, 39].

More recently, diffusion models have emerged as a powerful alternative to GANs, offering improved stability and higher-fidelity generation [33]. Denoising Diffusion Probabilistic Models (DDPMs) [45], introduced in 2020, model data generation as a reverse diffusion process: starting from Gaussian noise, the model iteratively denoises the input to produce a sample from the target distribution. Unlike GANs, which can suffer from mode collapse (limited output diversity), diffusion models provide probabilistic sampling and better coverage of diverse data modes. In RS, diffusion-based approaches are gaining traction for tasks like cloud removal, where they can generate realistic optical images by conditioning on SAR inputs.

GenAI's application in remote sensing, particularly for multimodal data fusion, leverages these models' ability to bridge domain gaps. For instance, GANs and diffusion models can synthesize cloud-free optical imagery from SAR, preserving structural details while enhancing interpretability. However, challenges such as training instability in GANs and high computational costs in diffusion models persist, motivating ongoing research into hybrid architectures.

1.5 SAR-to-optical image translation

Optical imagery provides rich spectral information and can often be interpreted without expert knowledge, but it is highly sensitive to atmospheric conditions such as cloud cover, which frequently limits its usability. In contrast, SAR imagery offers all-weather, day-and-night, cloud-penetrating capability, though its complex backscatter characteristics, speckle noise, and lack of color make interpretation challenging even for experts [34]. Moreover, while two objects with identical structures may appear different in optical imagery due to their spectral responses, they can appear similar in SAR imagery, reflecting SAR's emphasis on structural rather than spectral properties [41]. Bridging this gap, SAR-to-optical image translation generates synthetic optical-like, cloud-free images from SAR data, combining the interpretability of optical imagery with the robustness of SAR. Defined as an image-to-image translation (I2I) task, this process is particularly valuable for applications that depend on consistent cloud-free optical information, including land-cover classification, disaster monitoring, and vegetation analysis [9].

The domain gap between SAR and optical imagery, however, poses significant challenges. SAR images exhibit speckle noise due to coherent interference, geometric distortions from side-looking geometry, and intensity-based representations that differ fundamentally from the reflectance-based multispectral bands of optical sensors [27]. Acquiring perfectly co-registered SAR–optical pairs is also not trivial, as spatial and temporal alignment must be ensured. According to the First Law of Geography, the closer the distance, the greater the correlation between ground objects, and the shorter the time interval, the smaller the change between features [19]. Achieving such conditions in practice is difficult. Furthermore, SAR and optical imaging principles differ fundamentally, causing certain land features (e.g., roads, playgrounds, airport runways) to appear differently in terms of spectral reflectance versus SAR backscatter. These discrepancies complicate the establishment of accurate mappings between the two modalities [19]. As a result, advanced translation methods are needed to preserve textures, colors, and edges in the generated optical images and to ensure reliable cloud removal.

Before the emergence of generative AI, SAR-to-optical translation relied on heuristic and classical approaches. Examples include pseudo-colorization of SAR channels or polarization composites, which improved interpretability but did not resemble true optical imagery. Multisensor fusion was also common, for instance combining SAR with prior cloud-free optical images through intensity–hue–saturation (IHS) transforms or wavelet-based methods [46]. While such approaches provided partial solutions, they depended on handcrafted features and often required costly machine learning pipelines, limiting their scalability and accuracy in addressing cloud contamination. Recent advances in deep learning have revolutionized this process: Generative Adversarial Networks (GANs) [42] and, more recently, diffusion models [45] have enabled direct end-to-end learning of mappings between SAR and optical domains. These generative approaches better capture pixel distributions and feature correlations, making them far more effective for generating cloud-free optical equivalents. More details on these methods are presented in Section 1.4.

Most existing studies on SAR-to-optical translation have concentrated on reconstructing the visible RGB bands of Sentinel-2 products, with some extending into the near-infrared (NIR) domain. Only a limited body of work has addressed the reconstruction of the complete multispectral range of all 13 bands. In the literature, the former is typically referred to as SAR-to-optical translation, whereas the latter is denoted as SAR-to-multispectral (SAR-to-MS) translation. For the purposes of this thesis, the term *optical* is used in its broader sense, encompassing the full spectral domain. Accordingly, SAR-to-optical will be used herein to denote translation tasks irrespective of the number of bands involved. This thesis specifically investigates the reconstruction of the complete set of 13 Sentinel-2 bands, with a central research question examining whether all bands can be reliably reconstructed and to what extent. A further question concerns whether a model trained on global datasets can generalize effectively to regional data, and how fine-tuning with region-specific samples may enhance performance. In this context, multispectral GAN-generated images must not only be visually realistic but also reproduce the radiometric resolution and spectral signatures of natural optical data, ensuring their usefulness for quantitative remote sensing and effective cloud removal.

In this light, SAR-to-optical translation can be regarded as a cloud removal strategy that bridges the gap between robust SAR acquisitions and interpretable optical imagery, and this thesis explores its potential by extending the task to the full multispectral domain of Sentinel-2 data.

1.6 Application and Relevance to KIWA Project

The KIWA project [5] *(German: KI-basierte Waldüberwachung – Engl: AI-based Forest Monitoring)* addresses the growing ecological challenge of forest degradation and wildfire risk in Central Europe. Climate extremes, prolonged droughts, and pest infestations are severely affecting

[5] https://www.kiwa-projekt.de

coniferous forests, making them increasingly vulnerable to fire. Wildfires not only destroy ecosystems but also release the carbon previously sequestered by forests, thereby accelerating climate change. Current monitoring methods, such as aircraft patrols and stationary watchtowers, are resource-intensive, costly, and limited in performance [47].

To overcome these challenges, KIWA integrates artificial intelligence with advanced remote sensing technologies, particularly drones equipped with computer-vision systems, to improve early wildfire detection. The project's broader objectives include delivering high-resolution environmental data, providing decision support to emergency services, and supporting climate-resilient, biodiversity-rich forest management. KIWA thus exemplifies an AI "lighthouse" initiative with the ambition to serve as a transferable blueprint for forest monitoring systems across Germany and internationally [48].

Similar to the general challenges faced in remote sensing applications, the KIWA project requires gapless observation capabilities. For instance, recent KIWA-related research highlights the need for automated methods to delineate burned areas (BAs) and assess wildfire risks using remote sensing data [18]. These approaches primarily rely on optical indices such as NDVI, NBR, NDWI, and WFI, which require cloud-free multispectral imagery. However, cloud cover and wildfire smoke often make it difficult to obtain a continuous historic record of the areas of interest, which is critical in emergency services. As the authors state, *"excluding multi-temporal approaches per se for our KIWA workflow is not an option and not intended".*

In this context, the contribution of this thesis—translating SAR data into optical-like images using generative AI models—offers a direct benefit to KIWA. Once validated, the proposed methods can be integrated in the project workflow to enhance the spatial and temporal coverage of forest monitoring, even under cloudy or adverse weather conditions. By extending the availability of optical-equivalent data in near real-time, these approaches can improve the robustness of KIWA's burned-area mapping workflows, simplify mono-temporal analyses, and provide more reliable support for decision-making processes in emergency services. This integration has the potential to enhance KIWA's operational efficiency and transferability, supporting its mission to deliver automated, scalable, and accurate wildfire monitoring solutions.

2 Methodology

2.1 Datasets

2.1.1 SEN12-MS

This thesis relies exclusively on the SEN12MS dataset [49], curated by Schmitt et al.. SEN12MS is a large-scale, globally distributed benchmark explicitly designed to advance research in multimodal Earth observation and deep learning. It comprises 180,662 georeferenced image triplets, each consisting of (i) dual-polarized Sentinel-1 synthetic aperture radar (SAR) data in VV and VH polarization (σ^0 backscatter values in decibel scale), (ii) full Sentinel-2 multispectral imagery spanning all 13 bands, and (iii) MODIS land cover maps derived from the MCD12Q1 product and resampled to 10 m resolution. Each triplet is stored as a 256 × 256 pixel GeoTIFF at 10 m ground sampling distance, corresponding to a spatial coverage of approximately 2.56 × 2.56 km per patch.

The Sentinel-1 component originates from ground-range-detected (GRD) products acquired in interferometric wide swath (IW) mode. These data were radiometrically calibrated and orthorectified against SRTM or ASTER digital elevation models to ensure accurate geolocation. The Sentinel-2 imagery was curated using a cloud-free mosaicking workflow on Google Earth Engine: within each region of interest (ROI), multiple observations collected during a given meteorological season of 2017 were composited such that cloud-contaminated pixels were systematically excluded. This procedure ensured that every ROI is represented by seasonally consistent, nearly cloud-free multispectral data. Finally, the MODIS land cover maps were used to generate categorical reference layers; however, due to their relatively coarse native resolution (500 m), they are subject to spatial inaccuracies even after upsampling.

Importantly, all triplets underwent manual verification by a remote sensing expert. This revision step ensured that each patch is free from major artifacts, severe registration errors, or residual cloud contamination, thereby guaranteeing the dataset's quality and usability for machine learning tasks.

The ROIs were sampled globally across all inhabited continents and four meteorological seasons of 2017 to maximize spatial and temporal diversity. Nevertheless, it should be noted that the ROI selection was not purely random. In practice, locations were chosen to avoid large

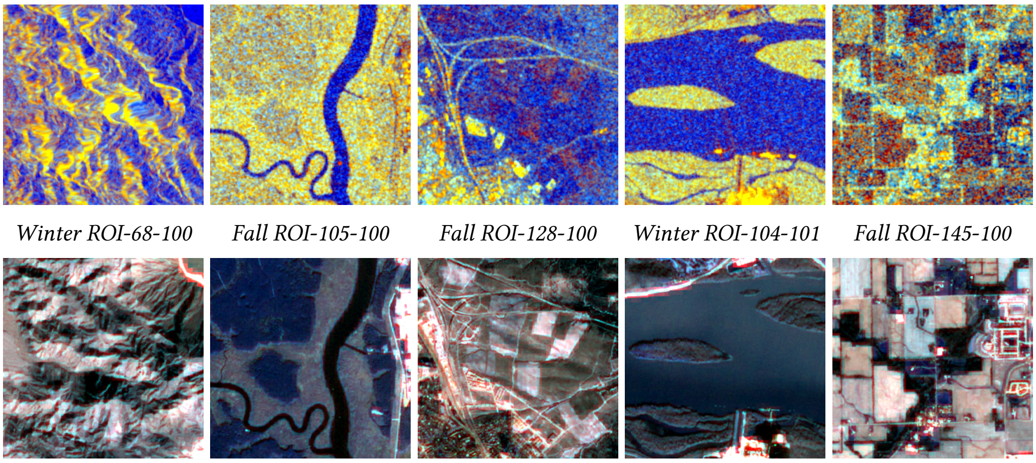

Figure 2.1: Sample pairs from the SEN12MS dataset. Top row: Sentinel-1 SAR patches (R: VV, G: VH, B: VV/VH). Bottom row: corresponding Sentinel-2 multispectral patches (only RGB bands).

homogeneous areas such as deserts or oceans and to ensure inclusion of diverse land cover classes. While this design improves the dataset's representativeness for a wide range of applications, it may introduce a bias toward heterogeneous landscapes and thus does not fully capture the true global distribution of land cover types.

For the purpose of this thesis, which addresses translation from SAR to multispectral optical imagery, only the Sentinel-1 and Sentinel-2 modalities are employed. The MODIS land cover products included in SEN12MS are disregarded, as they are not directly relevant to the translation task.

2.1.2 SEN12 datasets Family

SEN12MS is part of a broader line of datasets developed to foster multimodal remote sensing research. Its direct predecessor, SEN1-2 [50], curated by the same research group, contained approximately 282,000 paired patches of Sentinel-1 VV data and Sentinel-2 RGB composites. While groundbreaking in bridging SAR and optical domains, SEN1-2 lacked georeferencing, full spectral coverage, and multi-polarization SAR, limiting its applicability for remote sensing research beyond proof-of-concept image translation.

SEN12MS addressed these limitations by introducing full multispectral coverage, dual-polarized SAR, geocoded products, and auxiliary land cover labels, making it a comprehensive multimodal benchmark. Building upon this foundation, the dataset family has since been extended. SEN12MS-CR [51] added temporally matched cloudy and cloud-free Sentinel-2 imagery alongside Sentinel-1 data, enabling the development and benchmarking of cloud removal methods under realistic atmospheric conditions. Subsequently, SEN12MS-CR-TS [52] expanded the concept into the temporal domain, providing year-long multimodal time series with 30 co-registered Sentinel-1 and Sentinel-2 acquisitions per ROI. This evolution reflects a progression from simplified SAR–optical pairs, to globally diverse multimodal data, to temporally rich resources designed for time-series analysis and robust cloud removal.

A comparsion of these different datasets is provided in Table 2.1. In this thesis, however, the focus remains on the SEN12MS dataset, leveraging its multimodal SAR and multispectral imagery for the study of SAR-to-optical translation.

Table 2.1: Comparison of datasets in the SEN12 family.

Aspect	**SEN1-2** [50]	**SEN12MS** [49]	**SEN12MS-CR** [51]	**SEN12MS-CR-TS** [52]
Year released	2018	2019	2021	2022
Main purpose	Proof-of-concept SAR–optical translation	Multimodal learning and data fusion	Cloud removal with real cloudy/clear pairs	Multi-temporal cloud removal (sequence models)
Modalities	S1 (VV), S2 (RGB)	S1 (VV,VH), S2 (13 bands), MODIS LULC	S1 (VV,VH), S2 (13 bands; cloudy & cloud-free)	S1 (VV,VH), S2 (13 bands; cloudy & cloud-free time series)
Georeferencing	Not georeferenced	Fully georeferenced	Fully georeferenced	Fully georeferenced
Spatial sampling	Global patch pairs (282k)	180,662 patch triplets across 2017 seasons	169 ROIs; >100k patch triplets	53 ROIs; 30 time steps per ROI
Temporal coverage	Single time-point	Seasonal (2017)	Seasonal with paired cloudy/clear	Year-long time series (2018)
Patch size	256×256 px	256×256 px	256×256 px	256×256 px
Notable limitations	RGB only; VV only; no geocoding	MODIS labels are coarse (upsampled)	Mono-temporal pairs (no full time series)	Fewer ROIs; large storage ($\sim$2 TB)

2.2 Models

2.2.1 Pix2Pix Model

The image translation task in this thesis is addressed using the *pix2pix* framework, introduced by Isola et al. [53]. Pix2pix is based on the concept of *conditional generative adversarial networks* (cGANs), which extend the original GAN formulation by conditioning both the generator and discriminator on an input image. In this setup, the generator G learns to map an input image x to an output image y, while the discriminator D learns to distinguish between real image pairs $\{x, y\}$ and synthesized pairs $\{x, G(x)\}$. This adversarial objective enforces that generated outputs are not only realistic but also structurally consistent with the given input.

Formally, the cGAN loss is defined as:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x}[\log(1 - D(x, G(x)))]. \tag{2.1}$$

To encourage fidelity to the target image, the adversarial loss is combined with an ℓ_1 reconstruction loss:

$$\mathcal{L}_{\ell_1}(G) = \mathbb{E}_{x,y}[\|y - G(x)\|_1]. \tag{2.2}$$

The final objective is then:

$$G^* = \arg\min_G \max_D \ \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{\ell_1}(G), \tag{2.3}$$

where λ balances realism and reconstruction accuracy. Following Isola et al., $\lambda = 100$ is typically used.

Generator architecture. The generator is implemented as a *U-Net* encoder–decoder [54]. Unlike a plain encoder–decoder, U-Net introduces skip connections between corresponding downsampling and upsampling layers, allowing low-level spatial details from the input to directly propagate to the output. This design is particularly effective in tasks where the input and output share spatial structures, as in SAR-to-optical translation.

Discriminator architecture. The discriminator follows a *PatchGAN* design, which classifies local $N \times N$ patches of an image as real or fake instead of operating on the entire image [53]. This approach emphasizes high-frequency correctness and enforces local realism, while the ℓ_1 loss ensures global structural coherence. The original work demonstrates that a patch size of 70×70 provides a good trade-off between quality and efficiency.

Optimization. Training alternates between updating D to improve its ability to classify real versus fake pairs, and updating G to fool D while minimizing the ℓ_1 distance to the target. The Adam optimizer [55] with learning rate 2×10^{-4} and momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$ is typically employed. Dropout is used at both training and inference time to introduce stochasticity, though in practice outputs remain largely deterministic.

Relevance to this work. The pix2pix framework provides a principled and general-purpose solution for image-to-image translation tasks. In the context of this thesis, it is employed to learn mappings from Sentinel-1 SAR inputs to Sentinel-2 multispectral optical outputs. The combination of adversarial and reconstruction losses, together with the U-Net generator and

PatchGAN discriminator, makes pix2pix particularly suitable for producing sharp, realistic, and structurally aligned multispectral predictions.

2.3 Evaluation Metrics

The effectiveness of SAR-to-optical image translation depends not only on the choice of translation models but also on the methods employed for quality assessment. Image Quality Assessment (IQA) serves two key purposes: (i) to objectively evaluate the quality of results produced by different models, and (ii) to guide the optimization of network architectures and algorithms [56].

In [56], five IQA metrics—SSIM, FSIM, MSE, LPIPS, and DISTS—were compared through image restoration experiments to identify suitable measures for SAR-to-optical translation. Their results showed that SSIM, MSE, and LPIPS consistently aligned with human perception, converged reliably, and effectively captured both structural and textural details, whereas FSIM often failed to capture fine details and DISTS exhibited instability. Consequently, SSIM, MSE, and LPIPS were recommended as complementary metrics for pixel-level fidelity, structural similarity, and perceptual quality. Nevertheless, as summarized in Table 2.2, SSIM, PSNR, and SAM remain the most widely used indicators in SAR-to-optical translation, fusion, and cloud removal tasks, while LPIPS and MSE appear far less frequently [19].

TODO: specify which metrics will be used and why

Metric	References	Frequency
Structural Similarity Index Measurement (SSIM) [57]	[10, 30, 33, 35, 39, 41, 58–63]	12
Peak Signal-to-Noise Ratio (PSNR) [64]	[10, 30, 31, 35, 37, 39, 58–63]	12
Spectral Angle Mapper (SAM) [65]	[8, 31, 35, 41, 58–62, 66]	11
Fréchet Inception Distance (FID) [67]	[30, 33, 39, 61, 62, 66]	6
Root Mean Square Error (RMSE)	[8, 10, 31, 41, 58]	5
Learned Perceptual Image Patch Similarity (LPIPS) [68]	[10, 35, 39, 63, 66]	5
Mean Absolute Error (MAE)	[31, 58]	2
Mean Square Error (MSE)	[37, 60]	2

Table 2.2: Common evaluation metrics for SAR-to-optical and cloud removal tasks.

SSIM The Structural Similarity Index (SSIM) [57] measures perceptual similarity by comparing local patterns of luminance, contrast, and structure between two images. Unlike pixel-wise errors, it models human visual sensitivity to structural distortions [39, 61], which is crucial for evaluating translated images. For two images x and y, SSIM is defined as

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{2.4}$$

where μ_x, μ_y are means, σ_x^2, σ_y^2 variances, and σ_{xy} the covariance. Values close to 1 indicate strong structural similarity. By focusing on local patterns of pixel intensities and their structural relationships, SSIM better reflects perceptual fidelity compared to raw pixel-difference metric

PSNR The Peak Signal-to-Noise Ratio (PSNR) quantifies the distortion between a reconstructed image and its reference. PSNR is directly related to the Mean Squared Error (MSE), measuring pixel-level fidelity by comparing the residual error to the maximum possible signal intensity. For two images x and y, PSNR is defined as

$$\mathrm{PSNR}(x, y) = 10 \cdot \log_{10} \left(\frac{MAX^2}{\mathrm{MSE}(x, y)} \right), \tag{2.5}$$

with

$$\mathrm{MSE}(x, y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2, \tag{2.6}$$

where x_i and y_i denote the pixel values of the generated and reference images, N is the total number of pixels, and MAX is the maximum pixel intensity (typically 255 for 8-bit images).

Higher PSNR values indicate lower distortion and better image quality, as they imply that the reconstructed image more closely approximates the reference. Despite its popularity for tasks such as denoising and compression, PSNR is limited by its purely pixel-wise formulation and often correlates weakly with human visual perception [39].

SAM The Spectral Angle Mapper (SAM), originally proposed by Kruse et al. [65] in 1993, is widely employed in remote sensing to evaluate the spectral fidelity of reconstructed images. SAM regards the spectrum of each pixel as a high-dimensional vector and quantifies similarity by measuring the angle between the generated and reference spectral vectors. For two spectral vectors x and y, SAM is defined as

$$\mathrm{SAM}(x, y) = \arccos \left(\frac{\langle x, y \rangle}{\|x\|_2 \cdot \|y\|_2} \right), \tag{2.7}$$

where $\langle x, y \rangle$ denotes the dot product and $\| \cdot \|_2$ is the Euclidean norm.

SAM is typically expressed in degrees, with smaller values indicating higher spectral similarity and less distortion. Since it only considers the direction of the spectral vectors and not their magnitude, SAM is invariant to changes in illumination, making it particularly suitable for remote sensing and multispectral image analysis [35]. In practice, the global SAM score is computed as the average angle across all pixels in the image.

LPIPS The Learned Perceptual Image Patch Similarity (LPIPS) metric was proposed by Zhang et al. [68] to provide a perceptual measure of image similarity that better aligns with human visual judgment. It compares feature activations from pretrained convolutional networks, thereby capturing high-level semantics and perceptual realism. For two images x and y, LPIPS is defined as

$$\text{LPIPS}(x, y) = \sum_{l} w_l \cdot \|f_l(x) - f_l(y)\|_2, \tag{2.8}$$

where $f_l(\cdot)$ denotes the feature representation in the l-th layer of the network and w_l is a learned weight.

By measuring differences in a deep feature space rather than raw pixel intensities, LPIPS reflects perceptual similarity and visual realism. Lower LPIPS values indicate that the generated image is closer to the reference in terms of human-perceived quality [10,39], making this metric particularly useful for evaluating the naturalness of translated images.

FID Introduced in 2018 by Heusel at el. [67], the Fréchet Inception Distance (FID) is a perceptual metric that evaluates the realism of generated images at the distributional level. Instead of comparing images pixel by pixel, FID measures the distance between the feature distributions of generated and reference images, extracted by a pretrained Inception network. Let (μ_r, Σ_r) and (μ_g, Σ_g) denote the mean and covariance of the reference and generated feature distributions, respectively. FID is defined as

$$\text{FID} = \|\mu_r - \mu_g\|_2^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}\right). \tag{2.9}$$

Lower FID values indicate closer alignment between generated and real image distributions. While LPIPS assesses pairwise perceptual similarity, FID captures distributional alignment, making the two metrics complementary.

Evaluation Protocol The evaluation of SAR-to-optical translation performance was conducted using PSNR, SSIM, and SAM, complemented by a perceptual metric (LPIPS or FID). PSNR quantifies pixel-level fidelity, SSIM assesses local structural similarity, and SAM measures spectral consistency across all bands, which is critical in multispectral applications. To additionally capture perceptual realism beyond pixel-wise statistics, a deep feature–based perceptual score was employed, with LPIPS enabling pairwise comparisons and FID providing distributional similarity. For outputs with more than three bands, perceptual metrics were computed on a fixed RGB composite for both reference and prediction, and this limitation was explicitly acknowledged. This combination of metrics provides a comprehensive assessment covering spatial fidelity, structural integrity, spectral accuracy, and perceptual quality.

Table 2.3: Summary of evaluation metrics for SAR-to-multispectral translation.

Metric	**Aspect Evaluated**	**Advantages**	**Limitations**
PSNR	Pixel-level fidelity via mean squared error ratio	Simple, widely used, interpretable in terms of noise/distortion	Correlates weakly with human perception; sensitive to pixel shifts
SSIM	Structural similarity (luminance, contrast, texture)	Captures perceptual structure better than PSNR; patch-based	Still intensity-based; limited correlation with perceptual realism
SAM	Spectral fidelity across bands	Invariant to illumination; critical for multispectral data integrity	Ignores spatial/structural context; only reflects spectral angle
LPIPS	Perceptual similarity using deep features (pairwise)	Aligns well with human judgment; sensitive to high-level semantics	Requires pretrained CNN; limited to 3-channel inputs unless adapted
FID	Distributional similarity in feature space	Evaluates realism of entire image sets; widely adopted in generative models	Requires large sample size; sensitive to preprocessing; assumes Gaussian feature distributions

Bibliography

[1] C. Elachi and J. van Zyl, “Introduction,” in *Introduction to the Physics and Techniques of Remote Sensing*. John Wiley & Sons, Ltd, 2021, ch. 1, pp. 1–18. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119523048.ch1

[2] C. Toth and G. Jóźków, “Remote sensing platforms and sensors: A survey,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 115, pp. 22–36, 2016, theme issue 'State-of-the-art in photogrammetry, remote sensing and spatial information science'. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271615002270

[3] E. Chuvieco, “Introduction,” in *Fundamentals of Satellite Remote Sensing: An Environmental Approach*. CRC Press, 2020. [Online]. Available: https://www.taylorfrancis.com/books/mono/10.1201/9780429506482/fundamentals-satellite-remote-sensing-emilio-chuvieco

[4] European Space Agency (ESA). (2024) Copernicus: Sentinel missions. Accessed: 3 September 2025. [Online]. Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus

[5] European Space Agency. (2024) Copernicus: Sentinel missions. Accessed: 3 September 2025. [Online]. Available: https://sentinels.copernicus.eu/missions

[6] European Space Agency (ESA). (2025) Sentiwiki – copernicus sentinels. Accessed: 6 September 2025. [Online]. Available: https://sentiwiki.copernicus.eu/web/

[7] Z. Wang, L. Zhao, J. Meng, Y. Han, X. Li, R. Jiang, J. Chen, and H. Li, “Deep learning-based cloud detection for optical remote sensing images: A survey,” *Remote Sensing*, vol. 16, no. 23, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/23/4583

[8] C. Grohnfeldt, M. Schmitt, and X. Zhu, “A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images,” in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 1726–1729.

[9] 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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620301398

[10] J. Ning, L. Xie, J. Yin, and Y. Liu, “Cloud removal advances: A comprehensive review and analysis for optical remote sensing images,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 15 914–15 930, 2025.

[11] Y. Zhang, B. Guindon, and J. Cihlar, “An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images,” *Remote Sensing of Environment*, vol. 82, no. 2, pp. 173–187, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425702000342

[12] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1956–1963.

[13] 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.

[14] F. Xu, Y. Shi, P. Ebel, L. Yu, G.-S. Xia, W. Yang, and X. X. Zhu, “Glf-cr: Sar-enhanced cloud removal with global-local fusion,” 2022. [Online]. Available: https://arxiv.org/abs/2206.02850

[15] 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.

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

[17] J. N. Mvogo, W. A. V. Noumsi, and P. B. Wirba, “Exploration of machine learning techniques for cloud removal and gap filling on sentinel-2 time series images for better exploitation in far north cameroon,” *Discover Applied Sciences*, vol. 7, no. 8, p. 843, 2025. [Online]. Available: https://doi.org/10.1007/s42452-025-07026-w

[18] P. Hofmann, N. Trofanisin, and S. Wöllmann, “Automatic delineation of burned forest areas from satellite imagery to analyze and manage wildfires,” in *2024 14th International Conference on Advanced Computer Information Technologies (ACIT)*, 2024, pp. 766–771.

[19] Q. Xiong, G. Li, X. Yao, and X. Zhang, “Sar-to-optical image translation and cloud removal based on conditional generative adversarial networks: Literature survey, taxonomy, evaluation indicators, limits and future directions,” *Remote Sensing*, vol. 15, no. 4, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/4/1137

[20] T. Toizumi, S. Zini, K. Sagi, E. Kaneko, M. Tsukada, and R. Schettini, “Artifact-free thin cloud removal using gans,” in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3596–3600.

[21] J. Li, Z. Wu, Z. Hu, J. Zhang, M. Li, L. Mo, and M. Molinier, “Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 373–389, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620301787

[22] Q. Yang, G. Wang, Y. Zhao, X. Zhang, G. Dong, and P. Ren, "Multi-scale deep residual learning for cloud removal," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 4967–4970.

[23] D. Ma, R. Wu, D. Xiao, and B. Sui, "Cloud removal from satellite images using a deep learning model with the cloud-matting method," *Remote Sensing*, vol. 15, no. 4, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/4/904

[24] R. Jaisurya and S. Mukherjee, "Aglc-gan: Attention-based global-local cycle-consistent generative adversarial networks for unpaired single image dehazing," *Image and Vision Computing*, vol. 140, p. 104859, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885623002330

[25] Y. Yan, Y. He, N. Su, G. He, and H. Fu, "Pnbt-cr: A cloud removal method for ship detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.

[26] H. Ye, H. Xiang, and F. Xu, "Cycle-gan network incorporated with atmospheric scattering model for dust removal of martian optical images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.

[27] 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.

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

[29] L. Abady, M. Barni, A. Garzelli, and B. Tondi, "GAN generation of synthetic multispectral satellite images," in *Image and Signal Processing for Remote Sensing XXVI*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, L. Bruzzone, F. Bovolo, J. A. Benediktsson, F. Bovenga, C. Notarnicola, N. Pierdicca, and E. Santi, Eds., vol. 11533, Sep. 2020, p. 115330L.

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

[31] F. N. Darbaghshahi, M. R. Mohammadi, and M. Soryani, "Cloud removal in remote sensing images using generative adversarial networks and sar-to-optical image translation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–9, 2022.

[32] M. Zhang, J. Xu, C. He, W. Shang, Y. Li, and X. Gao, "Sar-to-optical image translation via thermodynamics-inspired network," 2023. [Online]. Available: https://arxiv.org/abs/2305.13839

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

[34] X. Bai and F. Xu, "Sar to optical image translation with color supervised diffusion model," 2024. [Online]. Available: https://arxiv.org/abs/2407.16921

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

[36] M. Wang, S. Hu, Y. Song, and Y. Shi, “Sar-decr: Latent diffusion for sar-fused thick cloud removal,” *Remote Sensing*, vol. 17, no. 13, 2025. [Online]. Available: https://www.mdpi.com/2072-4292/17/13/2241

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

[38] 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.

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

[40] D. A. Abuhani, I. Zualkernan, R. Aldamani, and M. Alshafai, “Generative artificial intelligence for hyperspectral sensor data: A review,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 6422–6439, 2025.

[41] P. Liu, J. Li, L. Wang, and G. He, “Remote sensing data fusion with generative adversarial networks: State-of-the-art methods and future research directions,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 2, pp. 295–328, 2022.

[42] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” 2014. [Online]. Available: https://arxiv.org/abs/1406.2661

[43] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” 2014. [Online]. Available: https://arxiv.org/abs/1411.1784

[44] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” 2020. [Online]. Available: https://arxiv.org/abs/1703.10593

[45] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” 2020. [Online]. Available: https://arxiv.org/abs/2006.11239

[46] W. Zhang and M. Xu, “Translate sar data into optical image using ihs and wavelet transform integrated fusion,” *Journal of the Indian Society of Remote Sensing*, vol. 47, no. 1, pp. 125–137, 2019. [Online]. Available: https://doi.org/10.1007/s12524-018-0879-7

[47] Technische Hochschule Deggendorf. (2025) Zentrum für Angewandte Forschung (ZAF). Accessed: Sep. 19, 2025. [Online]. Available: https://zaf.th-deg.de/

[48] KIWA Project. (2025) KI-basierte Waldüberwachung – AI-based Forest Monitoring. Accessed: Sep. 19, 2025. [Online]. Available: https://www.kiwa-projekt.de/eng/home

[49] 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.

[50] 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.

[51] P. Ebel, A. Meraner, M. Schmitt, and X. X. Zhu, "Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5866–5878, 2021.

[52] P. Ebel, Y. Xu, M. Schmitt, and X. X. Zhu, "Sen12ms-cr-ts: A remote-sensing data set for multimodal multitemporal cloud removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[53] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2018. [Online]. Available: https://arxiv.org/abs/1611.07004

[54] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597

[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017. [Online]. Available: https://arxiv.org/abs/1412.6980

[56] J. Zhang, J. Zhou, M. Li, H. Zhou, and T. Yu, "Quality assessment of sar-to-optical image translation," *Remote Sensing*, vol. 12, no. 21, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/21/3472

[57] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, pp. 600 – 612, 05 2004.

[58] X. Xiang, Y. Tan, and L. Yan, "Cloud-guided fusion with sar-to-optical translation for thick cloud removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.

[59] R. Liu, S. Meng, Y. Peng, and X. Tian, "Transfusion-cr: Two-phase sar-to-optical translation and deep feature fusion for cloud removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–11, 2024.

[60] C. Li, X. Liu, and S. Li, "Transformer meets gan: Cloud-free multispectral image reconstruction via multisensor data fusion in satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[61] W. Zhao, N. Jiang, X. Liao, and J. Zhu, "Hvt-cgan: Hybrid vision transformer cgan for sar-to-optical image translation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–17, 2025.

[62] Y. Chen, Z. Zhu, Y. Huang, P. Wang, B. Huang, and M. D. Mura, “Msf: A multi-scale fusion generative adversarial network for sar-to-optical image translation,” in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2024, pp. 9058–9061.

[63] Z. Guo, J. Liu, Q. Cai, Z. Zhang, and S. 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*, vol. 17, pp. 14 454–14 470, 2024.

[64] A. Tanchenko, “Visual-psnr measure of image quality,” *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 874–878, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047320314000091

[65] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, “The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data,” *Remote Sensing of Environment*, vol. 44, no. 2, pp. 145–163, 1993, airbone Imaging Spectrometry. [Online]. Available: https://www.sciencedirect.com/science/article/pii/003442579390013N

[66] S.-H. Kim and D. Chung, “Conditional brownian bridge diffusion model for vhr sar to optical image translation,” *IEEE Geoscience and Remote Sensing Letters*, vol. 22, p. 1–5, 2025. [Online]. Available: http://dx.doi.org/10.1109/LGRS.2025.3562401

[67] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” 2018. [Online]. Available: https://arxiv.org/abs/1706.08500

[68] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” 2018. [Online]. Available: https://arxiv.org/abs/1801.03924