



Reconstruction of vegetation indices using multi-source images and deep learning - A case of study in the Netherlands.

SANTIAGO ANDRES HIDALGO SULCA
Enschede, The Netherlands, July 2023

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfillment of the requirements for the degree of Master of Science in Spatial Engineering.

SUPERVISORS:
Dr. M. Khodadadzadeh
Dr.ir. R. de By

THESIS ASSESSMENT BOARD:
Prof.Dr. R. Zurita Milla
Dr. C. Paris (External Examiner, ITC)

DISCLAIMER

This document describes work undertaken as part of a program of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the author's sole responsibility and do not necessarily represent those of the faculty.

ABSTRACT

The projected increase in the global population by 2050, coupled with the impacts of climate change, presents a significant challenge of meeting global food demand while minimizing the risk of hunger. To achieve sustainable development and address the second Sustainable Development Goal of zero hunger, advancements in technology, policy, and governance are crucial. The European Union's Common Agricultural Policy (CAP) introduced the "greening" initiative in 2013 to promote sustainable agricultural productivity. However, the effectiveness of these measures in improving biodiversity and ecosystem services has been limited due to various challenges, including heterogeneous management rules, lack of spatial targeting, and costly field inspections.

To overcome these challenges, the utilization of Earth Observation data, particularly from the Sentinel-1 and Sentinel-2 satellites, has emerged as a promising solution. For example, the Sen4CAP project aims to develop algorithms and products based on Sentinel-1 and Sentinel-2 time series data to monitor agricultural activities at the parcel level, thereby supporting the greening policy. While vegetation indices (VIs), such as the normalized difference vegetation index (NDVI), derived from Sentinel-2 data, are commonly used to analyze vegetation properties, cloud cover poses a significant obstacle. Conversely, the use of Sentinel-1 (SAR) data provides advantages such as weather independence but requires interpretation due to signal complexity. Combining the strengths of both sensors has shown promising results in tasks such as mowing detection, crop monitoring, and crop mapping, although challenges persist.

Deep learning approaches, including generative adversarial networks (GANs), have gained popularity in remote sensing applications by demonstrating their potential in reconstructing missing data. A Multi-Temporal Conditional Generative Adversarial Network (MTcGAN) approach that combines SAR and optical data has been proposed to reconstruct VIs. This method considers two sensing times, utilizing SAR data from t_1/t_2 and optical data at t_1 to simulate optical data at t_2 . Additionally, studies employing GANs for image reconstruction have shown improved performance across different crop types.

The experiments conducted in this study involve the reconstruction of VIs using optical and SAR data in the Flevoland province in the Netherlands. Experiment A focuses on reconstructing optical bands at different times and subsequent VIs, while Experiment B aims to directly reconstruct VIs using VIs and SAR data. These experiments are compared with similar research conducted in the field.

The evaluation of results is performed using metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), along with visual evaluation metrics. In Experiment A, the RGB bands yielded higher evaluation metric values compared to the NIR band, subsequently affecting the calculation of VIs derived from them. Visual evaluation of RGB compositions demonstrated the model's potential to generate data under cloud cover in unseen data, albeit with certain limitations and mispredictions. Experiment B, conducted with a smaller training sample, exhibited unsatisfactory performance.

Notably, variations in the size of the training dataset, sensing interval between t_1 and t_2 , and pre-processing techniques were observed among reference studies. Reducing the sensing interval has the potential to enhance performance while using data from different locations but within a close sensing time could also yield improved results. Although the study area lacked cloud-free images during the more developed phenology stages, using data from partially clouded images remains a viable option, with the extent of feasibility yet to be determined. Consequently, the VIs did not exhibit significant differences between crops, with NIR predictions still proving to be the least accurate. However, patches of cropland exhibited fewer mispredictions, indicating that contextual information about crops within predominantly cropland areas enhances prediction accuracy.

ACKNOWLEDGEMENTS

I would like to begin by expressing appreciation to my thesis supervisors Mahdi and Rolf, for their invaluable guidance, expertise, and continuous encouragement. Their insightful feedback, constructive criticism, and dedication to my research have significantly shaped this thesis. I am truly grateful for their mentorship and the knowledge I have gained under their supervision.

I would also like to acknowledge ITC for the valuable knowledge and skills I obtained during this master's, as well as for the financial assistance provided by ITC spatial engineering scholarship has been crucial in enabling me to focus on my research without the burden of financial constraints. I am deeply grateful for their investment in my education and their belief in my potential.

I thank my parents Adriana and Carlos, my little brother José, and the rest of my beloved family for their steady support throughout my academic journey. Their love, encouragement, and belief in me have been essential in every aspect. I am forever grateful for their sacrifices and the opportunities they have provided me.

Lastly, I extend my thanks to my spatial engineering family Arunima, Aulia, Jafeth, and Sean for all the great moments we had in this journey we shared together. To Masis, Dennis, Carlos, and Raphaël, great friends that made me feel I was at home during my stay in the Netherlands.

TABLE OF CONTENTS

1.	Introduction.....	1
1.1.	Background.....	1
1.2.	Problem statement and research gap	3
1.3.	Wickedness in the problem	4
2.	Research Objectives	5
2.1.	Main objective.....	5
2.2.	Sub objectives.....	5
3.	Study area	6
4.	Literature review	7
4.1.	SEN4CAP products.....	7
4.2.	Vegetation indices	8
4.3.	Crop phenology	9
4.4.	Deep learning	9
4.5.	Generative Adversarial Networks (GANs).....	11
5.	Methodology.....	14
5.1.	Methodology overview.....	14
5.2.	Cloud cover assessment and crop vegetation growth stages recognition	15
5.3.	Data pre-processing	17
5.4.	Multi-Temporal Conditional Generative Adversarial Network MTcGAN	19
5.5.	Evaluation metrics.....	21
6.	Results and discussion.....	23
6.1.	Benchmark dataset	23
6.2.	Training	26
6.3.	Evaluation metrics.....	27
6.4.	Visual evaluation.....	31
6.5.	Limitations	36
7.	Conclusions.....	38
8.	Recommendations	40
	Appendix 1: BBCH	41
	Appendix 2: Input data.....	42

LIST OF FIGURES

Figure 1 Study area and target crops	6
Figure 2 Typical CNN architecture	10
Figure 3 GAN architecture	12
Figure 4 Cloudiness assessment and further Sentinel data collection	14
Figure 5 Experiments overview flowchart	15
Figure 6 Cloudiness over the whole Sentinel-2 scene in 2021	15
Figure 7 Cloudiness over the study area in 2021	16
Figure 8 Cloud percentage in Sentinel-2 whole scene images	16
Figure 9 Crop growing stages and cloud-free Sentinel-2 images timeline 2021	17
Figure 10 Data collected: a) RGB Sentinel-2 in t1, b) VV Sentinel-1 in t1, c) VH Sentinel-1 in t1, d) RGB Sentinel-2 in t2, e) VV Sentinel-1 in t2, f) VH Sentinel-1 in t2	18
Figure 11 Ground truth vegetation indices. a) NDVI t1, b) GNDVI t1, c) EVI t1, d) SAVI t2, e) NDVI t2, f) GNDVI t2, g) EVI t2, h) SAVI t2	18
Figure 12 MTcGAN flowchart	19
Figure 13 Generator CNN architecture	20
Figure 14 Discriminator CNN architecture	21
Figure 15 Histograms of input data	24
Figure 16 Pairs of RGB compositions for t1 and t2	25
Figure 17 Training loss	26
Figure 18 PSNR training dataset	27
Figure 19 PSNR validation dataset	27
Figure 20 SSIM training dataset	27
Figure 21 SSIM validation dataset	27
Figure 22 PSNR column chart per band in specific crops	29
Figure 23 SSIM column chart per band in specific crops	30
Figure 24 RMSE column chart per band in specific crops	30
Figure 25 Ground truth and simulated patches per band	31
Figure 26 Sentinel-2 real (left) and simulated (right) cloud penetration	33
Figure 27 Vegetation indices calculated with ground truth data and simulated bands	34
Figure 28 Vegetation indices calculated with ground truth data and directly simulated vegetation indices	34

LIST OF TABLES

Table 1 SEN4CAP products and compliance with CAP.....	7
Table 2 SEN4CAP products validation in the Netherlands.....	8
Table 3 Vegetation indices	8
Table 4 Sensing time of Sentinel-1 and Sentinel-2 pairs.	17
Table 5 Evaluation metrics	21
Table 6 Experiments type in comparable research	23
Table 7 Sensing dates of comparable research.	25
Table 8 Number of parameters in the model.....	26
Table 9 Train, test, and validation datasets number of patches in comparable research.....	27
Table 10 Experiment A: evaluation metrics testing dataset per band.....	28
Table 11 Experiment A: comparison of evaluation metrics with Reference A-I.....	28
Table 12 Experiment A: Comparison of evaluation metrics with Reference B-I.....	28
Table 13 Experiment A: Comparison of evaluation metrics with Reference C-I.....	28
Table 14 Evaluating metrics on reconstructed VIs calculated from predicted data.....	29
Table 15 Evaluation metrics in different crops	29
Table 16 Experiment B evaluation metrics	30
Table 17 Ground truth and reconstructed RGB compositions comparison for homogeneous and heterogenous areas	32
Table 18 Ground truth and reconstructed RGB compositions comparison.....	35

1. INTRODUCTION

1.1. Background

By 2050, the world's population is expected to reach 9.7 billion people (UN, 2019). Considering climate change effects, global food demand and risk of hunger are expected to increase up to 62% and 30% respectively (van Dijk et al., 2021). Particularly, crop demand tends to increase up to 110% by then (Tilman et al., 2011). As global food production is an integral part of sustainable development, improvements in technology, policy, and governance are imperative to achieve zero hunger, the second of the Sustainable Development Goals (SDGs) (UN, 2015).

The Netherlands is the second-largest agriculture exporter in the world, right behind the United States of America, which is 270 times larger (Viviano, 2017). Agriculture and horticulture represent around 10% of the Dutch economy and employment, approximately €65 billion are received from exports yearly by this industry (LNV, 2021). The surface of the Netherlands is used 54% by farmland, corresponding to 2.2 million hectares (CBS, 2021). Dutch agriculture is reaching ecological limits as in the whole global food system, the nitrogen problem and climate change are concerning issues among the different stakeholders (Hillsdon, 2019).

The European Union's Common Agricultural Policy (CAP) is a significant initiative that strives to enhance sustainable agricultural productivity while securing a decent standard of living for EU farmers. In 2013, as income support for farmers, the “greening” initiative was introduced (European Commission, 2013) aiming to give direct payments to the farmers based on adopting three measures: maintaining permanent grassland, crop diversification, and maintaining an ecological focus area (EFA). The Integrated Administration and Control System (IACS) manages the budget designated for the payments, and through paying agencies of each EU member state, the activity on the parcels is evaluated.

In an assessment of the impacts of the reform of CAP-2013 from an ecological-economic perspective (Hristov et al., 2020), the authors indicated that the effectiveness of the “greening” initiative and EFA has limitations to improving biodiversity and ecosystem services. For instance, broad detail in measures decreases the potential, some farmers already had 5% of their land covered by nitrogen-fixing crops so any change was produced. The heterogeneity of management rules in different locations, like permitting uncropped field margins tilling and use of herbicides. The weight factor for margins does not have an ecological motivation and the contribution in improvements for greening is not achieved. The lack of spatial targeting allowed the farmers to allocate the EFA in the less productive areas and therefore there was not a significant change. Additionally, land property adjustment also allowed the farmers to reduce EFA per property. Finally, field inspections to decide on payments, which is costly, time-consuming, and prone to errors (European Court of Auditors, 2020).

Since 2017, Sentinel-1 (Synthetic Aperture Radar) (SAR) and Sentinel-2 (optical sensor) have provided free high-resolution images, with the potential to monitor agricultural activities, the spatial and temporal resolution of these permit the monitoring over even heterogeneous areas characterized by small farm holders (Campos-Taberner et al., 2019). The legal framework of the CAP of 2013 was reformed in 2020 allowing ‘checks by monitoring’, the reform promotes the use of Earth Observation (EO) for crop monitoring and evaluation of cross-compliance, particularly encouraging the use of Sentinel missions (European Union, 2018).

Sen4CAP is a project that aims to develop algorithms and products based on Sentinel-1 and Sentinel-2 time series to monitor activities of interest for compliance of the CAP at the parcel level defined geometrically in the Land Parcel Identification System (LPIS). The products developed to support the greening policy consist of crop type maps, grassland mowing detection, vegetation status indicator, and agricultural practices monitoring; vegetation indices and SAR-derived time series are fundamental inputs for the development of these products (Sen4CAP, 2020). Six pilot EU members are part of the Sen4CAP service adopted by different paying agencies, including the Netherlands with the Netherlands Enterprise Agency (RVO). Spain, Belgium, and Malta officially apply Sen4CAP for checks in some paying agencies, and Denmark for the whole country (European Court of Auditors, 2020).

Vegetation indices (VIs) quantify specific properties of vegetation by analysing the reflectance of vegetation in different wavelengths. Sentinel-2 with its good spectral resolution, enables the calculation of VIs, which are sensitive to factors like growth stages, seasons, biomass, soil conditions, canopy structure, and further environmental variables (Huang et al., 2021). The normalized difference vegetation index (NDVI) has become the most popular one because of its robustness in the examination of vegetation properties (Huang et al., 2021). Nonetheless, one of the major challenges for the application of Sentinel-2 for VIs is the weather conditions. Clouds can obstruct the satellite's view of the Earth's surface and create a gap in the time series needed for accurate crop monitoring. van der Wal et al. (2013) point out that clear optical imagery during the crop growing season in the Netherlands is available only 20% of the time due to cloud cover. Nonetheless, Pfeifroth et al. in 2018 analysed the trends and variability of surface solar radiation based in satellite data in Europe, they found a positive trend in eastern and north-western Europe, and concluded it is due to a change in the cloud cover.

The use of microwaves in collecting data for Sentinel-1 with C-Band enables the system to be unaffected by weather conditions or daylight and allows the acquisition of data measuring backscattered signal in amplitude and phase. The intensity of backscattering differs on wavelength, polarization, and incidence angle. Due to the Sentinel-1 characteristics, it makes it an appropriate sensor for agricultural purposes, it is sensitive to surface parameters related to surface roughness, geometry, and dielectric properties (Sener et al., 2021). However, the complex interaction of the signal with the surface, the variability of soil moisture, and the intrinsic speckle of SAR make interpretation difficult (Campos-Taberner et al., 2019).

There are advantages and limitations when using a single-source sensor to develop the Sen4CAP products. In this context, an accuracy of 77 % of mowing events detection has been achieved using NDVI time series at regional scales (Kolecka et al., 2018). The main challenge still being cloud cover, according to the authors, cloud masking is critical and a sudden drop in NDVI can be confused as a mowing event and overestimate the detection, also during favourable weather conditions for grass growth detections can be skipped. The combined use of Sentinel-1 and Sentinel-2 achieved an 85% detection rate and a 73% precision (F1-score 79%) for mowing detection through Europe (De Vroey et al., 2022). In the case of crop type mapping, it has achieved a quite good overall accuracy of 94% in Spain. Nevertheless, the lack of cloud-free Sentinel-2 images could be of relevance in this task for different areas (Campos-Taberner et al., 2019). Crop monitoring using Sentinel-1 only was explored with different crops in the Netherlands to identify key dates for crop production (Khabbazan et al., 2019). However, harvest detection was particularly difficult to identify for potato and sugar beet, and therefore needed to be supported with NDVI data.

In a review of the reconstruction of missing information in remote sensing (Shen et al., 2015), the authors traditionally classified the methods into spatial, spectral, and temporal methods. These methods have advantages and disadvantages, in general, spatial-based methods are suitable when only few pixels are missing and are unable to accurately restore the underlying spatial patterns of distinct objects (Shen et al., 2015). Spectral-based methods utilize multispectral data to reconstruct the missing components, they are

most effective when the data is acquired under optimal conditions, such as cloud-free observations and properly calibrated sensors (Gao et al., 2020). Although the multitemporal approach can yield improved outcomes in cases of significant missing data, it proves ineffective when sudden changes occur and relies heavily on the availability of cloud-free images (Gao et al., 2020).

SAR and optical remote sensing exhibit fundamental disparities in their imaging principles, thus they are not directly correlated (Li et al., 2020). Remote sensing image fusion consist in using an algorithm that combines two or more images from different sensors to obtain a new one in a synergistic way (Kulkarni & Rege, 2020). SAR's inability to provide spectrally resolved measurements poses a significant challenge in ensuring the accuracy and reliability of the obtained spectrum by image fusion (Xiong et al., 2021).

The use of deep learning approaches for remote sensing applications has increased in recent years, particularly for digitally removing clouds because of its strong nonlinear fitting ability, also the combination with SAR auxiliary data presents a big advantage over optical auxiliary data (Gao et al., 2020) such as the use of Landsat-8 data in the Sentinel-2 time series in Sen4CAP methods. There has been significant interest in utilizing generative adversarial networks (GANs) for reconstructing missing data in this field, it has shown great promise in generating images with high-frequency features (X. Zhang et al., 2021).

A Multi Temporal Conditional Generative Adversarial Network (MTcGAN) approach using both SAR and optical data at different times has been proposed by He & Yokoya, (2018) to generate cloud-free optical images and track vegetation changes. It consists of using Sentinel-1 and Sentinel-2 images at t1 and Sentinel-1 data at t2 to predict Sentinel-2 bands at t2. Dumeur et al. (2021) utilized this method to explicitly address cropland changes after wildfire events where crops presented changes between both acquisition times. Sener et al. (2021) experimented with GAN to generate VIs for crop classification, simulated and real VIs were different for the different crops, which suggests the model can perform differently in diverse crops.

The novelty in this research lies in the generation of Sentinel-2 bands and reconstruction of VIs from them, as well as the direct generation of VIs using MTcGAN method. The evaluation of the results encompasses both the entire model performance and its specific application to three crops for direct human consumption: sweet maize, potato, and sugar beet.

1.2. Problem statement and research gap

The agricultural sector in the Netherlands, one of the world's leading producers, is facing significant challenges due to climate change. Extreme weather events such as wet or dry periods account for 70% of yield anomalies over the past 27 years and it is expected to persist (Van Oort et al., 2023). Biodiversity in the is highly threatened, intense agricultural production, urbanization, pollution, and use of chemical fertilizers have rapidly modified land use and put 40% of the species of the Netherlands on the Red List (Sanders et al., 2019). Moreover, the nitrogen crisis, which contributes to climate change has further exacerbated these challenges.

The CAP 2013 established the “greening” initiative aiming to improve biodiversity and reduce the environmental impact of agriculture, nonetheless the expected outputs could not be reached due to the lack of spatial information (European Court of Auditors, 2020). Sen4CAP came out after the CAP 2020 reform, aiming to develop products useful in agricultural monitoring and management, concretely the greening measures aim to promote environmentally sustainable practices and enhance ecological performance.

Sen4CAP products heavily rely on optical data time series, which may be limited by factors such as cloud cover and atmospheric conditions, the current method to fill missing data is a temporal resampling method that heavily depends on cloud-free images. Also, the use of Landsat-8 images as auxiliary data still

must deal with the cloud cover. Van der Wal et al. (2013) concluded that clear optical imagery during the crop growing season in the Netherlands is available only 20% of the time due to cloud cover. Although cloudiness tends to decrease in the next years (Pfeifroth et al., 2018), there is a research gap regarding the potential benefits of predicting VIs using alternative data sources, such as synthetic aperture radar (SAR) imagery as auxiliary data, and how it can improve the accuracy and usefulness of Sen4CAP products. The effectiveness of using a deep learning approach, specifically a Generative Adversarial Network (GAN), with SAR data to predict optical vegetation indices and enhance Sen4CAP products has not been broadly explored. This research gap necessitates investigating the potential improvements and benefits of such an approach. Additionally, CAP requires reliable and precise monitoring methods to ensure effective and reliable implementation.

1.3. Wickedness in the problem

Wickedness in a problem considers two dimensions: knowledge and stakeholders. The lack of knowledge and disagreement among the different stakeholders makes a wicked problem. The environmental problems and their consequences have been widely studied; the policies built around this can increase the discomfort for some stakeholders. CAP 2013 aimed to reduce the environmental impact in Europe from the agriculture industry. Academia proved this reform not to be producing the expected outputs due to the lack of the spatial component, farmers obtained benefits without any improvement in biodiversity or environmental conditions (Hristov et al., 2020). CAP 2020 addressed this by promoting the use of earth observation and geoinformation science through the Sen4CAP project. This research considers the wickedness of the problem from the knowledge gap in the use of generative deep learning methods for improving the density of optical data for the implementation of CAP.

2. RESEARCH OBJECTIVES

2.1. Main objective

The main objective of this research is to reconstruct vegetation indices using optical and SAR imagery data with a deep-learning model.

2.2. Sub objectives

1. To prepare a benchmark dataset that can be used to train, validate, and test the proposed method.
 - Which data is available in the study area given cloud cover and crop growing stages?
2. To generate optical bands with a GAN model for SAR-optical translation.
 - How do generated bands differ from ground truth data?
 - Can the generated bands be utilized for vegetation indices calculations?
3. To generate vegetation indices with a GAN model for SAR-optical translation directly.
 - How different are the generated vegetation indices from ground truth data?
4. To evaluate the performance of the fusion models for different crops.
 - Does the model perform better for specific crops?

3. STUDY AREA

The study area is the Flevoland province, a reclaimed area of 2,412 km² with an average elevation of 1m, which was drained in 1957 primarily for agricultural purposes. The data used to map the target crops for human consumption was obtained from Pdok, Basisregistratie Gewaspercelen (BRP)¹, the geometry of the agricultural plots is based on the Agricultural Area of the Netherlands (AAN)², the user annually updates the crop growing under their domain, the reference date is May 15 of 2021. Particularly, potatoes, sugarbeets, and maize for human direct consumption are considered as target crops for this study. Figure 1 shows the study area, location in the Netherlands and target crops.

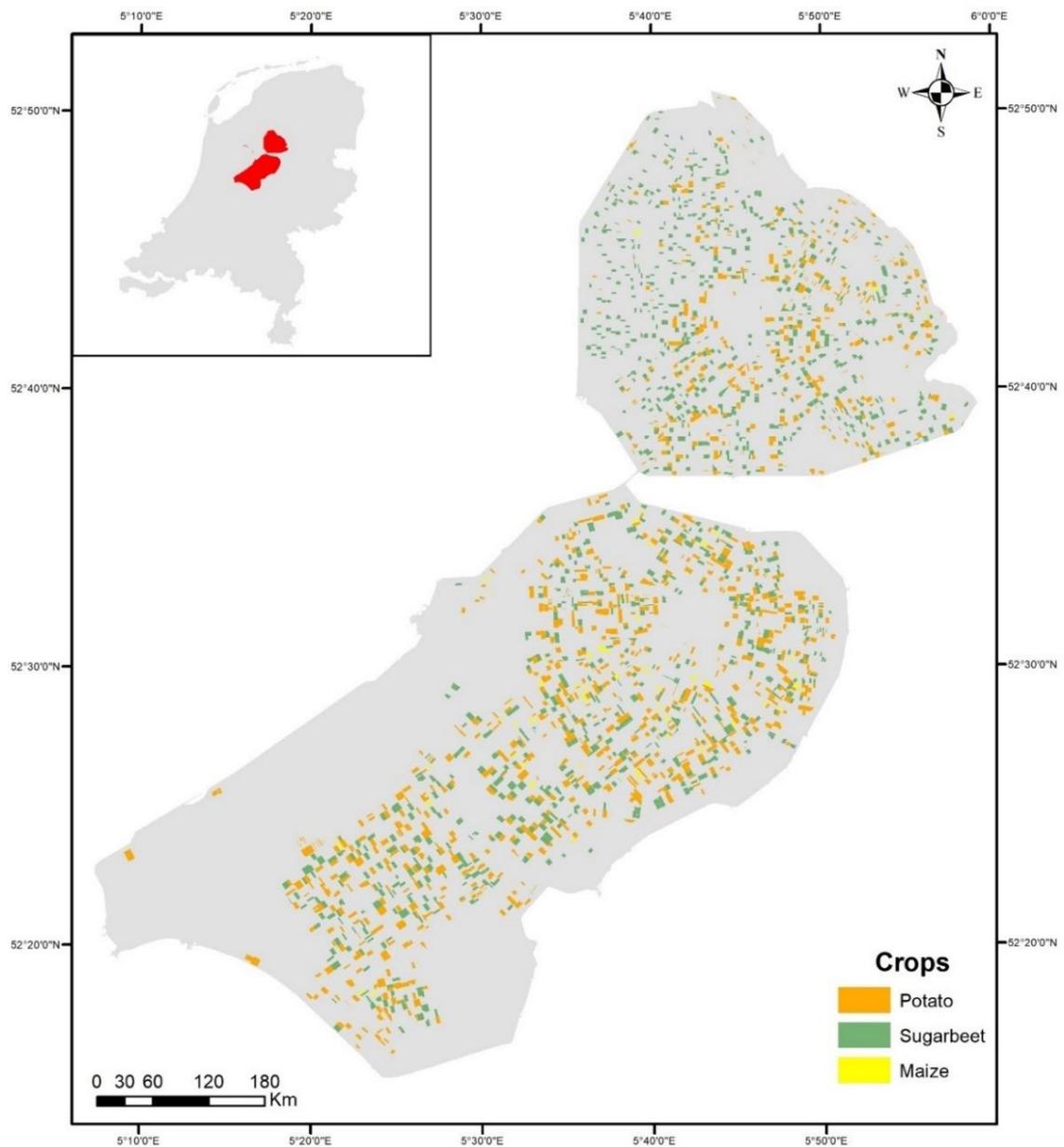


Figure 1 Study area and target crops

¹ <https://www.pdok.nl/introductie/-/article/basisregistratie-gewaspercelen-brp->

² <https://www.pdok.nl/introductie/-/article/agrarisch-areaal-nederland-aan->

4. LITERATURE REVIEW

4.1. SEN4CAP products

The products developed by SEN4CAP, and the corresponding measure adopted in the greening policy are detailed in Table 1. From the technical documents of SEN4CAP³. The algorithms related to these products are based on time series analysis, and the inputs corresponding to the feature extraction from Sentinel-2 are bands B2 (blue), B3 (green), B4 (red), B5 (Vegetation red edge), B6 (Vegetation red edge), B7 (Vegetation red edge), B8 (NIR), B11 (SWIR), B12 (SWIR), and the vegetation indices: NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index), SAVI (Soil-adjusted vegetation index), and EVI (Enhanced Vegetation Index). As well as LAI (Leaf Area Index), FAPAR (Fraction of Absorbed Photosynthetically Active Radiation), and FCOVER (Fractional Vegetation Cover) that are derived from the optical bands and in situ data. Currently, Sen4CAP products may rely on time resampling techniques for reconstructing optical missing data, with Sentinel-2 or Landsat 8 as auxiliary imagery, to address limitations in optical data availability. The data extracted from Sentinel-1 backscatter, and coherence for VV and VH polarizations, as well as VV/VH ratio.

Table 1 SEN4CAP products and compliance with CAP

		Measure		
		Crop diversification	Permanent grassland	Ecological focus area (EFA)
Product	Cultivated crop-type map	Indicate at the farm level number of crops if more than 2 or 3 crops, percentage distribution adheres to the regulatory thresholds.	Mapping of "grassland" or "permanent grassland".	The total area of catch crops, nitrogen-fixing crops, and land lying fallow.
	Grassland mowing product	Number of crops at the farm, considering the exclusion of grassland	Permanent grassland identification.	Considers the exclusion of annual grasslands and the inclusion of permanent grasslands in the "catch crops" and/or "nitrogen-fixing crops".
	Vegetation status indicator	Identification of the number of crops at the farm level.		Identification of crop types, crop calendars, and agricultural practices for catch crops, nitrogen-fixing crops, and land lying fallow
	Agricultural practices monitoring product	Occurrence of crop harvest.	Identification of events such as mowing that may indicate non-compliance with the definition of permanent grassland.	Obtention of data on the existence of catch crops, nitrogen-fixing crops, and land lying fallow at both the parcel and farm levels.

Table 2 summarizes the validation of the SEN4CAP products in the Netherlands in 2018 and 2019, also obtained from the technical documents of SEN4CAP. For both years the overall accuracy is quite high,

³ <http://esa-sen4cap.org/content/technical-documents>

the user's accuracy has variations as being the minimum 25% and 51.8% for the producer's accuracy. The grassland mowing detection has low accuracy in 2018 and improved in 2019, the accuracy is homogeneous across all parcel sizes. Harvest detection of the main crop is high and consistent for both years, the values in Table 2 indicate the detection within the first week after harvest, 5% and 10% of the parcels could not be detected in both years respectively. Catch crop monitoring presents high accuracy when based on NDVI, that is the value in Table 2, while if based on backscatter the accuracy reduces to 66.4%.

Table 2 SEN4CAP products validation in the Netherlands

Year	Product	Overall accuracy (%)
2018	Crop type map	94.95
	Grassland mowing detection product	53
	Harvest detection of the main crop	78
	Catch crop monitoring	83.3
2019	Crop type map	97.39
	Grassland mowing detection product	77
	Harvest detection of the main crop	82

4.2. Vegetation indices

Vegetation indices are essential in precision agriculture and crop monitoring because they offer a simple and dependable evaluation of the state and wellness of crops (Candiago et al., 2015). Typically, this involves identifying areas within a field that is not performing well or experiencing stress and concentrating resources in those zones.

A total of 519 vegetation indices have been created and are available in the Index Database (IDB, 2023). Most vegetation indices that use multispectral sensors are computed by considering the reflectance measurements of both red and near-infrared light to determine the level of photosynthetic activity or greenness in vegetation (Fernandez-Figueroa et al., 2022). The VIs time series that Sen4CAP utilize are NDVI and EVI as input for their products, also SAVI and GNDVI are included aiming to see how predictions would be in other VIs, formulas are shown in Table 3.

Table 3 Vegetation indices

Vegetation Index	Abbreviation	Formula
Normalized difference vegetation index	NDVI	$NDVI = \frac{NIR - R}{NIR + R}$
Enhanced vegetation index	EVI	$EVI = 2.5 * \frac{NIR - R}{NIR + 6 * R - 7.5 + B}$
Green-normalized difference vegetation index	GNDVI	$GNDVI = \frac{NIR - G}{NIR + G}$
Soil-adjusted vegetation index	SAVI	$SAVI = 1.5 * \frac{NIR - R}{NIR + R + 0.5}$

Despite the widespread use of NDVI as a vegetation index since it was first proposed by Rouse et al., (1974), several other indices have been developed to address its limitations and dependence on a single index. The primary objective of these alternative indices is to mitigate potential shortcomings and offer a

more extensive evaluation of vegetation. For instance, EVI (Enhanced vegetation index) was developed to address the limitations of NDVI by considering the non-linear relationship between reflectance and vegetation coverage and minimizing the impact of soil background and atmospheric errors (Huete et al., 2002). GNDVI (Green-normalized difference vegetation index) (Gitelson et al., 1996) is like NDVI, but instead of using the red band, it uses the green band and has the potential to be a more efficient method than NDVI for identifying alterations in vegetation caused by environmental factors such as nutrient deficiencies, disease, or water stress (Sankaran et al., 2015). Lastly, SAVI (Soil-adjusted vegetation index) (Huete, 1988) is a vegetation index developed to address soil brightness that can affect NDVI and is useful in areas with mixed vegetation.

4.3. Crop phenology

Accurate prediction of crop monitoring for different purposes requires the identification of critical dates such as emergence, closure, and harvest during the growing season (Khabbazan et al., 2019). Emergence refers to the appearance of shoots and leaves above the ground, while the closure is marked by the meeting of leaves from adjacent rows, these dates are crucial inputs for crop monitoring (MacKerron & Waister, 1985).

Khabbazan et al. (2019) used the BBCH scale (Meier et al., 2009) to estimate visually the growth stages of these crops in Flevoland in research to prove only SAR data for crop monitoring purposes. However, the authors found that backscattering alone was sometimes inadequate in detecting harvest due to disturbances in soil and vegetation properties, and therefore NDVI data was needed to monitor the crops. For instance, harvest detection for maize, sugar beet, and potato becomes challenging during autumn months due to elevated soil moisture and SAR backscattering becomes spurious. Therefore, other data sources or techniques may be necessary to improve the accuracy of harvest detection for these crops.

4.4. Deep learning

Remote sensing data acquisition involves different modalities such as optical and SAR, each with unique geometry and content. However, deep learning techniques allow for the fusion of data from diverse sources to harmonize information in a synergistic manner (X. X. Zhu et al., 2017). Traditionally, missing information in remote sensing data has been reconstructed using spatial, spectral, and temporal linear models (Shen et al., 2015). These models heavily rely on data from a single source, which introduces complexities when dealing with complex scenarios and large geographical areas. Moreover, the dependence on cloud-free images further intensifies the challenges associated with their application. Deep learning offers an opportunity to address these challenges (Q. Zhang et al., 2018) by leveraging large amounts of data and learning complex patterns, resulting in more accurate and reliable cloud-free simulated images.

Deep learning is a subfield of machine learning that involves training Artificial Neural Networks (ANNs) with several hidden layers to learn representations of data. These representations are hierarchical in nature, with each layer capturing increasingly abstract features of the input (Kamilaris & Prenafeta-Boldú, 2018). The success of deep learning can be attributed to its ability to automatically learn features from raw data without the need for explicitly designed features entered by humans (Schmidhuber, 2015).

ANNs are modelled after the structure of biological neurons, and their behaviour is simplified and formalized mathematically (Schmidhuber, 2015). A neuron is an activation function, depending on the value stored, neurons in the next layers will be activated, for pattern recognition purposes; the connections between every pair of neurons between layers are called weights (Jain et al., 1996). The process consists of applying a weight to each input and passing the sum of weighted inputs plus a bias through an activation

function (commonly Sigmoidal, Tanh, RELU, etc.) to produce an output, this is called forward propagation (Janiesch et al., 2021).

A loss function is defined to obtain the deviation between the predicted and target value after forward propagation, to update the parameters of the network the loss function must be minimized because the network is often initialized with random values, this is called backpropagation. The derivative of the loss function constitutes the partial derivatives following the chain rule: loss function, prediction, activation function, weighted sum plus bias, and finally weights. This derivative is a gradient that would lead to an update of the parameters of the network (Janiesch et al., 2021). As the size of the training set increases, the duration required to perform a single gradient step becomes unreasonably lengthy. To address this issue, stochastic gradient descent (SGD), and ADAM are used as optimization algorithms in the training phase of deep neural networks (Glorot & Bengio, 2010a).

Convolutional neural networks (CNNs) (Y. LeCun et al., 1998) are a type of neural network that excels at processing images and other spatial data that permit the reduction of the number of parameters maintaining pattern recognition. They use convolutional layers to extract local features from the input and pooling layers to reduce the dimensionality of the output. Like other neural networks, they are trained by minimizing a loss function through backpropagation (Isola et al., 2016).

The predefined architecture of a deep CNN is crucial to capture the patterns desired according to the goal of the network (Simonyan & Zisserman, 2014). The operations of feature extraction typically include an input layer, convolutional layer, nonlinear activation, and Max-pooling as shown in Figure 2. Convolution considers the different channels (bands) of the input image, a bank of filters based on weights is applied to all the channels, each filter of this bank has the same number of layers as in the input image, and the number of output images corresponds to the number of filters in the filter bank; the output of this process is called feature maps. Additional hyperparameters to consider are zero padding which adds a border of zeros of the image, and stride, which indicates the spatial interval the kernel moves during convolution. Max pooling is a spatial aggregation process that reduces the dimension, of the image while preserving features.

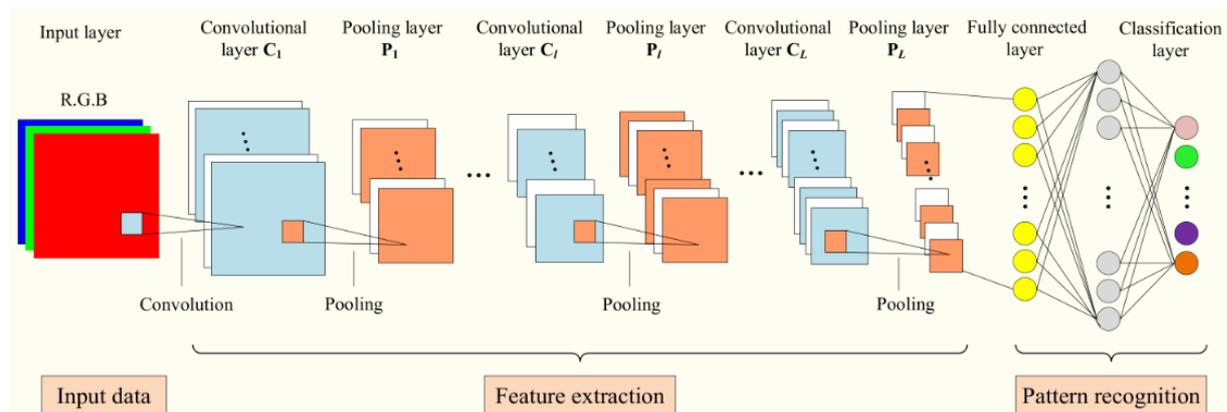


Figure 2 Typical CNN architecture (Luo et al., 2022)

It has been understood that normalizing the input data of neural networks, specifically setting the mean to zero and variance to one, is advantageous for training the network effectively because a similar scale can help the gradient descent converge more quickly toward the local minima (Y. A. LeCun et al., 2012). This idea is extended to the intermediate layers of a network using Batch normalization (BN) (Ioffe & Szegedy, 2015). It is a technique that commonly used in CNNs to deal with the internal covariate shift problem, which refers to the issue of the distribution of input data changing within the layers during training, and to normalize the activations within a mini batch, ensuring they have zero mean and unit variance. Moreover, BN makes it possible to do the training with larger learning rates. Additionally, Bjorck et al. (2018) indicate

that conventional weight initialization methods may not be suitable for networks with numerous layers unless BN is employed to enhance the network's resilience against poorly conditioned weights.

During the training of deep neural networks or dense networks as CNNs, the vanishing/exploding gradient problem appears, slowing down the training process and diminishing learning effectively, and therefore hampering convergence (K. He et al., 2015). This has been widely addressed with weight initialization methods, normalization techniques like batch normalization, gradient clipping, and different activation functions like ReLU (Glorot & Bengio, 2010b). Activation functions like sigmoidal and hyperbolic saturate to infinity with large values and for small values, it becomes zero, small gradients impede significant changes to the weights during backpropagation; non-saturating functions can deal with this problem like ReLU (Aggarwal, 2018).

A deep network can overfit when the training error decreases while the testing error increases. This means that the gradient descent approaches an absolute minimum as it is a gradient-based learning, and the generalization ability of the network disappears. Regularization is a method to make the network converge, it is done by early stop learning, L^2 weight decay adding a term to the loss function which is the square norm of the error as a function, and dropout, which randomly drops out (i.e., set to zero) a fraction of the output units or activations in a layer during each training iteration (Lemberger, 2017).

Degradation is also a problem that occurs when training dense, deep, or CNNs, as the depth of a neural network increases, there comes a point where the network's accuracy saturates and then starts to degrade rapidly. Unexpectedly, the common assumption that adding more layers to a deep model should improve its capacity to learn and generalize, is not always correct. Shallow neural networks tend to have a large decrease in the loss within the first few epochs of training, this means the model converges too fast without good learning. With CNN, little decrease in the loss is expected with more epochs. The problem here is after going through some layers, the initial input becomes random noise because it was initialized with random noise, and during backpropagation the gradients are scrambled, both produce not very meaningful learning, and the loss is reduced very slowly (K. He & Sun, 2014).

Employing skip connections or residual connections in architectures like Deep Residual Learning (ResNet) is how this problem has been addressed with the use of residual connections, also known as skip connections. K. He et al., (2015) introduced the ResNet architecture, which revolutionized deep learning by introducing residual connections, the idea is to group up layers of the network into so-called residual blocks, and the data flow through and around. Within a block, the data flows normally, which consists of a series of convolutional layers, followed by batch normalization and non-linear activation like ReLU. But, a new type of connection is created, and the output of the block is added to the input of the block, so two paths to follow are present now. Adding the input makes the block to figure out what the input contains, making forward propagation more efficient. Additionally, backpropagation becomes efficient, gradients have two paths to follow, between blocks and throughout the network, this means any layer in the network has a relatively shorter path by which loss gradients can arrive and update what that layer is computing with less scrambled gradients.

4.5. Generative Adversarial Networks (GANs)

To overcome the cloud removal problem from images, a deep learning method called Generative adversarial networks (GANs) (Goodfellow et al., 2014) has been proposed. It consists of a generator and a discriminator, both CNNs. The generator's role was to simulate images, while the discriminator's job was to distinguish between the generated and real content from a training dataset starting with random noise as illustrated in Figure 3. The strength of GANs relies precisely on the adversarial training process between the generator and the discriminator. As the generator improves its ability to produce more convincing samples, the discriminator is forced to become more discerning. This back-and-forth

competition pushes both components to improve iteratively, leading to the generation of highly realistic samples.

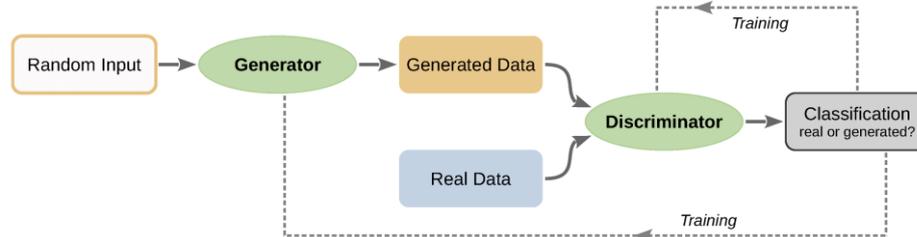


Figure 3 GAN architecture (Abedi et al., 2022)

Conditional GAN (cGAN) (Mirza & Osindero, 2014) is a variation of GAN. The generator takes conditional information as input to generate samples, and it operates as guidance or constraint for generating images. The discriminator in cGANs also receives conditional information along with the generated or real samples to aid in distinguishing between them. This conditional setup improves the stability and control of the generated samples. Cycle-Consistent Adversarial Networks (CycleGAN) (J. Y. Zhu et al., 2017) is an unsupervised method that uses unpaired datasets. Xiong et al. (2021) compared different types of supervised and unsupervised GANs for cloud removal in Sentinel-2 images, for this experiment, the author utilized images from different geographical areas as well as different intervals of sensing time, the best performance was obtained in MCcGAN and MTcGAN.

Most of the research conducted on Sentinel-1 and Sentinel-2 data has focused on using them as a mono-temporal pair, where only one pair of Sentinel-1 SAR and Sentinel-2 optical data is used as input (Grohnfeldt et al., 2018). However, a multitemporal approach can improve the results rather than using a mono-temporal pair (Xiong et al., 2021). When using a mono-temporal pair, the training phase involves learning the complex relationship between the two data sources. However, during the inference phase, Sentinel-1 data is used to simulate Sentinel-2 data. This approach assumes that Sentinel-1 can distinguish objects in the same way as Sentinel-2 does. Unfortunately, this assumption can result in low accuracy since backscatter values can be the same for distinct objects (L. Wang et al., 2019).

To obtain cloud-free Sentinel-2 images using the cGAN, a supervised Generative Adversarial Network, it is necessary to provide a pair of Sentinel-1 and Sentinel-2 images as input (Xiong et al., 2021). In this context, Gao et al. (2020) attempted to include the corrupted optical images as supplementary input data together with the pairs of Optical-SAR images as an effective approach to enhance the accuracy of the simulated images saving patterns from the uncontaminated images. However, it remains uncertain how much information can be effectively learned when the corrupted optical images are heavily obscured by a significant amount of cloudiness (Xiong et al., 2021). This also implies that using corrupted data as input to simulate optical data from a different time, when it is not available would not be possible (Bermudez et al., 2018).

MTcGAN (W. He & Yokoya, 2018) has the potential to utilize both SAR and optical data, it involves information from SAR and cloud-free optical data taken at a specific time (t_1), while also incorporating SAR data acquired when the area of interest is covered by clouds at a different time (t_2). By linking SAR and optical data at t_1 , it becomes possible to generate a cloud-free optical image at t_2 using the SAR data from the same time. The resulting image is expected to capture the details present in the SAR data (Dumeur et al., 2021).

The evaluation of GANs continues to be a topic of uncertainty, despite their impressive ability to model intricate distributions (Grnarova et al., 2018). The authors point out that unlike the traditional likelihood-based models, inspecting the loss curves of the generator and discriminator alone is insufficient to determine the performance of GAN. It is mainly because of the min-max game nature of the model,

which can result in convergence issues and presence of oscillations in the learning curves making it impossible to detect an early stop. Often, various evaluation metrics are utilized to measure the similarity between real and generated data, nonetheless, there is no consensus on which one is the most appropriate one (Gnarova et al., 2018). Evaluations like Peak signal-to-noise ratio (PSNR) and Structural similarity index (SSIM) are utilized to evaluate the performance of GANs using multisource remote sensing data (Dumeur et al., 2021; W. He & Yokoya, 2018; Xiong et al., 2021).

5. METHODOLOGY

5.1. Methodology overview

There are two distinct experiments conducted: Experiment A involves simulating optical bands to reconstruct VIs, while Experiment B directly simulates VIs themselves. Figure 4 illustrates the process to obtain the input data for both experiments, which is detailed in chapters 5.2 and 5.3. Figure 5 indicates how both experiments are conducted through MTcGAN that is explained in detail in Chapter 5.4. Finally, the results are assessed with evaluation metrics as well as visual evaluation in the chapters 5.5 and 5.6. Next to the processes in both figures it is illustrated the API, software, or programming language utilize for that specific task. Due to the friendly visual interface, Google Earth Engine (GEE) was utilized to obtain the Sentinel data and vegetation indices and ArcGIS for band visual inspection and further stacking. MTcGAN process was conducted in the ITC-CRIB platform using Python.

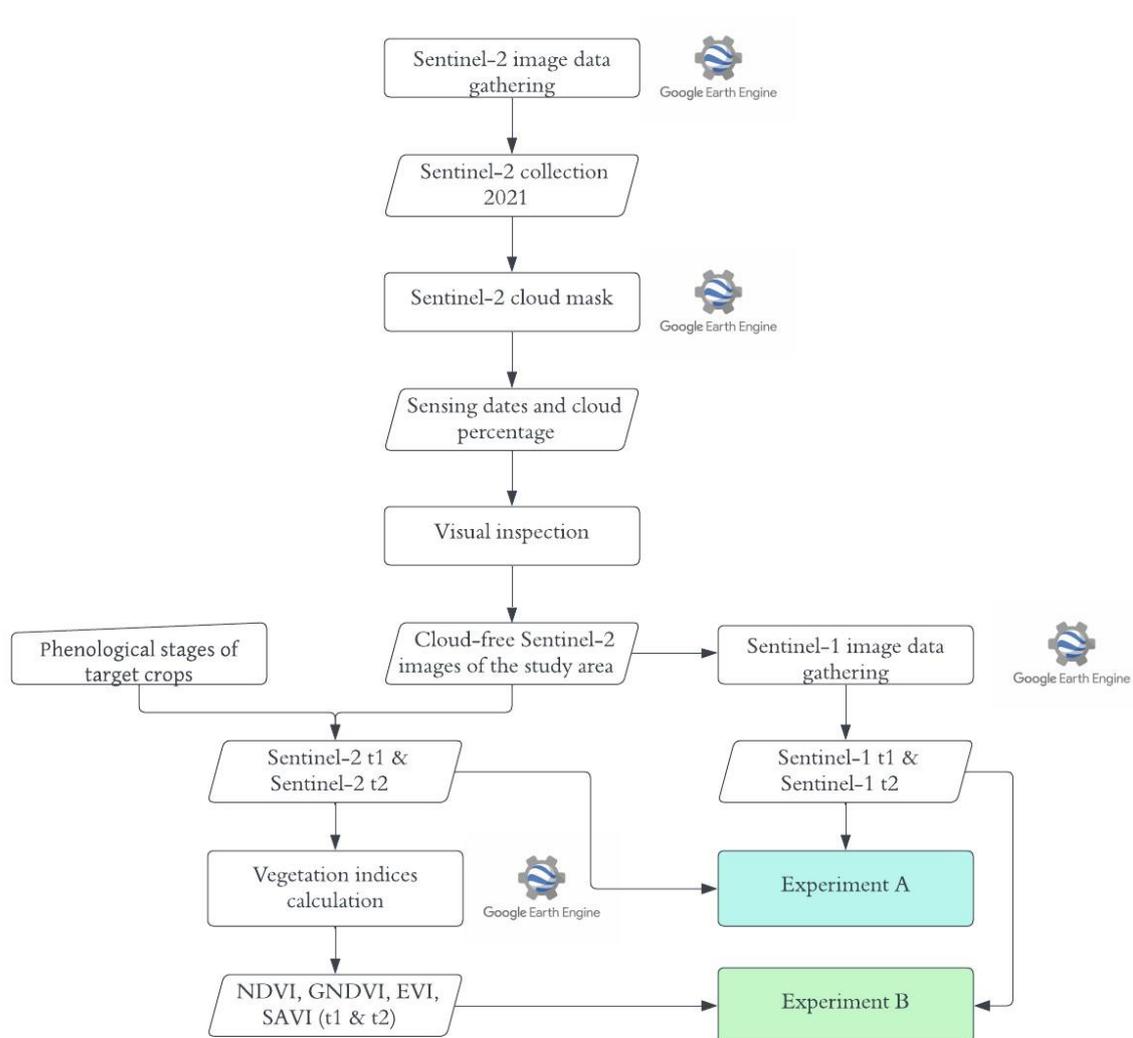


Figure 4 Cloudiness assessment and further Sentinel data collection

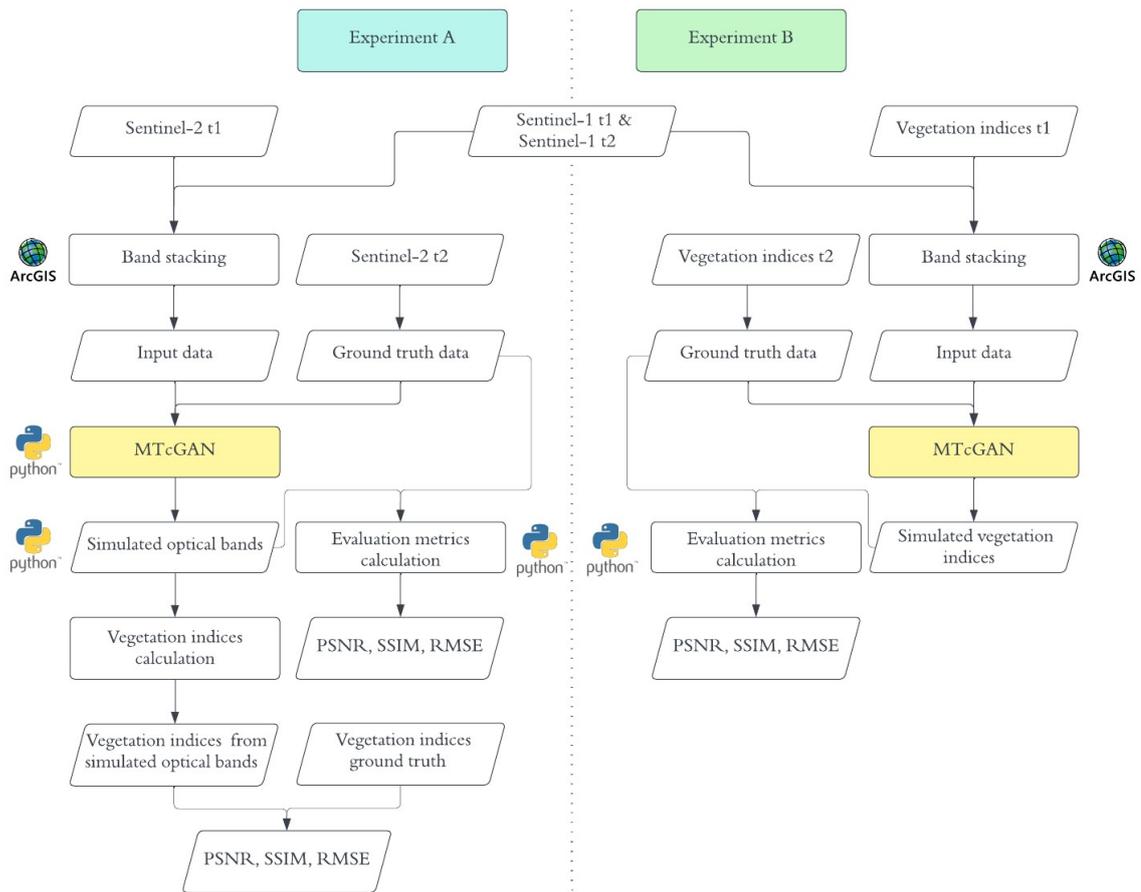


Figure 5 Experiments overview flowchart

5.2. Cloud cover assessment and crop vegetation growth stages recognition

The Sentinel-2 collection of the year 2021 obtained from Google Earth Engine (GEE), the cloud cover was assessed with the code of Gärtner (2020)⁴, for the whole scene as shown in Figure 6, as well as for the specific study in Figure 7. It uses cloud mask information QA60, obtaining the percentage of clouds, cirrus, and cloud-free area in both cases.

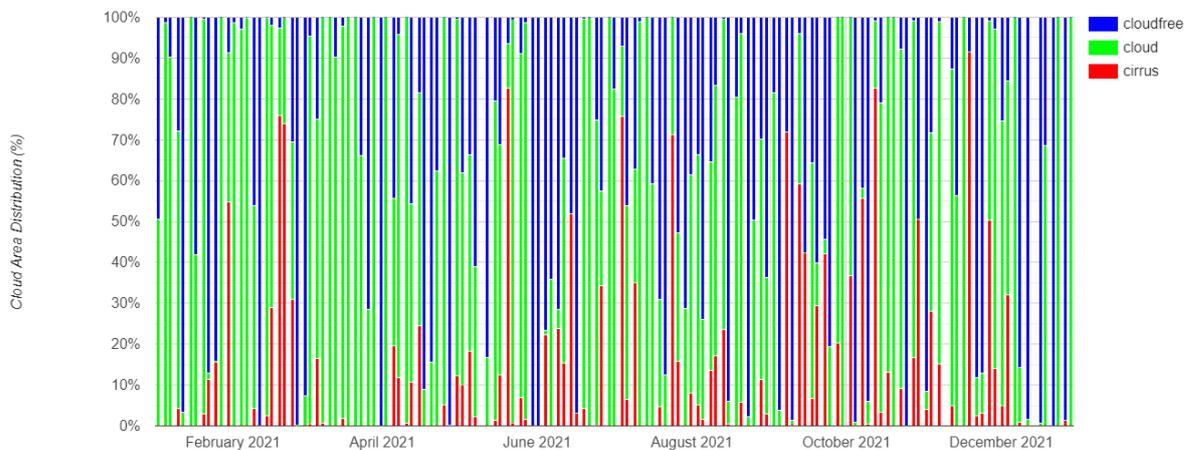


Figure 6 Cloudiness over the whole Sentinel-2 scene in 2021

⁴ <https://philippgaertner.github.io/2020/08/percent-cloud-cover/>

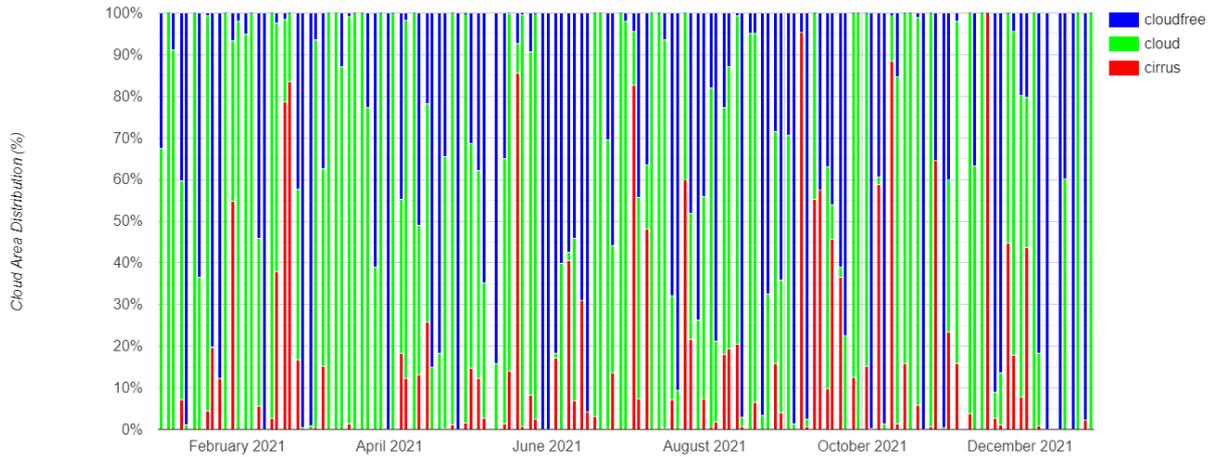


Figure 7 Cloudiness over the study area in 2021

Additionally, the relatively low reliability of this cloud masking method that Sentinel-2 itself provides has been reported (Boogaard, 2021; Chamnan, 2021). As shown in Figure 8, the number of cloud-free images that can be used for the cGAN supervised method considering only the study area is low, images with less than 10% cloudiness were double-checked by visual inspection and then it was determined which of them are appropriate for MTcGAN.

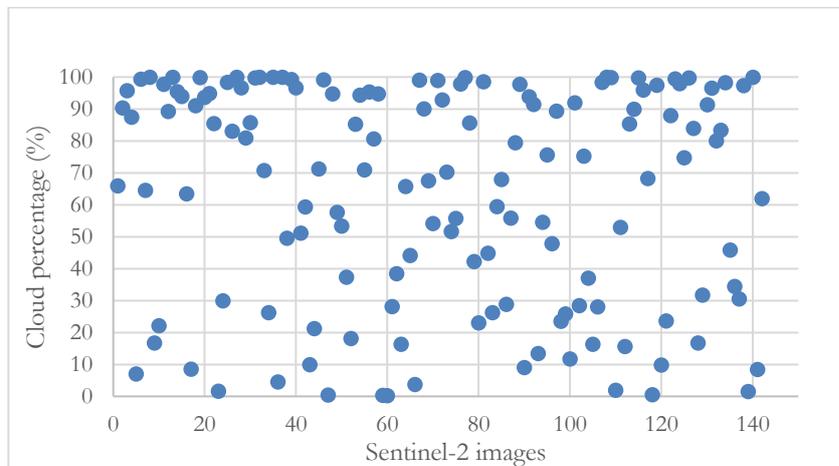


Figure 8 Cloud percentage in Sentinel-2 whole scene images

Vegetation stages in the study area were collected from the study of (Khabbazan et al., 2019) for the year 2017, each growth stage for maize, potato, and sugar beet is represented with a scale commonly used in agriculture and horticulture to assess the development and phenological stages of plants from leaf development to senescence, it was developed by BBCH "Biological Federal Research Centre for Agriculture and Forestry, Federal Office for Plant Varieties, and Chemical Industry" in English (see Appendix 1). Finally, the BBCH data is compared with the cloud-free images in a timeline, as illustrated in Figure 9, with the objective of choosing the fully cloud-free images. This selection process is a requirement for the MTcGAN technique, as explained in more detail in Chapter 5.4.

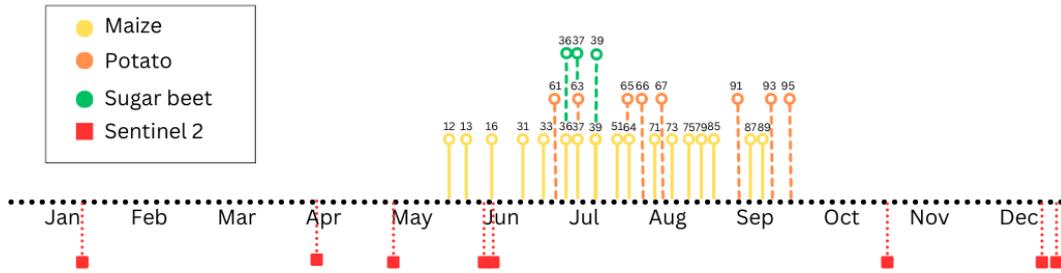


Figure 9 Crop growing stages and cloud-free Sentinel-2 images timeline 2021.

5.3. Data pre-processing

The images were downloaded from GEE; the datasets offered by this platform include pre-processing. The data of Sentinel-1 was acquired in Interferometric Wide swath mode (IW), the Ground Range Detected (GRD) product, obtaining the bands VV (vertical transmit and vertical receive polarization) and VH (vertical transmit and horizontal receive polarization) in ascending orbits, with a spatial resolution of 10m and transformed into backscatter coefficient (σ^0) in decibels (dB), it is pre-processed with orbit correction, GRD border noise removal, thermal noise removal, application of radiometric calibration values and terrain correction, additionally speckle filter was applied (Qiu et al., 2004). On the other hand, for Sentinel-2 data, Level-2A orthorectified atmospherically corrected surface reflectance dataset is chosen. This dataset has been already atmospherically corrected. The bands Red (R), Green (G), Blue (B), and Near-infrared (NIR) with the spatial resolution of 10m were selected for this study. The same spatial resolution is used to avoid errors by resampling. The projection in this dataset corresponds to WGS84 UTM 31N. Table 4 indicates the sensing date for both sensors in t1 and t2, with no more than two days difference for each time.

Table 4 Sensing time of Sentinel-1 and Sentinel-2 pairs.

Input data	Sensing date
Sentinel-1 t1	25/04/2021
Sentinel-2 t1	27/04/2021
Sentinel-1 t2	31/05/2021
Sentinel-2 t2	30/05/2021

Figure 10 presents the collected optical and SAR data, and Figure 11 illustrates the VIs at different times obtained from the optical data that are used for the direct VIs simulation. As the specific goal is to see the results for different crops, it would be more beneficial to focus mainly on cropland rather than including information about other land covers like built-up areas or water as is the case for this study area. By narrowing down the scope to specific crops, the GAN can generate data specifically tailored for crop related tasks. This involves carefully selecting training data representative of the target crop class, which can improve the results of modelling (L. Zhang et al., 2016). A 3km buffer over the province area was applied to clip the image, aiming not to lose information from the borders and reduce the number of patches needed for training the model. Additionally, data is rescaled to the range of [-1,1] because the last layer of the generator is Tanh in MTcGAN and this activation function ensures the generated data is bounded the same as real optical data (Radford et al., 2015).

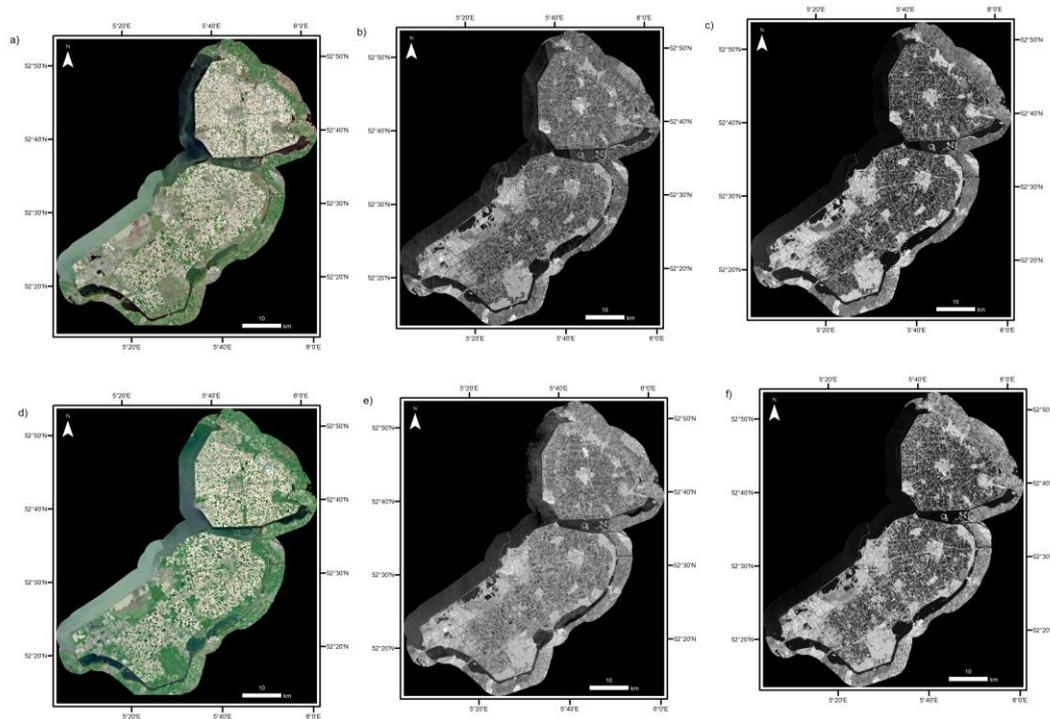


Figure 10 Data collected: a) RGB Sentinel-2 in t1, b) VV Sentinel-1 in t1, c) VH Sentinel-1 in t1, d) RGB Sentinel-2 in t2, e) VV Sentinel-1 in t2, f) VH Sentinel-1 in t2.

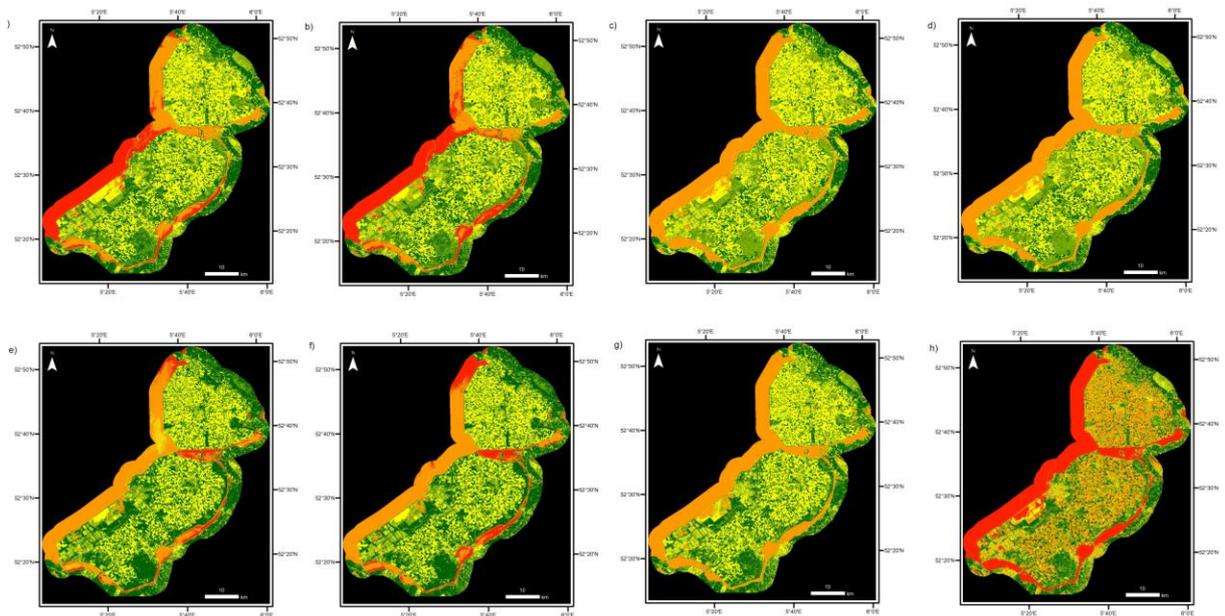


Figure 11 Ground truth vegetation indices. a) NDVI t1, b) GNDVI t1, c) EVI t1, d) SAVI t2, e) NDVI t2, f) GNDVI t2, g) EVI t2, h) SAVI t2

After obtaining the paired images and stacking the bands into two arrangements, they were split into patches (see Appendix 2) with the library GeoPatch⁵ in Python, with a patch size of 256, and a stride of 256 (without overlapping). The distribution of the data for training, testing, and validation is 70%, 15%, and 15% respectively as conventionally in deep learning.

⁵ <https://github.com/Hejarshahabi/GeoPatch>

The data used to map the target crops for human consumption was obtained from Pdok, Basisregistratie Gewaspercelen (BRP). The geometry of the agricultural plots is derived from the Agricultural Area of the Netherlands (AAN). Users have capability to annually update the crop type under their domain. The reference date for the dataset is May 15 of 2021.

5.4. Multi-Temporal Conditional Generative Adversarial Network MTcGAN

The approach used in this research corresponds to the MTcGAN by He & Yokoya (2018) with the aim of simulating an optical image at t_2 . It uses as input Sentinel-1 at t_2 and t_1 , as well as Sentinel-2 at data at time t_1 at the same geographic location. The code developed by Dumeur et al. (2021) for this approach can be found publicly⁶.

The cGAN technique is broadened from the traditional GANs (Goodfellow et al., 2014) and deep convolutional GAN (DCGAN) (Radford et al., 2015). The key idea of a GAN is to train a generative network (G) that captures the input data features, while a discriminative network (D) that calculates the probability that comes from input data rather than G . Basically, the generator is trying to fool with the data simulated and the discriminator tries not to be fooled. The training goal of G is to maximize the probability of the discriminator in making a mistake, whereas the training goal of D is to maximize the probability of correctly classifying of real and fake images. A cGAN, does not initialize the generator with random noise but with additional information (SAR at t_1/t_2 and optical at t_1) that successfully proved to solve image-to-image translation challenges (Isola et al., 2016).

As shown in Figure 12, the input corresponds to the concatenated optical bands in t_1 with the VV and VH polarization SAR images in t_1 and t_2 . The output of the generator together with the Sentinel-2 patches at t_1 are called input-fake pairs and the optical data at t_1 and t_2 are called real-input pairs. Typically, 50 percent of each goes through the discriminator to train this sub-network to classify correctly fake or real data. Loss is calculated for each sub-network due to the goal of each one is different, it uses the same objective function, but the gradients differ. Additionally, to the generator loss, L1 (Mean Absolute Error loss) is added (Isola et al., 2016).

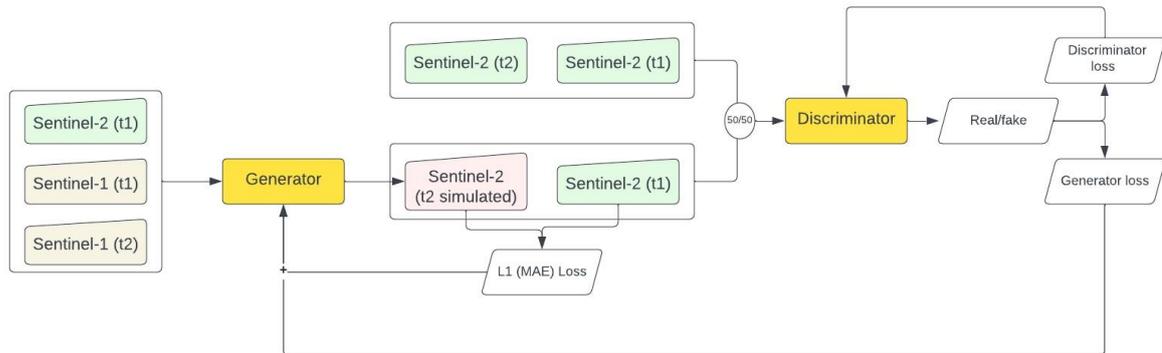


Figure 12 MTcGAN flowchart

The images used are denoted, as in (Dumeur et al., 2021), by:

$x_{t_1}^{S1}$ is a Sentinel-1 image at t_1

$x_{t_2}^{S1}$ is a Sentinel-1 image at t_2

$x_{t_1}^{S2}$ is a Sentinel-2 image at t_1

$x_{t_2}^{S2}$ is a Sentinel-2 image at t_2

⁶ URL: https://github.com/irisdum/cGAN_sent2_sim

The loss function in this cGAN comes from a binary cross entropy loss function, and for MTcGAN is defined:

$$L_{cGAN}(G, D) = E \log \left(D(x_{t_1}^{S2}, x_{t_2}^{S2}) \right) + E \log \left(1 - D \left(x_{t_1}^{S2}, G(x_{t_1}^{S1}, x_{t_2}^{S1}, x_{t_1}^{S2}) \right) \right)$$

The first term represents the discriminator's predictions on real data, real because both are Sentinel-2 images, $D(x_{t_1}^{S2}, x_{t_2}^{S2})$, it indicates the probability of images are indeed real Sentinel-2 images. The output of the generator is denoted by $G(x_{t_1}^{S1}, x_{t_2}^{S1}, x_{t_1}^{S2})$, which represents a fake Sentinel-2 image and $D(x_{t_1}^{S2}, G(x_{t_1}^{S1}, x_{t_2}^{S1}, x_{t_1}^{S2}))$ is the probability that real image and fake image are Sentinel-2 images. E is the expectation from the many patches.

$L_{cGAN}(G, D)$ is a minimax function, which means that G wants to minimize this expression, and a D that wants to maximize it. The samples go through the discriminator to make predictions by ascending the gradient to maximize the cost function:

$$\nabla E \log \left(D(x_{t_1}^{S2}, x_{t_2}^{S2}) \right) + E \log \left(1 - D \left(x_{t_1}^{S2}, G(x_{t_1}^{S1}, x_{t_2}^{S1}, x_{t_1}^{S2}) \right) \right)$$

Since $D(x_{t_1}^{S2}, x_{t_2}^{S2})$ only depends on the discriminator and the real data, the derivatives with respect to the generator are all zero. The generator parameters are updated by descending the gradient:

$$\nabla E \log \left(1 - D \left(x_{t_1}^{S2}, G(x_{t_1}^{S1}, x_{t_2}^{S1}, x_{t_1}^{S2}) \right) \right)$$

Additionally, earlier methods have discovered advantages in combining the objective of a GAN with a conventional loss function, such as the L2 (Euclidean) distance (Pathak et al., 2016), but it produced blurry results. Therefore, using L1 to avoid reconstruction artifacts and sharper images from the generator (Dumeur et al., 2021).

$$L_{cGAN}(G, D) + \lambda E \|x_{t_2}^{S2} - G(x_{t_1}^{S1}, x_{t_2}^{S1}, x_{t_1}^{S2})\|_1$$

where $\|x_{t_2}^{S2} - G(x_{t_1}^{S1}, x_{t_2}^{S1}, x_{t_1}^{S2})\|_1$ is the L1 norm, and $\lambda > 0$ is a hyperparameter. Ghamisi & Yokoya (2018) recommend setting λ to 100 for image reconstruction desirable accuracy.

The generator CNN architecture (W. He & Yokoya, 2018) is defined by nine ResNets blocks and six non-ResNets blocks as presented in Figure 13, where n are the number of filters and k is the kernel size. Except for the last layer, non-ResNets convolutional layers are composed of a convolutional layer (Conv), followed by Batch Normalization (BN), and ReLU as activation function. The training images were not subjected to any pre-processing steps other than scaling to fit within the range of the tanh activation function, which is between -1 and 1. To ensure that the output image maintains the desired scale range, the final layer of the model utilizes a Conv and tanh activation function configuration. Each Resnet block Conv-BN-ReLU-Drop (dropout)-Conv-ReLU.

The hyperparameters of the model are the same as in (W. He & Yokoya, 2018), being learning rate 0.0002, dropout rate 0.5, 200 epochs, batch size 1, batch normalization with momentum 0.99, and mini-batch stochastic gradient descent with Adam solver to train the model.

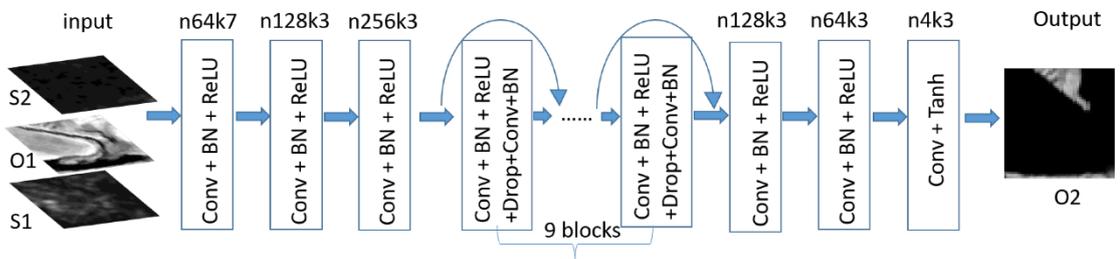


Figure 13 Generator CNN architecture (W. He & Yokoya, 2018)

The discriminator architecture (W. He & Yokoya, 2018), shown in figure 14, corresponds to a PatchGAN discriminator developed by Isola et al. (2016). By examining local patches instead of the entire image, PatchGAN can capture fine-grained details and local structure, providing more detailed feedback to the generator. It consists in five layers that classifies if each patch on real or fake where each cell contains a value between 0 and 1, the stride of the first three layers is 2 and for the other two layers 1. The outputs of the discriminator for each patch are usually combined to generate an overall score or prediction. This requires fewer parameters, runs faster, and can be applied to large images (Isola et al., 2016).

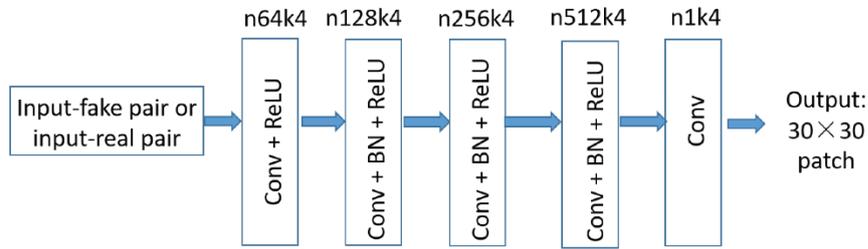


Figure 14 Discriminator CNN architecture (W. He & Yokoya, 2018)

5.5. Evaluation metrics

A process that involves images may introduce a loss of information and quality, the quality of a processed image can be assessed by subjective methods such as human visualization, and by objective methods using quantitative criteria (Horé & Ziou, 2010).

GAN evaluation and comparison, specifically regarding the images produced by GANs, is a difficult task (Grnarova et al., 2018). One of the reasons for this difficulty is the absence of a direct measure of likelihood, which is commonly used in similar probabilistic models, state-of-the-art GANs demonstrate limitations of subjective evaluation in accurately assessing their quality (Shmelkov et al., 2018). Consequently, previous studies have relied mainly on subjective visual assessment when evaluating images generated by GANs.

He & Yokoya (2018) and Xiong et al. (2021) applied Peak Signal Noise Ratio (PSNR) and Structural Similarity Index (SSIM) for each band between the real and simulated images at t2 as quantitative measure. The average of each evaluation metric per band is obtained to assess quality of the image (W. He et al., 2016). Additionally, Xiong et al. (2021) used Root-mean-square-error (RMSE) and Coefficient of determination (R2) per band in addition to the previous metrics, as summarized in Table 5.

Table 5 Evaluation metrics

Metric	Measure	Formula
Root-mean-square-error (RMSE)	Difference between the Sentinel-2 simulated and real images.	$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}$
Peak signal-to-noise ratio (PSNR)	Quality of the image considering pixel values.	$PSNR = 20 \log \left(\frac{MAX}{RMSE} \right)$
Structural similarity index (SSIM)	Measure the structural similarity between the real and simulated images by comparing luminance, contrast, and structure.	$SSIM = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)}$

Where:

y_i is the real value of the pixel in Sentinel-2,

\hat{y}_i is the predicted pixel value,

N is the total number of pixels from all the patches,

MAX is the max pixel value, as the pixel values range is between -1 and 1 MAX is 2,

μ_y and $\mu_{\hat{y}}$ are the mean of real and predicted values respectively,

σ_y^2 and $\sigma_{\hat{y}}^2$ are the variance of real and predicted values respectively,

$\sigma_{y\hat{y}}$ is the cross-correlation of real and predicted values,

C_1 , C_2 and C_3 are constants.

MSE measures the magnitude of the prediction errors, lower values of MSE indicate better model performance, as they represent smaller prediction errors. As MSE decreases towards zero, PSNR tends to increase towards infinity, therefore a higher PSNR value corresponds to a better image quality in a greyscale image, and a low PSNR value suggests significant numerical differences between the real and simulated images (Horé & Ziou, 2010). SSIM measures the similarity between real and simulated images considering the loss of correlation, brightness distortion, and contrast distortion, it ranges between -1 and 1, and a value closer to 1 indicates a better-simulated image (Z. Wang et al., 2004). If the simulated and ground truth images are similar, it is expected to have high values for PSNR and SSIM (Dumeur et al., 2021).

6. RESULTS AND DISCUSSION

Results are analyzed and compared with the research of Dumeur et al. (2021) that from now on is called Reference A, W. He & Yokoya (2018) research which from now on is called Reference B, and Xiong et al. (2021) called Reference C. The references explored optical data generation using SAR, they differ in the datasets; locations or differences between sensing times are different.

References and experiments in this research are matching with a number depending on the type of experiment that was conducted as shown in Table 6, types are:

I: training with patches of same location sensed between t1 and t2.

II: training with patches from multiple locations sensed between t1 and t2.

III: training with patches same from location and different intervals of sensing times.

IV: training with patches from multiple locations with different sensing times.

Table 6 Experiments type in comparable research

	I	II	III	IV
Experiment A				
Experiment B				
Reference A				
Reference B				
Reference C				

The three references utilize MTcGAN with the same location sensed between t1 and t2, so A-1, B-1, C-1, and D-1 are going to be compared, also Reference C utilizes MCcGAN as well as additional unsupervised methods and compare them. Reference A-I, B-I, B-II, and C-IV utilize bands R, G, B, and NIR. Reference C-I and C-III only R, G, and B. Metrics on NDVI calculated from predicted images are indicated in C-IV.

The findings in Reference B-II indicate better results in visual and quantitative evaluation compared to B-I. It provides evidence that using more complex data, to use different locations allows the model to generate better images. Nevertheless, the same sensing time interval is utilized so time transfer is not widely explored. On the other hand, findings in Reference C-II indicate that keeping the sensing time interval narrow produces better results, the wider one was three weeks, and images can still be successfully reconstructed. Reference C-IV proved spatial and temporal transference, the longer time interval used was one month using 14 locations for training and 5 locations for testing which were able to recover more details than C-IV. Here, best performances of MTcGAN or MCcGAN were not conclusive, for the rest of experiments of C MCcGAN showed better in evaluation metrics.

An aspect that differentiates experiments from references is the image pre-processing. The references downloaded images from Copernicus Open Access Hub and pre-processed in SNAP, also used Sentinel-2 Level-1C orthorectified top-of-atmosphere reflectance images. Thus, it is challenging to compare results across references because data is different and pre-processing is not always clearly described (Dumeur et al., 2021). Also, the data in the experiments was normalized between -1 and 1, references do not provide information about standardizing or rescaling.

6.1. Benchmark dataset

The input data collected distribution is presented in Figure 15 for Sentinel-1 and Sentinel-2 in t1 and t2, as well as the distribution of VIs.

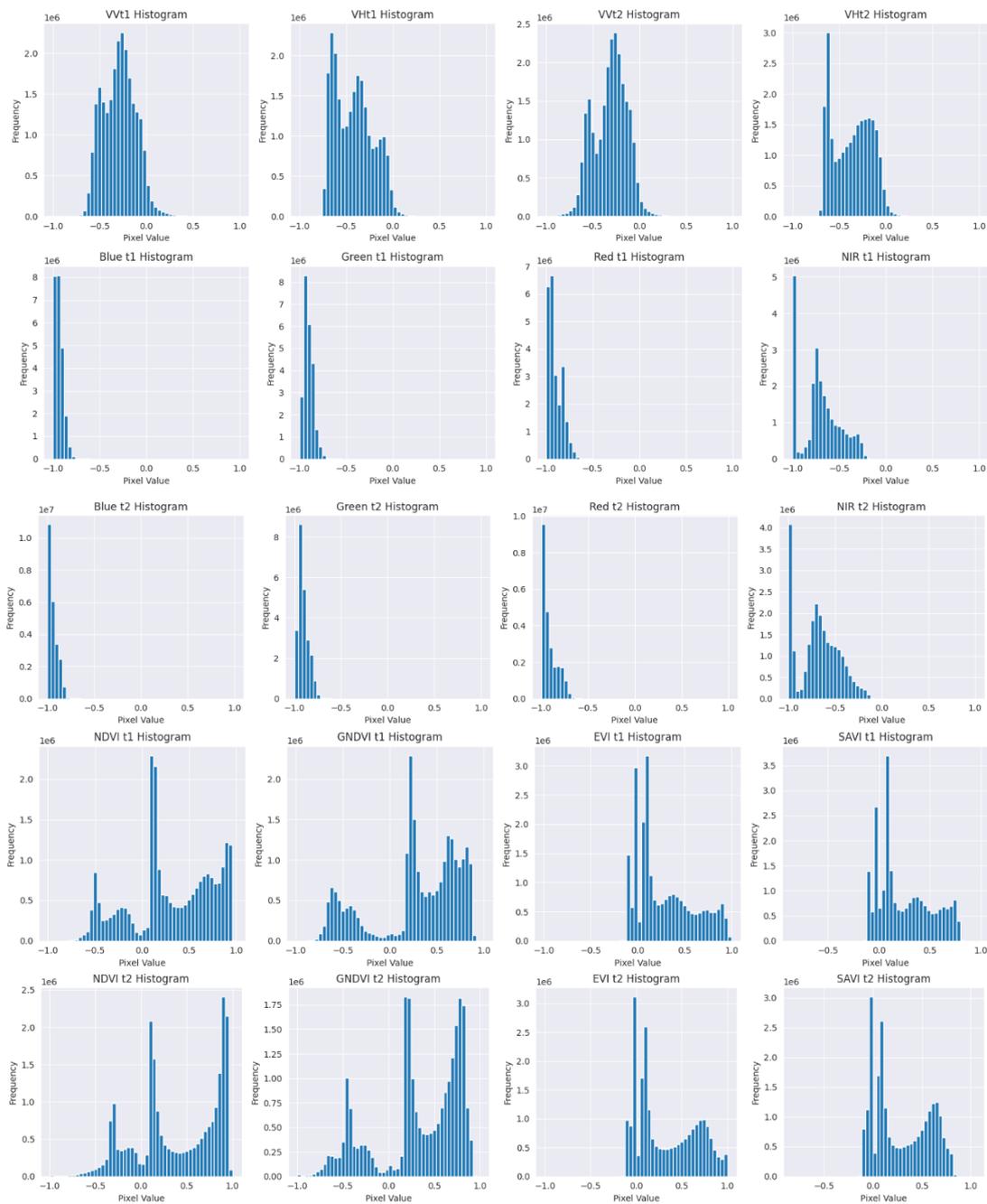


Figure 15 Histograms of input data

The size of the image is 7452 x 6534 pixels, and the total number of patches generated is 672, nonetheless, some of them contain full or partially NaN values, and these are discarded to avoid errors in the model. The final number of patches is 303, some patches are shown in Figure 16, each patch is a NumPy file named as `#_img.npy` after the corresponding number during patch generation. In Figure 13 random patches of the Sentinel-2 at t2 are posed, and in Figure 14, RGB compositions of optical data at t1 and t2 where the changes in the surface can be appreciated. From visual inspection, a greener scene for crops is perceived in t2, which coincides with the observable first growing stages of crops in the study area.

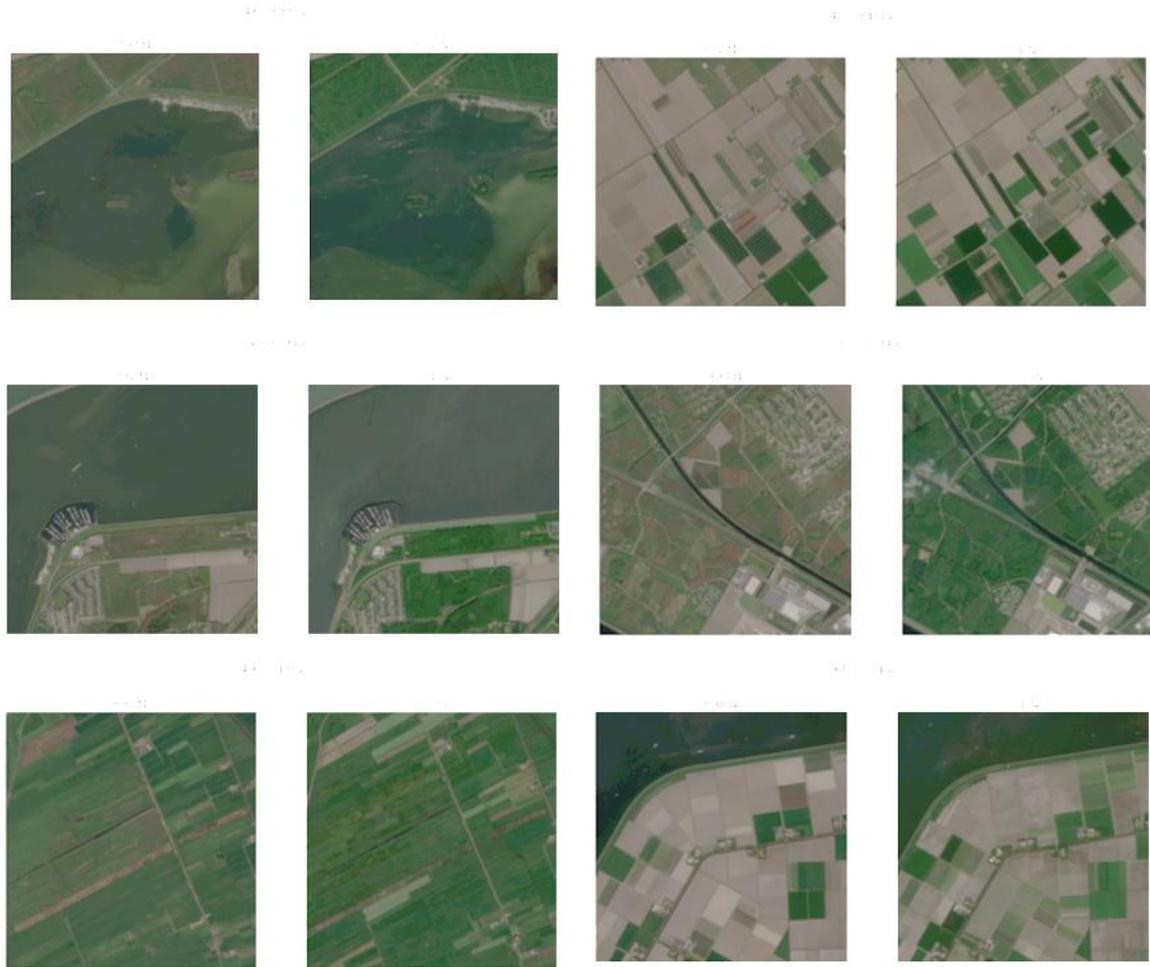


Figure 16 Pairs of RGB compositions for t1 and t2

Experiments A and B sensing dates t1 and t2 have a difference of 35 days. Reference A has a difference of around two months, both using MTcGAN in the same geographic location. Reference B similarly did the same in different locations with a sensing date difference of one month, the experiment with better results is going to be used for the comparisons. Reference C conducted several experiments with unsupervised GANs, as well as MCcGAN, which they developed, and MCcGAN using different sensing dates to prove its temporal transferability, as well as training the model with patches from images around the globe and testing in different locations as well to prove spatial transferability. The best metrics obtained in Reference C were with the closest date, but comparisons are going to be made with a sensing time difference of around one month. Sensing dates ± 2 days in the same location are indicated in Table 7.

Table 7 Sensing dates of comparable research.

	Sensing t1	Sensing t2
Experiment A and B	26/04/2021	31/05/2021
Reference A-I	06/11/2019	29/01/2020
Reference B-I	11/11/2017	08/12/2017
Reference C-I	01/09/2018	05/10/2018

6.2. Training

The training was performed using the Geospatial Computing Platform CRIB of the University of Twente with the computer unit 72 vCPU Intel x86-64, 768 GB RAM, single NVIDIA RTX A4000 GPU. The number of parameters, trainable and not trainable is summarized in Table 8, which indicates the overall complexity and size of the model, and the hyperparameters are kept as indicated in Chapter 5.

Experiment A and Experiment B were trained both with 200 epochs. At first, both models were trained with a reduced tile of training samples of 70 before running the model with the full set of training samples to understand if the model was working. Nevertheless, the limitation of memory allocation in the remote machine was a problem that was constantly coming across. Experiment A was completed with the whole dataset, and Experiment B was kept with 70 training samples. The training time was 7 hours for Experiment A and 1 hour for Experiment B.

Table 8 Number of parameters in the model

Sub-model	Total params	Trainable params	Non-trainable params
Generator	11,407,748	11,397,252	10,496
Discriminator	3,433,541	3,431,491	2,050

Figure 17 shows the generator loss curves in both experiments, learning a generative model is a difficult task due to the minimax nature of the objective function, and the involvement of neural networks as players in the process makes them not intuitive, loss curves usually present oscillations and deciding when to stop the training is difficult to tell (Grnarova et al., 2018). Nonetheless, there is a decreasing behavior, which indicates the generator is producing better samples, but the oscillating behavior indicates also that the updating between the generator and discriminator is not improving anymore and reached a limit. Experiment A seems to be a pattern in the behaviour of both sub-models and for Experiment B it is finding difficult to find consistent patterns.

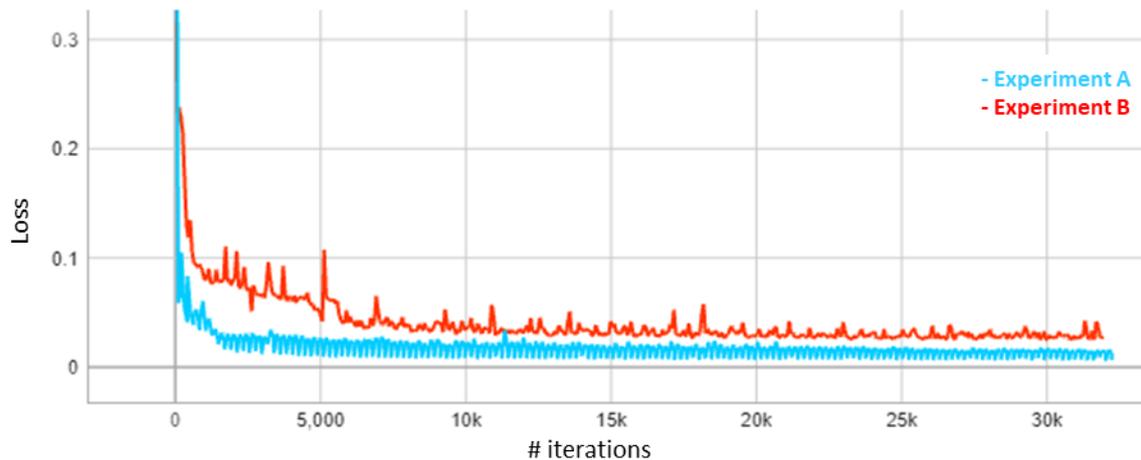


Figure 17 Training loss

In Table 9, the train, validation, and testing number of patches for each reference is indicated. For experiments A and B, the percentage of data was distributed to data 70% for train, 15% for test, and 15% validation as conventional in deep learning. Reference A distributed the data 80%, 15%, and 5% which can be considered unusual, also they conclude the model has a generalization ability because of the similarity of metrics on the three datasets. Reference B only separates data in training and testing patches.

Reference C mentions a validation dataset for spatial transfer, which corresponds to Reference C-IV, but not for tuning hyperparameters purposes.

Table 9 Train, test, and validation datasets number of patches in comparable research

	Experiment A and B	Reference A	Reference B	Reference C
Train	303	496	1188	6018
Test	45	95	165	669
Validation	45	32	-	-

6.3. Evaluation metrics

The comparisons were done considering the same metrics reported in the three references and at different levels (datasets, bands, or overall). In Reference A, Dumeur et al. (2021) report the metrics as overall for all the bands, also they conclude that as metrics for training, testing, and validation are similar which they claim the generalization ability was achieved.

For Experiment A, Figure 18 and Figure 19 show PSNR along training for training and validation datasets respectively, the curve does not seem to converge and presents oscillations around 30dB all along the training, it performs similarly in the validation dataset. For Experiment B, PSNR has similar oscillating behaviour, but it seems to deteriorate on the validation dataset. Figure 20 and Figure 21 SSIM approximates 1 in Experiment A, which indicates a high similarity between generated and real data. Experiment B presents strong oscillations on the training dataset.

PSNR mainly indicates similarity in pixel values, on the other hand SSIM is a metric more correlated with human vision because its calculation depends on whole patch values. The high values suggest that the model might be overfitting, this will be discussed using the test dataset in Sections 6.3.1 and 6.3.2. and discussed in Chapter 7.

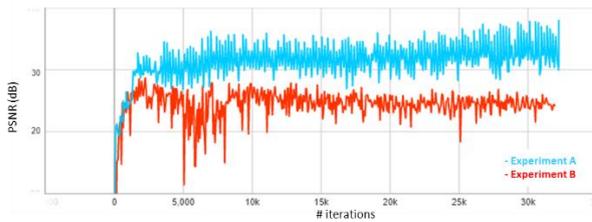


Figure 18 PSNR training dataset

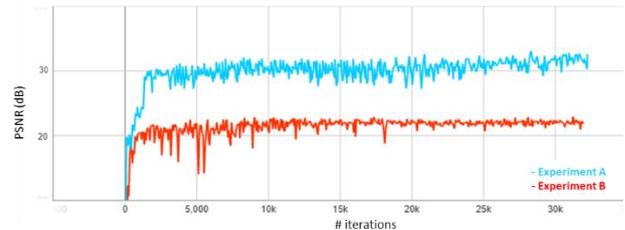


Figure 19 PSNR validation dataset

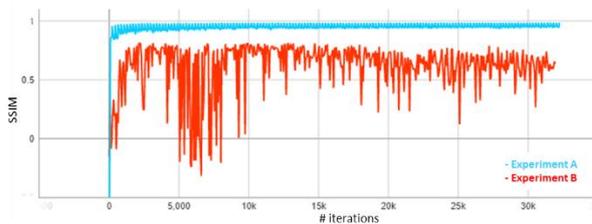


Figure 20 SSIM training dataset

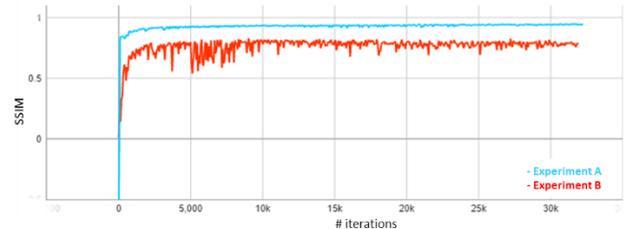


Figure 21 SSIM validation dataset

Grnarova et al. (2018) indicate the need for a convergence metric to validate a GAN, given loss curves present oscillations, deciding when to stop the training and understanding its performance of it is not plausible. Another validation alternative is suitable, SSIM has a convergence behaviour with smaller oscillations in Experiment A, but not in PSNR. The smaller amount of data in Experiment B is a major cause of the low performance of the model. Predictions on unseen data can be tricky, even if PSNR seems

not to converge for both experiments and for SSIM in Experiment B, the values it oscillates around are comparable with other studies.

6.3.1. Experiment A

The evaluation metrics per band on the testing dataset are presented in Table 10. The values are in the range of values of the amplitude in the PSNR and SSIM curves. The predictions of the NIR band have less similarity with the ground truth one, this suggests that the model could predict better for RGB bands, and using predictions to calculate VIs would not have much similarity with ground truth VIs. NIR had a different distribution compared to the RGB bands, the model learns better from them because there are three of them.

Table 10 Experiment A: evaluation metrics testing dataset per band.

	Blue	Green	Red	NIR
PSNR	25.061	27.758	29.382	17.89
SSIM	0.780	0.857	0.881	0.483

Table 11 indicates the metrics for each dataset, Reference A has similar values among datasets, and Experiment A has an expected value according to the metric curves. Numerically, Reference A has higher values which indicates a better similarity for real and generated data. The architecture of the GAN is the same, they differ in pre-processing, dataset amount, and distribution on training and validation/testing.

Table 11 Experiment A: comparison of evaluation metrics with Reference A-I

	Experiment A	Reference A-I		
	Test	Train	Test	Val
PSNR	25.023	41.8	41.8	40.4
SSIM	0.750	0.983	0.982	0.975

Reference B reports their results on an overall value for the image, in Table 12 they are compared with the average of the metrics of all the bands in the testing dataset. Numerically, Experiment A is lower than Reference B, also both are lower than the metrics in Reference A. Equally, the same differences exist among Experiment A and References B and C.

Table 12 Experiment A: Comparison of evaluation metrics with Reference B-I

	Experiment A	Reference B-I
PSNR	25.023	32.32
SSIM	0.75	0.911

Comparison in Table 13 excludes NIR, RGB seems to have consistent values, for visual evaluation it should give good results, but references experiments obtained higher values.

Table 13 Experiment A: Comparison of evaluation metrics with Reference C-I

	Experiment A			Reference C-I		
	Blue	Green	Red	Blue	Green	Red
PSNR	25.061	27.758	29.382	30.779	32.984	33.74
SSIM	0.780	0.857	0.881	0.8916	0.909	0.887

As expected, metrics indicate low similarity between ground truth VIs calculated from reconstructed bands because of the low performance on the NIR band as shown in Table 14. Additionally, it is compared with Reference C-IV which is the only research that utilized reconstructed bands to calculate VIs, it is evident that the VIs were not reconstructed well. Anyhow, from all the VIs, SAVI seems to have higher values, but not significantly better to accurately reconstruct that VI.

Table 14 Evaluating metrics on reconstructed VIs calculated from predicted data

	NDVI	GNDVI	EVI	SAVI	NDVI Reference C-IV
PSNR	12.543	12.943	12.879	15.903	24.741
SSIM	0.174	0.199	0.178	0.196	0.757
RMSE	0.483	0.461	0.490	0.353	0.116

References do not report metrics per band except by C, which even when they report NDVI metrics, NIR metrics are missing. The availability of simulated images for the following applications is unknown if each band is not evaluated also separately, specially NIR.

6.3.2. Specific crops

The quantitative metrics for the three target crops potatoes (P), sugar beets (S), and maize (M) per band calculated on unseen data are shown in Table 15, values for PSNR and SSIM are higher than the values of the overall image. This suggests that the contribution to the error is less by patches that contain these crops. Also, the NIR still performs worse than the other bands as shown in the metric histograms in Figures 22, 23, and 24. Notwithstanding, the phenology illustrated in Figure 9, it is expected that these metrics correspond to the early stages of crop growth, as is visualized in Section 6.4.3.

Table 15 Evaluation metrics in different crops

	Potatoes				Sugar beets				Maize			
	Blue	Green	Red	NIR	Blue	Green	Red	NIR	Blue	Green	Red	NIR
PSNR	30.328	29.087	25.266	19.433	30.218	28.964	25.153	19.099	30.587	29.343	25.652	18.790
SSIM	0.890	0.870	0.778	0.513	0.889	0.869	0.776	0.518	0.894	0.875	0.788	0.500
RMSE	0.062	0.071	0.110	0.220	0.063	0.073	0.112	0.230	0.060	0.070	0.106	0.239

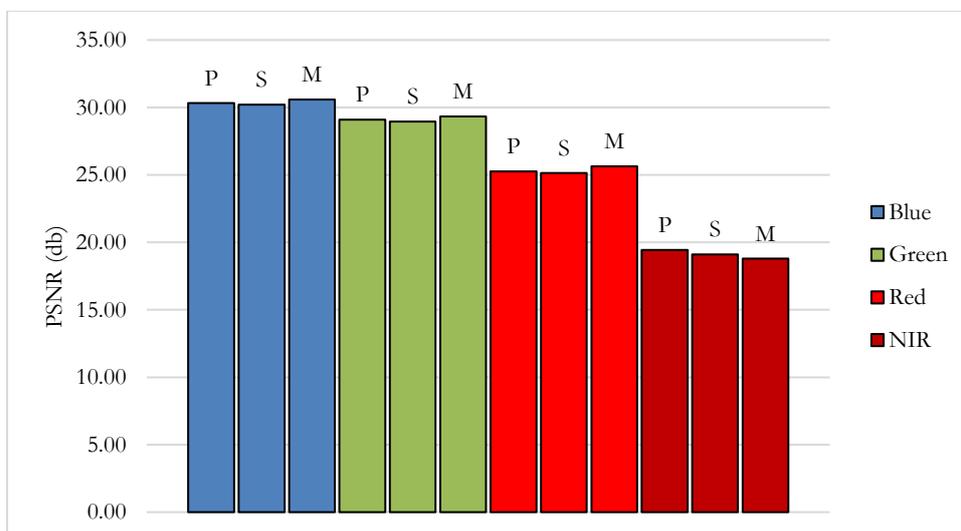


Figure 22 PSNR column chart per band in specific crops

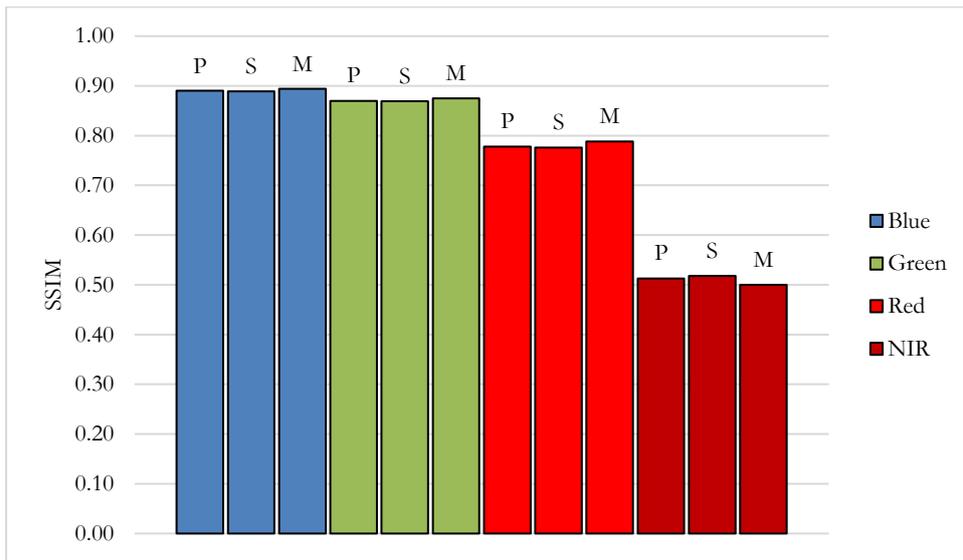


Figure 23 SSIM column chart per band in specific crops

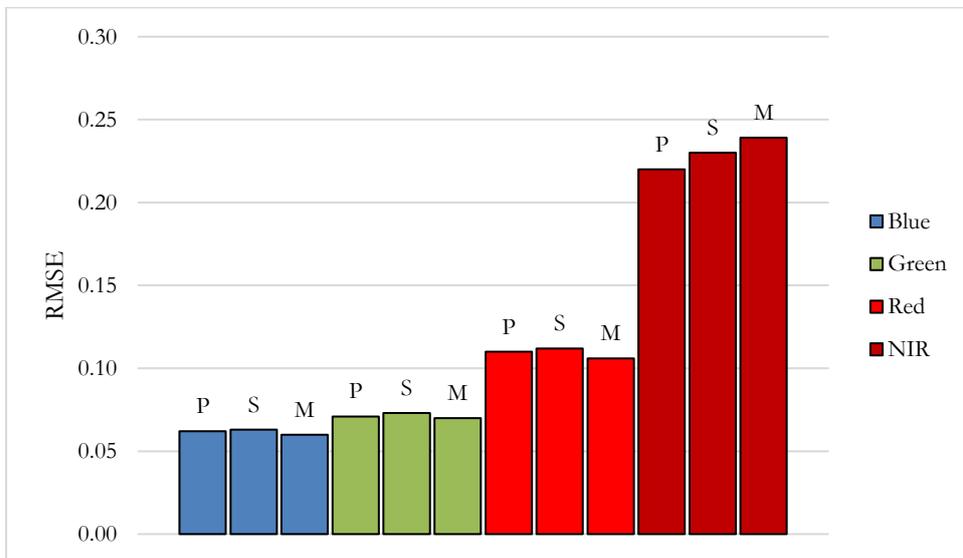


Figure 24 RMSE column chart per band in specific crops

6.3.3. Experiment B

Table 16 shows VIs reconstructed directly; the performance of the model is not good enough because of the reduced training dataset. PSNR values are lower than experiment A and SSIM has negative values which indicates negative correlation. Still, it is interesting to see how data was reconstructed under these conditions in visual inspection.

Table 16 Experiment B evaluation metrics

	NDVI	GNDVI	EVI	SAVI
PSNR	10.213	10.138	9.258	12.667
SSIM	-0.114	-0.147	-0.095	0.255
RMSE	0.657	0.639	0.746	0.446

6.4. Visual evaluation

6.4.1. Experiment A

In Figure 25, ground truth and simulated bands are presented, among ground truth data, they present certain differences in brightness and contrast, and different patches within a Sentinel-2 image correspond to diverse land cover types or terrain characteristics. Vegetation, water bodies, urban areas, and bare soil can exhibit different reflectance properties, leading to variations in the observed brightness and contrast across patches. Generated patches seem to be brighter than real ones, this difference could arise due to variations in the modelling as the rescaling during pre-processing, also the quantitative evaluation it was established has room for improvement.

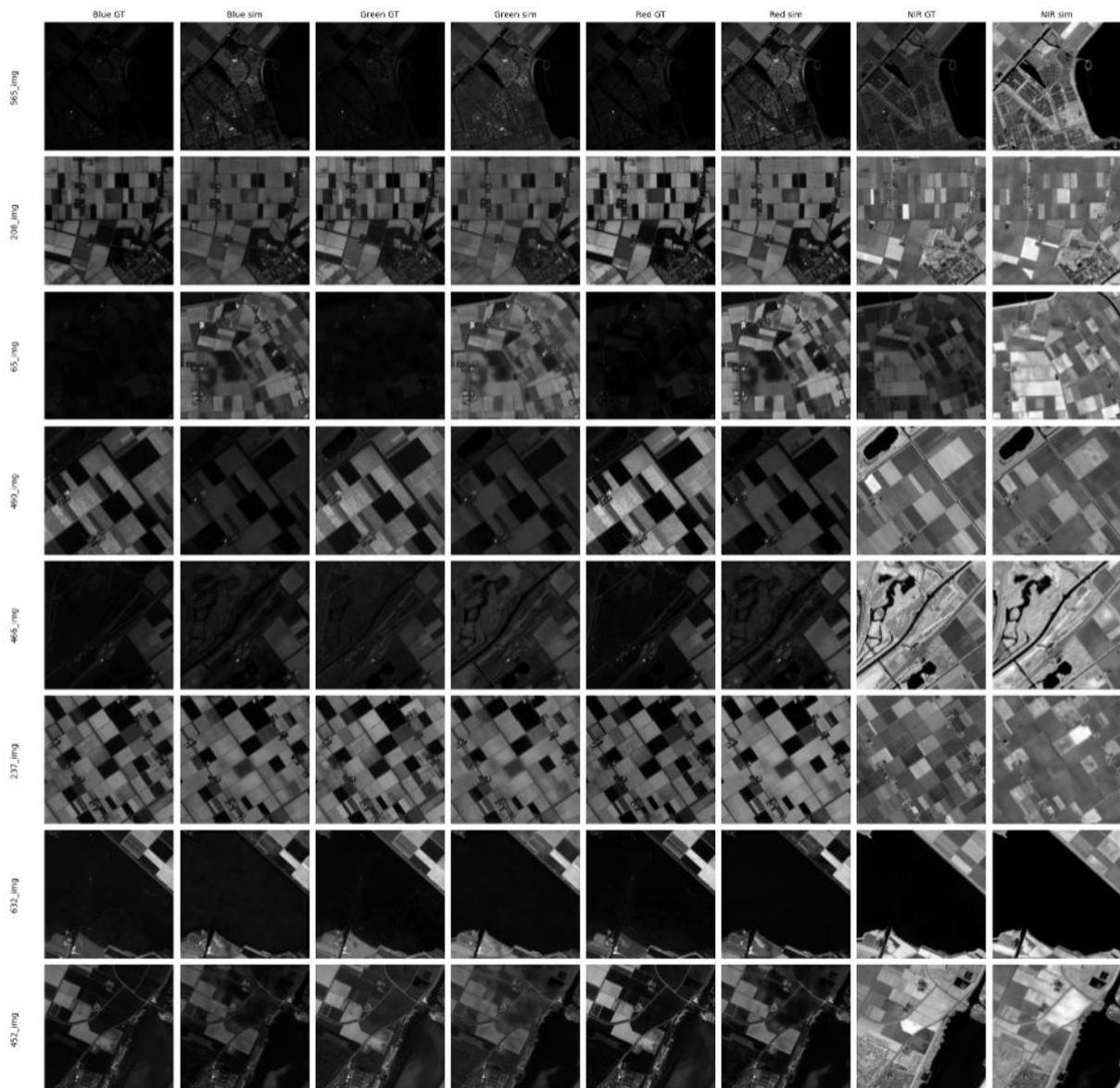
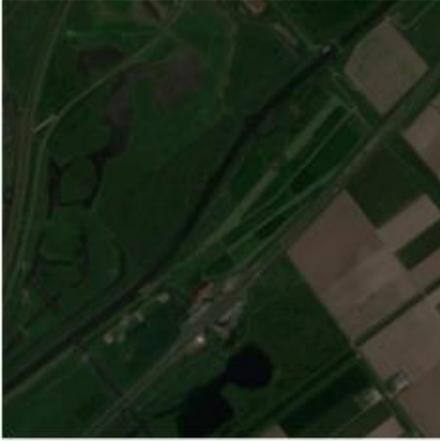


Figure 25 Ground truth and simulated patches per band

Table 17 includes Ground truth and reconstructed RGB images which are compared in different scenarios of heterogeneous or homogeneous parcels in a patch as well as the presence of other land cover.

Table 17 Ground truth and reconstructed RGB compositions comparison for homogeneous and heterogenous areas

Ground truth RGB composition	Reconstructed RGB composition	Description
		<p>These patches contain heterogeneous vegetation and built-up area, edges of similar vegetation are not defined in the simulated data, and the greenness seems to be attenuated or non-existent as shown in the red circles. Built-up on the other hand shows consistency with ground truth.</p>
		<p>These patches contain more homogeneous parcels and some infrastructure. The model predicts with difficulty on homogeneous parcels, edges are blurry, and greenness is attenuated or wrongly predicted in some parcels. Infrastructure seems to be better predicted. Brown parcels with a darker brown in some of them might suggest irrigation, which is not captured in the simulated data.</p>

		<p>These patches contain heterogeneous parcels, infrastructure, and a water body. Vegetation presents the same problems as previous patches, although the greenness seems to be conserved.</p>
		<p>These patches contain homogeneous vegetation and water bodies, the greenness seems to be attenuated.</p>

When the optical dataset has some small clouds, as it can be seen in Figure 26, an optical simulation with some cloudiness, may feasibly reconstruct the image under clouds. Nonetheless, the patches with the presence of clouds are minimal, concluding to what extent clouds can be removed still need test.



Figure 26 Sentinel-2 real (left) and simulated (right) cloud penetration

As expected, VIs are not well reconstructed for reasons mentioned in the quantitative evaluation and illustrated in Figure 27. SAVI seems to be better simulated, and this coincides with the results of the evaluation metrics, the soil brightness correction factor L might be crucial, PSNR values were higher, and it was the only one with a positive SSIM. But this is not conclusive due to the performance of the model.



Figure 27 Vegetation indices calculated with ground truth data and simulated bands

6.4.2. Experiment B

Simulated VIs are illustrated in Figure 28 in greyscale to highlight distortion and deterioration. Due to the reduced dataset, VIs were not successfully reconstructed. As mentioned in evaluation metrics and in visual evaluation on Experiment A, SAVI seems to reassemble better, but not enough to consider it reconstructed.

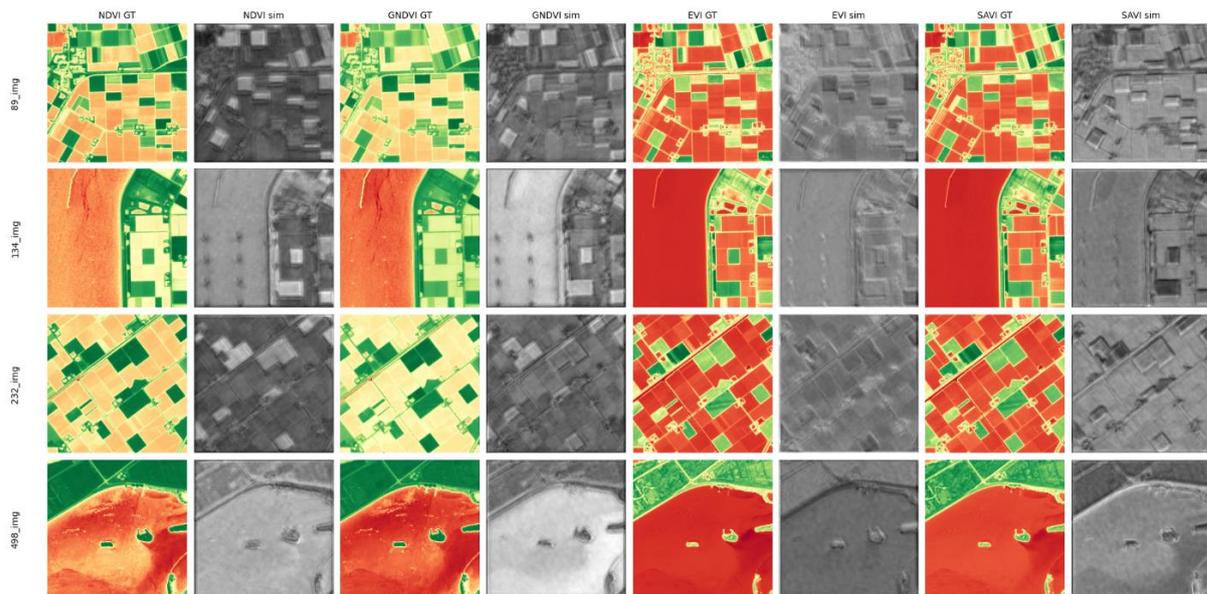


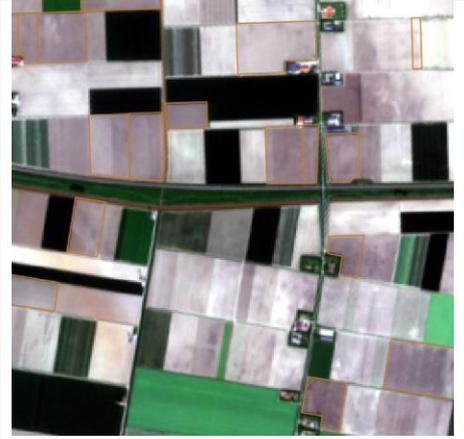
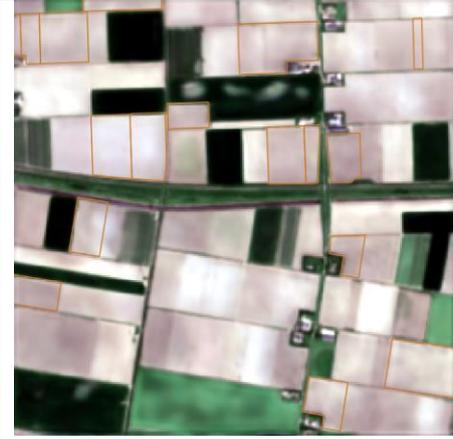
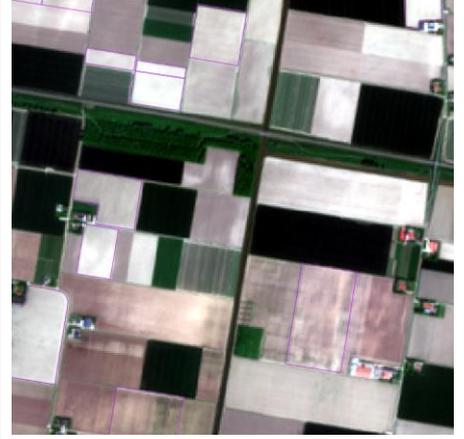
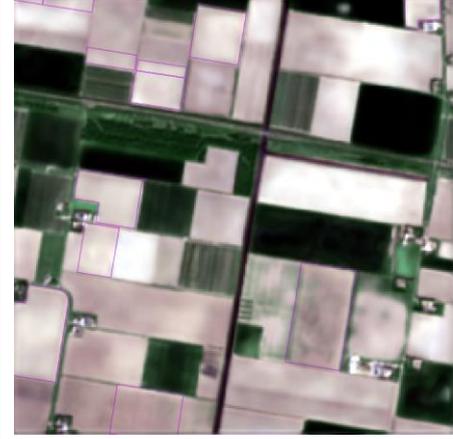
Figure 28 Vegetation indices calculated with ground truth data and directly simulated vegetation indices

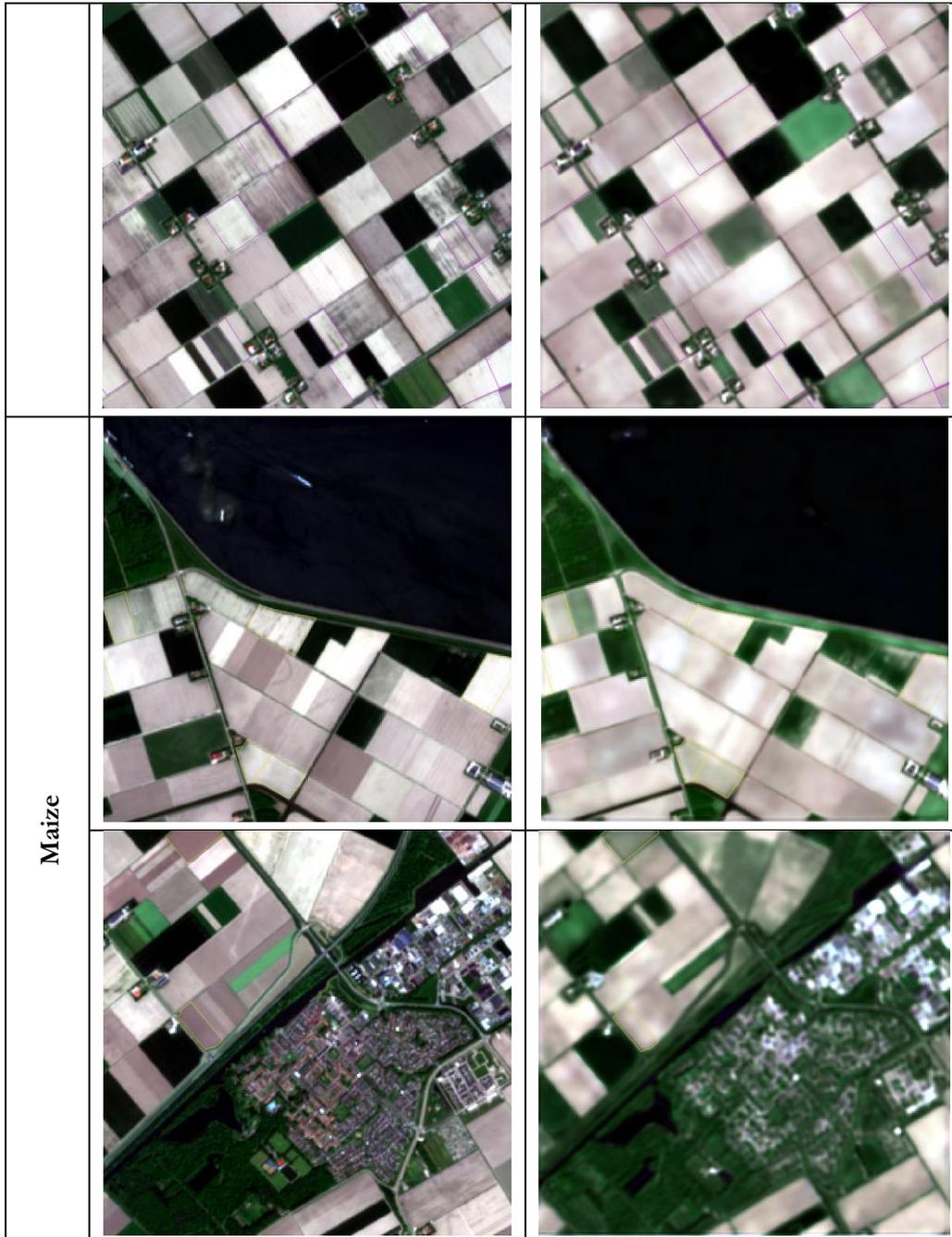
6.4.3. Specific crops

Table 18 contains real and simulated RGB compositions where the target crops are predominant, as seen in Figure 9, during t2 the first phenological stages of maize are there leaf development in the middle of t1 and t2 considered for the model. Potatoes and sugar beets' first observations correspond to flowering at the beginning of July. Therefore, green vegetation is not clearly observable.

An important observation is that in these patches, most of the land cover is cropland, and it seems to present fewer wrong predictions, like green fields shown brown and vice versa. Considering that most of the study area is cropland, the model learns better from the representativeness of the crops, wrong predictions might correspond to crops produced in lower quantities, thus the model cannot learn better from them.

Table 18 Ground truth and reconstructed RGB compositions comparison

Crop	Ground truth RGB composition	Reconstructed RGB composition
Potatoes		
		
Sugar beets		



6.5. Limitations

The comparison of the experiments of this project with the reference research would be more suitable if data collection and pre-processing were the same. In the case of Sentinel-2 data, images downloaded from GEE at Level-2A with a predefined atmospheric correction might introduce different noise than if downloaded from Copernicus Open Hub and processed with SNAP. For Sentinel-1, most of the pre-processing steps are the same, although the choice of speckle filter is not detailed; neither do we know whether the image was acquired in ascending or descending orbit. Additionally, normalization, rescaling and standardization were not deeply described in the references, so this also could be different between the two data situations.

The lack of GPU in the personal laptop and the sometimes-saturated CRIB due to an overwhelming number of users were important limitations. The compilation sometimes was suddenly stopped or could

not be initialized, which led to several failed attempts. Also, utilizing a larger number of patches would have improved the results, especially for Experiment B.

7. CONCLUSIONS

To conclude this research, the research questions are answered:

Which data is available in the study area given cloud cover and crop growing stages?

MTcGAN, a supervised method requires pairs of patches, which means cloud-free Sentinel-2 images in t1 and t2 are needed. The data utilized in this research was obtained at the beginning of May for t1 and the end of May for t2, this sensing interval encompasses only maize during leaf development, which is an early phenology stage, and for potatoes and sugar beets, observable growth above the surface is not detected during this period. Images around the beginning of July may be ideal to capture crop data, eight images were found in the whole year, but any around that desirable time.

It is possible, of course, to use data from other locations, spatial transfer for MTcGAN was proved in reference research, which means that cloud-free data from the whole image and from different locations can be included in the training dataset. Nonetheless, the challenge remains in time transfer, which proved that the longer the difference between t1 and t2 the lower the model's performance, and how good cloud masking can admit patches with reliable cloud-free data, especially because these masking algorithms might admit mistakes. Consequently, the main challenge would be to find cloud-free Sentinel-2 images around the world at t1 and t2 to reduce that error. Finally, using partially clouded patches is possible, but it is unknown to what extent this can influence the model performance.

How generated bands differ from ground truth data?

Red, green, and red bands obtained superior similarity rather than NIR. The distribution from the three bands is similar, thus the model has better chances to learn how to generate data with this distribution. The model showed similar values for training and unseen data which indicated the model has a moderate generalization ability, on the other hand, PSNR and SSIM values obtained were smaller than values reported in the reference research, which indicates that the model is capable of reconstructing images, there is still room for improvement. Given that the same architecture was used for the experiments and reference research, what is left is the difference in size between datasets. Likewise, when trained in different locations performance also changes, different images offer different contextual information, giving more representativeness of the training data.

Can the generated bands be utilized for vegetation indices calculations?

Based on the literature review, it is possible with moderate accuracy, the similarity with actual vegetation indices is not as high as bands separated also utilizing several pair of images around the world for training providing of very extensive real data representativeness. In this research, the reconstruction of NIR was not outstanding thus calculations were not accurate. Nevertheless, SAVI was the one with smaller errors, but not good enough to tell the vegetation index was reconstructed.

How different are the generated vegetation indices from ground truth data?

There is no research found related to the use of a GAN to generate vegetation indices directly. These results were inconclusive in this research due to the reduction in the training dataset carried out for memory allocation issues with the remote machine. Even though SAVI presented better results even with this scenario, but equally the result is not good enough to tell the vegetation index was reconstructed.

Does the model perform better in different crops?

In experiment A, based on the quantitative evaluation, it seems there is not a significant difference in reconstructed bands. Although, these results are based on very early phenological stages, which makes it not conclusive to tell if there is a difference in performance especially for potatoes and sugar beets. In the case of maize, there certain green areas are detected by visual evaluation. A desirable vegetation indices reconstruction accuracy was not achieved in any of both experiments.

8. RECOMMENDATIONS

One of the most important restraints for GAN is the lack of a convergence metric, as well as the non-consensus evaluation metric that could give a stopping criterion and hyperparameter tuning robust decision. Evaluation metrics is a broad study, and certainly in the field of remote sensing have not been deeply studied. The use of Frechet Inception Distance (FID) that models features from hidden layers for the simulated and ground truth data would be suitable for supervised GANs. Also, mean spectral angle is a common metric that can be implemented.

The performance of the model highly depends on diverse training data in a narrow sensing time between t_1 and t_2 , using data from different locations also increases the performance of the method. Using partially clouded images is acceptable, but it must be acknowledged that this can create uncertainty in the model. Using cloud-free images would reduce the uncertainty on this, and as images from different locations improve results then this is complementary. Although this can be time-consuming and computationally costly.

Due to the expected decrease of cloud cover in north-western Europe, the temporal methods to complete vegetation indices time series for Sen4CAP would be more robust. Nonetheless, as data is retrieved every 5 days in Sen4CAP products, a GAN method can provide a near-to-real-time warning of non-compliance with CAP during long periods of cloudiness.

APPENDIX

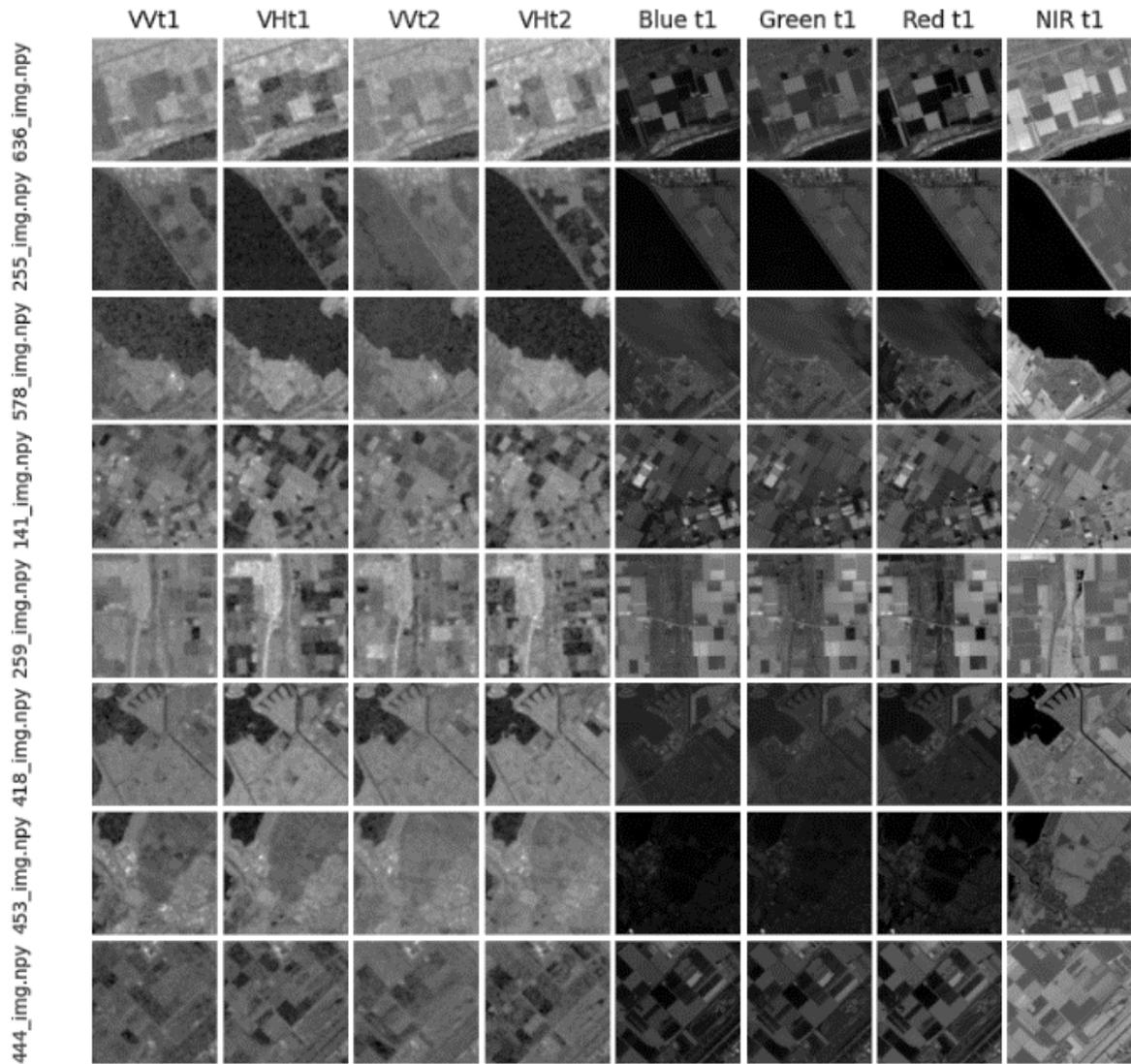
APPENDIX 1: BBCH

Source: (Meier, 2001)

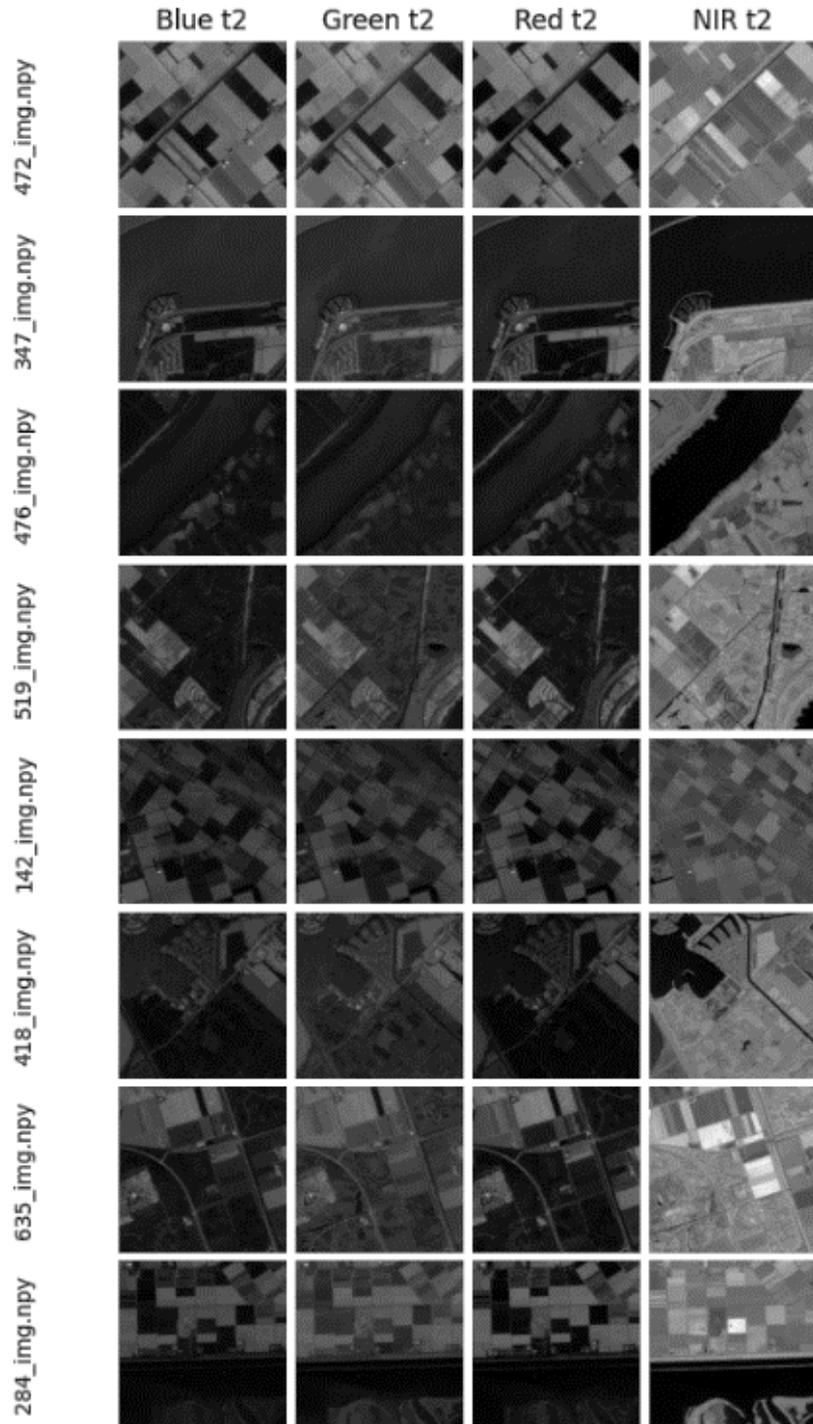
Crop	Growth stage	BBCH identification keys	Description
	Leaf development	12	Two leaves unfolded
		13	Three leaves unfolded
		1...	Stages continuous till
	Stem elongation	31	First node detectable
		33	Three nodes detectable
		3..	Stages continuous till
		39	Nine or more nodes detectable
	Inflorescence emergence, heading	51	Beginning of tassel emergence: tassel detectable at top of stem
	Flowering, anthesis	64	Male: beginning of pollen shedding Female: tips of stigmata visible
	Development of fruit	71	Beginning of grain development: kernels at blister stage, about 16% dry matter
		73	Early milk
		75	Kernels in middle of cob yellowish white (variety-dependent), content milky, about 40% dry matter
		79	All kernels have reached final size
	Ripening	85	Dough stage: kernels yellowish to yellow (variety dependent), about 55% dry matter
87		Physiological maturity: black dot/layer visible at base of kernels, about 60% dry matter	
89		Fully ripe: kernels hard and shiny, about 65% dry matter	
Potatoes	Flowering	61	Beginning of flowering: 10% of flowers in the first inflorescence open (main stem)
		63	Thirty percent of flowers in the first inflorescence open
		65	Full flowering: 50% of flowers in the first inflorescence open
		66	Sixty percent of flowers in the first inflorescence open
		67	Seventy percent of flowers in the first inflorescence open
	Senescence	91	Beginning of leaf yellowing
		93	Most of the leaves yellowish
		95	Fifty percent of the leaves brownish
Sugar beets	Rosette growth	36	Leaves cover 60% of ground
		37	Leaves cover 70% of ground
		39	Crop cover complete: leaves cover 90% of ground

APPENDIX 2: INPUT DATA

Training patches



Ground truth patches



LIST OF REFERENCES

- Abedi, M., Hempel, L., Sadeghi, S., & Kirsten, T. (2022). GAN-Based Approaches for Generating Structured Data in the Medical Domain. *Applied Sciences* 2022, Vol. 12, Page 7075, 12(14), 7075. <https://doi.org/10.3390/APP12147075>
- Aggarwal, C. C. (2018). Neural networks and deep learning. *Springer*, 10, 973–978.
- Bermudez, J. D., Happ, P. N., Oliveira, D. A. B., & Feitosa, R. Q. (2018). SAR TO OPTICAL IMAGE SYNTHESIS FOR CLOUD REMOVAL WITH GENERATIVE ADVERSARIAL NETWORKS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1(1), 5–11. <https://doi.org/10.5194/ISPRS-ANNALS-IV-1-5-2018>
- Bjorck, J., Gomes, C., Selman, B., & Weinberger, K. Q. (2018). Understanding Batch Normalization. *Advances in Neural Information Processing Systems*, 2018-December, 7694–7705. <https://arxiv.org/abs/1806.02375v4>
- Boogaard, E. L. (2021). *Remote sensing assessment of the impact of the 2018 and 2019 droughts on the forests in The Netherlands*.
- Campos-Taberner, M., García-Haro, F. J., Martínez, B., Sánchez-Ruiz, S., & Gilabert, M. A. (2019). A Copernicus Sentinel-1 and Sentinel-2 Classification Framework for the 2020+ European Common Agricultural Policy: A Case Study in València (Spain). *Agronomy* 2019, Vol. 9, Page 556, 9(9), 556. <https://doi.org/10.3390/AGRONOMY9090556>
- Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., & Gattelli, M. (2015). Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. *Remote Sensing* 2015, Vol. 7, Pages 4026-4047, 7(4), 4026–4047. <https://doi.org/10.3390/RS70404026>
- CBS. (2021). *How do we use our land? - The Netherlands in Numbers 2020*. <https://longreads.cbs.nl/the-netherlands-in-numbers-2020/how-do-we-use-our-land/>
- Chamnan, K. (2021). *YIELD MONITORING WITH SENTINEL-2: A FIRST ASSESSMENT FOR THE NETHERLANDS*.
- De Vroey, M., de Vendictis, L., Zavagli, M., Bontemps, S., Heymans, D., Radoux, J., Koetz, B., & Defourny, P. (2022). Mowing detection using Sentinel-1 and Sentinel-2 time series for large scale grassland monitoring. *Remote Sensing of Environment*, 280, 113145. <https://doi.org/10.1016/J.RSE.2022.113145>
- Dumeur, I., Chen, Y., Sur, F., & Zhou, Z.-S. (2021). *Sentinel-2 RGB and NIR bands simulation after fire events using a multi-temporal conditional generative adversarial network*. <https://hal.science/hal-03327421>
- European Commission. (2013). *CAP Reform—An Explanation of the Main Elements*. Brussels: European Commission, Technical Report No MEMO13/937.
- European Court of Auditors. (2020). *Using new imaging technologies to monitor the common agricultural policy : steady progress overall, but slower for climate and environment monitoring. Special report No 04*. <https://op.europa.eu/webpub/eca/special-reports/new-tech-in-agri-monitoring-4-2020/en/#para29>
- European Union. (2018). *Commission Implementing Regulation (EU) 2018/746 of 18 May 2018 amending Implementing Regulation (EU) No 809/2014 as regards modification of single applications and payment claims and checks*.
- Fernandez-Figueroa, E. G., Wilson, A. E., & Rogers, S. R. (2022). Commercially available unoccupied aerial systems for monitoring harmful algal blooms: A comparative study. *Limnology and Oceanography: Methods*, 20(3), 146–158. <https://doi.org/10.1002/LOM3.10477>

- Gao, J., Yuan, Q., Li, J., Zhang, H., & Su, X. (2020). Cloud Removal with Fusion of High Resolution Optical and SAR Images Using Generative Adversarial Networks. *Remote Sensing* 2020, Vol. 12, Page 191, 12(1), 191. <https://doi.org/10.3390/RS12010191>
- Gärtner, P. (2020). *How cloudy is my Sentinel-2 image collection? - The "QA60" band gives insights.* <https://philippgaertner.github.io/2020/08/percent-cloud-cover/>
- Ghamisi, P., & Yokoya, N. (2018). IMG2DSM: Height Simulation from Single Imagery Using Conditional Generative Adversarial Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 794–798. <https://doi.org/10.1109/LGRS.2018.2806945>
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3), 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)
- Glorot, X., & Bengio, Y. (2010a). Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh & M. Titterton (Eds.), *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Vol. 9, pp. 249–256). PMLR. <https://proceedings.mlr.press/v9/glorot10a.html>
- Glorot, X., & Bengio, Y. (2010b). *Understanding the difficulty of training deep feedforward neural networks* (pp. 249–256). JMLR Workshop and Conference Proceedings. <https://proceedings.mlr.press/v9/glorot10a.html>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 27). Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>
- Grnarova, P., Levy, K. Y., Lucchi, A., Perraudin, N., Goodfellow, I., Hofmann, T., & Krause, A. (2018). A domain agnostic measure for monitoring and evaluating GANs. *Advances in Neural Information Processing Systems*, 32. <https://arxiv.org/abs/1811.05512v2>
- Grohnfeldt, C., Schmitt, M., & Zhu, X. (2018). A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018-July, 1726–1729. <https://doi.org/10.1109/IGARSS.2018.8519215>
- He, K., & Sun, J. (2014). Convolutional Neural Networks at Constrained Time Cost. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June-2015, 5353–5360. <https://doi.org/10.1109/CVPR.2015.7299173>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- He, W., & Yokoya, N. (2018). Multi-Temporal Sentinel-1 and -2 Data Fusion for Optical Image Simulation. *ISPRS International Journal of Geo-Information* 2018, Vol. 7, Page 389, 7(10), 389. <https://doi.org/10.3390/IJGI7100389>
- He, W., Zhang, H., Zhang, L., & Shen, H. (2016). Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1), 178–188. <https://doi.org/10.1109/TGRS.2015.2452812>
- Hillsdon, M. (2019). *Dutch farmers plot a greener revolution by sowing biodiversity.* Reuters Events | Sustainable Business. <https://www.reutersevents.com/sustainability/dutch-farmers-plot-greener-revolution-sowing-biodiversity>
- Horé, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. *Proceedings - International Conference on Pattern Recognition*, 2366–2369. <https://doi.org/10.1109/ICPR.2010.579>

- Hristov, J., Clough, Y., Sahlin, U., Smith, H. G., Stjernman, M., Olsson, O., Sahrbacher, A., & Brady, M. V. (2020). Impacts of the EU's Common Agricultural Policy "Greening" Reform on Agricultural Development, Biodiversity, and Ecosystem Services. *Applied Economic Perspectives and Policy*, 42(4), 716–738. <https://doi.org/10.1002/AEPP.13037>
- Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, 32(1), 1–6. <https://doi.org/10.1007/S11676-020-01155-1/FIGURES/2>
- Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2), 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- IDB. (2023). *IDB - List of available Indices*. <https://www.indexdatabase.de/db/i.php?offset=2>
- Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *32nd International Conference on Machine Learning, ICML 2015*, 1, 448–456. <https://arxiv.org/abs/1502.03167v3>
- Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2016). Image-to-Image Translation with Conditional Adversarial Networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-January*, 5967–5976. <https://doi.org/10.48550/arxiv.1611.07004>
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), 31–44. <https://doi.org/10.1109/2.485891>
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695. <https://doi.org/10.1007/S12525-021-00475-2/TABLES/2>
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. <https://doi.org/10.1016/J.COMPAG.2018.02.016>
- Khabbazan, S., Vermunt, P., Steele-Dunne, S., Arntz, L. R., Marinetti, C., van der Valk, D., Iannini, L., Molijn, R., Westerdijk, K., & van der Sande, C. (2019). Crop Monitoring Using Sentinel-1 Data: A Case Study from The Netherlands. *Remote Sensing 2019, Vol. 11, Page 1887*, 11(16), 1887. <https://doi.org/10.3390/RS11161887>
- Kolecka, N., Ginzler, C., Pazur, R., Price, B., & Verburg, P. H. (2018). Regional Scale Mapping of Grassland Mowing Frequency with Sentinel-2 Time Series. *Remote Sensing 2018, Vol. 10, Page 1221*, 10(8), 1221. <https://doi.org/10.3390/RS10081221>
- Kulkarni, S. C., & Rege, P. P. (2020). Pixel level fusion techniques for SAR and optical images: A review. *Information Fusion*, 59, 13–29. <https://doi.org/10.1016/J.INFFUS.2020.01.003>
- LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K. R. (2012). Efficient backprop. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7700 LECTURE NO, 9–48. https://doi.org/10.1007/978-3-642-35289-8_3/COVER
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2323. <https://doi.org/10.1109/5.726791>
- Lemberger, P. (2017). *On Generalization and Regularization in Deep Learning*. <https://arxiv.org/abs/1704.01312v2>
- Li, Y., Fu, R., Meng, X., Jin, W., & Shao, F. (2020). A SAR-to-Optical Image Translation Method Based on Conditional Generation Adversarial Network (cGAN). *IEEE Access*, 8, 60338–60343. <https://doi.org/10.1109/ACCESS.2020.2977103>
- LNV. (2021). *Agriculture and horticulture*. <https://www.government.nl/topics/agriculture/agriculture-and-horticulture>

- Luo, H., Bo, L., Peng, C., & Hou, D. (2022). An Improved Convolutional-Neural-Network-Based Fault Diagnosis Method for the Rotor–Journal Bearings System. *Machines* 2022, Vol. 10, Page 503, 10(7), 503. <https://doi.org/10.3390/MACHINES10070503>
- MacKerron, D. K. L., & Waister, P. D. (1985). A simple model of potato growth and yield. Part I. Model development and sensitivity analysis. *Agricultural and Forest Meteorology*, 34(2–3), 241–252. [https://doi.org/10.1016/0168-1923\(85\)90024-3](https://doi.org/10.1016/0168-1923(85)90024-3)
- Meier, U. (2001). *Growth stages of mono-and dicotyledonous plants* BBCH Monograph Federal Biological Research Centre for Agriculture and Forestry.
- Mirza, M., & Osindero, S. (2014). *Conditional Generative Adversarial Nets*. <https://arxiv.org/abs/1411.1784v1>
- Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context Encoders: Feature Learning by Inpainting. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December*, 2536–2544. <https://doi.org/10.1109/CVPR.2016.278>
- Pfeifroth, U., Sanchez-Lorenzo, A., Manara, V., Trentmann, J., & Hollmann, R. (2018). Trends and Variability of Surface Solar Radiation in Europe Based On Surface- and Satellite-Based Data Records. *Journal of Geophysical Research: Atmospheres*, 123(3), 1735–1754. <https://doi.org/10.1002/2017JD027418>
- Qiu, F., Berglund, J., Jensen, J. R., Thakkar, P., & Ren, D. (2004). Speckle noise reduction in SAR imagery using a local adaptive median filter. *GIScience and Remote Sensing*, 41(3), 244–266. <https://doi.org/10.2747/1548-1603.41.3.244>
- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*. <https://arxiv.org/abs/1511.06434v2>
- Rouse, J. W. , Jr., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A*.
- Sanders, M., Henkens, R., & Slijkerman, D. (2019). *Convention on Biological Diversity - Sixth National Report of the Kingdom of the Netherlands*. <https://doi.org/10.18174/499170>
- Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., Miklas, P. N., Carter, A. H., Pumphrey, M. O., Knowles, R. R. N., & Pavek, M. J. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *European Journal of Agronomy*, 70, 112–123. <https://doi.org/10.1016/J.EJA.2015.07.004>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/J.NEUNET.2014.09.003>
- Sen4CAP. (2020). *Common Agricultural Policy Sen4Cap*. Project Background. <http://esa-sen4cap.org/node/1>
- Sener, E., Çolak, E., Erten, E., & Taskin, G. (2021). THE ADDED VALUE OF CYCLE-GAN FOR AGRICULTURE STUDIES. *International Geoscience and Remote Sensing Symposium (IGARSS), 2021-July*, 7039–7042. <https://doi.org/10.1109/IGARSS47720.2021.9553876>
- Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., & Zhang, L. (2015). Missing Information Reconstruction of Remote Sensing Data: A Technical Review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3), 61–85. <https://doi.org/10.1109/MGRS.2015.2441912>
- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. <https://arxiv.org/abs/1409.1556v6>
- Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), 20260–20264. https://doi.org/10.1073/PNAS.1116437108/SUPPL_FILE/PNAS.201116437SI.PDF

- UN. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development Transforming our world: the 2030 Agenda for Sustainable Development Preamble*.
- UN. (2019). *World Population Prospects 2019: Highlights (ST/ESA/SER.A/423)*. https://population.un.org/wpp/publications/files/wpp2019_highlights.pdf
- van der Wal, T., Abma, B., Viguria, A., Prévinaire, E., Zarco-Tejada, P. J., Serruys, P., van Valkengoed, E., & van der Voet, P. (2013). *Fieldcopter: unmanned aerial systems for crop monitoring services*. 169–175. https://doi.org/10.3920/978-90-8686-778-3_19
- van Dijk, M., Morley, T., Rau, M. L., & Saghai, Y. (2021). A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nature Food* 2:7, 2(7), 494–501. <https://doi.org/10.1038/s43016-021-00322-9>
- van Oort, P. A. J., Timmermans, B. G. H., Schils, R. L. M., & van Eekeren, N. (2023). Recent weather extremes and their impact on crop yields of the Netherlands. *European Journal of Agronomy*, 142, 126662. <https://doi.org/10.1016/J.EJA.2022.126662>
- Viviano, F. (2017). *National Geographic*. This Tiny Country Feeds the World. <https://www.nationalgeographic.com/magazine/article/holland-agriculture-sustainable-farming>
- Wang, L., Xu, X., Yu, Y., Yang, R., Gui, R., Xu, Z., & Pu, F. (2019). SAR-to-optical image translation using supervised cycle-consistent adversarial networks. *IEEE Access*, 7, 129136–129149. <https://doi.org/10.1109/ACCESS.2019.2939649>
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. <https://doi.org/10.1109/TIP.2003.819861>
- Xiong, Q., Di, L., Feng, Q., Liu, D., Liu, W., Zan, X., Zhang, L., Zhu, D., Liu, Z., Yao, X., & Zhang, X. (2021). Deriving Non-Cloud Contaminated Sentinel-2 Images with RGB and Near-Infrared Bands from Sentinel-1 Images Based on a Conditional Generative Adversarial Network. *Remote Sensing* 2021, Vol. 13, Page 1512, 13(8), 1512. <https://doi.org/10.3390/RS13081512>
- Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40. <https://doi.org/10.1109/MGRS.2016.2540798>
- Zhang, Q., Yuan, Q., Zeng, C., Li, X., & Wei, Y. (2018). Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8), 4274–4288. <https://doi.org/10.1109/TGRS.2018.2810208>
- Zhang, X., Qiu, Z., Peng, C., & Ye, P. (2021). Removing Cloud Cover Interference from Sentinel-2 imagery in Google Earth Engine by fusing Sentinel-1 SAR data with a CNN model. <https://doi.org/10.1080/01431161.2021.2012295>, 43(1), 132–147.
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision, 2017-October*, 2242–2251. <https://doi.org/10.1109/ICCV.2017.244>
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36. <https://doi.org/10.1109/MGRS.2017.2762307>