Masters Program in Geospatial Technologies

Developing Fictitious Country Maps through Generative AI Techniques

Aleksandra Jastrzębska

Dissertation submitted in partial fulfilment of the requirements for the Degree of *Master of Science in Geospatial Technologies*

NOVA IMS Information Management School
Universitat Jaume I
Westfälische Wilhelms-Universität Münster

Developing Fictitious Country Maps through Generative AI Techniques

Aleksandra Jastrzębska

Thesis submitted in partial fulfilment of the requirements for the Degree of Master of Science in Geospatial Technologies

Supervised by:

Prof. Dr. Jose Francisco Ramos Romero
Institute of New Imaging Technologies
Universitat Jaume I
Castellón de la Plana, Spain

Co-supervised by:

Prof. Dr. Benjamin Risse
Institute for Geoinformatics
University of Münster
Münster, Germany

Dr. Vicente Tang
NOVA Information Management School
Universidade Nova de Lisboa
Lisboa, Portugal

March 2025

Declaration of Academic Integrity

I hereby confirm that the thesis titled “Developing Fictitious Country Maps through Generative AI Techniques” is entirely my original work, completed with the guidance of my supervisors.
All sources used, including books, journals, handouts, unpublished manuscripts, and various internet resources, have been accurately cited.
This thesis has not been approved for any degree and is not currently being submitted for any other academic qualification.

Aleksandra Jastrzębska,
Castellón de la Plana, Spain
31.01.2025

Acknowledgements

First and foremost, I would like to express my gratitude to Leon Pielage.
His support, patience and guidance throughout this process have been invaluable. This thesis would not have been the same without his dedication and encouragement.

I am also grateful to all my official and unofficial supervisors, who guided me, provided important feedback and continuously pushed me to grow, both academically and personally.

And last but not least, a big thank you goes to my family, long-distance and on-site friends, who helped me get through everyday struggles — keeping me grounded when needed and lifting me up when necessary.

Abstract

This thesis explores the application of diffusion models to generate high-resolution maps for the fictional country of Carana, a scenario used by international peacekeeping organizations for training and strategic planning. The study addresses limitations of the manual creation of raster representations and limitations of existing imagery, which often lack the detail and adaptability required for various simulations.

To achieve this, a comprehensive framework was developed, beginning with the acquisition of Sentinel-2 satellite imagery and preprocessing the data into 64x64 pixel tiles. A diffusion model, based on the U-Net architecture, was adapted to process these tiles, with training conducted on high-performance computing resources. Validation of the generated maps was performed using histogram-based analysis and Fréchet Inception Distance (FID) scores, with additional assessments focusing on spatial coherence and color consistency.

While the framework successfully produced synthetic map tiles, challenges such as color mismatches and tile discontinuities highlighted areas for improvement. These challenges led to recommendations for future improvements, including testing various hyperparameters, advanced validation techniques and later on the integration of conditional generation using geospatial features.

Keywords: Carana, Diffusion models, Generative AI, Peacekeeping operations, Sentinel-2

Table of Contents

List of Figures:

1. Introduction:

Effective peacekeeping training and strategic decision-making require high-quality, realistic maps that accurately depict terrain, infrastructure, and environmental conditions. Carana, a fictional country developed for United Nations (UN) training exercises, serves as a scenario for simulating geopolitical conflicts, humanitarian crises and security operations. These simulations help military leaders, policymakers, and humanitarian workers analyze complex situations and develop strategic responses. However, the existing Carana map representations lack the necessary detail, adaptability, and realism to fully support these training exercises. This limitation highlights the need for a method capable of generating high-resolution, geospatially coherent synthetic maps that enhance the effectiveness of peacekeeping training.

One potential solution lies in GeoAI, an emerging field that integrates Geographic Information Systems (GIS) with Artificial Intelligence (AI) to address complex spatial challenges. GeoAI has been successfully applied in various domains, including urban planning, environmental monitoring, and disaster response, where AI-driven models improve the analysis, prediction, and visualization of geospatial data. Generative AI, a subset of AI that focuses on the creation of new, data-driven outputs, presents a promising approach for synthetic map generation, particularly through advanced machine learning techniques such as diffusion models.

This thesis explores the use of diffusion models to generate realistic, high-resolution maps for Carana. Diffusion models, originally developed for image synthesis and enhancement, have demonstrated significant potential in geospatial applications. By training a generative model on real-world satellite imagery, this research aims to produce synthetic maps that could be integrated into UN training simulations.
The proposed approach seeks to improve visual realism, spatial coherence, and adaptability in geospatial map generation, contributing to the broader application of AI in peacekeeping and crisis management.

1.1. Motivation and Problem Definition

Effective peacekeeping training and making strategic decisions require realistic, high-quality maps that accurately represent the terrain, infrastructure, and environmental conditions of conflicted regions. The fictional country of Carana[1], situated on the island of Kisiwa off the eastern coast of Africa, has been developed as a comprehensive training scenario by international organizations such as the United Nations. This setting serves as a valuable tool for modeling complex geopolitical situations, including ethnic conflicts, resource disputes, humanitarian crises and infrastructure planning. However, a major limitation in Carana's current application is the lack of detailed visual representations.

Current maps used in training simulations often suffer from insufficient detail and limited flexibility, which restrict being effective in dynamic operational planning. For peacekeeping personnel, military strategists and humanitarian workers, access to high-resolution, contextually accurate maps is crucial for simulating crisis response, resource distribution and security operations.

To address this challenge, this research explores the use of diffusion models for generating high-quality, synthetic geospatial data. Diffusion models have demonstrated remarkable success in image synthesis, gradually refining noise into detailed, high-resolution outputs, making them a promising tool for creating realistic maps (Dhariwal & Nichol, 2021). By utilizing this generative approach, the goal is to develop a diffusion model-based framework capable of producing synthetic, high-resolution maps for Carana. These maps would improve peacekeeping training, urban planning simulations, and disaster response exercises, offering a scalable and adaptable solution for geospatial visualization in strategic decision-making.

1.2. Objectives:

The goal of this thesis is to develop a diffusion model-based framework capable of generating high-resolution maps for the fictional country of Carana.

To achieve this goal, they have been divided into a set of objectives:

- Apply the data science project pipeline to develop the framework.
- Generate individual tiles and systematically assemble them into a coherent, high-resolution map.
- Validate and refine the model outputs by comparing generated maps to real-world samples.

[1] https://paxsims.wordpress.com/wp-content/uploads/2009/08/carana_long_version_english.pdf

1.3. Thesis outline

This thesis is structured into seven main chapters, each detailing a key aspect of the research process, from foundational concepts to implementation, results, and conclusions. The organization follows a logical progression, ensuring clarity in the development, training, and validation of the diffusion model-based framework for generating high-resolution maps of Carana.

Chapter 1: Introduction, provides an overview of the research, outlining the motivation and problem definition, the objectives of the study, and the structure of the thesis.

Chapter 2: Literature Review presents an analysis of relevant research, including background concepts in deep learning, particularly generative AI, a review of diffusion models applied in GIS, and an examination of the existing research gap. Additionally, it explores alternative generative models and other approaches to satellite imagery generation, providing a comparative analysis of different methodologies.

Chapter 3: Study Area describes the data sources and the geographical context of the study. It introduces Sentinel-2 satellite imagery, which serves as the training dataset, and Carana, the fictional training scenario used by international peacekeeping organizations.

Chapter 4: Methodology details the technical implementation of the study. It begins with data preprocessing, covering the acquisition, segmentation, and cleaning of Sentinel-2 RGB imagery. The processing and training pipeline is explained, describing the initial training on a local machine and the full-scale training on HPC Palma II. Moreover, the validation methods are described, then the chapter concludes with an explanation of the georeferencing process, which aligns generated maps with real-world coordinates.

Chapter 5: Results present the outcomes of the model training and evaluation, assembled to grid structures.

Chapter 6: Discussion and Future Work evaluates the limitations of the current approach, including data constraints, computational challenges, and color inconsistencies in generated maps. It also explores potential improvements, such as enhancing validation techniques, improving tile coherence and incorporating conditional generation using structured geospatial data.

Chapter 7: Conclusions summarizes the key contributions and findings of the study.

Chapter 8: Bibliography provides references to the scientific literature.

2. Literature Review

2.1. Background Concepts

Deep learning, a branch of machine learning rooted in artificial intelligence, has become an essential technology in contemporary geoscience. The integration of machine learning into Geographic Information (GI) systems has significantly increased the efficiency of spatial analysis and prediction capabilities. By the 2000s, machine learning gained popularity across various disciplines, including geography, where unsupervised learning methods made it possible to analyze large datasets. In recent years, the special branch of deep learning called generative models have advanced to generate limitless high-quality synthetic images (Brock et al., 2019; Dhariwal & Nichol, 2021; Razavi et al., 2019).

In geosciences, generative models such as Generative Adversarial Networks (GANs) have become increasingly popular, initially in unsupervised applications and later in supervised and semi-supervised frameworks (Mateo-García et al., 2021). While GANs currently lead in image generation quality, they face challenges such as limited diversity and training instability, which require carefully tuned hyperparameters and regularization techniques (Brock et al., 2019). In contrast, diffusion models, a class of likelihood-based generative models, have recently appeared as an alternative, demonstrating their ability to produce high-quality images (Sohl-Dickstein et al., 2015). These models offer several advantages, including stable training objectives, full distribution coverage, and easy scalability. Diffusion models rely on a unique framework that involves progressively removing noise from the data (Ho et al., 2020).

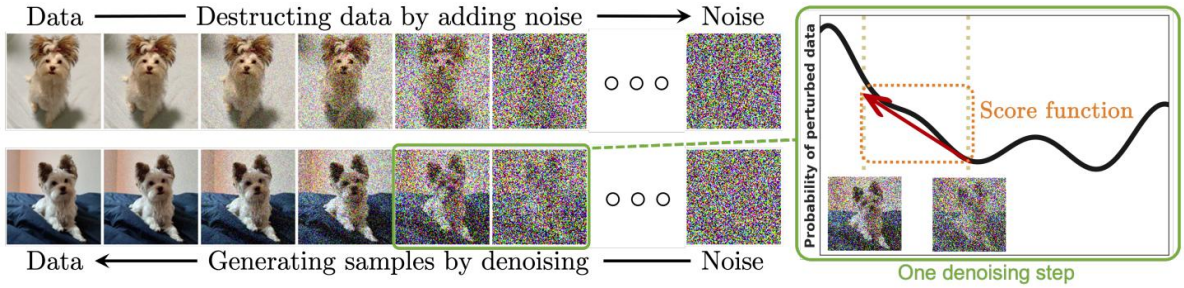

Figure 1: Process of adding noise and denoising (Yang et al., 2024)

This study builds upon the Denoising Diffusion Probabilistic Model (DDPM) proposed by Dhariwal and Nichol (Dhariwal & Nichol, 2021), which is an improve of the model primarily proposed by Ho, both based on the U-Net architecture and adapted here for the generation of satellite images (Ho et al., 2020).

Diffusion models are the a class of latent variable models, where the joint distribution defines the relationship between the original data x_0 and the sequence of latent variables $x_1, \ldots, x_T$. The reverse process aims to approximate this joint distribution by iteratively reconstructing x_0 from x_T. The forward process $q(x_{1:T} | x_0)$ is a Markov chain that gradually deforms the data x_0 by adding Gaussian noise at each step according to a predefined variance schedule. This ensures that x_T becomes nearly pure Gaussian noise after T steps. Importantly, the forward process is not learned; it is predefined and uses fixed Gaussian transitions to structure the model and generate training data for the reverse process (Ho et al., 2020).

In contrast, the reverse process $p_\theta(x_{t-1} | x_t)$ is the learned component of diffusion models and forms the core of their generative capability. Starting from pure noise $x_T \sim N(0, I)$, the reverse process iteratively removes noise step by step, reconstructing the original data x_0. Each reverse step is parameterized as a Gaussian distribution, where the mean $\mu_\theta(x_t, t)$ and variance $\Sigma_\theta(x_t, t)$ are predicted by a neural network, in this case U-Net. These predictions are optimized during training to minimize the inconsistency between the model's predicted noise and the actual noise added in the forward process.

The foundational work on diffusion models introduced a simplified training objective that focuses on directly predicting the noise added during the forward process. This loss function, referred to as the denoising objective, simplifies the variational lower bound (VLB) optimization:

$$L_{simple} = \mathbb{E}_{t, x_0, \varepsilon_0}\left[\left|\left|\varepsilon - \varepsilon_\theta(x_t, t)\right|\right|^2\right]$$

This formulation ensures computational efficiency while enabling the model to learn effective denoising transitions at every timestep. Building on this foundation, Nichol and Dhariwal proposed improvements to the training objective, incorporating a hybrid loss that combines the simplified objective with the VLB. This hybrid approach balances the model's ability to generate high-quality samples while improving its likelihood-based metrics (Ho et al., 2020).

The variational lower bound (VLB) serves as the base for training diffusion models, providing a mathematical framework to approximate the true log-likelihood of the data. Directly computing the log-likelihood is often intractable due to the high-dimensional nature of the latent variables. Instead, the VLB decomposes the problem into manageable terms, making the optimization possible. The first term ensures that the noisy state aligns with the Gaussian prior, enabling the model to effectively learn the data distribution. The subsequent terms involve comparing the reverse process with the forward posterior at each timestep. This stepwise evaluation ensures that the

reverse process can accurately map the noisy data back to its original state. Finally, the reconstruction loss ensures that the model can reliably predict from, which is the least noisy latent variable in the chain (Sohl-Dickstein et al., 2015).

Jonathan Ho's simplified objective sidesteps the need for calculating every component of the VLB by concentrating on noise prediction. While this approach is computationally efficient and leads to high-quality samples, it does not fully optimize the log-likelihood. Nichol and Dhariwal addressed this limitation by introducing a hybrid loss that combines the simplified objective with select components of the VLB.

A critical component of both approaches is the use of the U-Net architecture to parameterize the reverse process. The U-Net's encoder-decoder structure, augmented with skip connections, allows the model to preserve both local, like small, detailed patterns in an image and global features, like overall structure or context of an image during denoising. Nichol and Dhariwal further enhanced this design by integrating multi-head attention mechanisms (Vaswani et al., 2023), enabling the model to focus on relevant regions in the noisy data.

Another significant improvement introduced by Nichol and Dhariwal is the cosine noise schedule, which replaces the original linear variance schedule while adding the noise to the data. The cosine schedule adjusts the rate of noise addition, preserving more information in the early steps and reducing overfitting.

While diffusion models stand out in generating diverse and realistic outputs, they also pose challenges in terms of computational efficiency. The iterative nature of their sampling process can be computationally demanding, especially when working with large datasets like satellite imagery. However, their ability to maintain stable training and capture the full data distribution makes them a powerful choice for applications where quality and diversity are essential.

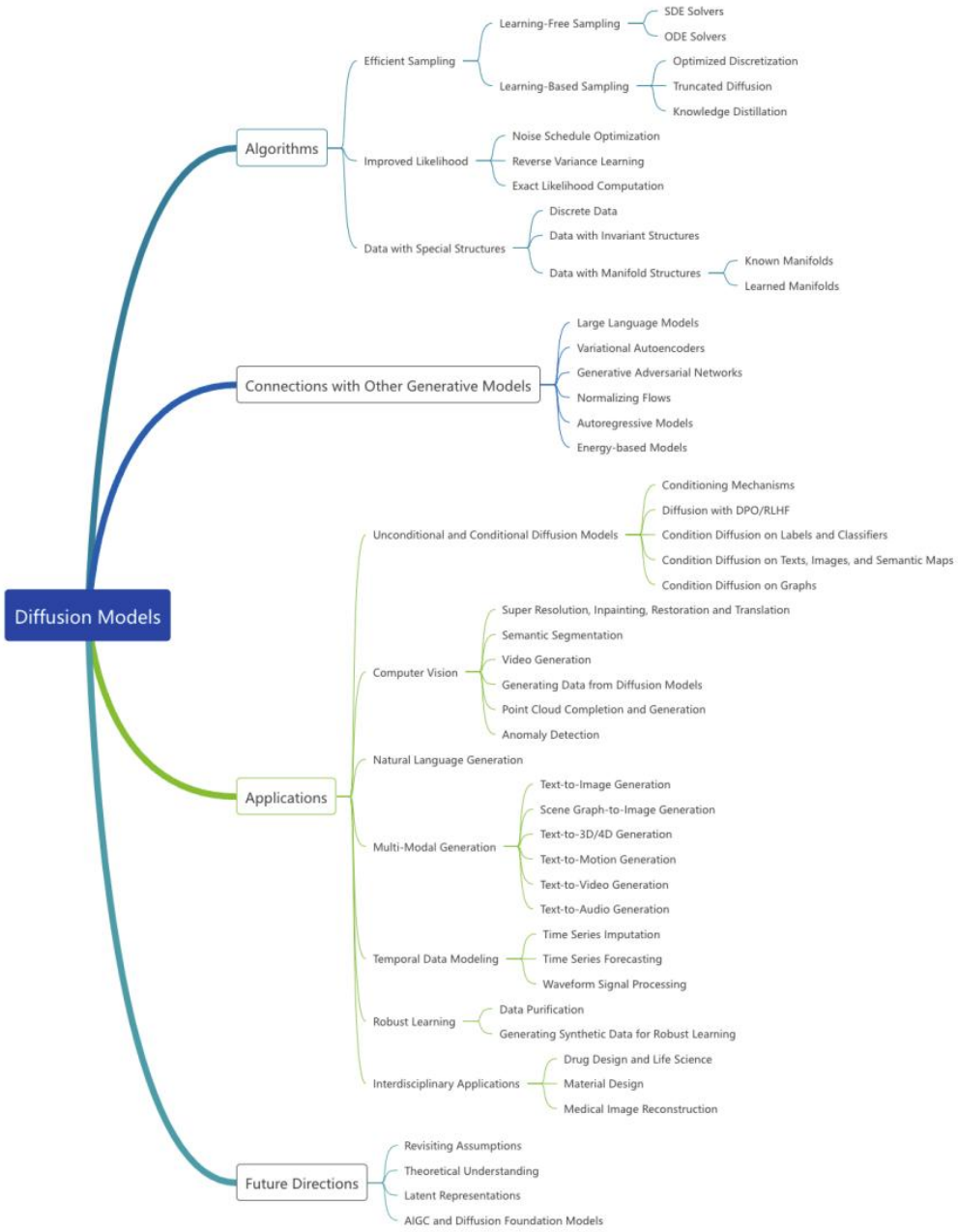

Figure 2: Variety of types of diffusion models (Yang et al., 2024)

2.2. Diffusion models in GIS

Diffusion models are emerging as powerful tools in GIS, particularly for precipitation nowcasting. Unlike traditional numerical weather prediction models, which rely on physics-based simulations, diffusion-based approaches generate probabilistic forecasts by refining noise into realistic weather patterns. Asperti et al. (2025) introduced Generative Ensemble Diffusion (GED), which produces multiple plausible rainfall scenarios using ERA-5 meteorological data, achieving a 25% error reduction compared to existing deep learning models (Asperti et al., 2025).

Beyond forecasting, diffusion models have been effectively applied in satellite image super-resolution, urban change detection, and land cover classification, significantly improving geospatial data quality and enabling more precise environmental monitoring. For instance, Luo et al. (2024) introduced SatDiffMoE, a diffusion-based algorithm that fuses sequential low-resolution satellite images to reconstruct high-resolution outputs, thereby improving applications in land crop monitoring and urban planning (Luo et al., 2024). In the realm of urban change detection, Sanguigni et al. (2023) demonstrated the utility of diffusion models in generating synthetic datasets for change detection tasks, facilitating urban replanning efforts (Sanguigni et al., 2023).

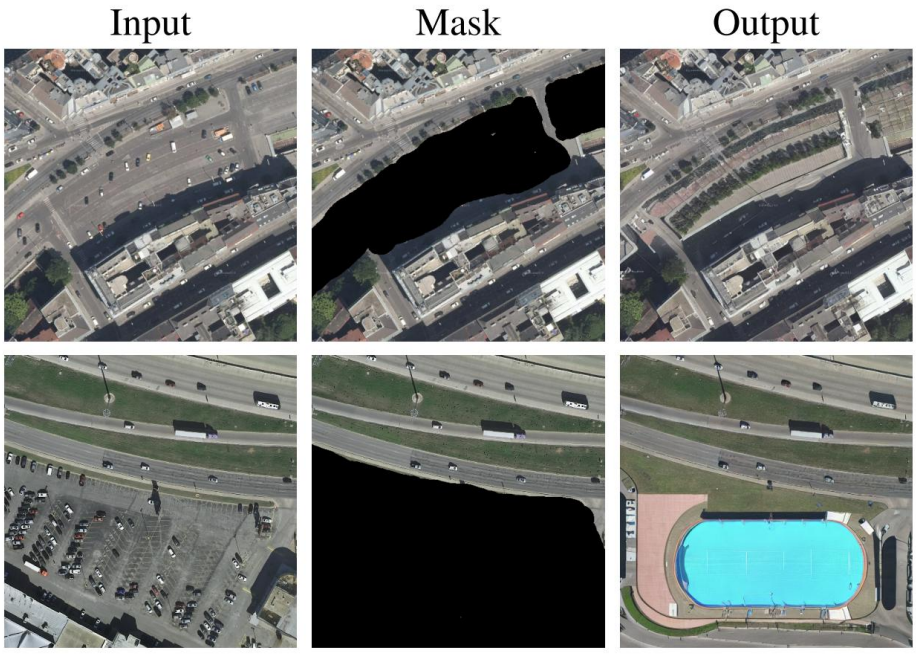

Figure 3: Example of urban replanning visualization (Sanguigni et al., 2023)

Additionally, Jiang et al. (2025) proposed a diffusion-based remote sensing image fusion method that improves classification accuracy by merging hyperspectral and LiDAR data, proving beneficial for land cover classification. Traditional fusion methods rely on handcrafted features or simple pixel-wise operations, but the diffusion model learns spatial feature distributions through iterative noise removal, leading to higher-quality fused images. Furthermore, this model remains effective even when tested on multispectral images (MSI) and synthetic aperture radar (SAR) data, showing its adaptability across different remote sensing applications (Jiang et al., 2025).

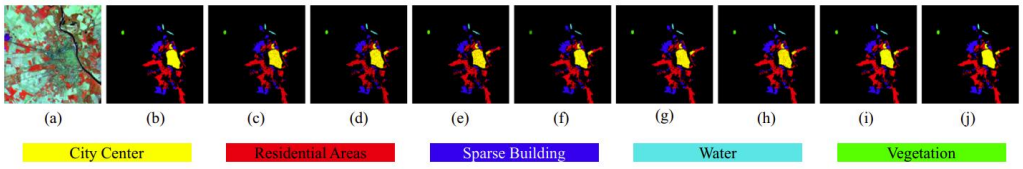

Fig. 9 Classification maps obtained by different methods on the grss-dfc-2007 dataset: (a) pseudocolor image, (b) ground truth maps, (c) TB-CNN (%90.44), (d) DeepCNN (%86.82), (e) FusAtNet (%89.15), (f) EndNet (%92.74), (g) DFINet (%93.13), (h) CALC (%92.15), (i) proposed (only MSI) (%96.46), (j) proposed (MSI-SAR pair) (%97.97)

Figure 4: Example of land use classification depending on used diffusion model (Jiang et al., 2025)

What is more, *The Map Diffusion model*, introduced by Przymus and Szymański (2023), represents a significant advancement in generative map synthesis, enabling text-promptable map generation trained on OpenStreetMap (OSM) data. Unlike conventional GIS-based approaches that rely on predefined datasets, this model allows users to generate maps by providing natural language descriptions, making it highly adaptable for urban planning, geospatial visualization and automated cartography.

Technically, the model is built on a latent diffusion framework similar to Stable Diffusion, where a denoising U-Net architecture progressively refines a noisy input into a realistic synthetic map tile. The training dataset consists of OSM-derived map tiles, each paired with automatically generated text prompts describing key geographic elements such as roads, buildings, and land use. For example, a prompt like "OSM from Warsaw, Poland, of a suburban area containing: 5 residential buildings, 2 primary roads, 1 park, and a river" generates a corresponding map tile with these features spatially arranged in a realistic manner. The model is conditioned on these structured textual descriptions, allowing users to specify geographical attributes, urban density, and terrain types, leading to accurate synthetic maps (Przymus & Szymański, 2023).

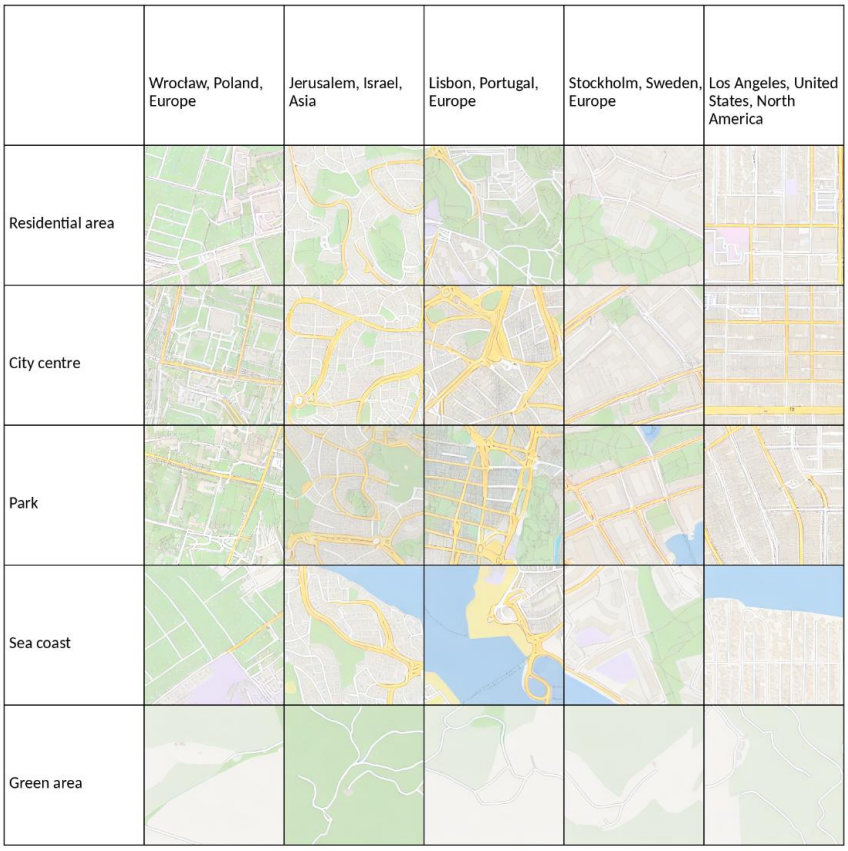

Figure 5: Example of generated samples from text prompts (Przymus & Szymański, 2023)

Similarly to the previous example, based on the OpenStreetMap data, Espinosa and Crowley (2023) introduced a novel approach to satellite image generation. Their method is a diffusion-based generative model, conditioned on cartographic data, to synthesize high-resolution satellite imagery that aligns with real-world geographic features. By employing a ControlNet-based architecture, their model changes space noise into detailed satellite images that accurately reflect the spatial structures of roads, buildings, and natural landscapes as represented in OpenStreetMap.
The dataset used for training consists of paired cartographic and satellite imagery from Mainland Scotland and the Central Belt, allowing the model to learn geospatial relationships and urban structures with high fidelity (Espinosa & Crowley, 2023).

Figure 6: Example of the OSM training data with real-world examples and the generated images (Espinosa & Crowley, 2023)

2.3. Research gap

2.3.1. Other generative models

In the field of generative modeling, several methods have been used in producing images and advancing the quality of synthetic images. These methods include GANs, ImageNet, Transformer-based architecture and VAEs, each contributing unique innovations and addressing specific challenges in image synthesis. This section provides an overview of these approaches, highlighting their principles, strengths, and limitations, with a focus on their relevance to generative tasks like high-resolution mapping and geospatial data modeling.

- Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) have shown significant promise in geosciences, improving land cover classification, urban planning and satellite image synthesis (Lütjens et al., 2021; Mateo-García et al., 2021).
They are designed to produce new data that resembles the input data. At their core, GANs consist of two competing neural network models: a generator and a discriminator. The generator's role is to create data that is identical from real-world data, while the discriminator's job is to differentiate between the generator's output and the actual data. During training, the generator learns to produce realistic data from noise by continuously trying to fool the discriminator. As a consequence the discriminator improves its ability to detect fakes. This adversarial process continues until the discriminator can no longer tell the difference between real and generated data, at which point the generator is considered to have learned enough to create similar data (Goodfellow et al., 2014).

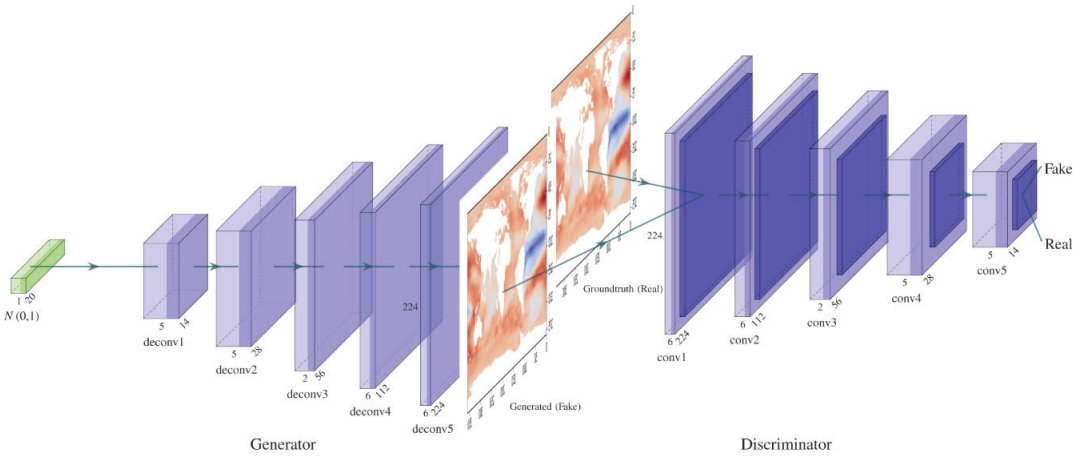

Figure 7: An example architecture of a generative adversarial model (Mateo-García et al., 2021)

However, key limitations remain in their application to fictitious map generation, particularly regarding training data requirements and spatial coherence.
Most GAN-based approaches rely on paired training datasets or domain adaptation techniques, making them less suitable for generating entirely new landscapes without real-world references (Goodfellow et al., 2014).
Another issue is spatial realism. GAN-generated maps often exhibit artifacts, disconnected features, and unrealistic land cover transitions, limiting their practical usability (Hughes et al., 2018). While *TileGAN* and similar approaches improve texture synthesis, maintaining coherence at multiple scales is still an open problem (Frühstück et al., 2019). Moreover, GAN training is computationally expensive, requiring substantial resources for high-resolution geospatial data (Zhang et al., 2023).
Beyond technical limitations, ethical concerns related to AI-generated maps, including data integrity, misinformation risks, and biases, require further attention (Zhao et al., 2021).

- ImageNet

ImageNet, a large-scale hierarchical image database, has played a major role in advancing image classification, object recognition, and deep learning applications (Deng et al., 2009). Built on WordNet's structured ontology, ImageNet provides millions of annotated images, which have supported the development of high-performance convolutional neural networks (CNNs) and generative models such as BigGAN (Brock et al., 2019) and StyleGAN. However, its relevance to geospatial AI and synthetic map generation remains limited due to fundamental differences in data structure, spatial relationships, and contextual dependencies.
One major limitation is that ImageNet primarily contains object-centered images, organized based on semantic categories, whereas datasets consist of continuous spatial features such as terrain, land cover, and road networks (Deng et al., 2009).

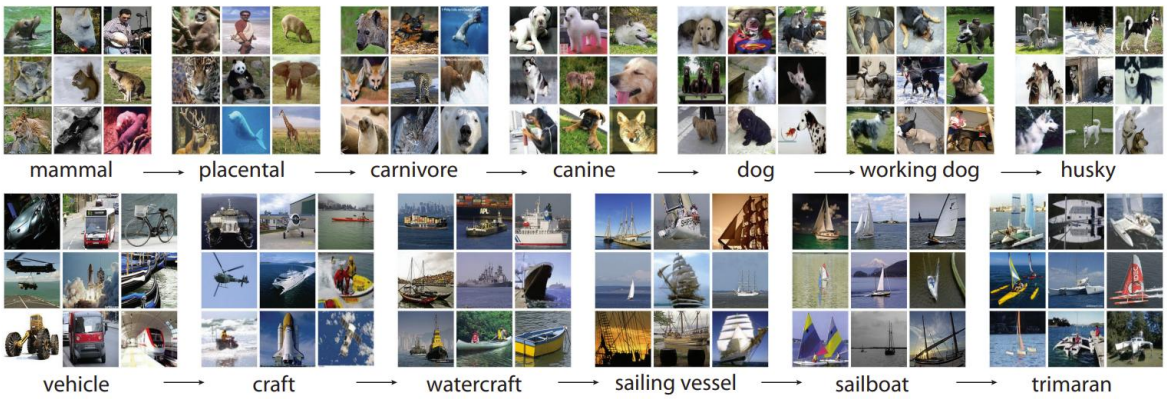

Figure 8: ImageNet's hierarchical object classification not suitable for continuous geospatial data, without major object (Deng et al., 2009)

Unlike traditional object classification tasks, where individual entities are distinct, geographic datasets require maintaining spatial coherence, making the direct application of ImageNet-trained models challenging. Although ImageNet's scale and diversity make it a valuable benchmark, its images do not reflect spatial hierarchies or geographic relationships, which are essential for generating synthetic landscapes.
Another challenge is the imbalance in ImageNet's class distribution. The dataset is densely populated with certain object categories (e.g., animals, vehicles) while lacking representation in environmental and geospatial domains. While domain adaptation and transfer learning techniques have been used in remote sensing, adapting ImageNet-based models for synthetic geospatial applications remains an open problem.
Given these constraints, this research explores diffusion models as an alternative, offering improved control over spatial structures, better consistency in generated imagery, and greater adaptability for synthetic geospatial applications.

- Transformer-based architecture

Transformer-based architectures have revolutionized deep learning by introducing self-attention mechanisms that effectively model long-range dependencies in data. Unlike convolutional neural networks (CNNs), which process information locally, transformers can capture both local and global relationships in a single step, making them particularly powerful for generative modeling (Vaswani et al., 2023). The transformer architecture replaces recurrence with a self-attention mechanism, allowing for parallelization and improved efficiency in sequence-based tasks. This breakthrough has led to significant advancements in natural language processing (NLP), with models such as BERT and GPT utilizing transformers to achieve state-of-the-art results (Vaswani et al., 2023). More recently, transformers have been adapted for vision tasks, including image generation, through models like Vision Transformers (ViTs) and text-to-image models such as DALL·E.

Figure 9: Example of DALL· E with a prompt "Create a map for the country which would have similar climate like Somalia"

These approaches tokenize images into sequences and apply self-attention to generate coherent and structured outputs. However, despite their success in text and vision applications, transformers pose challenges when applied to high-resolution satellite imagery, particularly due to their quadratic complexity in processing large images.

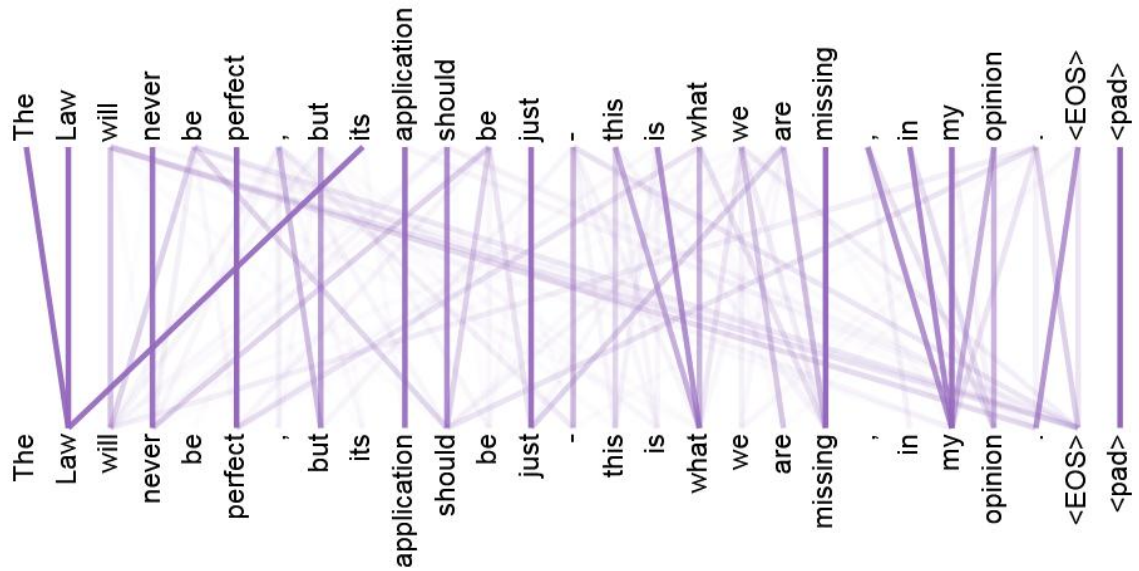

Figure 10: Transformer model constructs meaningful sentences using self-attention (Vaswani et al., 2023)

- Variational Autoencoders (VAEs)

Variational Autoencoders (VAEs) are widely used in generative modeling for their ability to learn continuous latent spaces, enabling smooth interpolations and structured data generation (Kingma & Welling, 2022). Unlike Generative Adversarial Networks (GANs), VAEs use probabilistic inference, improving sample diversity and representation learning (Rezende et al., 2014). However, their application in geospatial AI and synthetic map generation remains limited.

A key limitation is blurry outputs, caused by the variational reconstruction objective, which smooths features and reduces spatial sharpness, making it difficult to generate topologically consistent geographic structures (Yan et al., 2016).

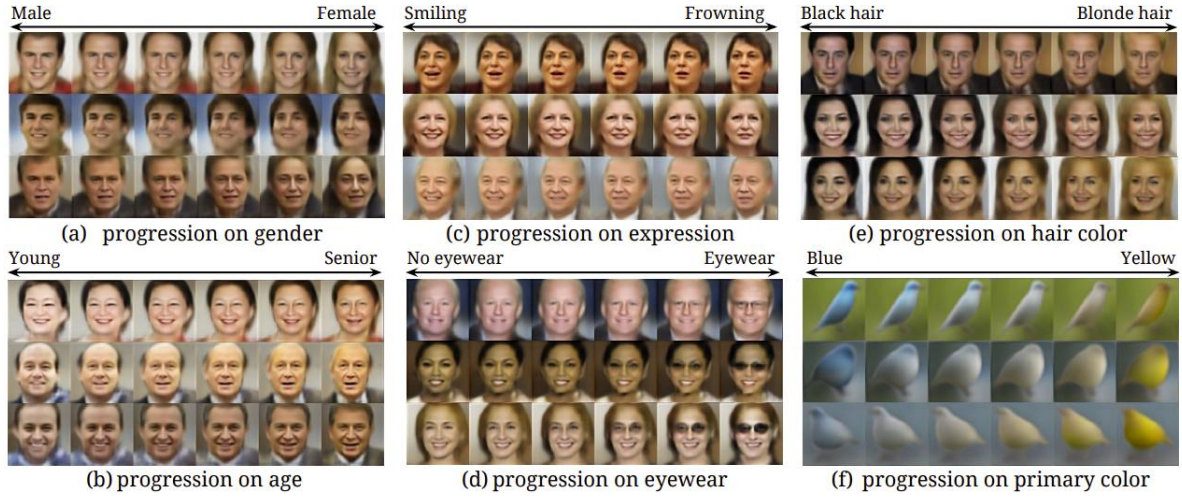

Figure 11: Example of blurry outputs from VAE (Yan et al., 2016).

Additionally, VAEs struggle with mode collapse and sample diversity, particularly in capturing complex geospatial patterns. While β-VAEs and Conditional VAEs improve latent space organization (Burgess et al., 2018), they have not yet achieved realistic spatial generation.

- Other examples based on diffusion models

Beyond image generation, diffusion models have been applied in video generation, where they capture temporal dynamics to produce coherent video sequences. In the field of natural language processing, these models have facilitated text generation tasks, improving the quality and diversity of generated content. Moreover, in the field of molecular design, diffusion models have been utilized to generate novel molecular structures with desired properties. (Yang et al., 2024)

Figure 12: Example of video generation with the text prompt (Yang et al., 2024)

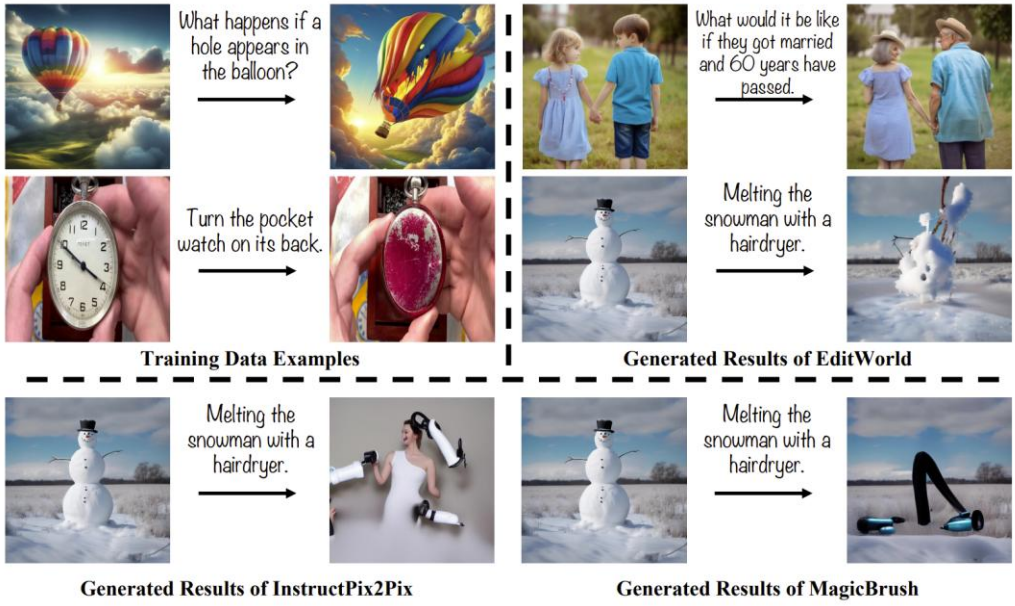

Figure 13: Comparing results of text-to-image outcome with several models based on diffusion models (Yang et al., 2024)

A comprehensive survey by Yang et al. (2024) categorizes the advancements in diffusion models into three key areas: efficient sampling methods, improved likelihood estimation techniques, and adaptations for data with special structures. The survey also highlights the potential of combining diffusion models with other generative frameworks to achieve improved results across other applications.

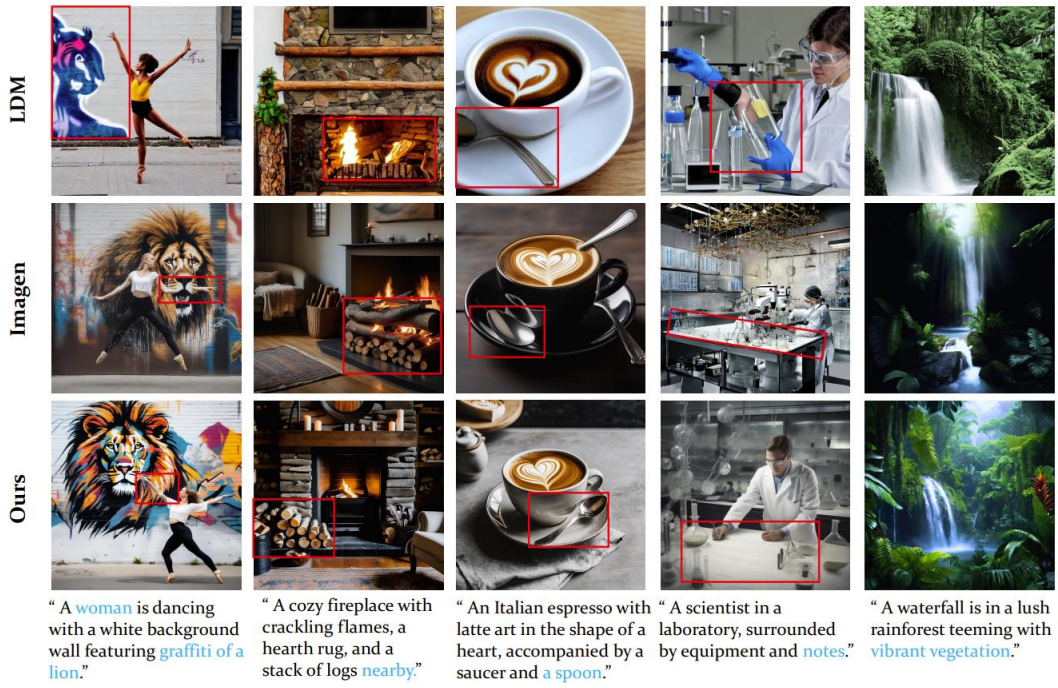

Figure 14: Incoherent samples generated by various generative frameworks (Yang et al., 2024)

2.3.2. Other approach for generating satellite imagery

Beyond generative AI, several traditional methods have been used in GIS and remote sensing to generate or improve satellite imagery. Image fusion techniques combine images from different sensors to create a single, more informative image, improving both spatial and spectral resolution. This approach is particularly useful for applications such as land cover classification and environmental monitoring. A comprehensive review and meta-analysis of image fusion methods in remote sensing is provided by Albanwan (Albanwan et al., 2024).

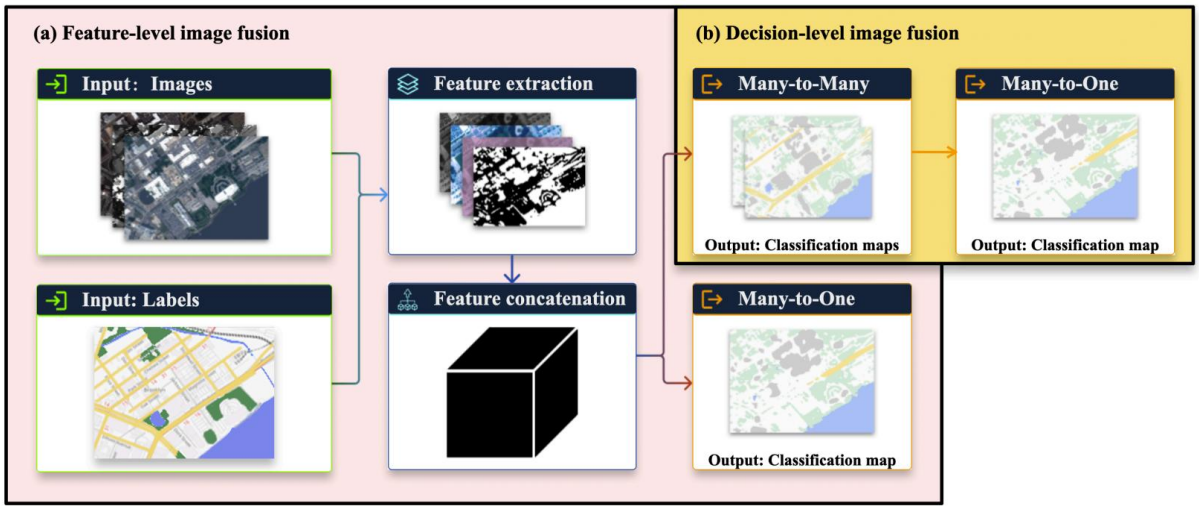

Figure 15: The combination of different sources of images (Albanwan et al., 2024)

3. Study Area

3.1. Training Data from Sentinel-2

Sentinel-2[2] is a project from the European Space Agency, focused on collecting high-resolution, multi-spectral images. The mission contains two twin satellites flying in the same sun-synchronous orbit, spaced 180 degrees apart from each other, allowing to revisit the same area with the frequency of 5 days. Sentinel-2 satellites carry a MultiSpectral Instrument (MSI) capable of capturing 13 distinct spectral bands. These vary from the visible (such as blue, green, and red) to the near-infrared and short-wave infrared parts of the spectrum. The spatial resolution differs depending on the bands. We can distinguish three groups: high-resolution, medium-resolution and low-resolution.

The highest spatial resolution is 10 meters, available in four bands covering visible light and near-infrared (NIR) wavelengths: blue (B2), green (B3), red (B4) and NIR (B8). The medium spatial resolution is 20 meters and it contains several red-edge bands (B5, B6, B7, B8A) and short-wave infrared bands (B11, B12). The lowest spatial resolution is 60 meters, primarily used for atmospheric correction, including aerosol band (B1), the water vapor band (B9) and the cirrus band (B10).

Sentinel Hub is a satellite imagery API service, allowing to download multi-spectral and multi-temporal Sentinel-2 data. From the Sentinel-2 data there is a selection of Sentinel-2 L1C and Sentinel-2 L2A. The Sentinel-2 L1C source is measured on top of the atmosphere (TOA) but the Sentinel-2 L2A[3] type of the measurement is Bottom of the atmosphere(BOA). The TOA product is corrected for radiometric and geometric distortions, while the BOA product includes these corrections and additionally removes atmospheric influences, such as haze and clouds, to provide clear surface reflectance data.

[2]https://sentiwiki.copernicus.eu/web/s2-products
[3] https://docs.sentinel-hub.com/api/latest/data/sentinel-2-l2a/#data-type-identifier-sentinel-2-l2a

3.2. Carana

The fictional country of Carana[4], situated on the island of Kisiwa off the eastern coast of Africa, has been developed as a comprehensive scenario for training and strategic exercises by international peacekeeping organizations, including the United Nations. Designed to simulate real-world geopolitical crises, Carana provides a realistic learning environment for senior leaders, military commanders, and peacekeeping personnel, preparing them for complex operational challenges. The scenario integrates detailed geographic, political, and socio-economic elements, making it an essential tool for modelling humanitarian crises, conflict resolution, and post-disaster response.

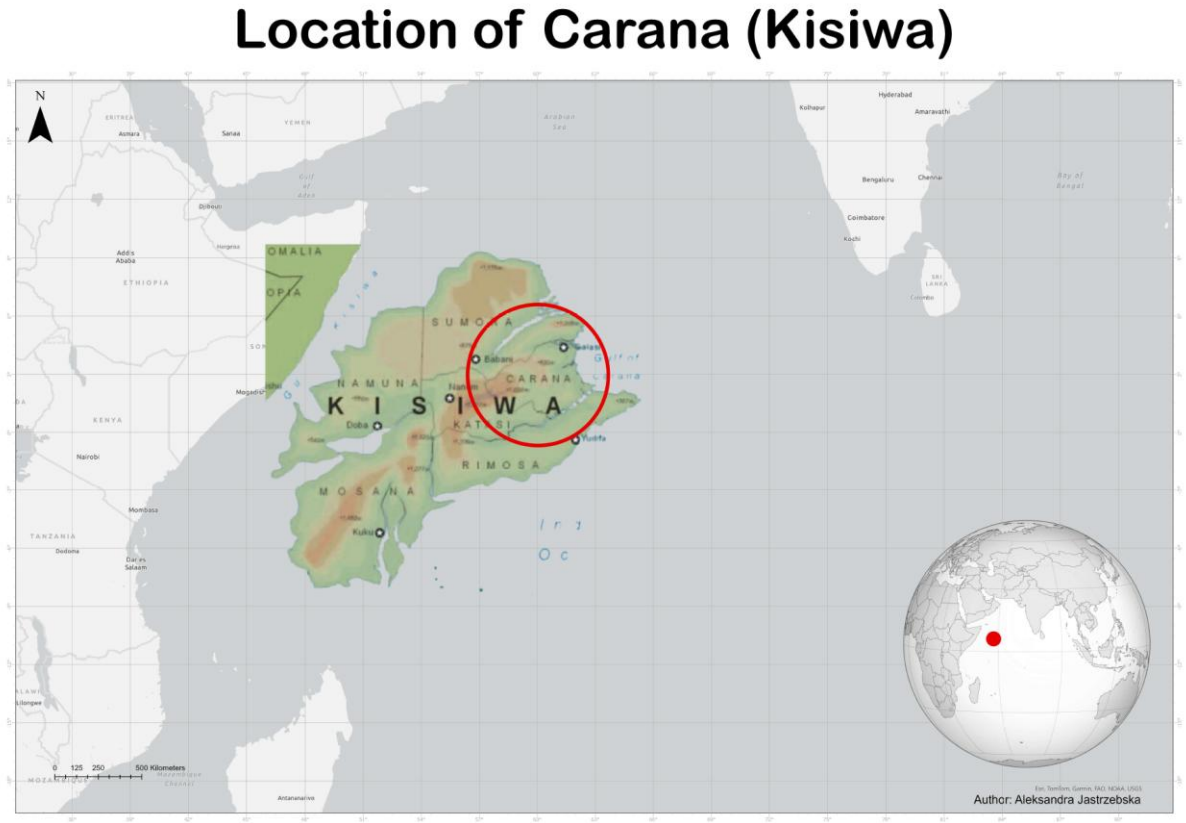

Figure 16: Location of Carana

Carana serves as a platform for hypothetical crisis simulations, covering a wide range of global security concerns. These include rising tensions over newly discovered mineral resources, civil wars, regional instability due to neighboring crises, and natural disasters such as floods and earthquakes. The training focuses on strategic decision-making, such as determining safe locations for refugee camps, assessing humanitarian aid distribution, and monitoring the movements of armed groups to protect civilians. Peacekeeping operations rely on dynamic mapping to analyze what

[4] https://paxsims.wordpress.com/wp-content/uploads/2009/08/carana_long_version_english.pdf

happened in a specific region, for example the day before, and how to respond effectively, which diffusion models may address in the future.

The training exercises incorporate both raster and vector data, depending on the operational needs. Base maps with lower resolution are generally used for big-picture analysis, such as assessing vegetation cover, flood extents, and overall terrain. However, high-resolution imagery is crucial for detailed tactical planning, including monitoring troop movements, assessing infrastructure damage, and ensuring the safety of humanitarian corridors. The ability to switch between different levels of imagery improves decision-making, particularly in high-stakes peacekeeping missions where real-time intelligence is important.

Carana's framework aligns with United Nations Sustainable Development Goals (SDGs)[5], particularly Goal 16 (Promote peaceful and inclusive societies, provide access to justice for all, and build effective, accountable, and inclusive institutions) and Goal 17 (Strengthen global partnerships for sustainable development). By providing a structured environment for multinational collaboration, Carana promotes joint training initiatives, helping peacekeepers, military strategists, and humanitarian workers develop the skills needed for conflict prevention, crisis management, and post-war reconstruction.

A key limitation of Carana's current utility is the quality and flexibility of its visual representations. Existing maps used in training often lack the necessary resolution, detail, and adaptability to fully support dynamic scenarios. Enhancing high-resolution, realistic map generation for Carana could significantly improve its effectiveness in peacekeeping simulations, disaster response planning, and urban development strategies. By integrating advanced AI-driven mapping techniques, including diffusion models, these training exercises could better reflect real-world terrain complexities, improving operational readiness for future peacekeeping missions.

[5] https://sdgs.un.org/goals

4. Methodology:

To better illustrate the methodology applied in this research, the workflow has been structured into stages, as depicted in the chart below. This diagram provides an overview of the taken steps, from acquiring and processing satellite imagery to training the diffusion model and validating the generated outputs.

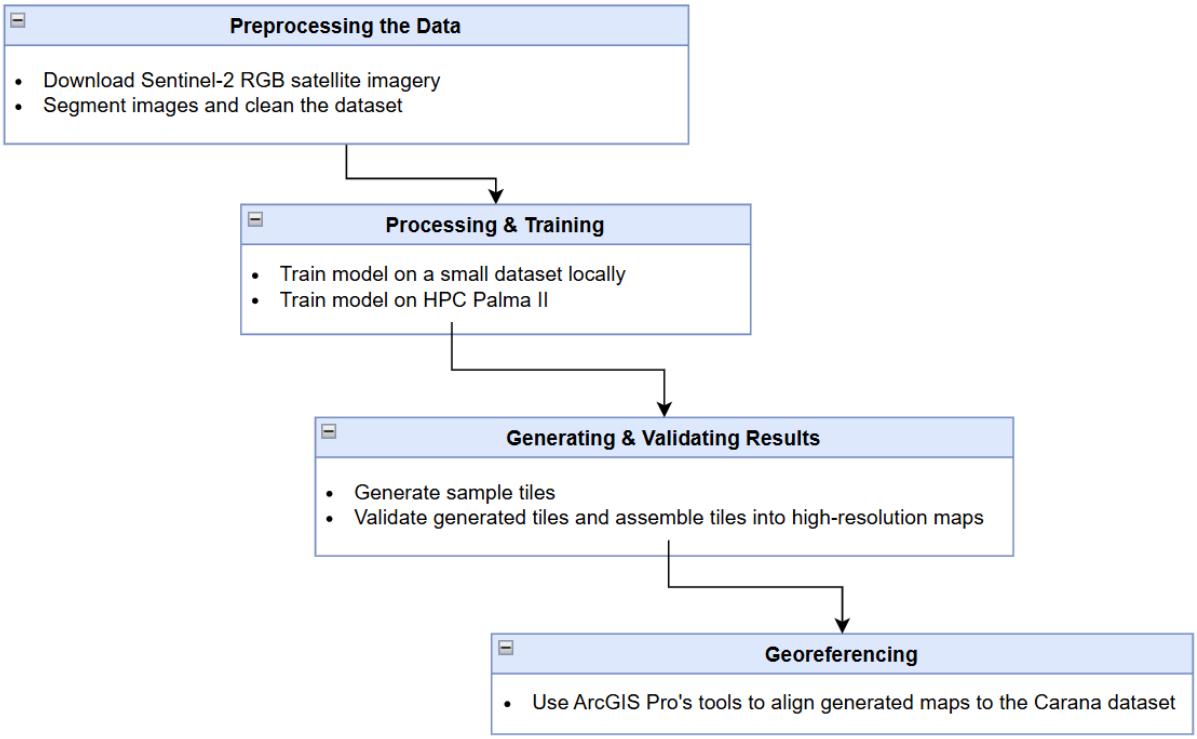

Figure 17: General workflow

4.1. Preprocessing the Data

4.1.1. Download Sentinel-2 RGB satellite imagery

Based on the project requirements, raster imagery was selected for the training dataset. The chosen images originate from regions in Ethiopia, Kenya, and Somalia, as their climatic conditions closely resemble those of Carana, ensuring that the training data contains similar environmental characteristics.

To acquire Sentinel-2 data, the SentinelHub platform was utilized, and a script was developed based on the Copernicus Data Space Ecosystem documentation to retrieve RGB imagery[6] for the specified area. The API request requires the following parameters: clientID, clientSecret, boundingbox, timeRange, maxCloudCoverage, collection/type, size and format.

[6]https://documentation.dataspace.copernicus.eu/APIs/SentinelHub/Process/Examples/S2L2A.html#true-color

The clientID and clientSecret are unique authentication tokens generated from an individual Sentinel Hub account. For the selected bounding box, coordinates were defined from 1°47'27.0"N to 5°38'08.7"N in latitude and 35°22'11.7"E to 45°36'19.4"E in longitude, covering regions in Ethiopia, Kenya, and Somalia.

For the timeRange, the entire year 2023 was selected to ensure a comprehensive dataset, while the maxCloudCoverage parameter was set to 5% to prioritize cloud-free terrain images, which served as a limiting factor for the available data. The dataset collection specified is Sentinel-2 L2A, providing atmospherically corrected surface reflectance data. The imagery was downloaded at a resolution of 1024 × 1024 pixels in TIFF format, ensuring easy geolocation as the TIFF metadata includes spatial coordinate information.

4.1.2. Segment images and clean the dataset

The algorithm requires training data in 64x64 pixel tiles, so a function was developed to systematically cut larger Sentinel-2 images into these smaller parts using Pillow.[7] Sentinel-2 imagery, however, often contains black or null pixels due to sensor gaps and detector misalignments, which needed to be addressed during preprocessing.

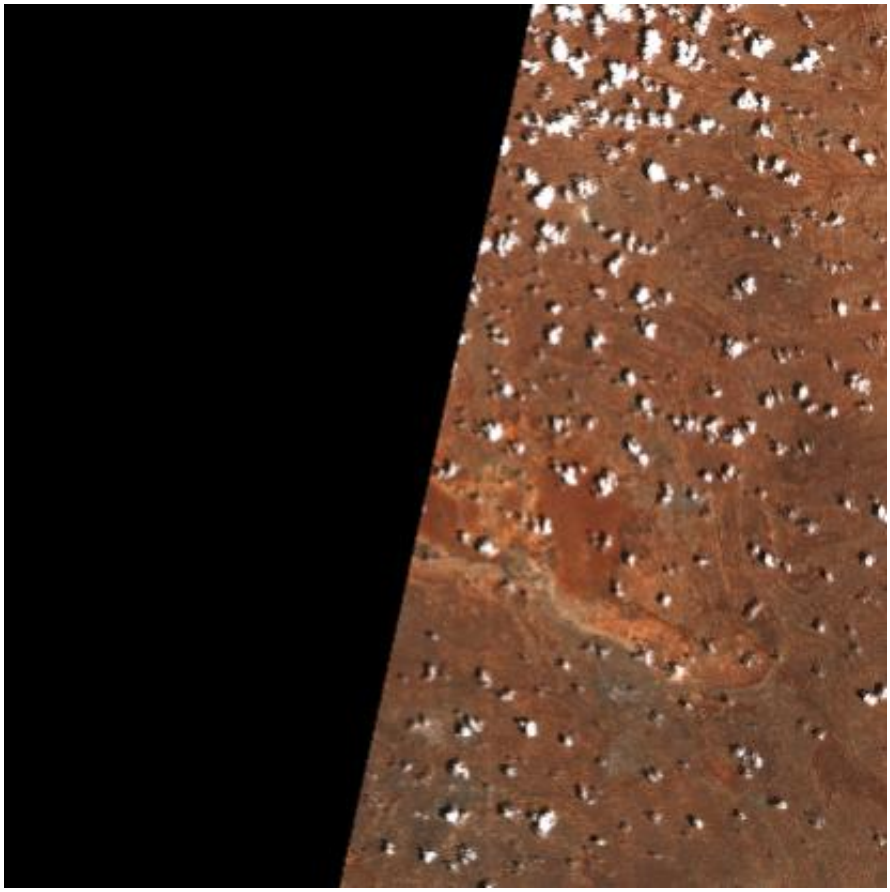

Figure 18: Example of downloaded Sentinel-2 image

[7] https://github.com/python-pillow/Pillow/?tab=readme-ov-file

Additionally, since Sentinel-2 captures images in 290 km-wide swaths, slight gaps or overlaps can occur where one strip ends and another begins. To minimize cloud interference, the maxCloudCoverage parameter was set to 5%, but clouds can still appear as white pixels in some tiles.

To address these issues, an additional function was implemented, incorporating a validation step to filter out tiles containing more than 0.05% black pixels or 0.05% white pixels. This filtering process resulted in a final dataset of 134,292 tiles, ensuring that only tiles with meaningful data were included, thereby contributing to a more accurate and reliable model.

Figure 19: Example of cut tile

4.2. Processing & Training

4.2.1. Train model on a small dataset locally

After the dataset was prepared and cleaned, the training process was initiated. Initially, training was conducted on a local machine using a small subset of 1000 tiles to test the model's functionality. During this phase, challenges arose in configuring the training to run on the GPU, which required troubleshooting and optimization. Once these issues were resolved and the model's requirements became clearer, it was evident that training on the full dataset would exceed the computational capabilities of the local machine. Consequently, the training was transitioned to a more powerful system to accommodate the larger dataset efficiently.

At the core of the implementation of this model provided by Leon Pielage (Pielage, n.d.), is a diffusion model based on the U-Net architecture, which is particularly suitable for image-to-image tasks such as generation and segmentation. In this case, U-Net is designed to handle input tiles of 64x64 pixels, which are progressively downsampled to a compressed size of 16x16 pixels in the bottleneck layer, with 64 feature channels at this stage. The choice of tile size balances efficiency and local detail preservation, enabling the model to focus on learning features in satellite imagery.

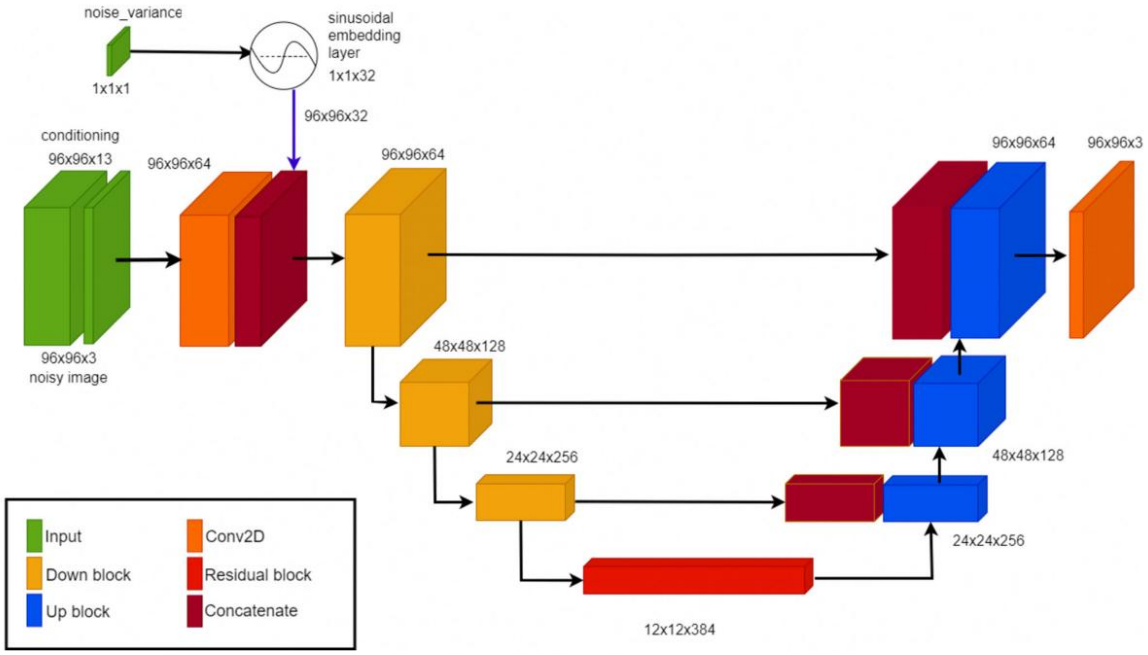

Figure 20: Visualization of the U-Net architecture (Asperti et al., 2024)

The U-Net follows a symmetric encoder-decoder structure. In the encoder (contracting path), the spatial resolution of the input is reduced through a series of downsampling operations, while the number of feature channels increases according to the specified scaling factors. In this case, the scaling factors ("dim_mults") are (1, 2), meaning the channels double at each downsampling step, starting from an initial dimension of 32. This results in feature maps of increasing abstraction as the model moves deeper into the network.

At the bottleneck, the model operates on the highly compressed 16x16 representation, which captures the most essential features of the input tile. The decoder (expanding path) then reconstructs the image back to its original 64x64 resolution by progressively upsampling the feature maps. This step is necessary to preserve important features to be able to recover them during the decoding process.

This U-Net implementation is heavily inspired by open-source materials from OpenAI[8], including the guided diffusion and CLIP projects. Modifications were made to tailor the model to satellite imagery, including specific configurations like the scaling factors, attention resolutions, and tile preprocessing. This design ensures the model's compatibility with the characteristics of the Sentinel-2 dataset.

To ensure that training goes well, there is a need for a framework, which manages all the processes. In this case configurations are managed through the *DiffusionConfigs* class, which specifies the parameters of the U-Net model, such as tile size, channel dimensions, detailed to the specific needs of satellite imagery.

The process of training the diffusion model relies on denoising diffusion probabilistic models (DDPM), which learn to generate data by progressively reversing a noise process. During training, the model learns to predict and remove added Gaussian noise at each step, gradually reconstructing meaningful data from random noise. This iterative denoising process enables the model to effectively learn the structure and patterns in the satellite imagery dataset.

The training process optimizes the diffusion loss, which measures how accurately the model predicts the noise added at each timestep. The training framework ensures that progress is tracked through checkpoints, samples and logs. Checkpoints store important information, including the model's learned weights, configurations (e.g., device type, batch size, learning rate, and tile size), and training progress.

During each epoch, the model generates sample outputs made up of 16 tiles. In the early epochs, these samples appear highly noisy, reflecting the model's initial state. As training progresses, the samples gradually improve, with the tiles becoming increasingly organized and resembling the patterns in the training dataset.

[8] https://github.com/openai/guided-diffusion/blob/main/guided_diffusion/unet.py

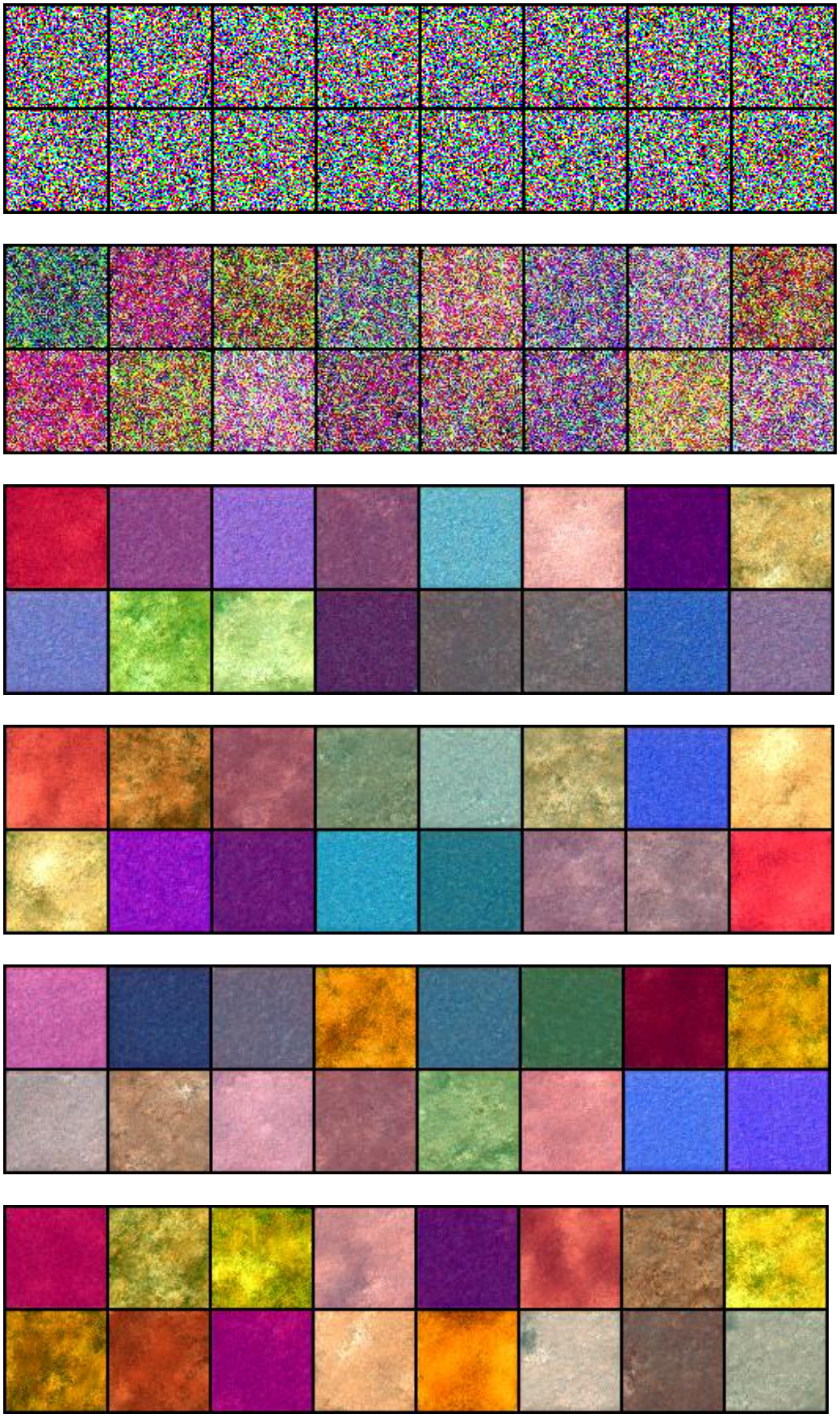

Figure 21: Examples of generated tiles at different training stages, starting from 0 epochs (first image), followed by 150 epochs, 300 epochs, 450 epochs, 600 epochs and 750 epochs.

4.2.2. Train model on HPC Palma II

The full-scale training was conducted on PALMA-II, a high-performance computing system at the University of Münster. Manufactured by MEGWARE, PALMA-II boasts 16,272 cores, 77,568 GB of memory, and 444 nodes, all powered by Intel Xeon Gold 6140 18C @ 2.30GHz (Skylake) processors. It features a GPFS storage capacity of 2.4 PB and uses a 100 Gbit/s Intel Omni-Path network for high-speed interconnectivity. Operating on Rocky Linux 9, the system achieves a LINPACK performance of Rmax: 800 TFlop/s and Rpeak: 1,277 TFlop/s.

To initiate the training process, it was essential to set a suitable computational environment. Since the initial development and testing were conducted on a local machine, the necessary libraries and dependencies were already configured. To replicate this setup on PALMA-II, the Conda environment was exported to a YAML configuration file, which enabled creation of an identical environment on the HPC system.

```
C:\Users\aljas>ssh ajastrze@palma-login.uni-muenster.de
*************************************************************************
*               Welcome to the Zen4 login-node of the                   *
*             Palma II HPC cluster running Rocky Linux 9                *
*                                                                       *
*      DO NOT START COMPUTATIONS ON THIS NODE, USE THE BATCHSYSTEM!     *
*                                                                       *
* Please read our documentation:                                        *
* https://confluence.uni-muenster.de/display/HPC                        *
*                                                                       *
*                                                                       *
*************************************************************************
*                                                                       *
*  DEFAULT FILESYSTEM QUOTAS/LIMITS:                                    *
*     - /home: 25GB (200k files)                                        *
*     - /scratch: 5TB (1M files)                                        *
*                                                                       *
*  To check your usage:                                                 *
*     mmlsquota --block-size auto home                                  *
*     mmlsquota --block-size auto scratch                               *
*                                                                       *
*************************************************************************
Last login: Fri Dec 20 16:07:42 2024 from vpn-pool1-pnt-31769.uni-muenster.de
(base) [ajastrze@r07m01 ~]$
```

Figure 22: Login-node to the Palma II HPC

Working within my allocated personal storage partition on PALMA-II, a dedicated folder was created to store all the necessary scripts, libraries and training data. After ensuring all files were successfully transferred, the training was initiated using a bash script. This approach ensured that the process ran independently of my active connection to the server, utilizing PALMA-II's resources efficiently and avoiding disruptions caused by potential network disconnections.

After successful training, the model produces checkpoint files, which as mentioned before store essential information needed to resume training, perform inference, or analyze model performance at different stages. These checkpoints contain the trained model weights, which include the optimized parameters (weights and biases) learned during training. Additionally, they store the optimizer state, preserving momentum estimates and learning rate values, allowing training to continue from the last saved state without losing progress.

Each checkpoint also includes training metadata, such as the number of epochs completed, training loss, and validation loss, providing a snapshot of the model's performance at the time of saving. While the model's architecture is defined separately in the script, checkpoint formats can also store this configuration.
After completing the training on the full dataset for 1500 epochs, the resulting files were generated. Later, these files were downloaded to the local machine to pursue the creation of individual tiles and the assembly of the final map.

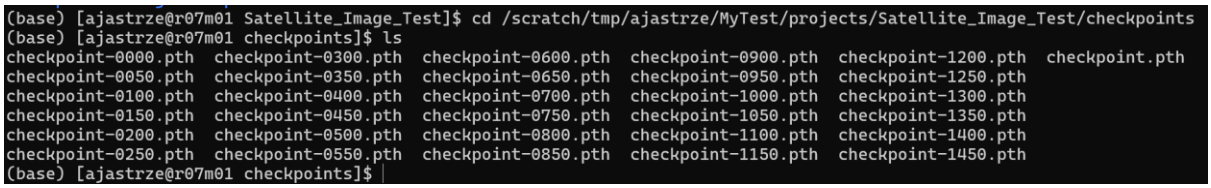

Figure 23: Screenshot of the directory containing the checkpoints

4.3. Validation

With the final checkpoint stored locally, the process of generation of individual tiles within the grid structure started, utilizing the data from the most recent checkpoint. For this process, the *sample_batch()* method from the *DiffusionTrainer* class was employed.

4.3.1. Histogram-Based Validation

To ensure the validity of the generated tiles, a script was developed that classifies tiles as either valid or invalid based on their histogram values. This approach allows for an automated filtering mechanism that evaluates each tile's color distribution and statistical consistency. The validation process involves computing key statistical metrics, specifically mean and standard deviation, for each color channel (Red, Green, and Blue). The computed values for a sample set of generated tiles are as follows:

- Red Channel: Mean = 111.63, Standard Deviation = 42.39
- Green Channel: Mean = 73.21, Standard Deviation = 26.57
- Blue Channel: Mean = 49.54, Standard Deviation = 19.36

Using these computed values, validity ranges were defined for each channel to establish acceptable thresholds for generated tiles:

- Red Mean Range: (35, 220)
- Green Mean Range: (20, 190)
- Blue Mean Range: (5, 80)

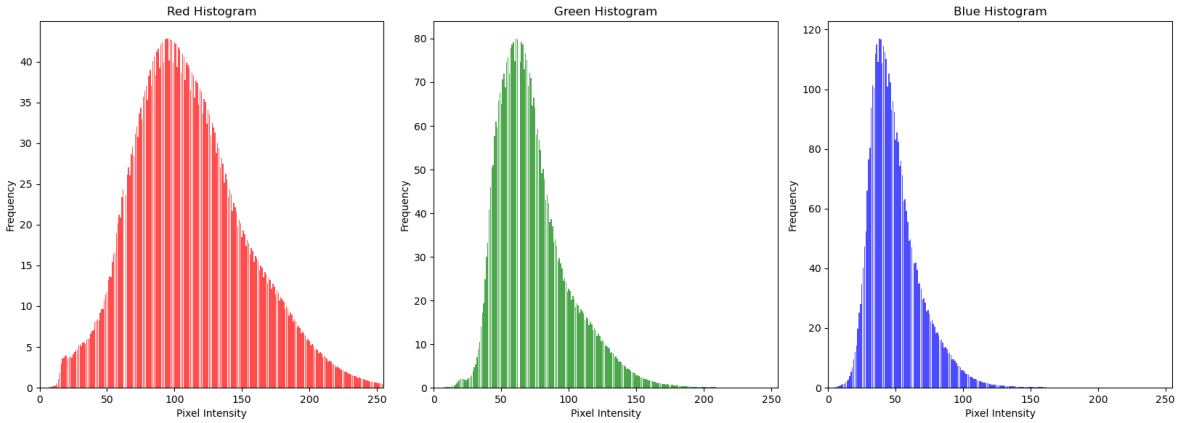

Figure 24: Plot of red, green and blue histograms calculated from the training dataset

Any tile with a mean value outside these predefined ranges is considered invalid and filtered out during the validation process. The effectiveness of this method is illustrated in the plots below, where the histogram distributions of valid and invalid tiles are displayed.

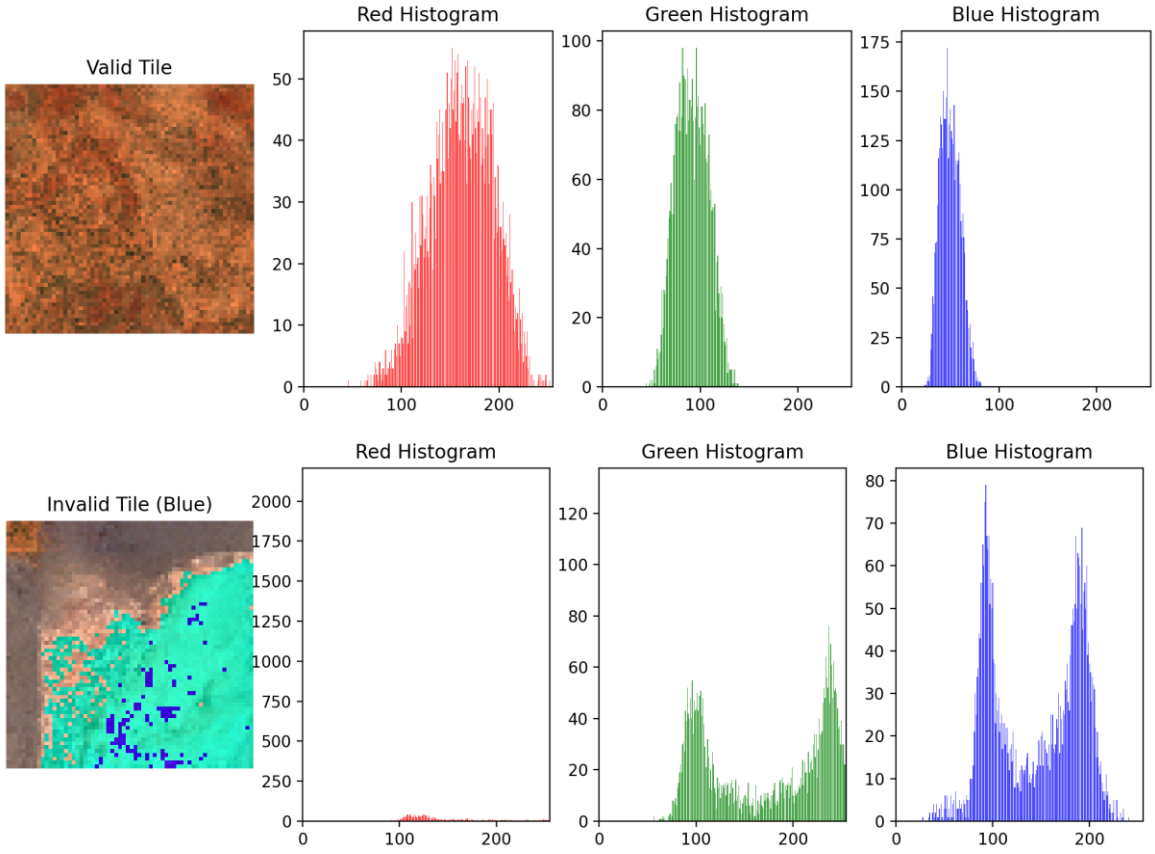

Figure 25: Example of validation classification based on histograms

4.3.2. Fréchet Inception Distance (FID)

To effectively validate the generated tiles, Fréchet Inception Distance (FID) was used, as it provides a more reliable measure of similarity between generated and real images compared to the Inception Score, despite certain limitations (Heusel et al., 2017). The FID computation was performed using the *pytorch-fid* package (version 0.3.0), but several challenges arose during the process. Since FID requires input images to be in PNG format, the generated tiles, originally in TIFF format, had to be converted before evaluation. Additionally, FID operates on 299×299 pixel images, whereas both the generated and training datasets consisted of 64×64 pixel tiles.
To meet this requirement, all images had to be rescaled before proceeding with the validation. Once the dataset was properly formatted, the FID score was calculated for the entire dataset. As shown in the figure below, the resulting score indicates that further improvements are needed to enhance the quality of the generated tiles.

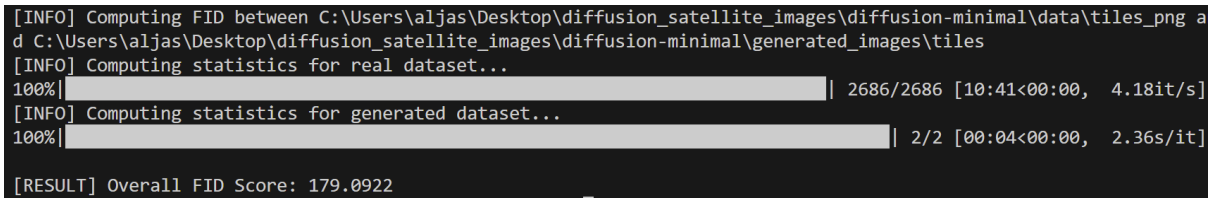

Figure 26: Results of computing FID

4.4. Georeferencing

To integrate the generated map into the provided geodatabase, the image was processed using ArcGIS Pro. Upon adding the PNG file to the project, the software flagged an *Unknown Coordinate System* warning, indicating the need for georeferencing to align the image with the reference vector dataset.
To ensure spatial consistency, the *Georeferencing* function was applied, adjusting the image's position relative to the reference data. This process was followed by the *Fit to Display* function to achieve a preliminary alignment.

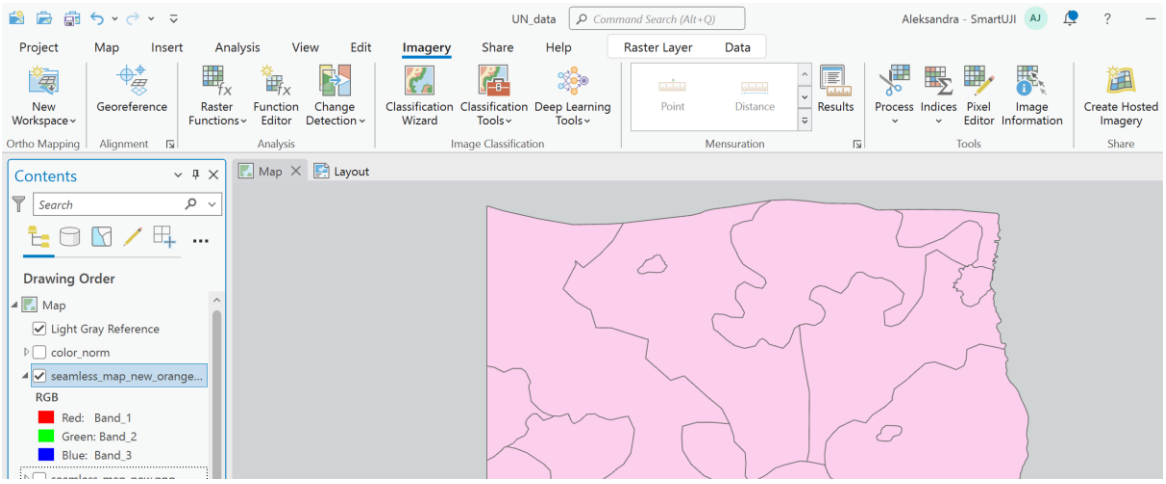

Figure 27: Usage of Georeference tool

Figure 28: Georeferenced map

To ensure full compatibility with the vector dataset, the *Clip Raster* tool was used to refine the output, resulting in a raster that seamlessly fits the designated area.

As improvements to the quality of the generated data continue, the processed tiles will progressively cover the entire vector dataset, enhancing the completeness and usability of the final map.

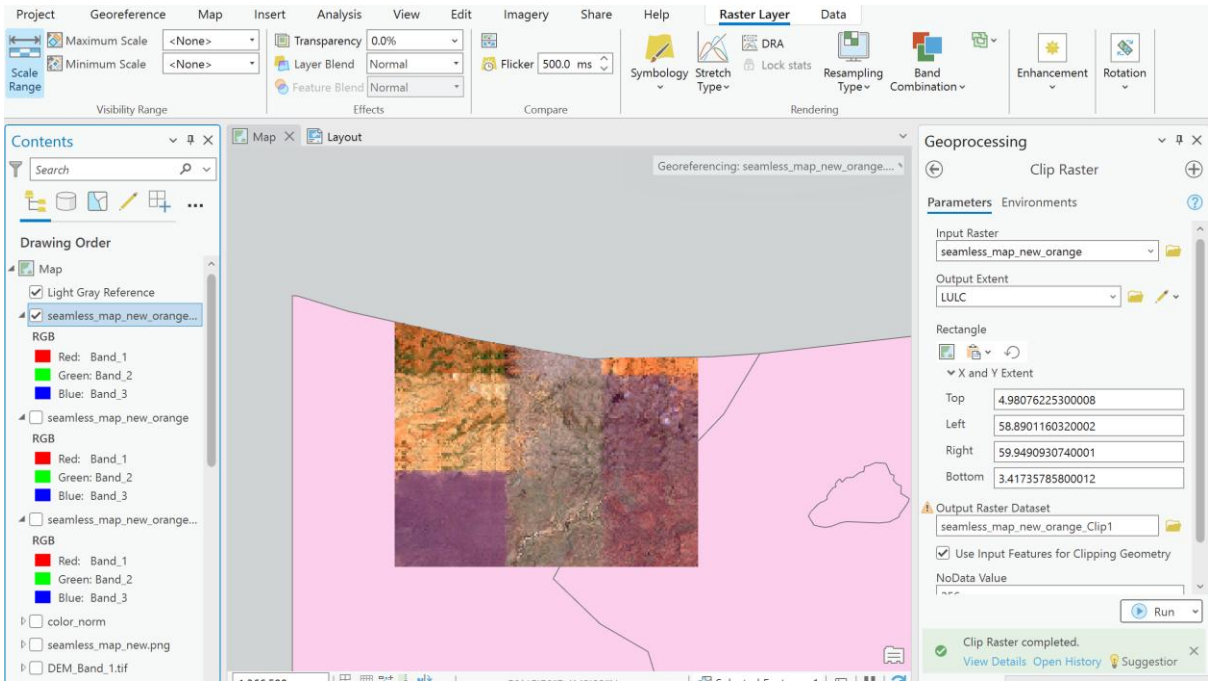

Figure 29: Clipped map

5. Results

The diffusion model-based framework developed in this study successfully generated synthetic 64×64 pixel tiles, capturing key spectral characteristics of Sentinel-2 satellite imagery. The model was trained to produce realistic map tiles, which could then be assembled into a continuous geospatial representation of the fictional country of Carana. The methodology followed a structured data science pipeline, integrating multiple validation steps to refine output quality.
The generated tiles were evaluated for their spatial coherence, color consistency, and structural integrity, ensuring they aligned with expected geospatial patterns. Several refinements, including histogram-based validation and Fréchet Inception Distance (FID) evaluation, were implemented to assess and improve the quality of the generated tiles. The results demonstrated that diffusion models could be leveraged to synthesize high-resolution geospatial imagery, offering potential applications in map generation for peacekeeping simulations.

However, while the model successfully generated individual tiles, assembling them into a seamless, high-resolution map introduced several challenges. Issues related to tile coherence, boundary blending, and color consistency required additional refinement. The following section discusses these challenges in detail, along with the techniques applied to improve the final outputs.

5.1. Challenges and Refinements

The biggest challenge in generating a seamless map was ensuring the coherence of the tile grid, both in terms of color consistency and boundary blending. Since the grid size was a parameter in the model, I could easily scale the map dimensions to meet the specific requirements of the project.
While the model successfully generated individual tiles, assembling them into a continuous map introduced several challenges with displayed boundaries of them and color transitions. The significant issue I encountered was the lack of coherence between neighboring tiles when assembling them into a grid structure. The *DiffusionTrainer* class includes a function to sample individual tiles, where the generation of each tile relies on the border information of the previously generated tile. The problem arises because the generated tiles are stored as a tensor during generation, and when assembling them into a grid, the dependency on the previous tile's border may not align properly. For example, a tile generated to match the left and top border of the previous tile might belong to a different row in the grid, causing visible discontinuities, as illustrated in the figure below.

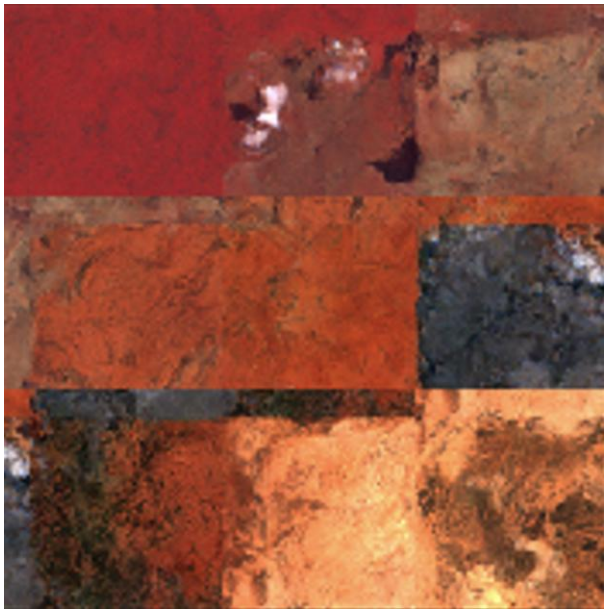

Figure 30: Visualization of grid-based tile generation, where each tile's left and top border influences the next tile

To address this challenge, when generating new tiles, the model incorporated an 8-pixel overlap from the left and top neighboring tiles to create smoother transitions. However, this approach led to repetitive patterns where certain areas appeared duplicated, breaking the overall coherence. To address these issues, I implemented several improvements. First, I applied histogram-based validation, as described in the previous chapter, to ensure each tile remained within an expected color range. Second, instead of directly copying the overlap regions, I introduced a weighted blending technique, where the transition between overlapping regions was gradually merged using a linear interpolation factor (alpha).

This approach significantly reduced visible seams and allowed the tiles to blend more naturally while preserving distinct textures. Finally, an additional blending step was performed after placing each tile into the final grid, further minimizing artificial boundaries. As a result, the final generated maps show enhanced color uniformity and smoother transitions.

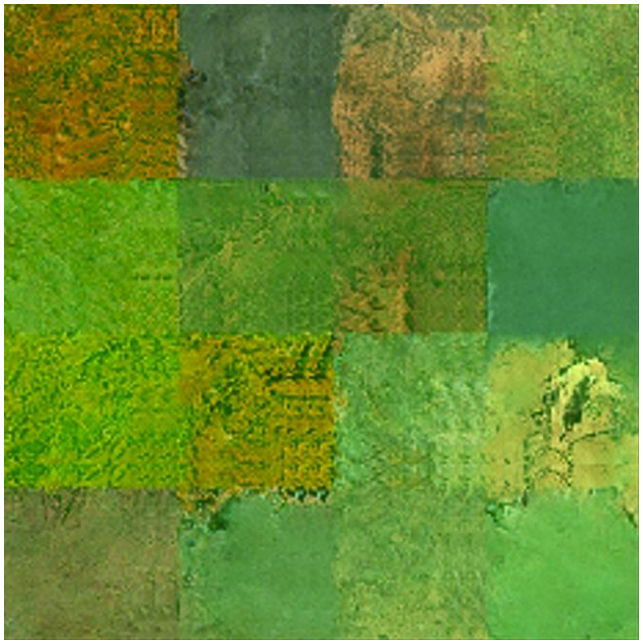

Figure 31: A 4x4 grid with a predominantly green color palette

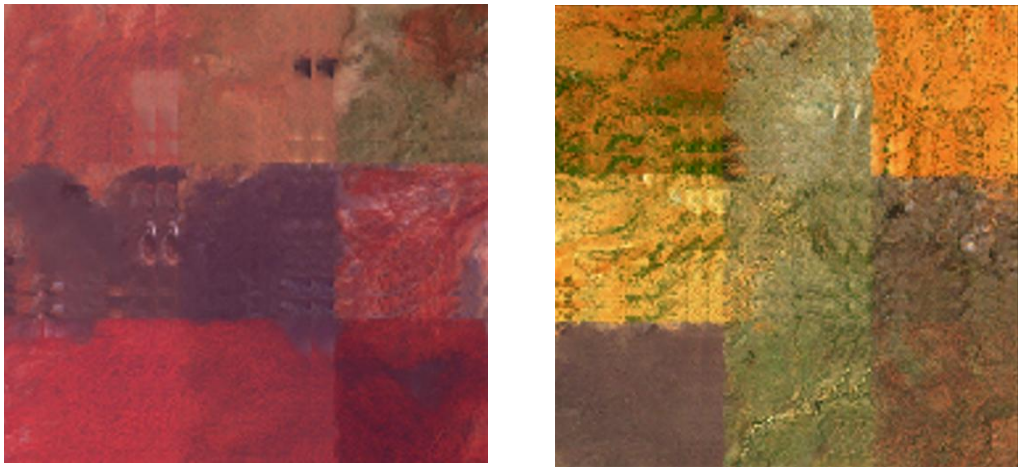

Figure 32: 3x3 grids with a predominantly red and brown colors

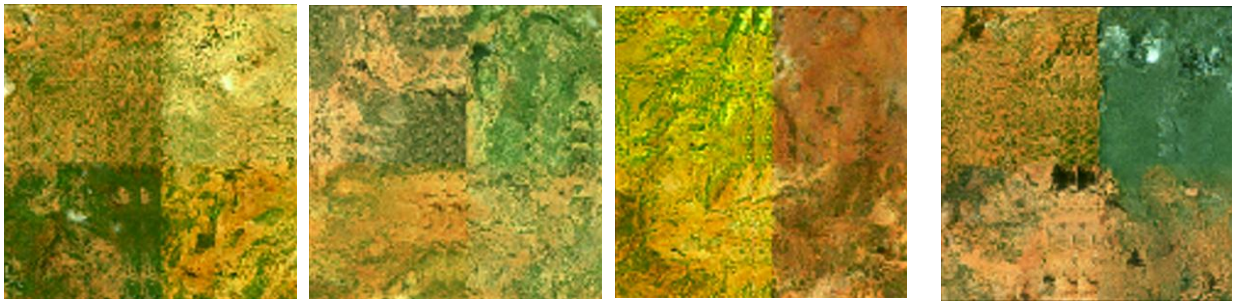

Figure 33: Examples of 2x2 grid

6. Discussion and future work

6.1. Limitations

One of the primary limitations of this study is the lack of labeled geospatial features in the training dataset. Unlike datasets that contain explicit annotations for rivers, roads, vegetation and other land cover types, the training data used in this project does not provide such feature-based segmentation. This absence makes it impossible to implement conditional generation, similar to the approach used by Szymański et al. (Przymus & Szymański, 2023), where OpenStreetMap (OSM) data serves as a structural guide for generating corresponding satellite imagery.
Ideally, a dataset like the one used in "Generate Your Own Scotland" (Espinosa & Crowley, 2023) would allow for feature-aligned conditional image synthesis, where satellite images are generated based on existing OSM features. However, in the case of Carana, the only available geospatial data consists of Digital Elevation Mode ls (DEM) and a limited set of vector features, making this type of conditional generation unrealistic. Another limitation is related to color consistency in generated tiles. While the model successfully produces synthetic tiles, the generated outputs are not constrained to the color distributions of the training dataset, sometimes resulting in unexpected variations, for example yellow or purple samples, unrealistic for a satellite imagery.

What is more, an important limitation was the presence of clouds in the training dataset, which directly influenced the quality of generated outputs. Despite efforts to preprocess and clean the dataset, some Sentinel-2 tiles contained significant cloud cover, and as a result, the model occasionally generated synthetic tiles that included cloud-like patterns. This issue highlights the need for more strict cloud masking techniques during data preprocessing, ensuring that training samples consist only of clear-sky satellite imagery to improve the realism and usability of generated maps.

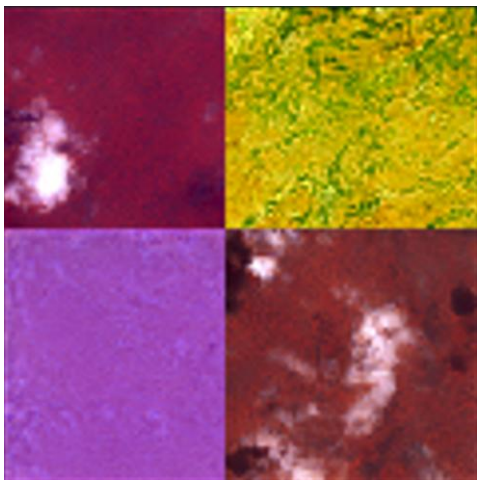

Figure 34: Example of generated tiles with unrealistic colors and cloud-like shapes

To address the challenge with cloud-like shapes, the data cleaning process was refined by gradually lowering the threshold for acceptable black and white pixel values from 50% to 0.05%. This stricter filtering approach ensured that only high-quality tiles were included in the training dataset, reducing the presence of shapes that could negatively impact model performance.

A major challenge in this study was the computational complexity associated with training a diffusion model on a large-scale dataset. Given the limited GPU resources available on a local machine, training was conducted on the HPC Palma II. However, this transition introduced additional complexities, as it required proficiency in high-performance computing (HPC) workflows, including job scheduling, data transfer, and storage management. Optimizing resource allocation, configuring the computing environment, and ensuring efficient data access demanded significant effort but ultimately enabled the successful execution of large-scale model training.

Furthermore, the sampling process, which was performed on a local machine, remained computationally intensive, requiring substantial processing time to generate high-resolution tiles. This highlights the need for more efficient inference strategies to facilitate faster generation without compromising output quality.

Another limitation came from the reliance on open-source Sentinel-2 satellite imagery, which offers a spatial resolution of only 10m per pixel. While Sentinel-2 data is widely used for remote sensing applications, this resolution is not optimal for generating highly detailed, high-resolution maps, particularly for applications requiring fine-grained geographic features.

6.2. Future work

Several improvements can be made to improve the performance of the model. One of them is better validation handling. Currently, when a tile is classified as invalid, it is simply discarded, and the model regenerates a new sample without any corrective learning. A more effective approach would be to integrate negative feedback into the training process, where invalid tiles are fed back into the model as negative samples to improve its ability to distinguish between realistic and unrealistic outputs. This would enable the model to gradually correct its generative capabilities and reduce the likelihood of producing invalid tiles over time.

What is more, adjusting key hyperparameters could significantly improve the initial coherence, color consistency, and structural integrity of the generated tiles. Potential refinements include increasing the latent space dimensionality to enable better feature extraction, utilizing a deeper network architecture with more scale levels for enhanced representation learning, and reducing the learning rate to allow for more gradual and stable convergence.

Since the generated maps are composed of multiple individual tiles, ensuring spatial consistency across tile boundaries remains a critical challenge. By refining model parameters and optimizing architectural choices, the overall coherence of the assembled maps can be improved, resulting in a more seamless integration of tiles and a higher-quality final output.

Additionally, improving FID scores and the visual quality of the generated tiles remains a priority. Although the model successfully produces high-resolution maps, further optimization in color consistency, texture sharpness, and structural accuracy could improve the realism of synthetic outputs. This could be achieved through better training data selection, higher-resolution imagery, or fine-tuning model hyperparameters to better capture geospatial patterns.

Finally, the future work could involve incorporating higher-resolution geospatial datasets than Sentinel-2 data with a 10m spatial resolution, for an improvement of the level of detail in generated maps. Using aerial imagery, commercial satellite data, or multi-source fusion techniques would allow for the creation of even more realistic, high-resolution maps.

7. Conclusions

The primary goal of this thesis was to develop a diffusion model-based framework for generating high-resolution maps of the fictional country of Carana. This objective was successfully pursued through a series of steps, beginning with acquiring a comprehensive understanding of diffusion models and their applications in GIS. This foundational knowledge led the implementation of the framework, which progressed into a structured data science project pipeline involving data collection, preprocessing, model training and validation.

The framework demonstrated its potential by producing synthetic map tiles that, despite some limitations, reflect significant progress toward generating realistic and high-resolution outputs. Validation methods, including histogram-based analysis and Fréchet Inception Distance (FID) evaluations, were employed to assess the quality of the generated maps. While the results highlight areas for improvement, such as tile coherence and color consistency, the overall framework offers a solid foundation for advancing map generation techniques.

This thesis not only achieves its stated objectives but also contributes to the broader field of geospatial AI by exploring the integration of diffusion models into GIS workflows. Ultimately, this research highlights the potential of generative AI in advancing geospatial visualization and enhancing peacekeeping operations.

8. Bibliography:

Albanwan, H., Qin, R., & Tang, Y. (2024). *Image Fusion in Remote Sensing: An Overview and Meta Analysis* (No. arXiv:2401.08837). arXiv. https://doi.org/10.48550/arXiv.2401.08837

Asperti, A., Merizzi, F., Paparella, A., Pedrazzi, G., Angelinelli, M., & Colamonaco, S. (2024). Precipitation nowcasting with generative diffusion models. *Applied Intelligence*, *55*(3), 187. https://doi.org/10.1007/s10489-024-06048-y

Asperti, A., Merizzi, F., Paparella, A., Pedrazzi, G., Angelinelli, M., & Colamonaco, S. (2025). Precipitation nowcasting with generative diffusion models. *Applied Intelligence*, *55*(3), 187. https://doi.org/10.1007/s10489-024-06048-y

Brock, A., Donahue, J., & Simonyan, K. (2019). *Large Scale GAN Training for High Fidelity Natural Image Synthesis* (No. arXiv:1809.11096). arXiv. https://doi.org/10.48550/arXiv.1809.11096

Burgess, C. P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., & Lerchner, A. (2018). *Understanding disentangling in $β$-VAE* (No. arXiv:1804.03599). arXiv. https://doi.org/10.48550/arXiv.1804.03599

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Dhariwal, P., & Nichol, A. (2021). *Diffusion Models Beat GANs on Image Synthesis* (No. arXiv:2105.05233). arXiv. https://doi.org/10.48550/arXiv.2105.05233

Espinosa, M., & Crowley, E. J. (2023). *Generate Your Own Scotland: Satellite Image Generation Conditioned on Maps* (No. arXiv:2308.16648). arXiv. http://arxiv.org/abs/2308.16648

Frühstück, A., Alhashim, I., & Wonka, P. (2019). TileGAN: Synthesis of Large-Scale Non-Homogeneous Textures. *ACM Transactions on Graphics*, *38*(4), 1–11. https://doi.org/10.1145/3306346.3322993

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks* (No. arXiv:1406.2661). arXiv. https://doi.org/10.48550/arXiv.1406.2661

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash

Equilibrium. *Advances in Neural Information Processing Systems*, *30*. https://papers.nips.cc/paper_files/paper/2017/hash/8a1d694707eb0fefe65871369074926d-Abstract.html

Ho, J., Jain, A., & Abbeel, P. (2020). *Denoising Diffusion Probabilistic Models* (No. arXiv:2006.11239). arXiv. http://arxiv.org/abs/2006.11239

Hughes, L. H., Schmitt, M., & Zhu, X. X. (2018). Mining Hard Negative Samples for SAR-Optical Image Matching Using Generative Adversarial Networks. *Remote Sensing*, *10*(10), Article 10. https://doi.org/10.3390/rs10101552

Jiang, Y., Liu, S., & Wang, H. (2025). Diffusion-based remote sensing image fusion for classification. *Applied Intelligence*, *55*(4), 247. https://doi.org/10.1007/s10489-024-06217-z

Kingma, D. P., & Welling, M. (2022). *Auto-Encoding Variational Bayes* (No. arXiv:1312.6114). arXiv. https://doi.org/10.48550/arXiv.1312.6114

Luo, Z., Song, B., & Shen, L. (2024). *SatDiffMoE: A Mixture of Estimation Method for Satellite Image Super-resolution with Latent Diffusion Models* (No. arXiv:2406.10225). arXiv. https://doi.org/10.48550/arXiv.2406.10225

Lütjens, B., Leshchinskiy, B., Requena-Mesa, C., Chishtie, F., Díaz-Rodriguez, N., Boulais, O., Piña, A., Newman, D., Lavin, A., Gal, Y., & Raïssi, C. (2021). *Physics-informed GANs for Coastal Flood Visualization* (No. arXiv:2010.08103). arXiv. https://doi.org/10.48550/arXiv.2010.08103

Mateo-García, G., Laparra, V., Requena-Mesa, C., & Gómez-Chova, L. (2021). Generative Adversarial Networks in the Geosciences. In *Deep Learning for the Earth Sciences* (pp. 24–36). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119646181.ch3

Pielage, L. (n.d.). *Diffusion Models in Dermatological Education: Flexible High Quality Image Generation for VR-based Clinical Simulations*.

Przymus, M., & Szymański, P. (2023). Map Diffusion—Text Promptable Map Generation Diffusion Model. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Advances in Urban-AI*, 32–41. https://doi.org/10.1145/3615900.3628787

Razavi, A., Oord, A. van den, & Vinyals, O. (2019). *Generating Diverse High-Fidelity Images with VQ-VAE-2* (No. arXiv:1906.00446). arXiv. https://doi.org/10.48550/arXiv.1906.00446

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). *Stochastic Backpropagation and Approximate Inference in Deep Generative Models* (No. arXiv:1401.4082). arXiv. https://doi.org/10.48550/arXiv.1401.4082

Sanguigni, F., Czerkawski, M., Papa, L., Amerini, I., & Saux, B. L. (2023). *Diffusion Models for Earth Observation Use-cases: From cloud removal to urban change detection*. https://doi.org/10.2760/46796

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). *Deep Unsupervised Learning using Nonequilibrium Thermodynamics* (No. arXiv:1503.03585). arXiv. https://doi.org/10.48550/arXiv.1503.03585

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (No. arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Yan, X., Yang, J., Sohn, K., & Lee, H. (2016). *Attribute2Image: Conditional Image Generation from Visual Attributes* (No. arXiv:1512.00570). arXiv. https://doi.org/10.48550/arXiv.1512.00570

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M.-H. (2024). *Diffusion Models: A Comprehensive Survey of Methods and Applications* (No. arXiv:2209.00796). arXiv. https://doi.org/10.48550/arXiv.2209.00796

Zhang, L., Rao, A., & Agrawala, M. (2023). *Adding Conditional Control to Text-to-Image Diffusion Models* (No. arXiv:2302.05543). arXiv. http://arxiv.org/abs/2302.05543

Zhao, B., Zhang, S., Xu, C., Sun, Y., & Deng, C. (2021). Deep fake geography? When geospatial data encounter Artificial Intelligence. *Cartography & Geographic Information Science*, *48*(4), 338–352. https://doi.org/10.1080/15230406.2021.1910075

Masters Program in Geospatial Technologies

Aleksandra Jastrzębska

Supported by:
Education and Culture
ERASMUS MUNDUS