POLITECNICO DI TORINO

Master's Degree in Data Science and Engineering

Master's Degree Thesis

Automated Landslide Detection from InSAR Sentinel-1 Acquisitions using Advanced Deep Learning Architectures

Supervisors

Prof. Paolo GARZA

Luca BARCO

Lorenzo INNOCENTI

Candidate

Davide BENOTTO

March 2026

Abstract

Natural disasters are extreme and sudden events that pose significant risks to human life and infrastructure. In the context of natural disaster management, the ability to rapidly map the affected areas is essential for risk assessment. This thesis investigates the use of deep learning techniques for landslide detection using InSAR data. The study is structured into two main parts. The first part focuses on the selection of landslide events and the construction of the corresponding dataset, following a processing pipeline that generates InSAR products from SAR data acquired by the Sentinel-1 mission. The dataset includes the following products: phase, coherence, intensity, displacement, unwrapped phase, elevation, and local incidence angle, combined with landslide ground truth and validity masks to account for missing or non-informative regions. The second part investigates deep learning–based models for pixel-level landslide mapping. Multiple architectures are evaluated: UNet, UPerNet, and SegFormer, combined with different encoder backbones. To address the class imbalance and spatial sparsity of landslide events, a set of strategies is introduced, including small-scale landslide oversampling, cropping-based augmentation, stratified sampling and label smoothing. Training is conducted using a masked loss formulation that excludes invalid pixels, while model performance is evaluated by selecting an optimal decision threshold during inference. The results provide a comparative analysis of different architectural and training strategies for InSAR-based landslide detection, showing that transformer-based architectures constitute a better solution and demonstrating the potential of InSAR data for automatic landslide recognition.

Table of Contents

List of Tables

List of Figures

Acronyms

CNN
Convolutional Neural Network

DEM
Digital Elevation Model

DInSAR
Differential Interferometric Synthetic Aperture Radar

ESA
European Space Agency

ESD
Enhanced Spectral Diversity

EO
Earth Observation

EW
Extra-Wide Swath

InSAR
Interferometric Synthetic Aperture Radar

IoU
Intersection over Union

IW
Interferometric Wide

NaN

Not a Number

POD

Precise Orbit Determination

POE

Precise Orbit Ephemerides

S1TBX

Sentinel-1 Toolbox

SAR

Synthetic Aperture Radar

SegFormer

Segmentation Transformer

SLAR

Side-Looking Airborne Radar

SLC

Single Look Complex

SM

Stripmap

SNAP

Sentinel Application Platform

SNAPHU

Statistical-cost, Network-flow Algorithm for Phase Unwrapping

TOPSAR

Terrain Observation with Progressive Scans SAR

U-Net

U-Shaped Network

UperNet

Unified Perceptual Parsing Network

ViT

Vision Transformer

WV

Wave

Chapter 1

Introduction

Natural disasters represent extreme and sudden events that pose significant risks to human life and infrastructure. In the context of disaster management, the ability to accurately map affected areas is essential for effective risk assessment. Traditionally, landslide inventory mapping has relied on the visual interpretation of optical satellite imagery. However, those sensors are limited by atmospheric conditions, such as cloud cover and poor illumination, which often coincide with the intense rainfall events that trigger landslides. To address this problem, the remote sensing community increasingly relies on Synthetic Aperture Radar (SAR), which maintains its effectiveness regardless of weather conditions and daylight.

Despite these advantages, SAR and interferometric SAR (InSAR) data are characterized by complex signal properties, that make their interpretation challenging and often dependent on expert analysis. As a consequence, analysing a vast amount of data is a slow process. Therefore, large-scale and rapid landslide mapping is a demanding and time-consuming process. In this context, research efforts have explored the use of machine learning and deep learning techniques to automate landslide detection from SAR-derived products.

Problem statement and approach

Building on these considerations, the automatic detection of landslides from interferometric products remains a challenging task. This difficulty arises from several factors, such as speckle noise, geometric distortions, atmospheric artifacts, and spatial sparsity of landslide events.

Existing studies focus on signal-level processing (e.g., phase unwrapping or noise mitigation) or on optical-based landslide mapping, while other works evaluate deep learning architectures for pixel-wise landslide detection from multi-channel InSAR data.

This thesis addresses the problem of automatic landslide detection from InSAR-derived products by investigating the suitability of modern semantic segmentation architectures for this domain. In contrast to multi-temporal InSAR monitoring approaches, this work focuses on rapid detection using only pre- and post-event acquisitions. For each event, multiple InSAR-derived features are combined to provide a rich and informative representation of the deformation signal.

The proposed methodology is structured into two main stages. First, a dedicated processing pipeline is defined to construct a dataset using Sentinel-1 SAR acquisitions. This pipeline generates phase, coherence, intensity, displacement, unwrapped phase, elevation, and local incidence angle bands, together with landslide ground-truth masks and validity masks used to exclude non-informative regions from training and evaluation. Second, deep learning architectures are trained and compared for pixel-level landslide segmentation. Particular attention is devoted to class imbalance, spatial sparsity, and invalid pixels through tailored sampling strategies, masked loss computation, and a label smoothing formulation designed to improve the detection of sparse deformation patterns.

Our contributions

The main contribution of this thesis lies in the formulation and evaluation of a deep learning framework for rapid landslide detection from InSAR data. The work formalizes landslide mapping as a semantic segmentation problem, relying exclusively on pre- and post-event acquisitions rather than long multi-temporal time series.

A multi-region dataset is constructed to assess model robustness under heterogeneous geomorphological conditions, moving beyond the single-area case studies that are more common in the literature.

Furthermore, the study introduces a tailored training and evaluation protocol designed to address the challenges of InSAR-based landslide mapping, including severe class imbalance, spatial sparsity of events, and the presence of non-informative or invalid pixels. Finally, both convolutional and transformer-based architectures are comparatively assessed, providing a consistent benchmark for architectural choices in this domain.

Document Structure

Chapter 2: Related Works Reviews the state of the art in landslide detection, covering traditional SAR methodologies, the transition to deep learning for Earth Observation, and previous research on neural networks applied to interferometric data. It also provides the theoretical background for the evaluated architectures: U-Net, UPerNet, and SegFormer.

Chapter 3: Synthetic Aperture Radar and Interferometry Introduces the fundamental principles of radar imaging, detailing signal properties, SAR geometry, and the foundations of Differential InSAR (DInSAR).

Chapter 4: Dataset Creation Describes the data acquisition process from the Sentinel-1 mission. It details the selection of study areas and the end-to-end InSAR processing chain implemented to generate the multi-channel stack of features.

Chapter 5: Methodology Presents the deep learning framework developed in this study. This includes dataset preprocessing, the specific implementation of model architectures, and the training strategies adopted to mitigate class imbalance, such as label smoothing and bucket sampling.

Chapter 6: Results Provides a comparative analysis of the experimental results. It evaluates model performance across different architectures and encoder backbones using quantitative metrics to identify the most effective solution for automated landslide recognition.

Chapter 7: Conclusions Summarizes the main findings of the research and suggests potential future developments for InSAR-based landslide detection.

Chapter 2

Related Works

2.1 Traditional SAR-based Landslide Detection Approaches

Synthetic Aperture Radar (SAR) remote sensing represents an important tool for landslide investigation, overcoming the limitations of optical sensors such as cloud occlusion and day-night dependency [1]. In the literature, detection strategies are generally categorized based on the specific signal component: amplitude (intensity) and phase.

Amplitude-based methods, such as Pixel Offset Tracking (POT), analyze changes in the backscattering intensity of the radar signal. A landslide event typically alters the surface roughness and consequently creates a contrast in the intensity image. These techniques are effective for detecting large, rapid displacements (in the order of meters) where the interferometric phase could be affected by decorrelation [2]. However, relying solely on amplitude often fails to capture slow and smaller movements, or deformations that do not drastically alter the surface texture.

On the other hand, phase-based methods, such as Differential SAR Interferometry (DInSAR), measure the phase shift between acquisitions to quantify ground displacement with millimeter precision. A problem of this method is represented by susceptibility to atmospheric artifacts. To mitigate it in long-term studies, Multi-Temporal InSAR algorithms were introduced, first with the Permanent Scatterer (PSI) [3] approach and later extended by algorithms such as SqueeSAR [4]. While these time-series methods represent an effective solution for monitoring, they are computationally expensive due to the higher number of acquisitions required (generally at least 15-20) and making them less suitable for rapid post-event mapping.

Historically treated as separate domains, phase and amplitude provide complementary information: the former quantifies displacement, while the latter highlights

surface changes. Modern Deep Learning workflows aim to integrate both components to enhance detection robustness.

2.1.1 Limitations of Manual Interpretation

Despite the high precision of InSAR measurements, the generation of landslide inventories still relies on visual interpretation by domain experts. For instance, *Rosi et al.* [5] show that identifying active landslides requires an intensive analysis of interferometric targets, combined with lots of expertise. While reliable, this manual approach is difficult to scale to large regions or rapid response scenarios [6].

Automating this process using deterministic algorithms or applying fixed thresholds on some parameters often lack generalization. This happens due to the complexity of the signal, that is affected by vegetation and atmospheric artifacts. This limitation highlights the necessity for data-driven strategies, capable of learning complex features.

2.2 Deep Learning for semantic segmentation in EO (Earth Observation)

The automated identification of landslides from Earth Observation (EO) data has historically relied on traditional Machine Learning (ML) algorithms, such as Random Forest and Support Vector Machines (SVM). While effective for pixel-based classification, these methods often fail to capture the spatial context of complex features. As shown by *Prakash et al* [7], Deep Learning (DL) approaches have outperformed traditional ML models, particularly in scenarios characterized by high variability in terrain and land cover.

Following this technological transition, Convolutional Neural Networks (CNNs) emerged as the dominant methodology. In this domain, the work of *Bragagnolo et al.* [8] proved the effectiveness of U-Net compared to traditional machine learning approaches.

Other studies have focused on architectural enhancements to handle deeper networks and improve gradient flow. For instance, *Qi et al.* [9] proposed a Residual U-Net (ResU-Net) that integrates ResNet blocks within the U-Net framework. This modification enhanced feature extraction capabilities in complex mountainous environments compared to the standard architecture.

2.3 Deep Learning for InSAR-based Landslide Detection

The application of Deep Learning to InSAR imagery represents a recent development compared to optical remote sensing, due to the complex nature of the interferometric signal. Consequently, the adaptation of semantic segmentation models to this domain required specific architectural strategies.

Deep Learning for InSAR signal processing A specific line of research has focused on addressing the physical limitations of the signal using encoder-decoder architectures. For instance, *Wu et al.* [10] proposed a dual-network approach (DDNet and PUNet) to detect and unwrap mining induced subsidence. *Kumar et al.* [11] demonstrated that a U-Net could effectively learn the mapping between wrapped and unwrapped phase. These studies are significant because they prove the capability to extract meaningful patterns even from noisy, wrapped interferometric data.

2.3.1 From CNNs to Vision Transformers for Landslide Mapping

In the domain of landslide detection, efforts have been directed towards adapting CNNs to identify active deformation directly. A representative study is done by *Chen et al.* [12] where a Deep Residual U-Net to segment active landslides is introduced. By incorporating residual blocks, they achieved superior performance compared to standard architectures.

However, pure CNN-based models are limited by their local receptive field. This limitation becomes relevant in geomorphological contexts, where deformation patterns extend over large spatial scales. As presented by *Abdallah et al.* in [13], the self-attention mechanism allows for a more effective distinction between deformation signals and atmospheric noise by analyzing the entire image context. This evolution highlights that hierarchical architectures—combining local feature extraction with global semantic understanding—represent the current state-of-the-art for robust landslide mapping.

2.4 Architectures

2.4.1 U-Net

U-Net [14] is an architecture originally proposed for biomedical image segmentation and represents a development in Fully Convolutional Networks (FCNs). Its design, shown in Figure 2.1 , is defined by a symmetric U-shaped structure consisting of a contracting path that decreases spatial dimension to capture relevant information, and a symmetrical expanding path, that upsamples and combines the learnt features. Skip connections allow the network to propagate context information to higher resolution layers, preserving spatial details that are lost during downsampling.

One of the main advantages of U-Net is the efficiency shown in scenarios characterized by limited training data. In fact, it relies on data augmentation that helps U-Net learn the necessary invariance and robustness required for complex mapping tasks without necessitating massive datasets. Furthermore, the network's flexibility regarding input dimensions and the number of input channels, allows it to process diverse information sources simultaneously.

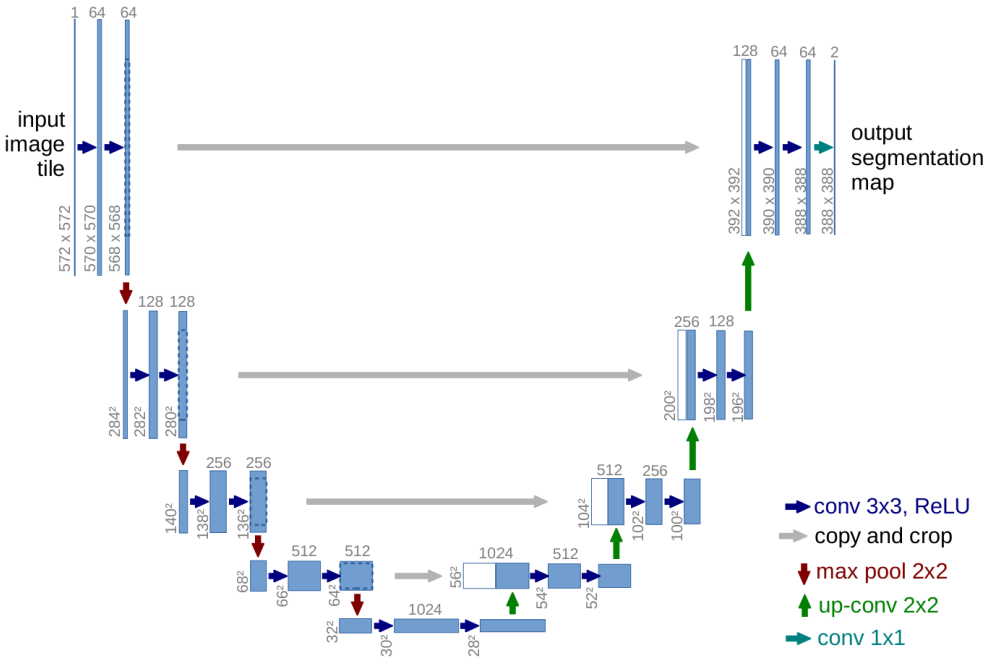

Figure 2.1: The U-Net architecture, illustrating the symmetrical contracting and expanding paths linked by skip connections [14].

Modern implementations integrate different encoders. An example is the use of ResNet [15], which introduces the concept of residual learning. Its main characteristic is the presence of residual blocks (Figure , which use shortcut connections to add the input of a layer to its output. This mechanism addresses the vanishing gradient problem, allowing the network to become much deeper without losing accuracy during training. This approach combines the localization capabilities of U-Net's decoder with the feature extraction of residual encoders, often representing a baseline for modern semantic segmentation studies.

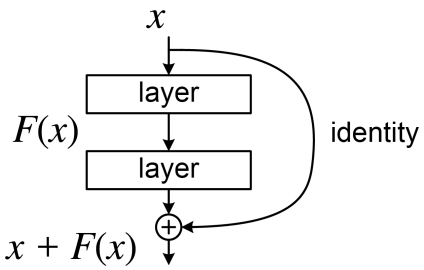

Figure 2.2: The ResNet residual block structure, illustrating the skip connection that enables residual learning [15].

2.4.2 UperNet

The Unified Perceptual Parsing Network (UPerNet) [16] is an encoder-agnostic framework designed to handle the complexities of dense prediction. It is based on the Feature Pyramid Network (FPN) [17], a top-down architecture with lateral connections that fuse high-level semantic features with lower-level details. To mitigate local ambiguities, UPerNet (Figure 2.3) incorporates a Pyramid Pooling Module (PPM) [18] on the last layer, which aggregates scene-level information through pooling operations.

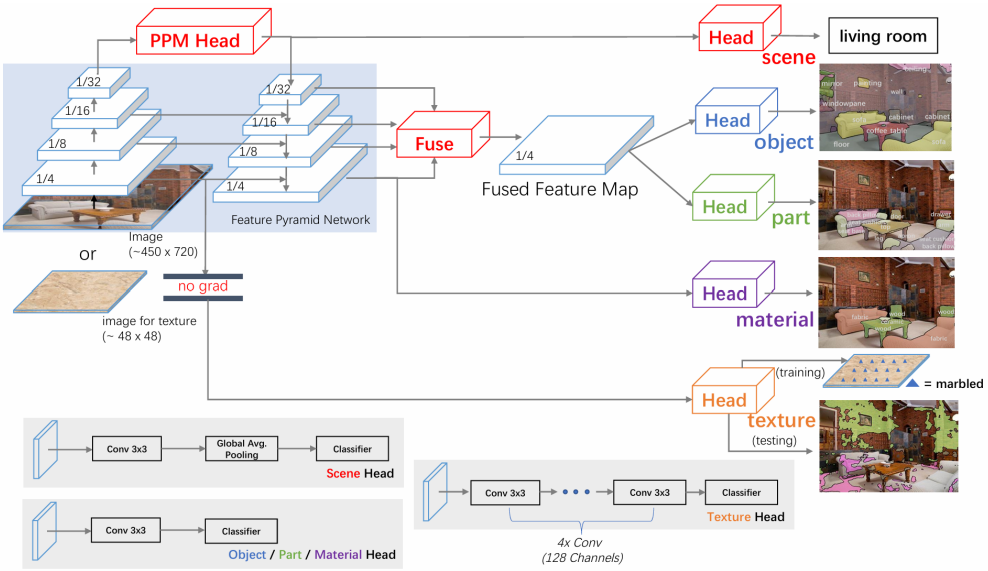

Figure 2.3: Overview of the UPerNet framework [16]

In modern implementations, ConvNeXt [19] is often adopted as the encoder. It represents a modernized convolutional architecture that adopts some design principles from Vision Transformers (ViTs), such as inverted bottlenecks and larger kernel sizes. Compared to traditional ResNet, ConvNeXt refines the standard residual block by optimizing the layer distribution and using fewer but more effective activation functions and normalization layers. In particular, it replaces the standard Rectified Linear Unit (ReLU) and Batch Normalization used in ResNet respectively with Gaussian Error Linear Unit (GELU) and Layer Normalization.

By integrating ConvNeXt within the UPerNet framework, the model achieves a balance between representational capacity and computational simplicity, providing an alternative to the transformer-based or traditional convolutional models.

2.4.3 Segformer

SegFormer [20] represents a shift toward purely Transformer-based architectures for semantic segmentation, designed to address the limitations of both traditional CNNs and Vision Transformers (ViTs). Its design (Figure 2.4) consists of a hierarchical transformer encoder paired with a lightweight decoder composed of Multilayer Perceptron (MLP) layers; the decoder only needs to aggregate multi-level features from the encoder to generate the final mask. This design ensures an effective trade-off between computational efficiency and segmentation accuracy.

Unlike previous models that adapted Vision Transformers (ViTs) designed for image classification, SegFormer introduces a Mix Transformer (MiT) backbone. It is

characterized by a hierarchical structure that produces multi-scale features, similar to convolutional networks, leveraging the global receptive field of self-attention. An important aspect of this architecture is the absence of explicit positional encodings. Instead, positional information is implicitly captured through Mix-FeedForward networks incorporating 3×3 depthwise convolutions, which embed local spatial context within the transformer blocks. This allows SegFormer to maintain consistent performance even when the input resolution during inference differs from that used during training.

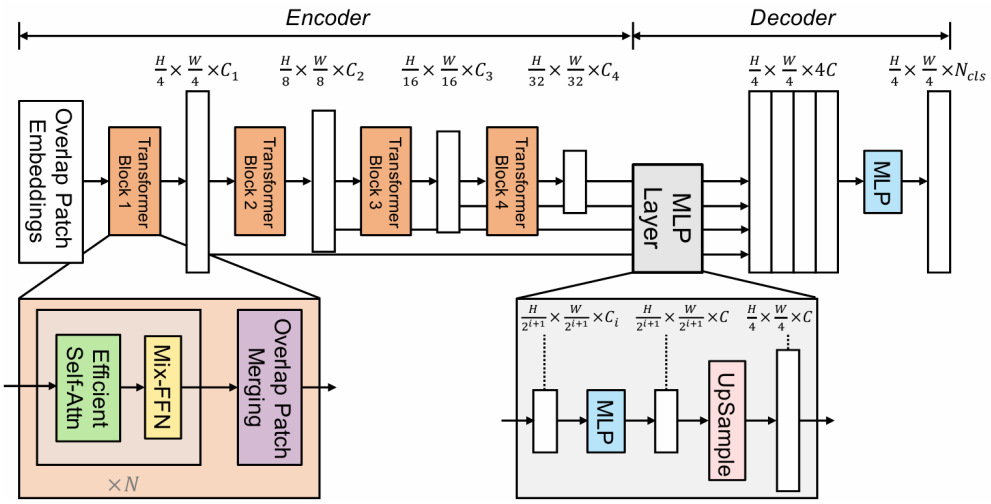

Figure 2.4: The SegFormer architecture, featuring a hierarchical transformer encoder and a lightweight All-MLP decoder [20].

Chapter 3

Synthetic Aperture Radar and Interferometry

This chapter provides a theoretical overview of Synthetic Aperture Radar (SAR) and Interferometric SAR (InSAR) principles that are strictly necessary to understand the data representation adopted in this thesis. Starting from the fundamental principles of radar imaging, the discussion explores the characteristics of SAR sensors and their advantages in Earth observation. In particular, this work focuses on single-event pre–post Differential InSAR (DInSAR) which enables the extraction of surface deformation patterns and constitutes the core data source for the landslide detection framework developed in this thesis.

3.1 Fundamentals of Synthetic Aperture Radar (SAR)

Synthetic Aperture Radar (SAR) is an active microwave imaging system that generates high-resolution, two-dimensional images of the Earth's surface [21, 22]. It can be considered the evolution of traditional Side-Looking Airborne Radar (SLAR), where the resolution along the flight path (azimuth) was strictly limited by the beamwidth of the physical antenna. SAR overcomes this limitation through signal processing: the phase and amplitude of the backscattered signals are stored and coherently combined to synthesize a virtual antenna that is significantly longer than the physical one. This technique allows SAR to achieve high azimuth resolution that is independent of the range distance, enabling the capture of fine details even from spaceborne platforms. Consequently, the invention of the SAR principle contributed to the development of airborne and spaceborne imaging radar systems over the last 40 years.

3.1.1 SAR Systems: Pros and Cons.

The primary motivation for employing SAR in Earth observation lies in its fundamental superiority over traditional optical sensors. In fact, SAR provides a set of key advantages with respect to optical sensors:

- Weather Independence: SAR utilizes microwaves, which can penetrate clouds, haze, fog, rain, and smoke; this makes it a powerful tool for monitoring regions covered by clouds or for emergency response during extreme weather events.
- Day-and-night operational capabilities: since SAR is an active imaging method, it operates independently of daylight, whereas optical sensors generally require sunlight.
- Flexible Resolution and Coverage: modifying parameters such as bandwidth and pulse repetition frequency, radar can be adjusted to capture high-resolution images of small areas or lower-resolution images covering vast regions.
- Coherent imaging: SAR is a coherent system, meaning it preserves both the amplitude (reflectivity) and the phase of the backscattered signal.

On the other hand, SAR images are affected by noise, geometric distortions and complex microwave interaction, making their radiometric and visual interpretation more challenging compared to optical imagery. Later in this thesis we will discuss how these challenges are addressed.

3.1.2 SAR Applications

Thanks to its all-weather, day-and-night acquisition capability and sensitivity to surface properties, SAR is adopted in a multitude of applications that go beyond conventional two-dimensional mapping. In particular, it is implemented in environmental monitoring tasks, such as tracking changes in ice and snow cover, detecting oil spills, or observing glacier flows. Among the many SAR applications, surface deformation monitoring and geohazard detection are particularly relevant for this work, as they directly relate to landslide processes and terrain instability.

3.2 SAR Geometry and Signal

The formation of a SAR image is based on a side-looking geometry, where the sensor moves along the flight direction (azimuth) while transmitting microwave pulses and receiving echoes in the perpendicular direction (slant range). Unlike optical systems, a SAR sensor fundamentally acts as a highly sophisticated stopwatch: its primary measurement is the two-way travel time of the radar pulse, which

determines the distance to the target. To ensure geographical consistency, each pixel is associated with geolocation information, mapping radar coordinates to a terrestrial system through geocoding.

3.2.1 The Resolution Cell

The resolution cell is the specific physical area on the ground that contributes energy to a single pixel in the SAR image. It represents a 3D "column" of space projected onto a 2D plane. The dimensions of this cell are defined by the system's bandwidth and antenna aperture, in particular the slant range resolution is determined by the former and the azimuth resolution is determined by the latter.

3.2.2 The Single Look Complex (SLC) Signal

The Single Look Complex (SLC) product represents a fundamental SAR data format, as it is the closest representation of the raw signal acquired by the sensor after. Unlike a standard photograph, every pixel z contains two pieces of information: a Real part (I, the In-phase component) and an Imaginary part (Q, the Quadrature component). This complex representation preserves the original phase relationships of the signal, which are essential for processing techniques such as interferometry. As a result, each pixel inherently contains both amplitude and phase information:

- Amplitude(A): it is the strength of the backscattered signal, calculated as $|z| = \sqrt{I^2 + Q^2}$. It represents the microwave reflectivity of the target, where bright pixels indicate strong returns and dark pixels indicate weak returns. This strength is determined by the target's physical properties, such as surface roughness, dielectric constant and geometry.

- Phase(ϕ). it is the angle of the complex number, calculated as $\arctan(Q/I)$ and expressed in radiants. It represents the difference in the phase of the radar signal reflected from the ground between the two acquisitions. From a physical perspective, the phase is directly proportional to the two-way travel path ($2R$) between the sensor and the target, following the relation:

$$\phi = \frac{4\pi}{\lambda} R \pmod{2\pi} \tag{3.1}$$

  where λ is the radar wavelength and R is the sensor-to-target distance. As illustrated in Figure 4.4, the phase represents the specific stage of the oscillating wave upon its return and it is "wrapped" in a range of $[-\pi, \pi]$, measuring the fractional part of the wave cycle. In a single SLC image, the phase alone looks like random noise and only becomes useful when comparing two different images to measure changes in that distance.

In learning-based approaches, amplitude and intensity are not only physical observables, but also act as informative features that encode surface roughness, material properties, and backscattering behavior, contributing to the discrimination between stable terrain and surface displacement. Overall, SAR SLC products provide a physically grounded, multi-dimensional signal representation of the observed scene.

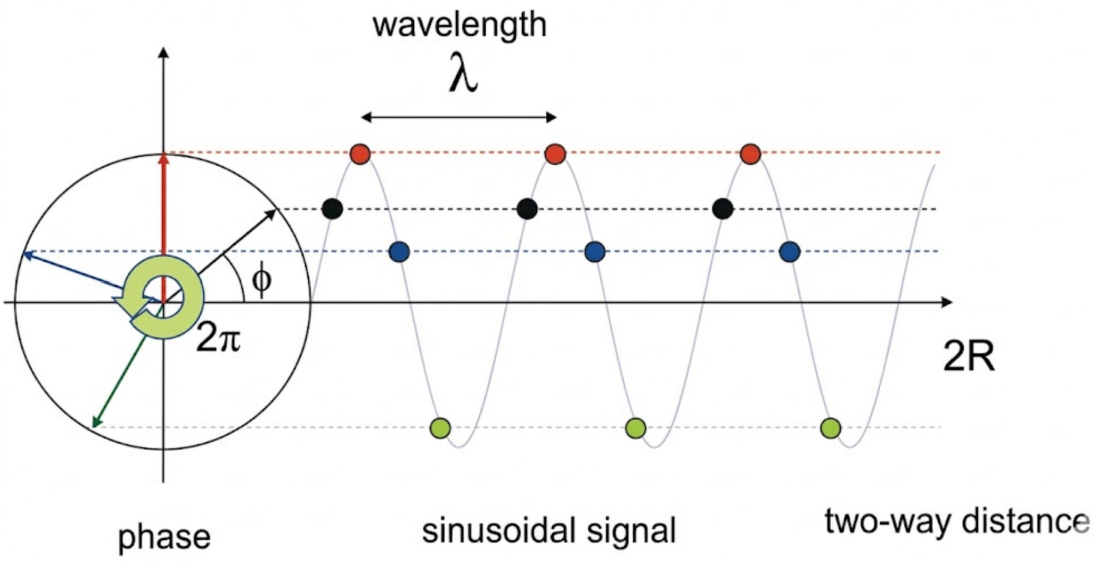

Figure 3.1: Sinusoidal function of ϕ [23]

3.2.3 Speckle

Because a resolution cell contains many "elemental scatterers" (such as leaves, branches), their individual echoes interfere. This interference causes a called "salt and pepper" noise known as speckle. In practical processing pipelines, speckle is mitigated through a technique called multilooking, leading to more stable signal representations that are better suited for subsequent analysis and learning-based modeling.

3.3 SAR Interferometry (InSAR)

InSAR (Interferometric Synthetic Aperture Radar) is a remote sensing technique that combines two or more complex-valued radar images to extract three-dimensional information or measure surface changes with high precision [24, 23, 25, 22]. By exploiting the phase of the coherent radar signal, InSAR transforms radar from a tool for viewing surface reflectivity into a quantitative instrument for surveying topography and motion. In fact, since the phase of each pixel contains distance information accurate to a small fraction of the radar wavelength, comparing two signals allows for the detection of minute path length differences with centimetric or even millimetric accuracy.

3.3.1 The Interferogram Generation.

The interferogram is the primary product of InSAR analysis. It is a visual and mathematical representation of the phase difference between two SAR images. It is generated by cross-multiplying, pixel by pixel, the first SAR image with the complex conjugate of the second. In particular, the result is a new complex image with:

- Interferometric Amplitude (ΔA): the product of the amplitudes of the starting images.
- Interferometric Phase ($\Delta\phi$): the difference between the phases of the starting images.

Image co-registration For the interferogram to be reliable, a critical step is image-to-image co-registration. This process ensures that the pixels from both images represent exactly the same resolution cell on the ground, achieving a precise one-to-one fit. If co-registration is inaccurate, the resulting interferogram will be dominated by noise, making it impossible to extract meaningful information.

3.3.2 Coherence and Interferogram Quality

Coherence (γ) is a fundamental parameter in InSAR processing that quantifies how similar the two SAR signals are for each pixel. This value is estimated by comparing the amplitude and phase information of the master and slave images within a small local window of pixels The coherence assumes values between 0 and 1. A coherence value close to 1 indicates high similarity and a stable target with minimal changes. In contrast, a value close to 0 suggests that the target has changed significantly, making the phase information dominated by noise and practically unusable.

Factors Affecting Coherence. The loss of coherence, known as decorrelation, can be caused by several factors:

- Temporal Decorrelation: Changes in the ground surface over time, such as vegetation growth, harvesting, or weather effects (rain, snow). The larger the temporal distance, the lower the coherence will be, as we can observe in Figure 4.3
- Spatial Decorrelation: Differences in the satellite's viewing angle between the two passes
- Processing errors: Inaccurate co-registration of the two SAR images.

In this work, coherence represents a discriminative feature that provides indirect information on surface changes, vegetation dynamics, and terrain disturbance, making it relevant for learning-based landslide detection.

Figure 3.2: Degradation of coherence over different temporal baselines. Areas become darker as scattering properties change over time, [26].

3.3.3 Geometric configurations

Interferometric SAR techniques are illustrated in and can be classified into two main categories based on their acquisition geometry:

- Across-Track Interferometry: uses two antennas separated by a spatial distance (baseline) to acquire data simultaneously. This configuration is primarily used for deriving terrain elevation and generating Digital Elevation Models (DEMs).

- Repeat-Pass (or Repeat-Track) Interferometry: involves a single antenna that illuminates the same area at different times (t_1, t_2). This configuration enables the detection of surface changes and ground deformation phenomena, such as landslides and terrain instability.

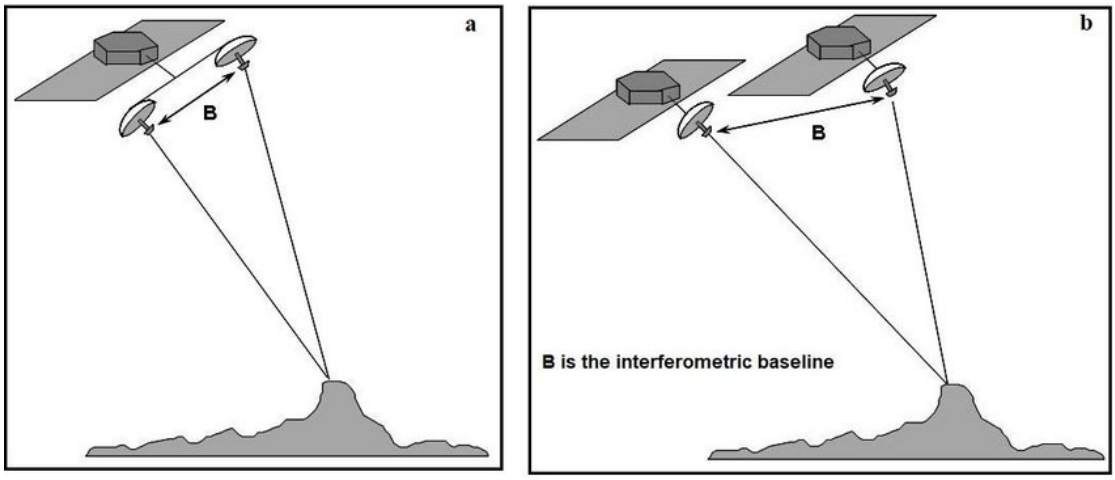

Figure 3.3: Comparison between InSAR acquisition modes: (a) Across-Track (Single-Pass) interferometry and (b) Repeat-Pass interferometry [27].

3.3.4 Differential InSAR (DInSAR).

Within the repeat-pass framework, Differential InSAR (DInSAR) is the specific approach used to isolate ground displacement. Depending on the number of acquisitions, deformation analysis can be performed either through multi-temporal approaches (Time-Series InSAR) or through single-event approaches based on pre–post image pairs. This thesis adopts a single-event repeat-pass DInSAR framework. Unlike time-series analysis, this method focuses on a single pair of acquisitions to detect changes occurred between two specific dates. The success of this approach depends on the ability to isolate the actual displacement from the total signal recorded by the radar. The most significant interference elements in the signal are the topography (the shape of the hills and valleys) and the Earth's curvature. In modern DInSAR processing, this is handled by using an external Digital Elevation Model (DEM). Once these elements are removed, the remaining information reveals the displacement of the ground that occurred between the two satellite passes. in Figure 3.4 is represents an example of the creation process that

translates the temporal difference between acquisitions into a spatial displacement map.

Phase Unwrapping. As previously described, the phase $\Delta\phi$ is only measured within the "wrapped" interval of $[-\pi, \pi]$. This means that the radar cannot directly distinguish between a displacement of half a wavelength and a displacement of multiple full wavelengths plus that same fraction. Visually, this creates the characteristic "fringe" pattern where values repeat every 2π radians. Phase Unwrapping is the mathematical process used to reconstruct the original, continuous phase signal from these wrapped values. By adding or subtracting the correct integer multiple of 2π to each pixel, the algorithm "unrolls" the phase, converting the cyclical interference fringes into a continuous values, as we can observe in Figure 3.5. Figure 3.4 shows an example of the final output of the entire process, SAR images selection to unwrapped phase: a displacement map where the phase has been fully unwrapped and converted into centimeters of movement.

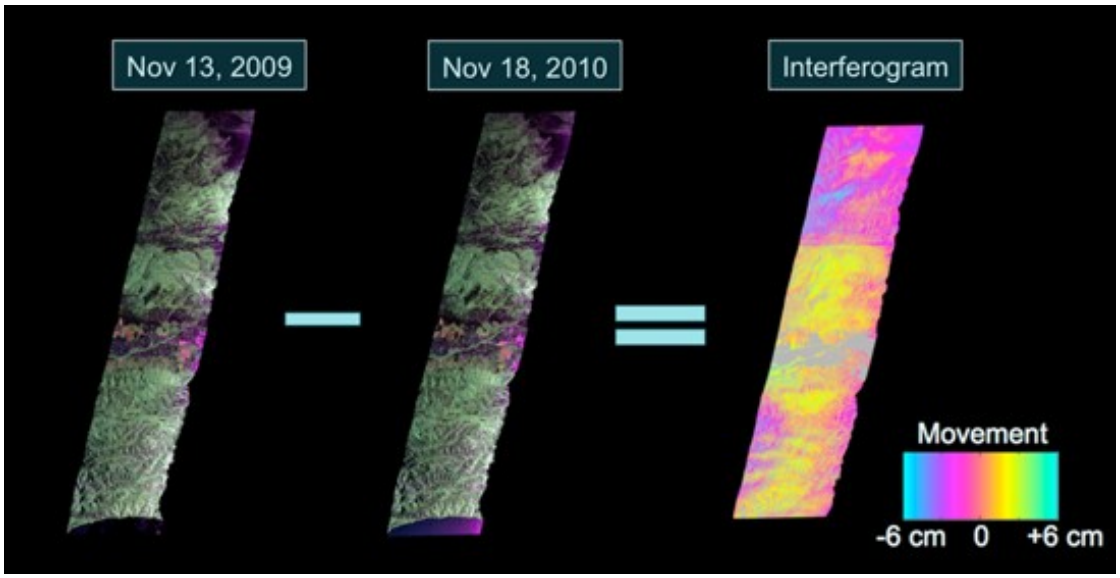

Figure 3.4: Final displacement map after DInSAR processing and phase unwrapping, representing ground deformation in centimeters, [26].

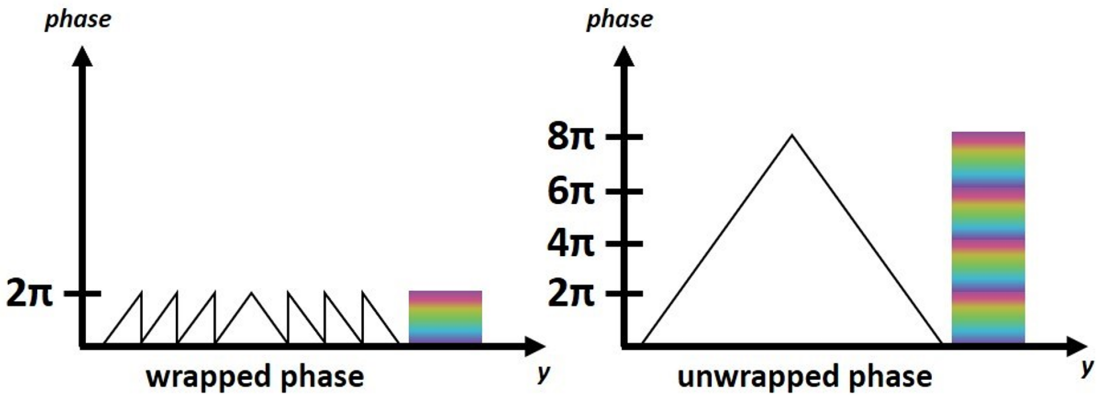

Figure 3.5: Comparison between wrapped phase (modulo 2π) and the continuous unwrapped phase profile. [28]

Chapter 4

Dataset Creation

4.1 The Copernicus Sentinel-1 Mission

The Sentinel-1 [29, 30] is a series of satellites of the Copernicus programme, an Earth Observation initiative coordinated by the European Union in partnership with the European Space Agency (ESA). The primary goal of Copernicus is to establish a modern, multi-disciplinary monitoring system that provides information for environmental management, disaster response, and climate change mitigation. Within this framework, Sentinel-1 serves as the dedicated radar instrument of the constellation, ensuring that high-quality imagery is available for a wide range of scientific and operational services. The architecture of the mission is based on a constellation of polar-orbiting satellites sharing the same orbital plane. This design was specifically chosen to provide a consistent and long-term data archive, which is essential for monitoring slow-moving phenomena like landslides over years or even decades. The mission began with the launch of Sentinel-1A in 2014, followed by Sentinel-1B in 2016. Although Sentinel-1B concluded its operations in 2022 due to a technical anomaly, the mission's continuity is currently supported by the Sentinel-1A and the newly launched Sentinel-1C (December 2024) and Sentinel-1D (November 2025). This sequence of satellites ensures uninterrupted access to radar data, maintaining a repeat cycle of 12 days for a single satellite and 6 days for the full constellation.

C-Band SAR Sentinel-1 is equipped with an active C-band Synthetic Aperture Radar (SAR) sensor, operating at a frequency of 5.4 GHz ($\lambda \approx 5.6$ cm). As described in the previous chapter, SAR offers significant advantages over passive optical sensors: it is completely independent of solar illumination and can penetrate cloud cover or rain, ensuring "all-weather" imaging.

Interferometric Wide, Bursts and Slice The Sentinel-1 mission offers four acquisition modes tailored to different operational needs: Stripmap (SM), Extra-Wide Swath (EW), Wave (WV), and Interferometric Wide (IW). The data utilized in this thesis was acquired in IW mode. This choice is dictated by the fact that IW is the primary acquisition mode of the mission, specifically optimized to satisfy most service requirements for environmental and land-cover monitoring. The technical advantage of this mode lies in the TOPSAR (Terrain Observation with Progressive Scans SAR) technique. The system achieves a nearly uniform signal quality and avoids common radar imaging artifacts. This results in more stable and homogeneous image quality, which is essential for the interferometric processing. Each Sentinel-1 image acquired in IW mode is organized into three sub-swaths (IW1, IW2, and IW3). Each sub-swath is further divided into several bursts—typically nine—which are small, overlapping snapshots that ensure continuous coverage. Furthermore, long data acquisitions are distributed as a sequence of slices within a single orbit. These components must be correctly selected and reassembled during the processing chain to generate a seamless displacement map.

4.2 Landslide Inventory

4.2.1 Study Area

The selection of the study areas was primarily driven by the availability of high-quality landslide inventories. As summarized in Table 4.1, the dataset covers several regions characterized by different geological and climatic conditions, providing a diverse range of landslide typologies. These events were identified through various inventories, which provided the location and timing of each movement. The pre- and post-event images were selected to bracket the entire period of activity reported in the inventories, ensuring that the total displacement was captured within the interferometric pair.

Region	Event Date	Pre- Image	Post- Image
Peshgor, Afghanistan	2018-07-12	2018-07-09	2018-07-21
Beijing, China	2023-07-28 – 07-31	2023-07-24	2023-08-05
Jishishan, China	2023-12-18	2023-12-14	2023-12-26
Jiuzhaigou, China	2017-08-08	2017-08-06	2017-08-18
Luding, China	2022-09-05	2022-08-28	2022-09-09
Maerkang, China	2022-06-10	2022-06-03	2022-06-15
Tiburon, Haiti	2021-08-14 – 08-17	2021-08-05	2021-08-17
Cugenang, Indonesia	2022-11-21	2022-11-19	2022-12-01
Lombok, Indonesia	2018-08-19	2018-08-11	2018-08-23
Pasaman, Indonesia	2022-03-04	2022-02-23	2022-03-07
Emilia-Romagna, Italy	2023-05-02 – 05-17	2023-04-28	2023-05-22
Liguria, Italy	2020-10-01 – 10-03	2020-09-26	2020-10-08
Iburi, Japan	2018-09-06	2018-09-05	2018-09-17
Michoacan, Mexico	2022-09-19	2022-09-13	2022-09-25
Mon State, Myanmar	2019-08-04 – 08-09	2019-07-31	2019-08-12
Thame, Nepal	2024-08-17	2024-08-10	2024-08-22
Porgera, Papua N. G.	2018-02-25	2018-02-19	2018-03-03
Cotobato, Philippines	2022-11-01	2022-10-22	2022-11-03
Goriška, Slovenia	2023-08-06	2023-07-28	2023-08-09
Milin, Tibet	2017-11-17	2017-11-11	2017-11-23

Table 4.1: Summary of the study areas and the Sentinel-1 image pairs used for the DInSAR processing, showing the event dates and the relative acquisition dates.

4.2.2 Data Harmonization

The ground truth for these events was initially collected in various vector formats, including Shapefiles (.shp), Keyhole Markup Language (.kml), and GeoJSON (.geojson) files. These files contain the polygons or points representing the spatial extent of the landslides as mapped by local authorities or geological surveys. To ensure compatibility with the InSAR processing products and the subsequent deep learning pipeline, all vector data were rasterized and converted into the NetCDF (.nc) format. This process ensures that the ground truth shares the same spatial reference system and grid geometry as the radar data. The final output of this harmonization is a set of binary masks, where a value of 1 represents a landslide pixel and 0 represents the background.

4.3 InSAR Processing Chain

The transformation of raw Sentinel-1 radar data into actionable displacement and coherence maps requires a sequence of mathematical and geometric operations. This procedure is designed to isolate the phase component related to ground movement from other contributions, such as topography and atmospheric noise. The methodology adopted in this work for processing Sentinel-1 image pairs follows the official guidelines and standards established by the European Space Agency (ESA) for the S1TBX (Sentinel-1 Toolbox) [28]. The processing is structured into several sequential stages, illustrated in Figure 4.1.

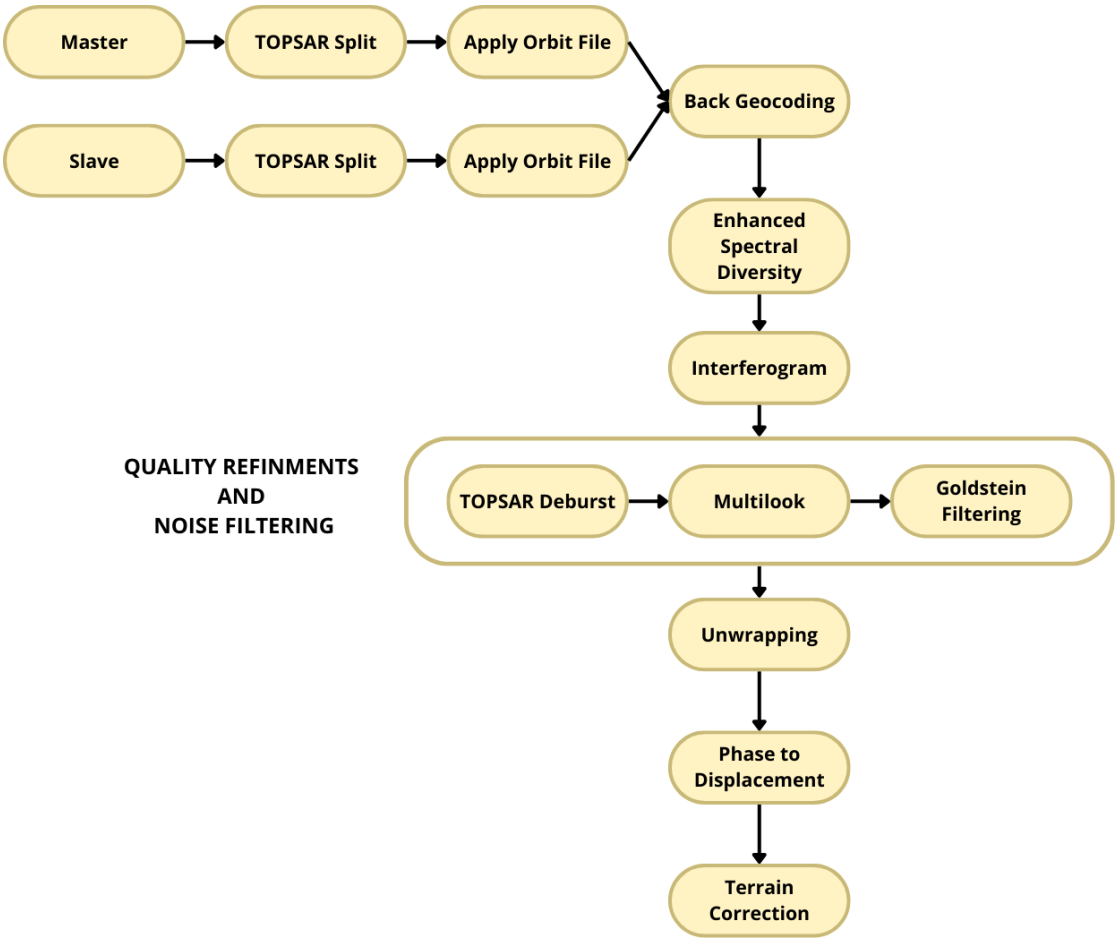

Figure 4.1: Detailed workflow of the InSAR processing chain.

Software envinronment and tools The processing chain was implemented using the Sentinel Application Platform (SNAP), the open-source architecture developed by ESA for exploiting Sentinel data [31]. To automate the workflow and ensure reproducibility across the entire landslide dataset, the esa-snappy Python interface was utilized [32]. This bridge allows for the programmatic execution of SNAP's C-Java operators within a Python environment, facilitating the handling of large-scale data arrays and integration with the deep learning preprocessing steps.

For the task of phase unwrapping (i.e. the process of converting the relative interferometric phase into absolute displacement) the SNAPHU (Statistical-cost, Network-flow Algorithm for Phase Unwrapping) software was employed [33]. Developed at Stanford University, SNAPHU is the industry-standard tool for solving the phase ambiguity through statistical cost functions, and it was integrated into the pipeline as an external plugin.

Data Acquisition The first stage of the pipeline constitues the acquisition of Sentinel-1 Single Look Complex (SLC) images. This is achieved through the OData API [34] available from the Copernicus Data Space Ecosystem. This approach allows the pipeline to automatically query and download specific image pairs based on orbit, slice number and and time windows.

TOPSAR Split Once we have SAR data, the first operator in the pipeline is the Terrain Observation with Progressive TOPSAR (Terrain Observation with Progressive) -Split. This step is fundamental for reducing the computational complexity, allowing for the selection of only the desidered sub-swath (IW1, IW2, or IW3) and the specific bursts that cover the area of interest. It is important to underline that a minimum of two bursts must be selected for each study area to ensures sufficient spatial coverage for the co-registration.

Applying Orbit File After the sub-swath selection, the Apply Orbit File operator is executed on the split product. This step is crucial for enhancing the geometric accuracy of the radar data. While the metadata of the original Sentinel-1 products contains preliminary orbital information, this operator restitutes the Precise Orbit Ephemerides (POE), which are calculated by the Copernicus Precise Orbit Determination (POD) service and made available days after the acquisition. By updating the orbit state vectors with these precise data, the pipeline ensures a more accurate co-registration of the image pairs, minimizing potential offsets during the interferogram formation.

Co-registration The co-registration process ensures that the primary (Master) and secondary (Slave) images are pixel-to-pixel aligned. In our pipeline, the pre-event image corresponds to the Master, while the post-event to the Slave.

The alignment is performed using the Back-Geocoding operator, which uses the orbital data and a Digital Elevation Model (DEM) to compensate for the topographic distortions. In particular, the Copernicus 30m Global DEM was utilized to achieve high-precision mapping between the two radar geometries. After this first step, the Enhanced Spectral Diversity (ESD) operator is applied. This step has the responsibility to refines the co-registration in the azimuth direction by analyzing the phase difference in the overlapping areas between consecutive bursts. The application of ESD increases the quality quality of the alignment required for the interferometric phase estimation.

Interferogram formation Once the co-registration is complete, the Interferogram Formation operator is invoked to compute the phase difference between the Master and Slave images. In this stage, the Copernicus 30m Global DEM is again utilized to simulate and remove the contribution of the Earth's curvature and topography from the interferometric phase. Specifically, the pipeline is configured with the "Subtract Flat-earth Phase" and "Subtract Topographic Phase" parameters set to true. By removing these components, the resulting Differential Interferogram isolates the phase variations primarily related to ground displacement and noise. Additionally, the "Output Elevation" parameter is enabled to maintain a consistent topographic reference for the subsequent geocoding steps. A critical output of this process is the coherence band, enabled via the "Include Coherence" parameter. As discussed in previous sections, coherence serves as a proxy for the reliability of the interferometric phase, with the resulting phase band characterized by values wrapped in the range $[-\pi, +\pi]$. Additionally, the **Intensity** band is included in the output products. While phase and coherence represent the geometric and correlation properties of the signal, intensity reflects the backscattering strength of the radar pulse from the ground. This band has a crucial role for identifying structural features, land cover, and the general morphology of the study area.

An example of the primary products generated in this phase—Coherence, Intensity, Wrapped Phase, and Elevation—is presented in Figure **??**, showing their different contributions to the characterization of the landslide area.

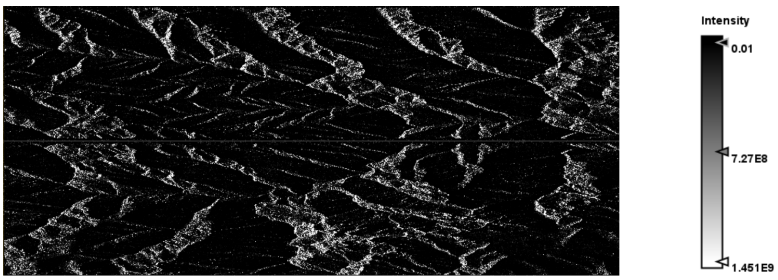

Figure 4.2: Intensity band

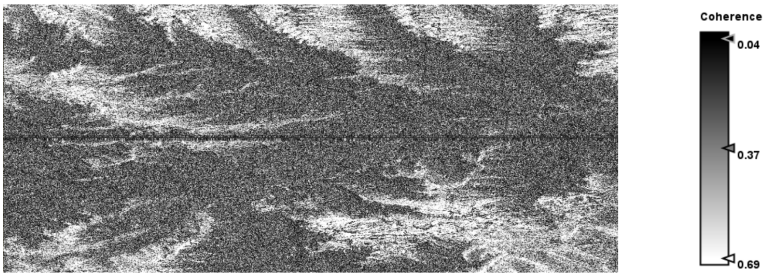

Figure 4.3: Coherence band

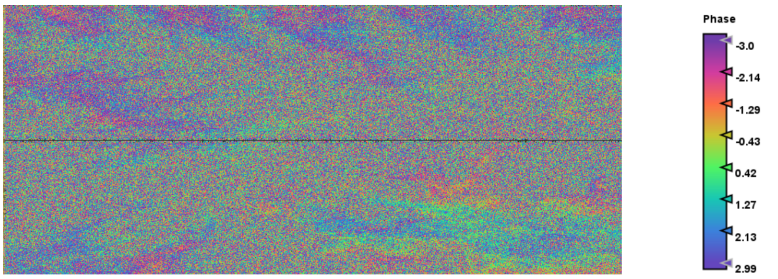

Figure 4.4: Wrapped Phase band

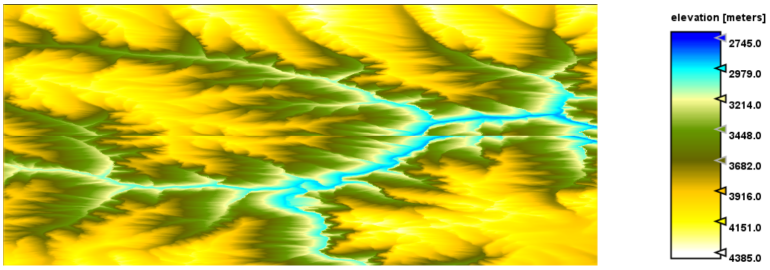

Figure 4.5: Elevation band

Figure 4.6: Visual representation of the primary InSAR products obtained after the interferogram formation step, China Markeang.

Post-interferometric Refinements Following the interferogram formation, a series of refinement steps are applied to improve the signal quality and reduce the data dimensionality before the phase unwrapping process.

- TOPSAR Deburst: since the initial processing is performed on individual bursts, the resulting image contains black separation lines (demarcation areas) as we can observe in Figure **??**. This operator merges the selected bursts into a single, continuous interferometric product, ensuring spatial continuity across the entire study area.
- Multilooking: it serves to reduce the speckle noise and to improve the interpretability of the image, combining multiple lines of pixels through a convolution operation within a small moving window. In this pipeline, an adaptive approach was implemented to maintain an optimal balance between noise reduction and spatial detail. Rather than using fixed parameters, the number of range looks is dynamically selected based on the image width. This ensures that the computational load is reduced and the signal-to-noise ratio is enhanced consistently across different study areas, regardless of their original dimensions.
- Goldstein Phase Filtering: the final refinement step is the Goldstein Phase Filtering. This adaptive frequency filter is specifically used to further reduce the noise associated with temporal and geometric decorrelation. The Goldstein filter facilitates the subsequent phase unwrapping, making the phase transitions clearer and more consistent.

The improvement in phase clarity and noise reduction achieved through these steps is visually demonstrated in Figure 4.9

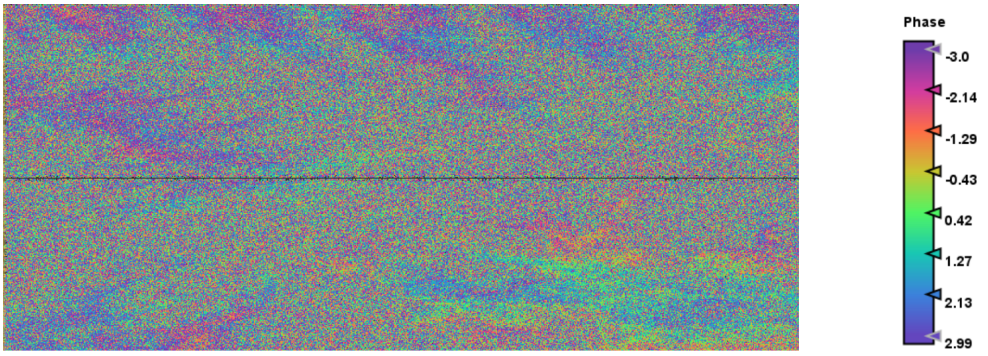

Figure 4.7: Wrapped Phase (Pre-refinement)

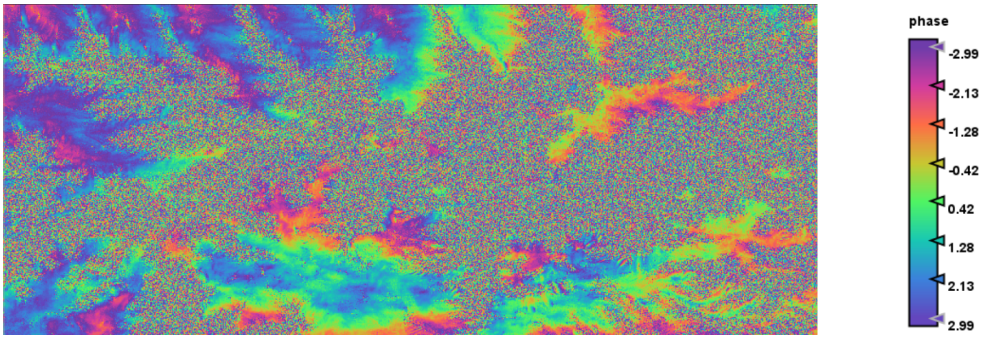

Figure 4.8: Wrapped Phase (Post-refinement)

Figure 4.9: Comparison of the wrapped phases, China Markeang: (a) before and (b) after the application of TOPSAR Deburst, Multilooking, and Goldstein Phase Filtering. The enhanced clarity and reduced noise in the post-refinement image can be easily seen; it will help improving the quality for subsequent phase unwrapping.

Phase Unwrapping Phase unwrapping is the process of recovering the absolute phase from the wrapped interferometric phase, which is otherwise constrained within the $[-\pi, +\pi]$ interval. As previously introduced, this stage requires the transition from the esa-snappy environment to the SNAPHU software. The workflow is structured into three phases:

- Export for SNAPHU: Using the SnaphuExport operator within the esa-snappy pipeline, the refined interferogram is converted into a format compatible with SNAPHU. During this step, the pipeline generates the necessary configuration files necessary for snaphu.
- Unwrapping Execution: The unwrapping is performed externally using the SNAPHU algorithm. By solving the phase ambiguity through statistical cost functions, the software converts the cyclic fringes into a continuous phase representation that is directly proportional to the ground displacement.
- Snaphu Import: Once the external processing is complete, the unwrapped phase is re-imported into the SNAP environment using the SnaphuImport operator. This step is essential to re-associate the unwrapped data with the original metadata and georeferencing information, allowing the subsequent processing stages to be executed once again via the esa-snappy library.

The substantial difference between the wrapped and unwrapped phase, which transforms cyclic fringes into a continuous displacement field, is visually highlighted in Figure 4.12.

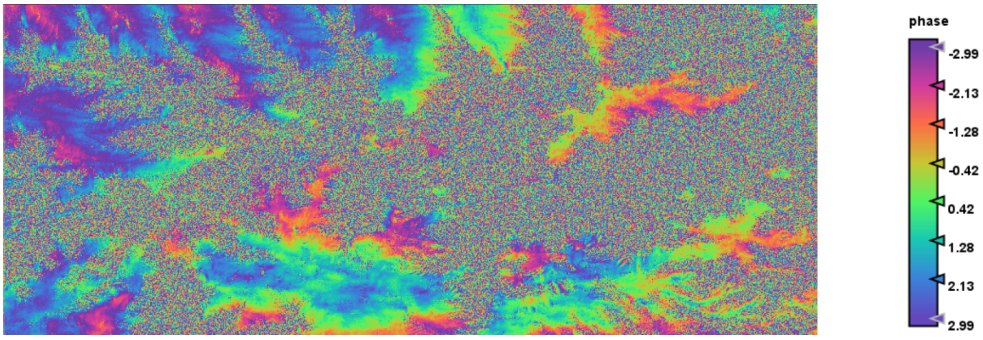

Figure 4.10: Wrapped Phase band

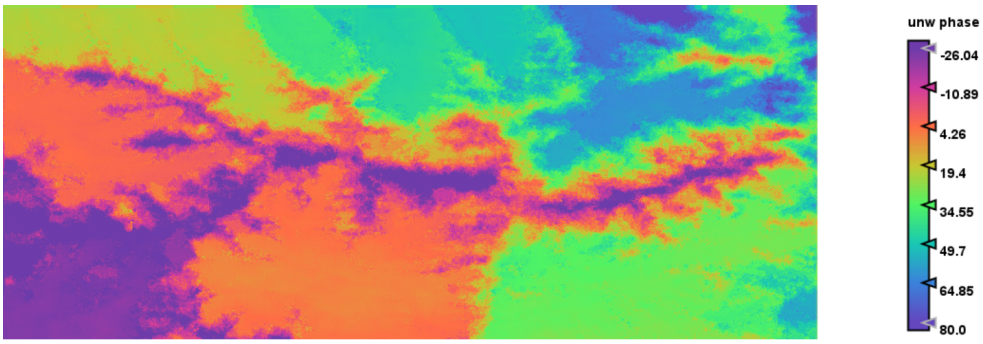

Figure 4.11: Unwrapped Phase band

Figure 4.12: Comparison of the interferometric phases, China Markeang: (a) Wrapped Phase, constrained within $[-\pi, +\pi]$ showing cyclic fringes, and (b) Unwrapped Phase, representing continuous displacement values after SNAPHU processing.

Phase to Displacement This stage of the process involve the conversion of the signal into physical units. The unwrapped phase, initially expressed in radians, is converted into absolute displacement measured in meters. To ensure data integrity, a BandMerge operation is included in this step, allowing the new "displacement" band to be stored alongside the previous generated bands.

Terrain Correction The radar data, initially acquired in the satellite's line-of-sight (slant-range) geometry, is affected by significant geometric distortions due to the Earth's topography. To correct these, the Terrain Correction operator is applied using the Copernicus 30m Global DEM. The process incorporates the following specific configurations:

- Pixel Spacing: The output resolution is set to 30 meters, a value chosen to align with the effective spatial resolution resulting from the adaptive multilooking process.
- Local Incidence Angle: The "saveLocalIncidenceAngle" parameter is enabled to provide critical information regarding the radar's illumination geometry for each pixel.

The final map-projected product is exported in NetCDF (.nc) format, providing a multi-dimensional structure for the deep learning training pipeline. The resulting geocoded products, specifically the displacement map derived from the unwrapped phase and the local incidence angle, are illustrated in Figure 4.15.

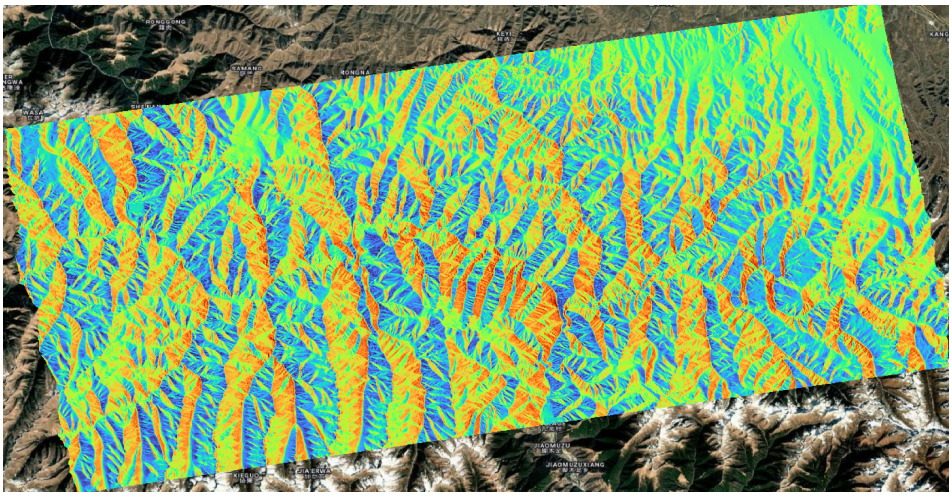

Figure 4.13: Local Incidence Angle band

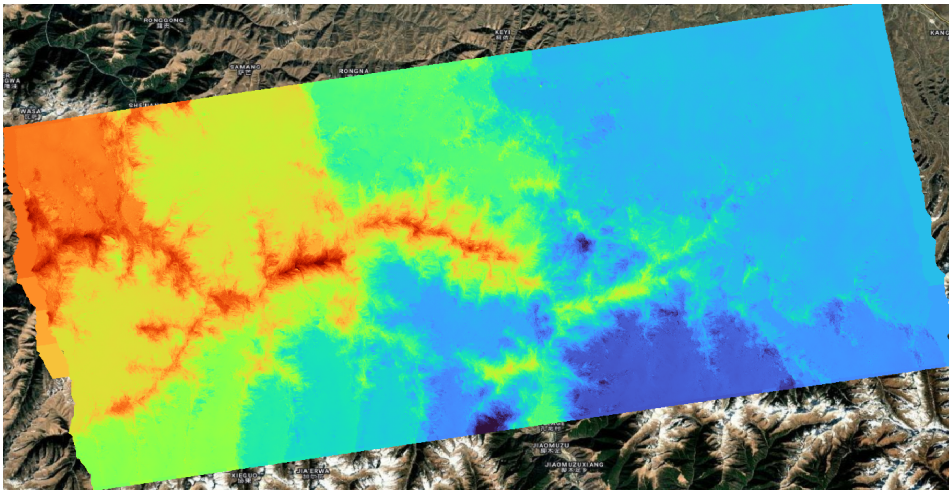

Figure 4.14: Geocoded Displacement Map (meters)

Figure 4.15: Final outputs after Terrain Correction: (a) the displacement map converted into meters and (b) the local incidence angle map. Both products are orthorectified and projected onto the geographic coordinate system, ready for integration into the deep learning model.

Chapter 5

Methodology

This chapter presents the methodological framework adopted in this thesis for landslide detection using deep learning and InSAR data. It describes the processing pipeline, that goes from raw data preparation and dataset construction, to model architecture design and training strategy. In the first section, the chapter details the preprocessing procedures used to transform InSAR products into smaller and easier to manage multi-bands patches. Then the architectural choices adopted for semantic segmentation are presented, and following the strategies implemented to address challenges typical of a landslide detection task. The main challenge presented and handled is the class imbalance, that is managed through smoothing, sampling and cropping strategies. The chapter concludes with the experimental setup and evaluation methodology, defining the data splitting and an optimal threshold selection used throughout the experimental analysis.

5.1 Dataset Preprocessing

This section describes the preprocessing steps required to transform the raw InSAR products into a better format for deep learning. The primary objective is to convert the images into a format that facilitates efficient training, while incorporating additional information (i.e. landslide ground truth and validity mask) to guide the model's learning process.

5.1.1 Data tiling and patches

Once the InSAR processing chain is complete and the seven physical bands (wrapped phase, coherence, intensity, elevation, displacement, unwrappaed phase, and local incidence angle) are generated, the final stage involves integrating the binary landslide mask into the same file. To ensure spatial correspondence between the

SAR images and the ground truth, the masks are aligned to the radar at the same spatial resolution. This operation creates a multi-channel stack where the ground truth becomes the eighth band of the product. Then, to suit the architectural requirements of the deep learning model and optimize computational efficiency, data is partitioned into non-overlapping patches of 256×256 pixels. This operation brings the total dataset size to 1098 patches.

Valid Mask Generation

For each patch, a valid mask has been calculated. It consists in a binary mask that serves to help the model to ignore parts of the data if a single band has a Nan pixels. Nan pixel can represent simply missing values or water (sea, lakes, rivers); in both the cases it must be excluded during the loss calculation: that's the main purpose of the masks. The generated valid mask constitute the ninth and last band of our patches.

In addition to the landslide ground truth, a Valid Mask is generated for each patch. It consists in a binary mask that identifies pixels that have missing values (NaN) in any of the input InSAR bands.

In our study these gaps occur in radar images for two primary reasons:

- Natural features, in particular water bodies such as lake, rivers or the sea.
- Geometric artifacts, that can occur in case of unstable radar signal.

The primary purpose of the Valid Mask is to act as a spatial filter during the loss calculation phase. By masking out these pixels, the model is prevented from learning from non-informative data, which would otherwise introduce significant noise and bias. The generated Valid Mask is integrated as the ninth and final band of the multi-channel patch. An example of is illustrated in Figure 5.1, which displays a representative sample of the 9-band multi-channel patches. In particular, it is visible how the valid mask is employed to exclude the sea surface during the loss calculation.

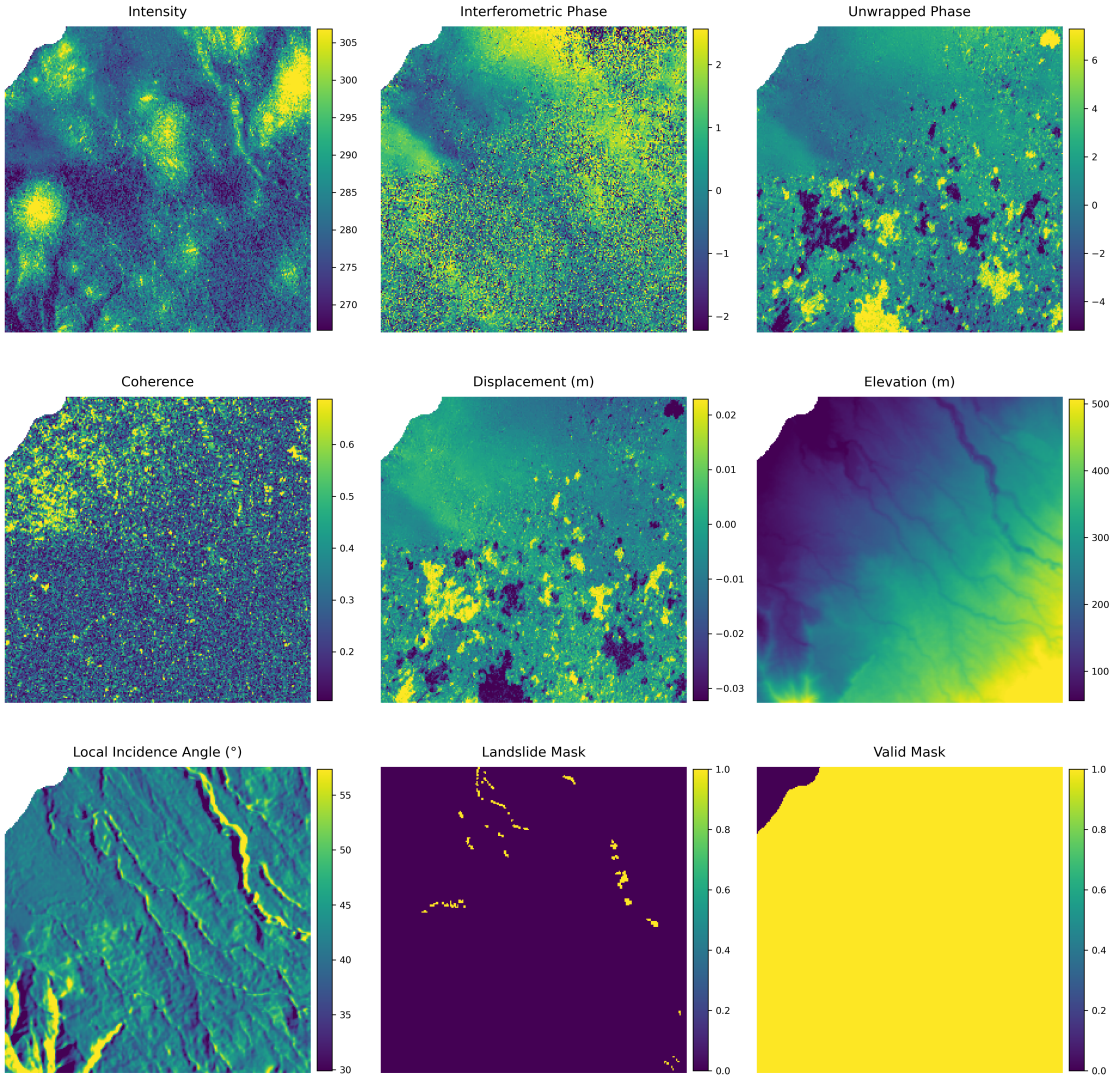

Figure 5.1: Lombok, Indonesia. The first seven channels represent the InSAR features, while the eighth and ninth channels provide the landslide ground truth and the valid mask, respectively.

5.2 Model Architectures

In this section we provide a brief overview of the deep learning architectures evaluated in the thesis. The following paragraphs focus on the architectural solutions implemented. Deeper theoretical considerations of these models are detailed in the Related Works chapter. The architectures were chosen to compare different approaches in semantic segmentation.

UNet with Resnet34 encoder

It constitutes the baseline and was used in the initial experiments, as it is widely adopted in landslide detection research. Its skip-connection structure effectively captures local features. In this configuration, the architecture employs a ResNet34 encoder, a convolutional backbone characterized by residual blocks that facilitate the training of deep networks.

UPerNet with ConvNeXt encoder

This architecture is included to evaluate whether a more sophisticated model can improve upon UNet. Using ConvNeXt as a backbone provides strong local feature representation while maintaining scalability and efficiency.

SegFormer with MiT encoder

SegFormer is implemented for its ability to model global context and its established effectiveness in semantic segmentation tasks. Like UNet, it uses a MiT backbone to combine the advantages of transformer-based encoders with a lightweight and versatile decoder.

5.3 Training Strategy

5.3.1 Class Imbalance

The statistical analysis of the generated dataset highlights a severe spatial sparsity of the target class. Even when considering only patches where a landslide is present, the *landslide* class remains a marginal component of the overall pixel distribution. Specifically, as shown in Figure 5.2a, the distribution of landslide pixels per patch is heavily skewed, with the vast majority of positive samples containing between 0 and 50 landslide pixels. This behavior is further confirmed by the relative analysis in Figure 5.2b, which shows that even within landslide patches, the fraction of pixels belonging to the landslide class remains consistently low, with most samples falling well below the 5% threshold. This structural sparsity will be handled and discussed in the next chapter.

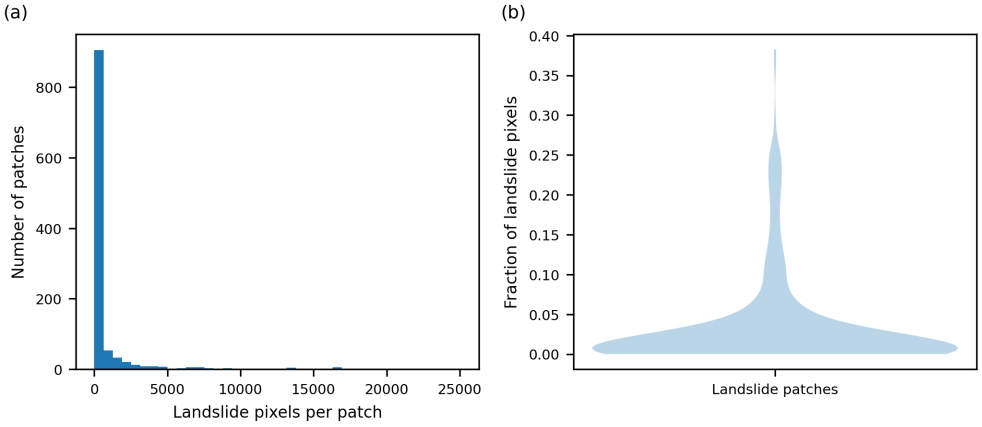

Figure 5.2: Analysis of spatial sparsity within landslide-positive patches: (a) highlights that most patches contain fewer than 50 landslide pixels, while (b) shows the fraction of landslide pixels in positive patches.

Cropping and Oversampling

To address the class imbalance described before and enhance the model's ability to detect small landslides, a cropping and oversampling strategy is implemented during the training phase. It consists in selecting patches containing landslides within the first 25th percentile (i.e. with fewer than 120 landslide pixels) among the landslide patches and, creating a new 128×128 sub-patch is extracted, centered on the bounding box of the landslide. This cropped area is then padded with zero values to restore the standard 256×256 resolution. To maintain the original context, a copy of the original full-scale patch is preserved in the dataset.

Label Smoothing

To improve the model's performance, particularly in detecting small landslide events, a label smoothing technique is applied to the ground truth masks. Instead of using discrete binary labels (0 or 1), the masks are converted into continuous target values. This process involves:

- Preserving core landslide pixels: Pixels definitively identified as landslides retain their target value of 1.0.
- Smoothing surrounding areas: Pixels in the immediate vicinity of a landslide are assigned continuous values between 0.0 and 1.0, following a Gaussian decay based on their spatial distance from the landslide boundary. This is achieved by computing the Euclidean distance transform from the landslide pixels and then applying a Gaussian function to these distances.

With this approach, loss function penalizes the model less severely for predicting values spatially close to the ground truth label. An example of a smoothed label map is illustrated in Figure 5.3.

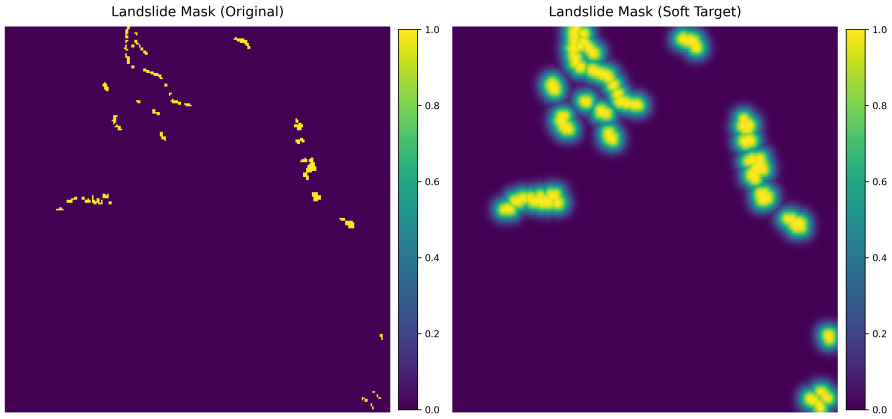

Figure 5.3: Lombok, Indonesia. Example of label smoothing applied to a binary landslide mask. (a) Original binary mask with discrete values (0 or 1). (b) Smoothed mask with continous values.

Bucket Sampler

To ensure that each training mini-batch contains a representative and consistent distribution of landslide sizes, a Bucket Sampling strategy is implemented. This technique helps the model to not see too many background-only patches and not to be biased toward a specific landslide scale during a single optimization step. The

dataset is partitioned into four distinct "buckets" based on the number of landslide pixels within each patch:

- No-landslide: Patches with 0 landslide pixels.
- Small: Patches with a low pixel count (i.e. following the 25th percentile criteria discussed before).
- Medium: Patches with landslide pixels from 25th to 75th percentile.
- Large: Patches containing extensive landslide features (i.e. top 25th percentile)

During training, the WeightedRandomSampler function from the PyTorch library is used to draw samples according to a predefined distribution $(0.1, 0.5, 0.3, 0.1)$. By assigning higher weights to small and medium landslides, the sampler forces the model to encounter these challenging cases more frequently, effectively balancing the proportion of classes within every batch and leading to more stable convergence.

Data Augmentations

To enhance the model's generalization capabilities and prevent overfitting, a data augmentation pipeline was applied to the training patches. The augmentations are implemented using the Albumentations library and restricted to geometric transformations, in particular Horizontal Flip, Vertical Flip and Random 90-degree Rotations.

The decision to limit the augmentation to geometric shifts is due to the context of SAR interferometry. These geometric operations allow the model to become invariant to the orientation of landslides and the specific slope direction, effectively increasing the diversity of the training set without altering radar signal characteristics.

Loss function

The optimization of the model is driven by the Binary Cross-Entropy (BCE) loss function. This choice is particularly effective when paired with the Label Smoothing strategy previously described, as it allows the model to process the continuous target values (soft labels) generated for the landslide boundaries, rather than being restricted to a rigid binary classification.

To ensure that the training process is not biased by unreliable data, the Valid Mask is integrated directly into the loss calculation. The loss is computed pixel-wise and then multiplied by the corresponding valid mask, effectively zeroing out the contribution of non-informative pixels (such as water bodies or radar shadows). This masking ensures that the gradient descent is guided exclusively by valid

land-based features, preventing the network from attempting to minimize error on noise or missing values.

The total loss L for a given patch can be formally expressed as:

$$L = -\frac{1}{N_v} \sum_{i \in \text{valid}} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

where $\hat{y}_i$ is the predicted probability, y_i is the soft target value, and N_v is the number of valid pixels in the patch.

5.3.2 Experimental setup and model evalution

Data splitting

To evaluate the model's performance, the dataset is partitioned into three independent subsets: Training (70%), Validation (20%), and Testing (10%).

The cropping and oversampling strategy described above are applied to the training set only after the initial split. This ensures that the validation and testing sets consist only of original tiles, providing a realistic scenario for model evaluation. After the oversampling of small-scale landslides, the training set was expanded from its initial count of 768 patches to a final total of 830 patches.

The final distribution of the patches across the different sets is summarized in Table 5.1.

Dataset Split	**Number of Patches**
Training (Pre-Oversampling)	768
Training (Post-Oversampling)	830
Validation	220
Testing	110

Table 5.1: Distribution of patches across the training, validation, and test sets.

Optimal Threshold Selection

The output of the deep learning architectures is a probability map, where at each pixel is assigned a value between 0 and 1 representing the likelihood of belonging to the landslide class. While a standard threshold of 0.5 is applied during the training phase, this default it's not optimal for the evaluation inference. To address this, a threshold optimization procedure is performed over the validation set across a set of potential values; the best threshold is identified over a set of potential values, selecting the one that maximizes the Intersection over Union (IoU) metric. Once

identified, this optimal threshold is fixed and kept also for the final testing phase on unseen data.

5.4 Evaluation Metrics

To quantitatively assess the performance of the segmentation models, we rely on standard pixel-level metrics. In the context of landslide detection, the "positive" class represents the landslide-affected pixels, while the "negative" class represents the background. Based on True Positive (TP), False Positive (FP) and False Negatives (FN) components, the following metrics are calculated:

- Precision: it measures the accuracy of the positive predictions, representing the fraction of predicted landslide pixels that are actually landslides.

$$Precision = \frac{TP}{TP + FP}$$

- Recall: it measures the ability of the model to find all the landslide pixels within the ground truth.

$$Recall = \frac{TP}{TP + FN}$$

- F1: it is the harmonic mean of Precision and Recall. It provides a single score that balances both metrics, which is particularly useful when dealing with the class imbalance typical of landslide datasets.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$

- Intersection over Union: it is the standard metric for semantic segmentation. It measures the overlap between the predicted segmentation mask (A) and the ground truth mask (B). It is a more stringent metric than the F1-score as it penalizes single-class errors more heavily.

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN}$$

Chapter 6

Results

This chapter presents the experimental results obtained from the landslide detection task using InSAR data. First, the hardware and software configurations and the training hyperparameters are detailed. Then, a comparative analysis is performed to evaluate different deep learning paradigms, from traditional CNNs to modern Transformer-based architectures. Finally, the chapter investigates the impact of Label Smoothing to improve detection performance on small-scale and sparse deformation patterns, providing both quantitative metrics and qualitative visual assessments.

6.1 Experimental Setup

Software and Hardware Environment The experimental framework is implemented in PyTorch, leveraging the *Segmentation Models PyTorch* (SMP) library for efficient instantiation of the segmentation architectures (SegFormer, U-Net, and UPerNet). All experiments were conducted on an NVIDIA GeForce RTX 2080 Ti GPU equipped with 11 GB of VRAM.

Input data were provided in NetCDF (.nc) format in order to preserve the physical properties of the interferometric products. Interferograms were divided into patches of size 256×256 pixels. A batch size of 32 was adopted to ensure a balance between memory constraints and training stability.

Training Strategy All models were trained for 100 epochs using the AdamW optimizer. The initial learning rate was set to 5e-5. A Cosine Annealing learning rate scheduler was employed to progressively decrease the learning rate following a cosine decay schedule.

6.2 Architectures Performance Comparison

To evaluate the effectiveness of different paradigms for InSAR-based landslide detection, a comparative analysis was conducted using U-Net with Resnet34 as traditional CNN baseline, UperNet with ConvNextTiny as modern convolutional architecture, and Segformer with Mit-b2 and Mit-b3 as transformer-based models. The quantitative results, evaluated on the unseen test set, are summarized in Table 6.1.

The CNN baseline (U-Net with ResNet34) achieved low performance due to the noisy nature of InSAR data, collapsing to an IoU of 0.042 and a low Recall of 0.056. This highlights the limitation of standard convolutional kernels, which struggle to separate broad phase deformation patterns from high-frequency atmospheric artifacts. On the other hand, the UperNet architecture, driven by the transformer-inspired ConvNeXt2-Tiny backbone, successfully outperformed the baseline metrics by leveraging large-kernel convolutions and complex multi-scale aggregation.

However, the attention-based paradigm demonstrated the highest effectiveness, with the SegFormer architecture establishing the optimal solution. Within this type of architectures, the results expose a trade-off between model capacity and prediction purity. The MiT-B2 encoder achieved the best overall balance, recording the highest IoU (0.129), F1-score (0.228), and Precision (0.187). Scaling the architecture up to MiT-B3 maximized the network's sensitivity—yielding the absolute highest Recall (0.338)—but this came at the cost of severe over-segmentation. The resulting drop in Precision (0.145) ultimately penalized the overall geometric accuracy (IoU 0.113), confirming that for the considered dataset and noise regime, the MiT-B2 configuration represents a "sweet spot" between global context understanding and spatial precision. These findings suggest that hierarchical attention mechanisms are particularly effective in modeling large-scale deformation patterns while preserving spatial coherence in highly noisy InSAR data.

The quantitative results are reflected by the qualitative visual analysis presented in Figure 6.1; in this example, the Porgera area in Papua New Guinea is analysed. It can be noted that the standard U-Net almost completely fails to identify the phenomenon, resulting in massive False Negatives (orange areas). UperNet manages to detect fragmented portions of the landslide but lacks spatial continuity. In contrast, both SegFormer models (mit-b2 and mit-b3) successfully capture the macroscopic shape of the phenomenon. The improved spatial coherence suggests that global context helps model predictions, increasing true Positives (green areas) and tracing the complex boundaries much more accurately.

Architecture	**Encoder**	**Params**	**IoU**	**F1**	**Recall**	**Precision**
U-Net	ResNet34	24.4M	0.042	0.081	0.056	0.142
UperNet	ConvNeXt2-T	37.1M	0.107	0.188	0.212	0.139
SegFormer	MiT-B2	**24.7M**	**0.129**	**0.228**	0.291	**0.187**
SegFormer	MiT-B3	44.6M	0.113	0.203	**0.338**	0.145

Table 6.1: Performance comparison of different architectures. "Params" column refers to the total encoder-decoder size.

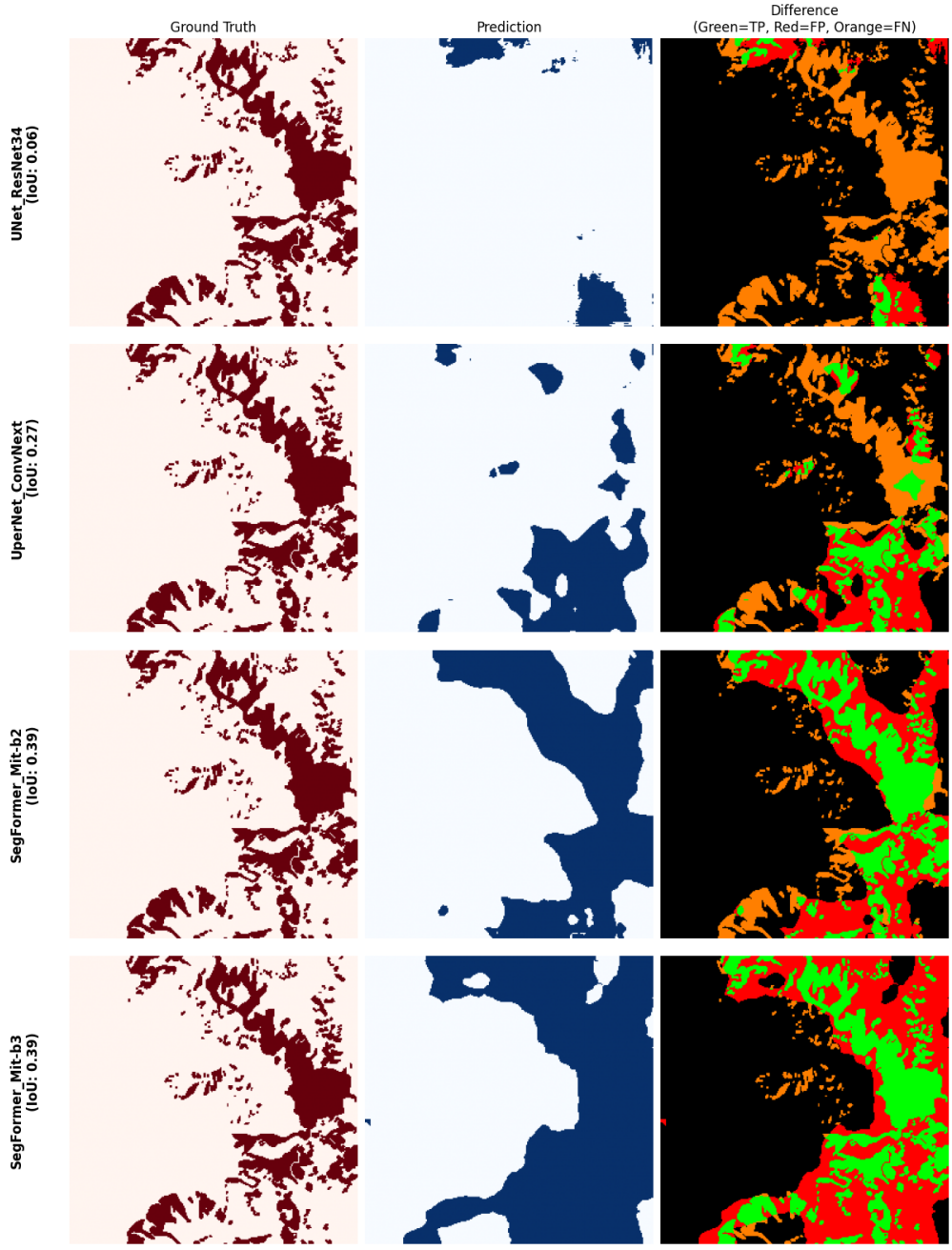

Figure 6.1: Papua New Guinea, Porgera: qualitative comparison of different architectures' predictions. In the Difference map: Green indicates True Positives, Red indicates False Positives, and Orange indicates False Negatives.

6.2.1 Label Smoothing

Following the comparative architectural analysis, the SegFormer MiT-B2 was selected as the reference baseline for further optimization, as it demonstrated a compromise between global context understanding, overall performance, and computational efficiency. Despite the observed superiority of the Transformer-based configurations, detecting small and highly sparse landslide phenomena remains a critical challenge. To address this, we applied Label Smoothing (discussed in 5.3.1) to the training process. By softening the hard target distributions, this technique helps the model to identify small deformation patterns that would otherwise be ignored.The quantitative impact of varying the smoothing factor (σ_{label}) is detailed in Table 6.2. Setting a moderate smoothing factor ($\sigma\_label = 3$) forces the model into a conservative regime. In fact, while this configuration achieves the highest absolute Precision (0.299), it penalizes a Recall drop to 0.249 compared to the baseline. On the other hand, increasing the smoothing factor to $\sigma_{label} = 5$ brings the best overall performance. This configuration achieved the highest geometric accuracy (IoU 0.175) and overall balance (F1-score 0.298). It also boosts the Recall to 0.332, indicating that the model successfully recovers a significant portion of the previously missed deformations, maintaining a good Precision (0.270). The quantitative trends are illustrated in the Figure 6.2, that represent the Emilia Romagna area, characterized by fragmented and sparse landslides. It's easy to observe that Segformer configuration with hard labels missed almost the entire phenomenon (massive False Negatives, in orange). In this case the $\sigma_{label} = 3$ configuration performed worse, predicting no deformation at all. Finally, the $\sigma_{label} = 5$ model successfully detects multiple true deformation spots (True Positives, in green). While this increased sensitivity inevitably generates a broader, over-estimated boundary (False Positives, in red), it proves that label smoothing factor can prevent the model from ignoring challenging and sparse landslides, making it a much more viable configuration for hazard screening.

Architecture	**Encoder**	σ_{label}	**IoU**	**F1**	**Recall**	**Precision**
SegFormer	MiT-B2	0	0.129	0.228	0.291	0.187
SegFormer	MiT-B2	3	0.157	0.271	0.249	**0.299**
SegFormer	MiT-B2	5	**0.175**	**0.298**	**0.332**	0.270

Table 6.2: Quantitative impact of Label Smoothing (σ_{label}) on the SegFormer MiT-B2 architecture.

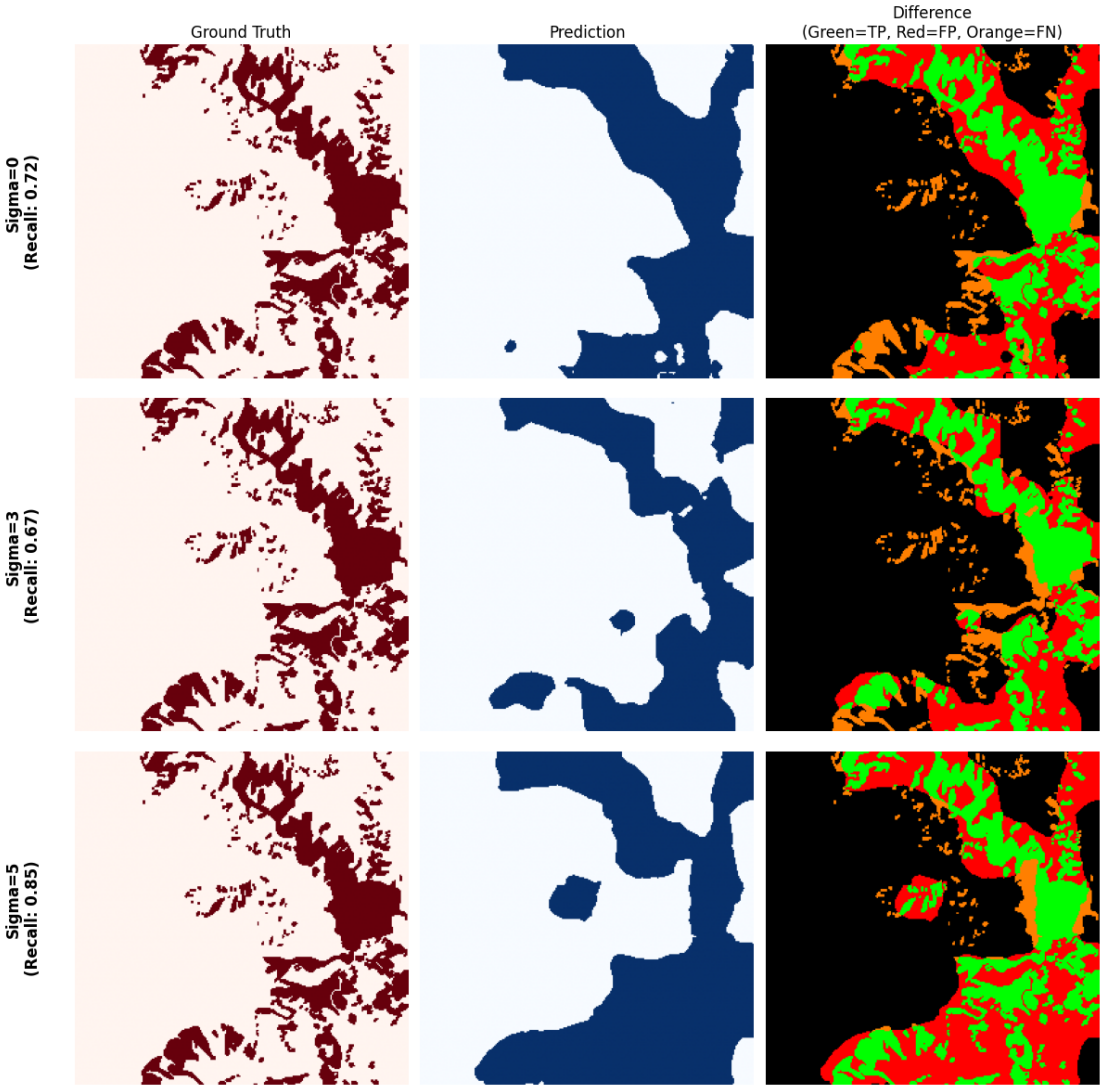

Figure 6.2: Papua New Guinea, Porgera: qualitative effect of Label Smoothing on a region with sparse landslides. In the Difference map: Green=True Positives, Red=False Positives, Orange=False Negatives.

Impact of Label Smoothing by Landslide Size

To understand the operational behavior of the models, the performance metrics were stratified according to the true pixel area of the landslides. The dataset was divided into three classes (Small, Medium, and Large) based on the 33rd and 66th percentiles of the landslide size distribution. The resulting analysis is presented in Figure 6.3.

Detecting small-scale ground deformations remains a fundamental challenge in Earth Observation. As shown in the charts, metrics for Small (< 156 px) and Medium (156-1043 px) landslides are close to zero across all configurations. This behavior is due to the highly decorrelated InSAR data, because small phenomena are more affected by background noise and atmospheric artifacts. However, the impact of Label Smoothing becomes evident. Applying a smoothing factor of 5, the absolute value remains low, while the Recall for small landslides goes from 0.004 to 0.022. This represents a more than five-fold relative improvement, suggesting that softening the target distribution helps the model to recover sparse deformation that the baseline ignores.

The stratified analysis also demonstrates that the global metrics presented in previous sections are driven by large landslides (> 1043 px). For these macro-phenomena, the signal is sufficiently extended to be consistently captured by the model, which is consistent with the ability of hierarchical attention-based architectures to model long-range spatial dependencies. In this context, the optimal smoothing configuration (factor of 5) further improves the baseline, maximizing the Recall (0.383) and the F1-Score (0.338), while maintaining the highest IoU (0.204). Conversely, the intermediate smoothing level (factor of 3) achieves a peak in Precision for large targets (0.406), but collapses on small and medium targets and loses overall sensitivity.

Ultimately, this analysis confirms that a higher label smoothing factor (level 5) is the most effective strategy to reduce the model's tendency to ignore small-scale deformations, while boosting the detection of large-scale deformations.

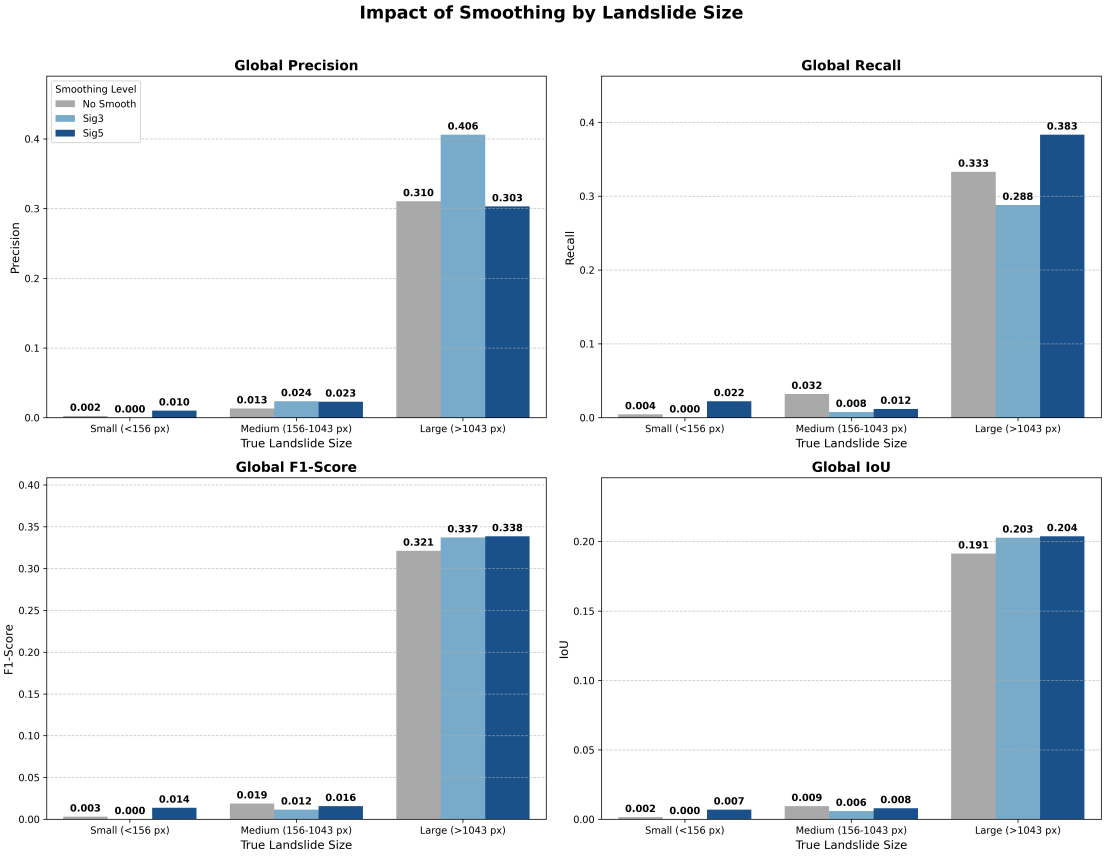

Figure 6.3: Performance metrics (Precision, Recall, F1-Score, and IoU) stratified by landslide size (Small, Medium, Large) calculated using the 33rd and 66th percentiles of the pixel area distribution.

Chapter 7

Conclusions

This thesis investigated the challenging task of automated landslide detection using Interferometric Synthetic Aperture Radar (InSAR) data through advanced deep learning architectures. The experimental results highlight the intrinsic complexity of the problem, particularly in the presence of highly decorrelated and noisy signals typical of small-scale deformations.

The empirical analysis showed that traditional convolutional architectures such as U-Net struggle to accurately reconstruct complex displacement patterns, likely due to their limited capacity to model long-range spatial dependencies. In contrast, hierarchical attention-based architectures demonstrated a superior ability to preserve spatial coherence and capture large-scale deformation structures, resulting in improved segmentation performance.

Overall, the study confirms the potential of combining multi-channel InSAR representations with modern deep learning architectures for automatic landslide recognition, while also emphasizing the limitations imposed by data sparsity and class imbalance.

7.1 Our Contributions

The contributions of this work are articulated across three primary pillars. First, a processing pipeline was developed to transform raw Sentinel-1 SAR acquisitions into a structured multi-channel InSAR dataset; by integrating phase, coherence, displacement, elevation, and local incidence angle, the models were provided with a physically meaningful representation of the geophysical processes under study. Second, an architectural benchmarking was conducted to evaluate the performance of diverse segmentation paradigms, including traditional CNNs, modern convolutional architectures, and transformer-based models. This analysis identified the

SegFormer configuration with an MiT-B2 backbone as the most balanced solution, capable of achieving superior spatial coherence and competitive performance across all evaluation metrics. Finally, the research investigated the impact of label smoothing as a strategy to mitigate the near-zero recall observed for small and sparse landslide events. The analysis indicated that a higher smoothing factor improves performance for small-scale phenomena while preserving robust results on large-scale deformations, providing a more favorable operational trade-off for hazard screening applications.

7.2 Future Works

While this study provides an empirical evaluation of deep learning architectures for InSAR-based landslide detection on different regions, several research directions remain open for further exploration. An ablation analysis should be conducted to quantify the specific contribution of each input channel (phase, coherence, displacement, elevation, and incidence angl) as well as to evaluate the individual impact of architectural components like backbone depth and decoder structure, which would provide deeper insight into the features and model elements that most strongly influence segmentation performance. Furthermore, extending the dataset with additional landslide events from different topographic and climatic regions could enhance statistical reliability and enable a more comprehensive evaluation of model generalization across heterogeneous geological contexts. Given the limited availability of labeled landslide inventories, future work could also explore the generation of synthetic interferograms, with the aim of complementing real data and improving robustness across diverse geomorphological settings. Finally, investigating alternative model architectures could be interesting to study whether further gains in spatial precision and noise robustness can be achieved for the task of landslide recognition.

Bibliography

[1] Fausto Guzzetti, Alessandro Cesare Mondini, Mauro Cardinali, Federica Fiorucci, Michele Santangelo, and Kang-Tsung Chang. «Landslide inventory maps: New tools for an old problem». In: *Earth-Science Reviews* 112.1 (2012), pp. 42–66. ISSN: 0012-8252. DOI: https://doi.org/10.1016/j.earscirev.2012.02.001. URL: https://www.sciencedirect.com/science/article/pii/S0012825212000128 (cit. on p. 4).

[2] Yusen Cheng, Hongli Pang, Yangyang Li, Lei Fan, Shengjie Wei, Ziwen Yuan, and Yinqing Fang. «Applications and Advancements of Spaceborne InSAR in Landslide Monitoring and Susceptibility Mapping: A Systematic Review». In: *Remote Sensing* 17.6 (2025). ISSN: 2072-4292. DOI: 10.3390/rs17060999. URL: https://www.mdpi.com/2072-4292/17/6/999 (cit. on p. 4).

[3] A. Ferretti, C. Prati, and F. Rocca. «Permanent scatterers in SAR interferometry». In: *IEEE Transactions on Geoscience and Remote Sensing* 39.1 (2001), pp. 8–20. DOI: 10.1109/36.898661 (cit. on p. 4).

[4] Alessandro Ferretti, Alfio Fumagalli, Fabrizio Novali, Claudio Prati, Fabio Rocca, and Alessio Rucci. «A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR». In: *IEEE Transactions on Geoscience and Remote Sensing* 49.9 (2011), pp. 3460–3470. DOI: 10.1109/TGRS.2011.2124465 (cit. on p. 4).

[5] Rosi et al. *The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: geomorphological features and landslide distribution*. 2018. DOI: 10.1109/36.898661 (cit. on p. 5).

[6] B. Bayer, A. Simoni, D. Schmidt, and L. Bertello. «Using advanced InSAR techniques to monitor landslide deformations induced by tunneling in the Northern Apennines, Italy». In: *Engineering Geology* 226 (2017), pp. 20–32. ISSN: 0013-7952. DOI: https://doi.org/10.1016/j.enggeo.2017.03.026. URL: https://www.sciencedirect.com/science/article/pii/S0013795216303787 (cit. on p. 5).

[7] Nikhil Prakash, Andrea Manconi, and Simon Loew. «Mapping Landslides on EO Data: Performance of Deep Learning Models vs. Traditional Machine Learning Models». In: *Remote Sensing* 12.3 (2020). ISSN: 2072-4292. DOI: 10.3390/rs12030346. URL: https://www.mdpi.com/2072-4292/12/3/346 (cit. on p. 5).

[8] L. Bragagnolo, L.R. Rezende, R.V. da Silva, and J.M.V. Grzybowski. «Convolutional neural networks applied to semantic segmentation of landslide scars». In: *CATENA* 201 (2021), p. 105189. ISSN: 0341-8162. DOI: https://doi.org/10.1016/j.catena.2021.105189. URL: https://www.sciencedirect.com/science/article/pii/S0341816221000485 (cit. on p. 5).

[9] Wenwen Qi, Mengfei Wei, Wentao Yang, Chong Xu, and Chao Ma. «Automatic Mapping of Landslides by the ResU-Net». In: *Remote Sensing* 12.15 (2020). ISSN: 2072-4292. DOI: 10.3390/rs12152487. URL: https://www.mdpi.com/2072-4292/12/15/2487 (cit. on p. 5).

[10] Zhipeng Wu, Teng Wang, Yingjie Wang, Robert Wang, and Daqing Ge. «Deep Learning for the Detection and Phase Unwrapping of Mining-Induced Deformation in Large-Scale Interferograms». In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–18. DOI: 10.1109/TGRS.2021.3121907 (cit. on p. 6).

[11] Sachin Vijay Kumar, Xinyao Sun, Zheng Wang, Ryan Goldsbury, and Irene Cheng. «A U-Net Approach for InSAR Phase Unwrapping and Denoising». In: *Remote Sensing* 15.21 (2023). ISSN: 2072-4292. DOI: 10.3390/rs15215081. URL: https://www.mdpi.com/2072-4292/15/21/5081 (cit. on p. 6).

[12] Ximing Chen, Xin Yao, Zhenkai Zhou, Yang Liu, Chuangchuang Yao, and Kaiyu Ren. «DRs-UNet: A Deep Semantic Segmentation Network for the Recognition of Active Landslides from InSAR Imagery in the Three Rivers Region of the Qinghai–Tibet Plateau». In: *Remote Sensing* 14.8 (2022). ISSN: 2072-4292. DOI: 10.3390/rs14081848. URL: https://www.mdpi.com/2072-4292/14/8/1848 (cit. on p. 6).

[13] Mahmoud Abdallah, Samaa Younis, Songbo Wu, and Xiaoli Ding. «Automated deformation detection and interpretation using InSAR data and a multi-task ViT model». In: *International Journal of Applied Earth Observation and Geoinformation* 128 (2024), p. 103758. ISSN: 1569-8432. DOI: https://doi.org/10.1016/j.jag.2024.103758. URL: https://www.sciencedirect.com/science/article/pii/S1569843224001122 (cit. on p. 6).

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV]. URL: https://arxiv.org/abs/1505.04597 (cit. on p. 7).

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV]. URL: https://arxiv.org/abs/1512.03385 (cit. on p. 8).

[16] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. *Unified Perceptual Parsing for Scene Understanding*. 2018. arXiv: 1807.10221 [cs.CV]. URL: https://arxiv.org/abs/1807.10221 (cit. on pp. 8, 9).

[17] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. *Feature Pyramid Networks for Object Detection*. 2017. arXiv: 1612.03144 [cs.CV]. URL: https://arxiv.org/abs/1612.03144 (cit. on p. 8).

[18] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. *Pyramid Scene Parsing Network*. 2017. arXiv: 1612.01105 [cs.CV]. URL: https://arxiv.org/abs/1612.01105 (cit. on p. 8).

[19] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. *A ConvNet for the 2020s*. 2022. arXiv: 2201.03545 [cs.CV]. URL: https://arxiv.org/abs/2201.03545 (cit. on p. 9).

[20] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*. 2021. arXiv: 2105.15203 [cs.CV]. URL: https://arxiv.org/abs/2105.15203 (cit. on pp. 9, 10).

[21] Richard Bamler. «Principles Of Synthetic Aperture Radar». In: *Surveys in Geophysics* 21 (Jan. 2000). DOI: 10.1023/A:1006790026612 (cit. on p. 11).

[22] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P. Papathanassiou. «A tutorial on synthetic aperture radar». In: *IEEE Geoscience and Remote Sensing Magazine* 1.1 (2013), pp. 6–43. DOI: 10.1109/MGRS.2013.2248301 (cit. on pp. 11, 15).

[23] Alessandro Ferretti, Andrea Monti-Guarnieri, Claudio Prati, and Fabio Rocca. «InSAR Principles: Guidelines for SAR Interferometry Processing and Interpretation». In: *ESA Publications* (Feb. 2007) (cit. on pp. 14, 15).

[24] Richard Bamler and Philip Hartl. «Synthetic Aperture Radar Interferometry.» In: *Inverse Problems* 14 (Jan. 1998) (cit. on p. 15).

[25] P.A. Rosen, S. Hensley, I.R. Joughin, F.K. Li, S.N. Madsen, E. Rodriguez, and R.M. Goldstein. «Synthetic aperture radar interferometry». In: *Proceedings of the IEEE* 88.3 (2000), pp. 333–382. DOI: 10.1109/5.838084 (cit. on p. 15).

[26] Kevin Jones. *Advanced SAR Training*. https://www.udemy.com/course/sar-course/ (cit. on pp. 16, 18).

[27] Ammar Jasim. *Study of Earthquake and flood hazards of Khanaqin-Abbasan NE Iraq using Remote Sensing and GIS Techniques*. Jan. 2021. DOI: 10.13140/RG.2.2.16883.86565 (cit. on p. 17).

[28] Michael Foumelis, Luis Veci, Nick Selmes, and Jose Manuel Delgado Blasco. *S1TBX TOPSAR Interferometry with Sentinel-1 Tutorial*. Version 2.0. European Space Agency (ESA). 2020. URL: https://step.esa.int/docs/tutorials/S1TBX%20TOPSAR%20Interferometry%20with%20Sentinel-1%20Tutorial_v2.pdf (cit. on pp. 19, 24).

[29] European Space Agency. *SentiWiki - Sentinel-1 Mission Overview*. https://sentiwiki.copernicus.eu/web/sentinel-1. Accessed: 2024-05-24. 2024 (cit. on p. 20).

[30] Ramon Torres et al. «GMES Sentinel-1 mission». In: *Remote Sensing of Environment* 120 (2012). The Sentinel Missions - New Opportunities for Science, pp. 9–24. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2011.05.028. URL: https://www.sciencedirect.com/science/article/pii/S0034425712000600 (cit. on p. 20).

[31] European Space Agency. *SNAP Online Help*. Accessed via ESA STEP platform. 2024. URL: https://step.esa.int/main/doc/online-help/ (cit. on p. 25).

[32] Senbox-org. *esa-snappy: SNAP Python Interface*. 2024. URL: https://github.com/senbox-org/esa-snappy (cit. on p. 25).

[33] Curtis W. Chen and Howard A. Zebker. *SNAPHU: Statistical-cost, Network-flow Algorithm for Phase Unwrapping - User's Manual*. Stanford University. 2002. URL: https://web.stanford.edu/group/radar/softwareandlinks/sw/snaphu/snaphu_man1.html (cit. on p. 25).

[34] European Space Agency. *Copernicus Data Space Ecosystem: OData API Documentation*. 2024. URL: https://documentation.dataspace.copernicus.eu/APIs/OData.html (cit. on p. 25).