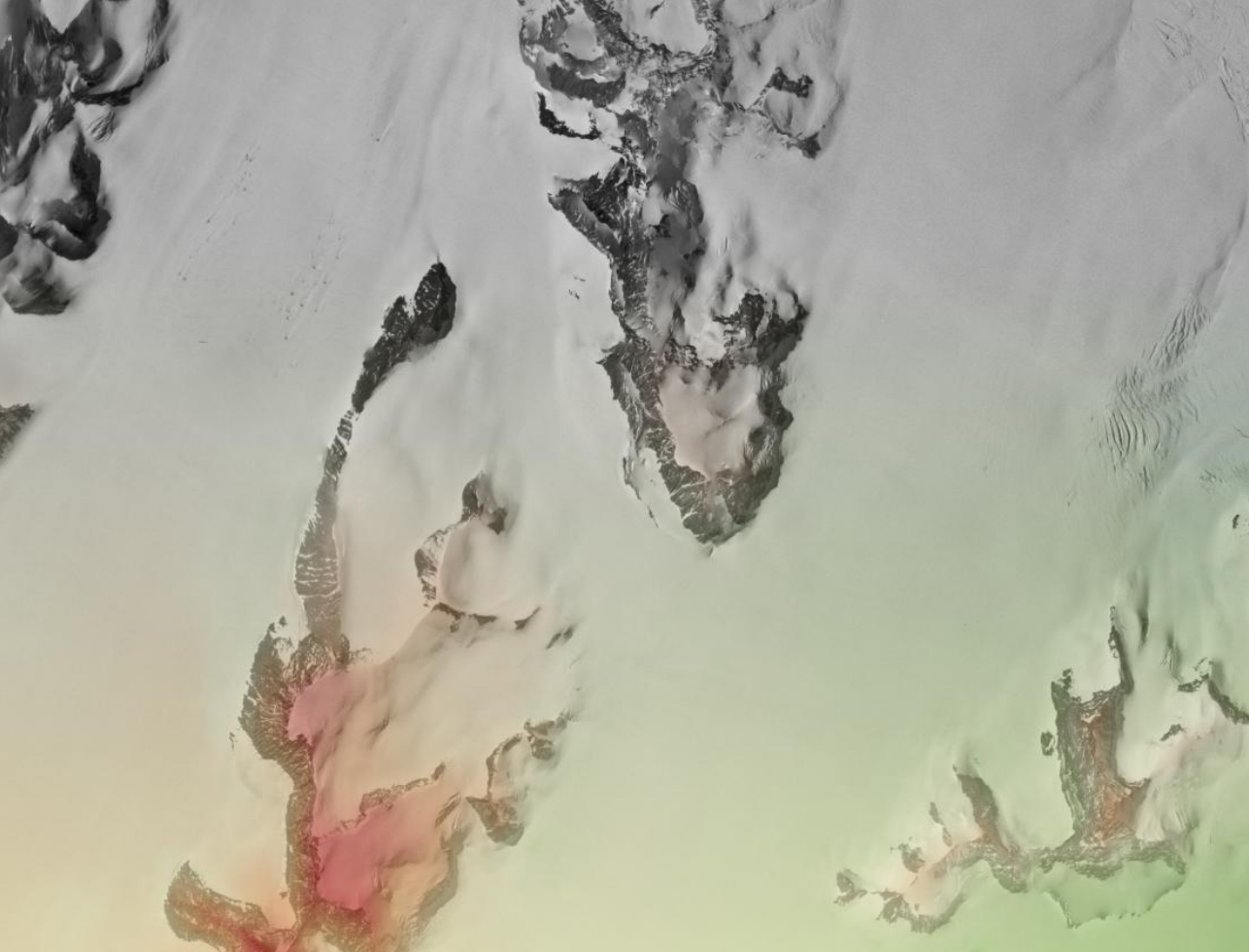


Antarctic Time Machine

3D Reconstruction of Glaciers
in the Antarctic Peninsula using
Historical Structure-from-Motion

Felix Dahle

Antarctic Time Machine

3D Reconstruction of Glaciers in the Antarctic Peninsula using Historical Structure-from-Motion

Antarctic Time Machine

3D Reconstruction of Glaciers in the Antarctic Peninsula using Historical Structure-from-Motion

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus, Prof.dr.ir. H. Bijl,

Chair of the Board for Doctorates

to be defended publicly on

Thursday, 5 March 2026 at 15:00 o'clock

by

Felix-Heinrich DAHLE

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Dr. R.C. Lindenbergh,	Delft University of Technology, *promotor*
Dr.ir. B. Wouters,	Delft University of Technology, *promotor*
Independent members:	
Prof.dr.ir. P.J.M. van Oosterom,	Delft University of Technology
Prof.Dr.-Ing. A. Eltner,	Dresden University of Technology, Germany
Dr. A.A. Bjørk,	Københavns Universitet, Denmark
Dr. S.F. Child,	University at Albany, USA
Prof.dr.ir. R.F. Hansen	Delft University of Technology

Keywords: Antarctica, Historical Imagery, Machine Learning, Tie point matching, Photogrammetry, Structure-from-Motion

Printed by: Ridderprint | www.ridderprint.nl

Cover: Historical Orthophoto and DEM based on images from the TMA archive, displaying Stubb Glacier on the Antarctic Peninsula.

ISBN 978-94-6384-912-8

An electronic copy of this dissertation is available at https://repository.tudelft.nl/.

It is well known that a vital ingredient of success is not knowing that what you're attempting can't be done.

— Terry Pratchett, *Equal Rites*

CONTENTS

SUMMARY

This thesis presents a fully automated framework for transforming mid-20th-century aerial photographs from the U.S. Trimetrogon Aerial (TMA) archive into geospatial datasets reconstructing past glacier elevations. Although the TMA imagery provides a unique and extensive record of Antarctic glacier conditions, it has remained largely underused due to its analogue format, degraded image quality, and the high manual effort typically required for processing. Previous efforts have relied heavily on manual digitisation and expert intervention. This work introduces a modular, end-to-end pipeline that automates the entire reconstruction process, combining semantic segmentation, metadata extraction, geo-referencing, and Structure-from-Motion (SfM) photogrammetry.

The workflow begins with semantic segmentation using a custom-trained U-Net model, which classifies pixels in degraded grayscale aerial images into six categories: snow, ice, water, rock, clouds, and sky. Despite limited training data (100 manually labelled images) and challenges such as low contrast and artefacts, the model achieves an overall accuracy of 73% and an F1-score of 71%. By masking out unusable regions such as sky and ocean, this step significantly improves the reliability of the photogrammetric reconstruction.

An automated metadata extraction module complements the segmentation by retrieving key parameters, including focal length, altitude, and fiducial marker positions, directly from the images. Using a combination of optical character recognition and computer vision techniques, it recovers essential information and estimates missing values by exploiting redundancy across flight series. This reduces the need for manual transcription and converts handwritten image annotations into structured digital formats.

The geo-referencing component establishes a spatial link between historical images and modern coordinate systems. It uses LightGlue, a recent deep-learning-based matching algorithm, along with a progressive tiling strategy adapted to the characteristics of historical imagery. By matching tie points between the TMA scans and Sentinel-2 satellite imagery, the system automatically generates ground control points (GCPs) with

positional accuracies of just a few meters, therefore dramatically improving upon the original, often kilometre-scale geolocation estimates.

In the final stage, the segmented images, extracted metadata, and GCPs are automatically passed to Agisoft Metashape, which is integrated into the processing pipeline via its Python API. This stage performs Structure-from-Motion photogrammetry to generate dense point clouds, orthophotos, and digital elevation models (DEMs) without user interaction. Applied across the Antarctic Peninsula, the pipeline successfully reconstructed 3D glacier surfaces for 49 glacier systems. Validation against the high-resolution Reference Elevation Model of Antarctica (REMA) shows median elevation differences of approximately 90 meters across full glacier extents and 76 meters in topographically stable areas.

While the outputs do not yet match the accuracy of fully manual processing, the developed system enables large-scale, repeatable reconstruction of historical glacier surfaces at a scale previously unattainable. By combining all components into a modular, end-to-end framework, this work makes the TMA archive broadly accessible for contemporary cryospheric research and extends observational baselines by over half a century. All code and workflows are openly available on GitHub, and the resulting data products, including semantic masks, metadata tables, geo-referenced image positions, and 3D glacier models, are publicly released to support further scientific use.

SAMENVATTING

Dit proefschrift presenteert een volledig geautomatiseerd raamwerk voor het transformeren van luchtfoto's uit het Amerikaanse Trimetrogon Aerial (TMA) archief uit het midden van de 20e eeuw in geospatiale datasets die de reconstructie van vroegere gletsjerhoogtes mogelijk maken. Hoewel de TMA-beelden een uniek en uitgebreid overzicht geven van de toestand van de Antarctische gletsjers, zijn ze grotendeels onderbenut gebleven vanwege het analoge formaat, de verminderde beeldkwaliteit en de hoge handmatige inspanning die meestal nodig is voor de verwerking. Eerdere inspanningen waren sterk afhankelijk van handmatige digitalisering en interventie door experts. Dit werk introduceert een modulaire, end-to-end pijplijn die het gehele reconstructieproces automatiseert door semantische segmentatie, metadata-extractie, georeferencing en Structure-from-Motion (SfM) fotogrammetrie te combineren.

De workflow begint met semantische segmentatie met behulp van een speciaal getraind U-Net model, dat pixels in gedegradeerde grijswaarden luchtfoto's classificeert in zes categorieën: sneeuw, ijs, water, rotsen, wolken en lucht. Ondanks beperkte trainingsgegevens (100 handmatig gelabelde beelden) en uitdagingen zoals laag contrast en artefacten, behaalt het model een algehele nauwkeurigheid van 73% en een F1-score van 71%. Door onbruikbare gebieden zoals lucht en oceaan af te schermen, verbetert deze stap de betrouwbaarheid van de fotogrammetrische reconstructie aanzienlijk.

Een geautomatiseerde metadata-extractiemodule vult de segmentatie aan door belangrijke parameters, zoals brandpuntsafstand, hoogte en fiduciaire markeerlijnposities, rechtstreeks uit de beelden te halen. Door gebruik te maken van een combinatie van optische tekenherkenning en computervisietechnieken wordt essentiële informatie teruggewonnen en worden ontbrekende waarden geschat door gebruik te maken van redundantie in de verschillende vluchtreeksen. Dit vermindert de noodzaak voor handmatige transcriptie en zet handgeschreven beeldannotaties om in gestructureerde digitale formaten.

De georeferentiecomponent legt een ruimtelijke link tussen historische beelden en moderne coördinatensystemen. Het maakt gebruik van LightGlue, een recent algoritme voor matching op basis van deep learning,

samen met een progressieve tiling-strategie die is aangepast aan de kenmerken van historische beelden. Door verbindingspunten tussen de TMA-scans en Sentinel-2 satellietbeelden te matchen, genereert het systeem automatisch grondcontrolepunten (GCP's) met een positienauwkeurigheid van slechts enkele meters, waardoor de oorspronkelijke, vaak kilometers grote geolocatieschattingen aanzienlijk worden verbeterd.

In de laatste fase worden de gesegmenteerde beelden, geëxtraheerde metadata en GCP's automatisch doorgegeven aan Agisoft Metashape, dat in de verwerkingspijplijn is geïntegreerd via zijn Python API. Deze fase voert Structure-from-Motion fotogrammetrie uit om dichte puntenwolken, orthofoto's en digitale hoogtemodellen (DEM's) te genereren zonder interactie van de gebruiker. Toegepast op het Antarctisch schiereiland reconstrueerde de pijplijn met succes 3D-gletsjeroppervlakken voor 49 gletsjersystemen. Validatie tegen het hoge-resolutie referentiehoogtemodel van Antarctica (REMA) toont mediane hoogteverschillen van ongeveer 90 meter over de volledige gletsjeruitbreidingen en 76 meter in topografisch stabiele gebieden.

Hoewel de resultaten nog niet de nauwkeurigheid van volledig handmatige verwerking evenaren, maakt het ontwikkelde systeem grootschalige, herhaalbare reconstructie van historische gletsjeroppervlakken mogelijk op een schaal die voorheen onbereikbaar was. Door alle componenten te combineren in een modulair, end-to-end raamwerk, maakt dit werk het TMA-archief breed toegankelijk voor hedendaags cryosferisch onderzoek en breidt het de basislijnen van observaties uit met meer dan een halve eeuw. Alle code en workflows zijn openlijk beschikbaar op GitHub, en de resulterende dataproducten, inclusief semantische maskers, metadatatabellen, geogerefereerde beeldposities en 3D gletsjermodellen, worden openbaar gemaakt om verder wetenschappelijk gebruik te ondersteunen.

ZUSAMMENFASSUNG

In dieser Arbeit wird ein vollautomatisches Framework für die Umwandlung von Luftbildern aus der Mitte des 20. Jahrhunderts aus dem US-amerikanischen Trimetrogon Aerial (TMA)-Archiv in Geodatensätze vorgestellt, die die Rekonstruktion früherer Gletscherhöhen ermöglichen. Obwohl die TMA-Bilder eine einzigartige und umfassende Aufzeichnung der Gletscherbedingungen in der Antarktis bieten, wurden sie aufgrund ihres analogen Formats, der schlechten Bildqualität und des hohen manuellen Aufwands, der typischerweise für die Verarbeitung erforderlich ist, bisher kaum genutzt. Bisherige Bemühungen stützten sich stark auf die manuelle Digitalisierung und das Eingreifen von Experten. Diese Arbeit stellt eine modulare, durchgängige Pipeline vor, die den gesamten Rekonstruktionsprozess automatisiert und semantische Segmentierung, Metadatenextraktion, Georeferenzierung und Structure-from-Motion (SfM) Photogrammetrie kombiniert.

Der Arbeitsablauf beginnt mit der semantischen Segmentierung mithilfe eines speziell trainierten U-Net-Modells, das Pixel in degradierten Graustufen-Luftbildern in sechs Kategorien klassifiziert: Schnee, Eis, Wasser, Felsen, Wolken und Himmel. Trotz begrenzter Trainingsdaten (100 manuell klassifizierte Bilder) und Herausforderungen wie niedrigem Kontrast und Artefakten erreicht das Modell eine Gesamtgenauigkeit von 73% und einen F1-Score von 71%. Durch die Ausblendung unbrauchbarer Regionen wie Himmel und Meer verbessert dieser Schritt die Zuverlässigkeit der photogrammetrischen Rekonstruktion erheblich.

Ein automatisiertes Modul zur Extraktion von Metadaten ergänzt die Segmentierung, indem es Schlüsselparameter wie Brennweite, Höhe und die Position von Referenzmarken direkt aus den Bildern abruft. Mithilfe einer Kombination aus optischer Zeichenerkennung und Computer-Vision-Techniken werden wesentliche Informationen wiederhergestellt und fehlende Werte geschätzt, indem die Redundanz in den Flugserien genutzt wird. Dies reduziert den Bedarf an manueller Transkription und konvertiert handschriftliche Bildkommentare in strukturierte digitale Formate.

Die Georeferenzierung stellt eine räumliche Verbindung zwischen historischen Bildern und modernen Koordinatensystemen her. Sie verwendet LightGlue, einen neueren, auf Deep Learning basierenden Matching-Algorithmus, zusammen mit einer progressiven Bildkachelung, die an

die Eigenschaften historischer Bilder angepasst ist. Durch den Abgleich von Verknüpfungspunkten zwischen den TMA-Scans und den Sentinel-2-Satellitenbildern generiert das System automatisch Bodenkontrollpunkte (GCPs) mit einer Positionsgenauigkeit von nur wenigen Metern und verbessert damit die ursprünglichen, oft kilometerweiten Geolokalisierungsschätzungen erheblich.

In der letzten Phase werden die segmentierten Bilder, extrahierten Metadaten und GCPs automatisch an Agisoft Metashape übergeben, das über seine Python-API in die Verarbeitungspipeline integriert ist. In dieser Phase wird eine Structure-from-Motion-Photogrammetrie durchgeführt, um dichte Punktwolken, Orthophotos und digitale Höhenmodelle (DEMs) ohne Benutzerinteraktion zu erzeugen. Bei der Anwendung auf der antarktischen Halbinsel hat die Pipeline erfolgreich 3D-Gletscheroberflächen für 49 Gletschersysteme rekonstruiert. Die Validierung anhand des hochauflösenden Referenz-Höhenmodells der Antarktis (REMA) zeigt mittlere Höhenunterschiede von etwa 90 Metern über die gesamte Gletscherausdehnung und 76 Metern in topographisch stabilen Gebieten.

Obwohl die Ergebnisse noch nicht die Genauigkeit einer vollständig manuellen Verarbeitung erreichen, ermöglicht das entwickelte System eine groß angelegte, wiederholbare Rekonstruktion historischer Gletscheroberflächen in einem bisher unerreichten Umfang. Durch die Kombination aller Komponenten in einem modularen, durchgängigen Rahmen macht diese Arbeit das TMA-Archiv für die zeitgenössische Kryosphärenforschung allgemein zugänglich und erweitert die Beobachtungsgrundlagen um mehr als ein halbes Jahrhundert. Der gesamte Code und die Arbeitsabläufe sind auf GitHub offen zugänglich, und die resultierenden Datenprodukte, einschließlich semantischer Masken, Metadatentabellen, georeferenzierter Bildpositionen und 3D-Gletschermodelle, werden zur Unterstützung der weiteren wissenschaftlichen Nutzung öffentlich freigegeben.

ACRONYMS

AGIC	Antarctic Geospatial Information Center
AP	Antarctic Peninsula
CNN	Convolutional Neural Network
DEM	Digital Elevation Model
FCN	Fully Convolutional Network
GCP	Ground Control Point
KP	Key Point
MAD	Median Absolute Deviation
MVS	Multi-view stereo
OCR	Optical Character Recognition
RANSAC	Random Sample Consensus
REMA	Reference Elevation Model of Antarctica
RMSD	Root Mean Square Difference
SfM	Structure-From-Motion
SIFT	Scale-Invariant Feature Transform
SURF	Speeded-Up Robust Features
TMA	Trimetrogon Aerial
TP	Tie Point
USGS	United States Geological Survey

1

INTRODUCTION

The Antarctic Peninsula (AP) is one of the fastest changing regions on our planet, and experienced a warming unparalleled anywhere else in the Southern Hemisphere in the late-20th century (Bracegirdle *et al.* 2024; Turner *et al.* 2016). Several ice shelves along its margins have retreated or disintegrated, leading to rapid dynamical changes of the glaciers on its mainland (Scambos *et al.* 2004; Shepherd, Ivins, Rignot, *et al.* 2018). Due to its relatively small scale and complex topography, a detailed picture of the mass balance of the AP was missing, until the advent of high-resolution satellite observations (e.g. in chronological order the Landsat-missions, TerraSAR-X, Sentinel-2A/B, and ICESat-2). While these data revealed that the Peninsula is losing ice rapidly, a longer, comprehensive perspective of the changes is still lacking.

To gain a meaningful long-term perspective on changes in this region, data spanning multiple decades rather than just recent years is essential (Meredith *et al.* 2019). However, obtaining such a dataset for the AP is challenging. Early satellite missions with adequate spatial resolution were limited, only beginning coverage of the region in the late 20th century (Nilsson, Gardner, and Paolo 2022; Otosaka *et al.* 2023).

To address the lack of a long-term data record, this thesis unlocks the potential of the Trimetrogon Aerial (TMA) archive, a vast archive of aerial photographs, obtained by the U.S. Navy since the 1940s. These images were acquired using a set of cameras pointing vertical, left and right, along ten thousands of kilometres spanning the Antarctic continent. Using the overlap between consecutive acquisitions in a Structure-From-Motion (SfM) photogrammetry approach, combined with recent developments in machine learning, I will derive digital elevation models of the AP, dating back more than five decades.

1

Generally, historical images offer a unique glimpse into the past and represent a largely untapped source of valuable data for a wide range of scientific applications. Over the past decades, many historical image archives around the world have been digitised and made publicly accessible (Cowley and Stichelbaut 2012; Kostrzewa 2024). Despite this progress, these archives remain significantly underutilized due to a range of challenges, including incomplete metadata, inconsistent image quality, and the labour-intensive nature of processing historical imagery. Yet, historical aerial images, particularly from the pre-satellite era, offer unique insights into environmental and geophysical changes that are otherwise unavailable. They provide essential data for extending temporal baselines in scientific studies and for examining long-term trends in remote and often inaccessible regions.

In recent years, several studies have successfully leveraged these archives for scientific research, such as those by Holmlund (2021) and Knuth *et al.* (2023). For the TMA dataset specifically, work has been conducted by, for example, North and Barrows (2024) and Zhao *et al.* (2017) for glaciers on the AP and Child *et al.* (2020) for the Byrd glacier in East Antarctica. Although the products and derived insights from these images are promising, much of their processing workflow was manual and time-consuming. Therefore, their research was limited in scale and only focused on single glaciers.

To address these challenges and expand the scope to more areas, this PhD research aims to develop an automated workflow that significantly reduces the amount of manual intervention required to create 3D products from historical aerial imagery with SfM. Using recent advances in machine learning, I aim to automate key steps in the workflow, making it more efficient and scalable.

1.1. Motivation

The motivation for this research stems from both methodological and application-oriented challenges. Existing SfM methods are primarily designed for contemporary, high-quality datasets and are not necessarily suited for processing historical aerial imagery. Existing methods for SfM often require significant manual intervention to account for issues such as varying image quality, inconsistent metadata, and differences in camera parameters. Therefore, there is a need to adapt and optimize these methods to handle the unique characteristics of historical data.

From an application perspective, the automated workflow developed in this thesis will be applied to the TMA archive to create accurate, geo-referenced historical 3D models of the AP. While previous studies have

used parts of this dataset to reconstruct individual glaciers, these efforts have been limited in scope and relied heavily on manual processing. This research will automatize this process and scale it to cover a larger portion of the dataset. The resulting 3D models will provide valuable insights into changes in elevation and mass balance between historical and modern times. This comparison will shed light on the long-term impacts of climate change on Antarctica's glaciers, including their dynamic responses to ice shelf weakening and disintegration.

1.2. BACKGROUND

This section provides an overview of the dataset used to develop the algorithm and outlines the key components of the applied methodology. A crucial step in this workflow is Tie Point (TP) matching, which plays a fundamental role in both geo-referencing and 3D reconstruction. Due to its importance, TP matching is examined first, followed by an analysis of geo-referencing. Lastly, the theoretical background of SfM is introduced, a process that generates 3D models from photographic imagery, allowing for the extraction of spatial information from the historical images.

1.2.1. TMA-ARCHIVE

The TMA-archive is an extensive collection of historical aerial photographs of Antarctica, taken by the U.S. Navy between 1946 and 2000. This archive contains approximately 333,000 frames of Antarctic photography, with annual records beginning in 1980. Images were taken in different phases, with the collection starting with Operation Highjump in 1946, followed by Operation Windmill in 1947, and continued with the first Operation Deep Freeze mission in 1955 (USGS 2018).

The images in this archive were taken along approximately 6,000 flight lines. During these flights, airplanes typically maintained a straight course at a constant altitude and speed, capturing images every few seconds to ensure significant overlap between consecutive frames. As shown in Figure 1.1, these flight paths cover vast areas of coastal Antarctica, particularly in the west and along the AP. Some sections of East Antarctica were also photographed, though inland coverage is limited. The sparse inland coverage is primarily due to the absence of distinguishable surface features in the interior, where images would mostly depict featureless snowfields.

1

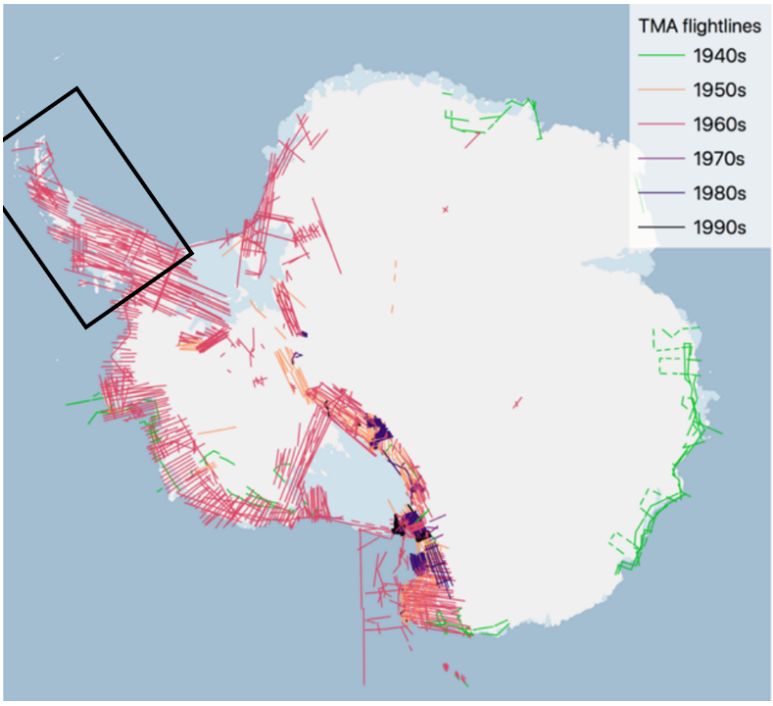


Figure 1.1: Overview of flight paths of the TMA archive with the majority of flights taken in the 1960s. The AP is marked with a black box.

Most of the images in the TMA archive are black-and-white photographs captured using a system of three cameras mounted together: two oblique cameras angled at 30 degrees and one nadir (vertical) camera. This setup, illustrated in Figure 1.2, provided comprehensive 180-degree horizon-to-horizon coverage. However, not all flights utilized the full camera setup. In some cases, only the vertical images are available, and technical malfunctions during flights further reduced the number of usable images.

Each image in the archive is assigned a unique identifier based on its filename. For example, an image labelled `CA021432V0042` represents a vertical image from flight path 2143, with `0042` indicating that it is the 42nd image taken along that flight path. The identifier `32V` specifies a vertical image, while `31L` and `33R` denote left and right oblique images, respectively.

In 2009, the United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center scanned the Antarctic film collection. In collaboration with the Antarctic Geospatial Information Center (AGIC) at the University of Minnesota, approximate latitude and

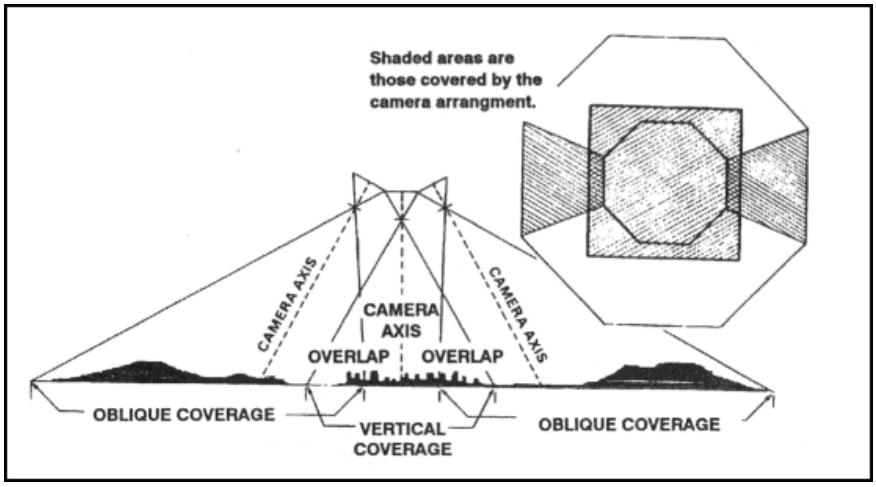


Figure 1.2: Setup of a TMA system, in which one camera is mounted in nadir view and two cameras oblique, offering a 180 degree coverage. Source: 550 Cord Software (2024).

longitude coordinates were assigned to the single-frame records, which were subsequently made publicly available (Polar Geospatial Center 2023). Additional metadata was also released, including camera reports and a shapefile containing flight paths and image positions, along with supplementary information such as the date of recording, flight direction, and altitude; though some details remain incomplete. Despite their age, many of the images in the archive retain high contrast and provide invaluable historical perspectives on Antarctic conditions, particularly where other satellite or aerial imagery is unavailable. Figure 1.3 shows examples of high-quality TMA images.

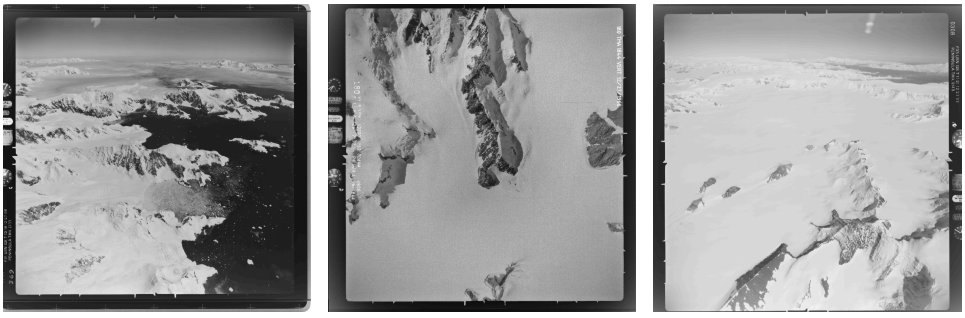


Figure 1.3: Example of TMA images with good quality

However, the TMA images also have some limitations. Many film sources in the U.S. Antarctic Resource Center (USARC) collection have been affected by vinegar syndrome degradation, compromising the quality of the film and subsequent scans. Additionally, during the digitisation process, scanning errors such as Newton rings and image rotation (sometimes by 180 degrees) were introduced. For the images themselves, text annotations on the images often interfere with automated feature detection processes, introducing irrelevant features. Other challenges include variable exposure levels, cloud cover, and vast expanses of featureless snowfields, which reduce the usability of some images. Even though additional metadata is existing and camera reports are available, it is often unclear which specific camera corresponds to a given report, complicating the accurate use of the images. Moreover, the provided geographic positions of the images are only approximate, with some deviations reaching several dozen kilometres, further limiting their accuracy in spatial analyses. Examples for images with suboptimal image quality are shown in Figure 1.4.

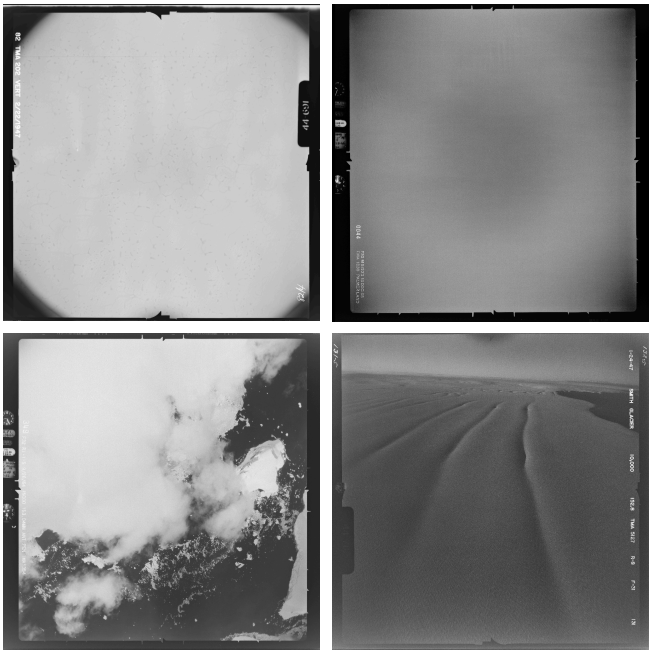


Figure 1.4: Example of images with suboptimal quality. Top left: No contrast; Top right: Scanning error; Bottom left: clouds, Bottom right: too dark because of too little exposure

1.2.2. TIE POINT MATCHING

Tie Point (TP) matching involves identifying corresponding points in images that capture the same scene from different angles, at different times, or using different sensors. This process is essential for multiple aspects of this thesis, particularly in Chapter 3, where it plays a key role in geo-referencing historical images by matching points between modern and historical imagery. In Chapter 4, TP matching is fundamental to the SfM workflow, ensuring the identification of corresponding points across different images of the same scene. The matching process consists of three main steps: feature detection, feature description, and feature matching.

Figure 1.5 illustrates an example of the TP matching process. In this example, two images are compared: both depict the same geometric pattern, but the second image is slightly rotated and shifted to the left. The objective of TP matching is to correctly identify corresponding

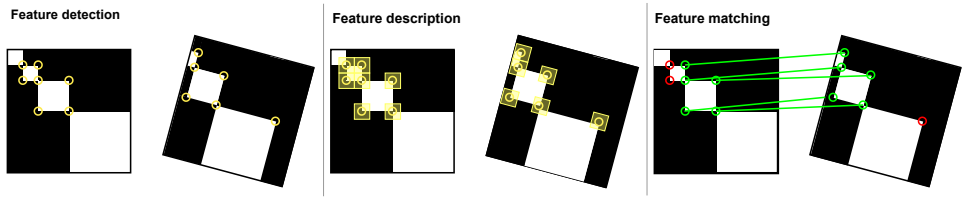


Figure 1.5: Schematic example for TP matching between two different images from feature detection (left), feature description (middle) to feature matching (right)

points between these two images, despite the differences in orientation and position.

The first step in the process is feature detection, which involves identifying all prominent points (the so called Key Points (KPs)) within each image. These points are typically corners, edges, or blobs: regions that exhibit rapid changes in colour or brightness. Classical feature detectors include the Harris Corner Detector (Harris and Stephens 1988), which identifies corners by analysing local intensity variations, and Scale-Invariant Feature Transform (SIFT) (Lowe 2004), which detects KPs by searching for extrema in a Difference of Gaussians (DoG) scale-space representation. Another widely used detector is Speeded-Up Robust Features (SURF) (Bay, Tuytelaars, and Van Gool 2006), which improves upon SIFT by using Hessian matrix approximations and integral images for faster computation. More recent approaches leverage Convolutional Neural Networks (CNNs) to generate heat maps that indicate the presence of KPs across the image. These heat maps enable the extraction of sparse KPs, providing a more robust and adaptable set of features compared to traditional hand-crafted detectors.

However, simply detecting the coordinates of KPs is not sufficient. The second step, feature description, is necessary to ensure that the detected points can be matched accurately between images. This process involves analysing the pixel distribution around each detected KP to generate a distinctive signature, or descriptor, that uniquely represents the feature. An effective descriptor should be invariant to changes in orientation, scale, and illumination, yet specific enough to distinguish between repeating patterns, such as a chessboard. Some algorithms, such as for example SIFT and SURF function as both feature detectors and descriptors. The first generates 128-dimensional feature vectors based on gradient histograms computed around each KP, while the latter produces 64-dimensional descriptors using Haar wavelets for faster processing. In contrast, more recent methods utilize graph neural

networks (GNNs) to create feature descriptors by learning a vector representation for each KP, allowing for more robust and adaptable feature matching.

The final step is feature matching, where descriptors from one image are compared to those in the other to identify corresponding points. Each KP should ideally be matched to a single point in the other image. Any unmatched points are treated as outliers. Robust filtering techniques, such as Random Sample Consensus (RANSAC) by Fischler and Bolles (1981), are frequently employed to filter out incorrect matches by simultaneous fitting and evaluating a transformation model to the matched points and removing outliers. Additionally, geometric constraints can be applied to refine the matching process. For instance, KPs that are close to each other in one image should remain spatially close in the other. Furthermore, consistency in orientation and scale can be enforced to improve accuracy. However, these constraints must be applied carefully in remote sensing applications, where images may undergo complex transformations due to varying sensor types, viewing angles, and environmental conditions. The left-over KPs, that are now matched between images are then called TPs.

Epipolar geometry, which describes the geometric relationship between two views of the same scene, is often employed to further enhance matching accuracy. This geometric constraint helps reduce the search space for matching points by limiting potential matches to points lying along corresponding epipolar lines in the images.

Traditional approaches to KP matching, such as SIFT and SURF, rely on handcrafted feature detectors and descriptors. These methods are robust and well-established but can struggle with images that have significant differences in appearance. In contrast, modern deep learning-based approaches use pre-trained models to learn both the features and the matching criteria. These models are capable of adapting to a wide variety of image types and transformations. While training these models requires substantial computational resources, the resulting matching process is typically fast and efficient.

In summary, TP matching is a multi-step process that forms the foundation for several key components of this research. By detecting, describing, and matching features across images, this process enables the extraction of spatial information from historical imagery, supporting both geo-referencing and 3D reconstruction workflows in this thesis.

1

Recent Advances in Learning-Based Tie point Matching

In recent years, TP matching has increasingly relied on machine learning, with a significant breakthrough being the combination of SuperPoint and LightGlue.
SuperPoint by DeTone, Malisiewicz, and Rabinovich (2018) is a self-supervised framework that detects and describes local KPs using a fully CNN. The network employs a shared backbone to extract a dense feature map from the input image, which is then processed by two parallel heads. The detector head converts the feature map into a raw score map that is transformed into a probability map via SoftMax-activation and upsampling; non-maximum suppression is subsequently applied to retain only the most distinctive KPs. Simultaneously, the descriptor head generates a dense descriptor map where each location is associated with an L2-normalized feature vector capturing the local appearance. Bilinear interpolation is used to extract precise descriptors at the detected KP locations. This process enables SuperPoint to robustly detect a rich set of features even for images that failed with other algorithms.
Building on this, LightGlue by Lindenberger, Sarlin, and Pollefeys (2023) is a matching algorithm employing a lightweight graph neural network that constructs a graph of KPs, where each node represents a KP and edges encode relationships based on spatial proximity and descriptor similarity. The network leverages an iterative message-passing mechanism that refines matching scores: during each iteration, nodes exchange information through attention-based layers, combining local descriptor similarities with global geometric context. This process incorporates both self- and cross-attention mechanisms, allowing the network to dynamically weigh and aggregate context from neighbouring KPs and enforce spatial consistency. Furthermore, by integrating uncertainty estimates and confidence weighting, LightGlue effectively suppresses outlier correspondences.
The combination of these two algorithms forms an efficient and accurate pipeline for TP matching that is capable of handling more challenging scenarios. This integrated approach has been extensively adopted in recent years, enabling SfM on more complex datasets, including the archive data used in this thesis.

1.2.3. Semantic Segmentation

Historical aerial images were often not captured with photogrammetric principles and applications in mind, which means they frequently include elements such as water bodies and clouds that can limit the quality of SfM reconstructions. For example, water surfaces, especially those with wave patterns, often display repetitive, high-contrast textures that can result in false TP-matches between images, even when no physical correspondence exists. Similarly, clouds are unstable in shape and position, making them unreliable for 3D reconstruction. To address these issues, semantic segmentation can be a valuable tool. By identifying and masking problematic regions such as water and clouds, it helps prevent erroneous TPs and improves the overall accuracy of the SfM process.

Semantic segmentation, as shown in Figure 1.6 is the process of classifying each individual pixel in an image into a semantic category. Usually, the semantic classes are defined in advance and tailored to the specific goals of the analysis. Unlike instance segmentation, which distinguishes between separate occurrences of the same object class, semantic segmentation treats all instances of a class uniformly. For example, all trees in the figure are labelled simply as “trees”, without distinguishing between individual trees.

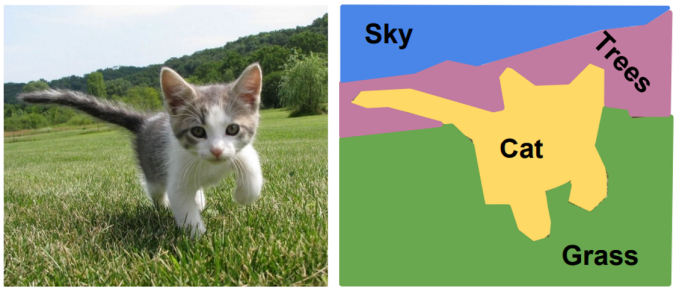


Figure 1.6: Example for semantic segmentation. Source: Li, Johnson, and Yeung (2017)

Earlier approaches to semantic segmentation often relied on traditional machine learning methods such as Random Forests or Support Vector machines (L. Wang *et al.* 2020). In recent years, however, deep learning-based methods have been used extensively. For a comprehensive overview of modern deep learning-based segmentation techniques, see Minaee *et al.* (2022). Among these methods, U-Net (Ronneberger, Fischer, and Brox 2015) has emerged as a widely used architecture, which is also used in Chapter 2. It leverages multiple CNNs in an

encoder–decoder structure, enabling accurate segmentation even with relatively limited training data.

Beyond improving the quality of TP-matching by masking parts of the images, semantic segmentation offers additional benefits. For instance, in the case of oblique historical imagery, the segmentation results can serve as a quality control indicator: if the class “sky”appears in the bottom part of an image, it may suggest an incorrect image orientation which needs to be corrected. Furthermore, segmentation can be used to filter out images that are unsuitable for SfM due to a lack of texture (e.g., scenes dominated by snow or cloud cover).

1.2.4. GEO-REFERENCING

Geo-referencing is the process of aligning the internal coordinate system of a digital map, aerial photograph, or other spatial dataset with a real-world geographic coordinate system. By establishing this correspondence, it becomes possible to integrate, compare, and analyse spatial data from multiple sources on a common, absolute basis. This alignment is critical because, without geo-referencing, it is not feasible to compare newly created spatial products (e.g. those created by SfM) with modern geospatial data, and the datasets would lack an absolute position on the earth’s surface, including the vertical component.

If geo-referencing an image, an approximate location of the image must be known beforehand. Typically, the workflow for geo-referencing begins with obtaining a reference image or map of the same area that already contains absolute geographic coordinates, such as satellite imagery. The next step involves identifying and marking corresponding Ground Control Points (GCPs) in both the target and reference images. This can be seen in Figure 1.7, in which the same points are identified in the non geo-referenced historical map and in a modern geo-referenced equivalent (OpenStreetMap). For optimal results, these points should be marked with the highest precision possible and distributed evenly across the image. Once a sufficient number of GCPs, three for a simple affine transformation, more for higher-order polynomial transformations, have been identified, a transformation matrix is computed. This matrix encapsulates the spatial relationship between the two coordinate systems and is then applied to the target image to align it with the absolute coordinate system of the reference data.

After the geo-referencing process is complete, its quality must be thoroughly evaluated. Quantitative analysis, such as calculating the difference between the transformed positions of the control points

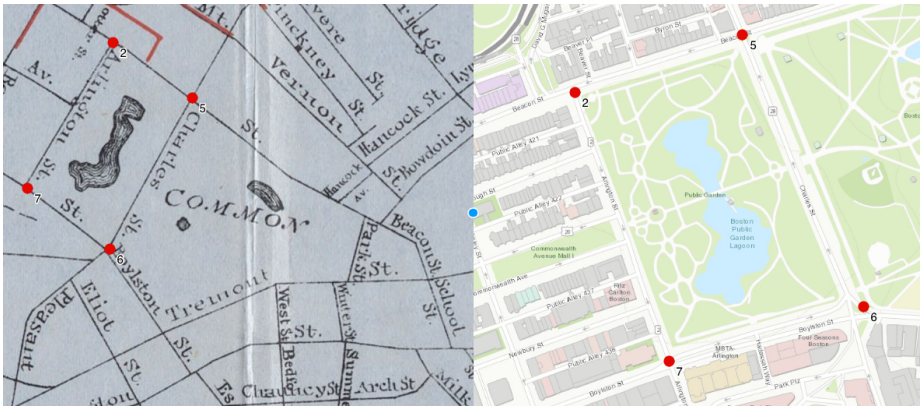


Figure 1.7: GCPs identified in a non geo-referenced map and it's modern geo-referenced equivalent. Source: Leventhal Map & Education Center (2022)

and their known absolute coordinates, provides a numerical measure of accuracy. However, low residuals alone do not guarantee correct spatial alignment, as errors may still occur outside the area covered by control points or due to unsuitable transformation models. Therefore, qualitative evaluation through visual inspection is essential. Although automated overlay and correlation techniques can supplement this evaluation, manual verification remains a critical step in detecting residual misalignments or systematic errors.

However, the process of geo-referencing comes with its own set of challenges. In regions where prominent and persistent structures, such as streets or landmark buildings like churches, exist, the process is generally straightforward because these features remain consistent over time. Conversely, in areas dominated by natural landscapes or when there is a considerable time gap between the available images, geo-referencing becomes more difficult. Natural features may change over time, and the lack of distinct landmarks complicates the identification of corresponding points between datasets.

1.2.5. STRUCTURE-FROM-MOTION

SfM is a widely adopted photogrammetric technique for reconstructing three-dimensional structures from sequences of two-dimensional images taken from different perspectives. Through a process of feature detection and matching, the algorithms estimate the relative positions of the camera and the 3D coordinates of points in the scene simultaneously, offering a flexible approach for reconstructing complex scenes.

When using SfM, it is important to understand camera parameters because they directly influence the accuracy and reliability of the reconstructed 3D scene. These parameters can be distinguished into internal (intrinsic) and external (extrinsic) parameters (Stankiewicz, Lafruit, and Domański 2018). Internal parameters are properties inherent to the camera, defining how it projects 3D world points onto the 2D image plane. They include focal length (the distance between the camera centre and the image plane), principal point (the intersection point between the optical axis and the image sensor), skew (the angle between the x and y pixel axes, typically zero for most modern cameras), and distortion coefficients (parameters accounting for lens distortions such as radial and tangential distortion). External parameters on the other hand, describe the camera's pose within the world coordinate system. This includes the camera's spatial position (defined by a translation vector) and its orientation (represented by rotation angles or matrices) relative to the scene.

For modern images, many camera parameters are typically embedded in the image metadata, such as EXIF data, providing convenient access to necessary information. However, historical images, like those from the TMA archive, are more challenging as parameters are often only partially documented or entirely missing. While some historical cameras may have calibration reports, these are rarely universally available, though occasionally certain parameters might be annotated directly on the images.

SfM can operate without complete prior knowledge of camera parameters, as missing values such as (relative) position, orientation, and focal length are estimated and refined during bundle adjustment (Westoby *et al.* 2012). Each additional unknown, however, increases the solution's uncertainty, whereas supplying known parameters improves the reconstruction quality and reduces ambiguity in the optimization process. Compared to traditional photogrammetry, which requires precise input data, SfM offers greater flexibility but often relies on a larger number of overlapping images to achieve similar accuracy (James and Robson 2012).

While both traditional photogrammetry and SfM can yield high geometric consistency, the spatial accuracy of the latter tends to be more variable due to its reliance on the quality and density of detectable features across images. SfM is particularly well-suited for this dataset, as the intrinsic (e.g. focal length or lens distortion) and extrinsic parameters (e.g. camera position or orientation) of the historical images are often only partially known. By refining these parameters throughout the process, it offers an effective approach to address the uncertainties inherent in the historical dataset.

It can also be applied retrospectively, utilizing images not originally intended for 3D reconstruction, as camera parameters can be estimated during processing. However, SfM has limitations: low-textured or repetitive-pattern scenes pose challenges for feature matching, and reconstructions in the absence of absolute geo-referencing information lack absolute scale and orientation, resulting in relative orientation and camera poses.

Numerous software implementations of SfM exist, including proprietary programs like *Agisoft Metashape*[1] and *Pix4D*[2], and open-source options such as *MicMac*[3] and *COLMAP*[4]. While initial efforts in this research explored *MicMac*, *Agisoft Metashape* was ultimately selected due to its efficient GPU-based processing, high reconstruction quality, and extensive Python API. This allowed for the development of numerous custom Python scripts to automate and improve various processes.

The basic principles of SfM remain consistent across different software implementations, involving a sequence of steps to achieve 3D reconstruction (Schönberger and Frahm 2016). The process begins with the **image acquisition**, a critical phase that significantly impacts the final model's quality. Images should capture the scene from various angles with substantial overlap (ideally between 60% and 80%) to ensure sufficient common features for TP matching across consecutive images. Scenes with rich texture, distinct points, and consistent lighting conditions are optimal, and using the same camera ensures uniformity in camera parameters, enhancing reconstruction quality. However, in this thesis, the historical images were captured several decades ago, independent of SfM requirements, limiting control over these factors.

The next step is **feature matching**, where TPs across images are identified (see Subsection 1.2.2). The historical imagery of these archive presents unique challenges, such as borders, artifacts and unwanted semantic classes that interfere with accurate TP matching, requiring masking of these features as described previously. Additionally, given the large number of images in TMA-datasets, matching every image pair is impractical. Using sequential image IDs or spatial footprints optimizes matching, ensuring only relevant image pairs are processed together.

With TPs established, **camera calibration and alignment** estimate the intrinsic and extrinsic camera parameters required for 3D reconstruction. The reconstruction begins by estimating the 3D structure and camera poses for an initial pair of cameras based on their relative pose. As

[1] https://www.agisoft.com/
[2] https://www.pix4d.com/
[3] https://github.com/micmacIGN/micmac
[4] https://colmap.github.io/

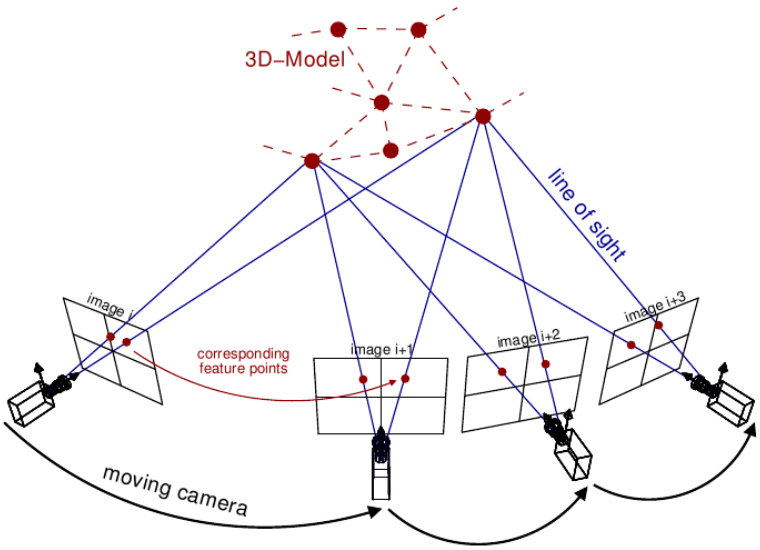


Figure 1.8: Basic principle of SfM: Incremental addition of cameras based on detected KPs for estimation of the 3D-structure, Source: Sweeney (2016).

can be seen in Figure 1.8, additional cameras are then sequentially incorporated, expanding the 3D structure as new parts of the scene become visible. At each step, bundle adjustment is applied to minimize reprojection error across all images, ensuring a globally consistent parameter set and preventing drift. This iterative refinement optimizes both camera parameters and scene structure, leading to a high-quality sparse point cloud that serves as the foundation for subsequent dense reconstruction.

The next stage involves **depth map generation**, which forms the foundation for dense 3D reconstruction. Depth maps are 2D images where each pixel encodes the distance from the camera to the corresponding point in the scene. These are computed using pixel-wise stereo matching across overlapping images, guided by the camera poses and intrinsic parameters derived from SfM. Once computed, depth maps are back-projected into 3D space using the camera's intrinsic and extrinsic parameters, producing dense point clouds. The fusion of depth maps from multiple views further refines the dense reconstruction, reducing noise and improving geometric accuracy.

As an optional step, but essential for comparison with modern data, **geo-referencing** aligns the spatial data from the 3D model with a known coordinate system for precise placement and analysis. This

is achieved by incorporating known camera positions or using GCPs, which are points with known spatial coordinates within the images. These methods convert the relative 3D positions generated by SfM into absolute 3D positions, ensuring accurate scaling, orientation, and placement.

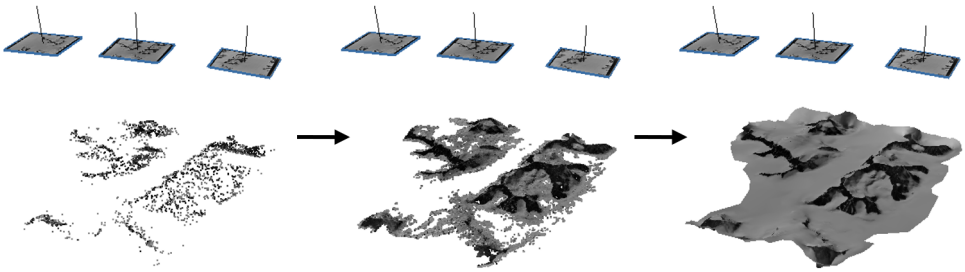


Figure 1.9: From TPs to dense point cloud to surface mesh (Example using *Agisoft Metashape* with three images from the archive: CA215132V0299, CA215132V0300, CA215132V0301)

The final step is **dense 3D reconstruction** using Multi-view stereo (MVS) techniques as can be seen in Figure 1.9. The TPs are refined to the dense point cloud to enable the creation of a continuous surface mesh, which can then be textured using the original images. This 3D model can be further processed to derive additional products, such as a Digital Elevation Model (DEM) or an orthophoto.

SfM has diverse applications across multiple disciplines. In the context of this research, two key areas of relevance are topographic mapping and glaciology: SfM is widely employed to generate high-resolution topographic models (Eltner *et al.* 2016), particularly in remote or inaccessible regions, where traditional surveying methods may be impractical. These models are crucial for geological analysis, environmental monitoring, and land-use planning. In glaciology, SfM enables the reconstruction of detailed three-dimensional glacier models, facilitating the quantification of volume changes and surface dynamics. This capability provides valuable insights into glacier evolution and the broader impacts of climate change.

Given its flexibility and compatibility with historical imagery, SfM is an invaluable tool for addressing the unique challenges posed by this project. The method's ability to work with incomplete datasets and reconstruct 3D structures from partially known camera parameters aligns perfectly with the constraints of this research.

1.3. RESEARCH OBJECTIVES

The overarching goal of this thesis is to develop and validate automated workflows for processing historical aerial imagery using SfM, with the aim of reconstructing past surface conditions and generating geospatial datasets suitable for long-term environmental change analysis. This work focuses on the TMA archive to unlock new opportunities for studying the historical evolution of the AP, particularly in terms of glacier dynamics and surface morphology. Given the substantial volume and heterogeneity of available imagery, this research emphasizes automation, scalability, and modularity to ensure that the proposed approach is broadly applicable to other archival datasets and regions. This leads to the following main research question:

How can historical aerial imagery of the Antarctic Peninsula be processed in an automated manner using Structure-from-Motion to generate geospatial datasets suitable for long-term glacier change analysis?

This overarching question is broken down into the following sub-questions, each addressed in the corresponding chapters of this thesis. The first two sub-questions focus on preparing the historical images, while the final question investigates their use in 3D reconstruction.

(Q1) *How effectively can semantic segmentation be applied to historical aerial imagery of the cryosphere?* **(Chapter 2)**

Many state-of-the-art algorithms for semantic segmentation are designed for modern imagery with colour information and scenes with higher contrast. However, historical aerial photographs such as those from the TMA archive differ significantly in quality and spectral characteristics. This chapter explores the feasibility of training a U-Net-based neural network to segment historical aerial images of Antarctica and evaluates the reliability of the resulting classifications.

(Q2) *How reliable can camera calibration metadata be extracted automatically from degraded historical aerial imagery?* **(Chapter 3)**

In contrast to modern digital datasets, historical aerial images typically lack comprehensive metadata, such as exterior orientation parameters (camera position and rotation), and often only contain incomplete internal calibration information (e.g., focal length or position of the fiducial marks). However, some of this information is visibly embedded in the scanned photographs themselves, including printed annotations, fiducial marks, or instrument readings (see Figures 1.3 and 1.4). This chapter explores the potential to automatically extract such metadata

using techniques like Optical Character Recognition (OCR)-based text recognition and pattern matching. A particular focus lies on exploiting the redundancy within image series from the same flight path to estimate missing or corrupted values, thereby improving the reliability and completeness of the recovered metadata.

(Q3) *To what extent is it possible to automatize geo-referencing of historical aerial imagery?* **(Chapter 4)**

Geo-referencing historical aerial images is typically a labour-intensive process that involves manually identifying GCPs with known absolute coordinates. This chapters examines the potential of modern TP matching techniques to automate the identification of such points, thereby reducing the need for manual intervention.

(Q4) *How feasible is the 3D reconstruction from historical aerial imagery using Structure-from-Motion?* **(Chapter 5)**

Building upon the previous steps, this chapter investigates the extent to which the SfM workflow can be automated to generate 3D models from historical aerial images. It assesses the feasibility of creating accurate reconstructions and derived products, aiming to improve accessibility and usability for broader scientific applications.

1.4. Scope and limitations

The scope of this study is defined by several parameters. The research focuses exclusively on data from the TMA archive, without incorporating additional historical datasets. The emphasis is on automating workflows to minimize manual intervention. Geographically, the study is confined to the AP, with a primary focus on applying SfM rather than conducting broader cryospheric analyses.

However, this approach is subject to several limitations. The TMA archive images suffer from inherent issues such as missing information, a sparse distribution of TPs, and the absence of precise positioning data, all of which constrain the accuracy of the derived products. The semantic segmentation presented in Chapter 2 will be imperfect, reflecting the inherent challenges associated with processing historical imagery. Similarly, the extraction of metadata from the historical images as described in 3, will not work all the time, as many images are under/over-exposed and not scanned correctly. The geo-referencing of historical images in Chapter 4 is expected to achieve a precision within a few meters, which falls short of the centimetre-level standards of

contemporary datasets. Furthermore, the DEMs generated in Chapter 5 will exhibit meter-scale precision, with variations in quality across regions; areas with denser TPs, such as rock outcrops, will yield more accurate results compared to sparsely covered regions like snowfields.

Despite these limitations, the study provides several key contributions that significantly enhance the usability of the TMA archive for future research. First, the derived semantic information allows researchers to filter images based on content, enabling more targeted analysis without the need to manually inspect each photograph. Second, image metadata, previously available only in handwritten form, are now extracted and stored in digital format, improving accessibility and integration into automated pipelines. Third, an accurate geo-referencing is applied to the images for the first time, positioning them within a consistent spatial framework. Finally, the generation of 3D surface models adds an entirely new dimension of data, offering elevation information that was previously unavailable. Together, these products not only support the specific goal of evaluating elevation change in the AP, but also transform the TMA archive into a richer and more accessible dataset for the broader scientific community.

1.5. THESIS ORGANIZATION

The content of this thesis is based on four journal articles, which together form Chapters 2 to 5. Each of these chapters is self-contained, featuring separate introductions and conclusions. Thus some overlap between chapters may be present.

Chapter 2 focuses on segmenting historical images into semantic classes, where each pixel is assigned one of six classes (snow, ice, water, rocks, clouds & sky). This chapter investigates the effectiveness of a neural network to add semantic information to historical scans from the TMA archive, assessing the feasibility and reliability of classifying such images without manual intervention.

Chapter 3 presents an automated workflow to extract metadata from historical images, including fiducial marks, altimeter readings, and textual annotations. The workflow combines computer vision and OCR techniques, with fallback estimation based on images from the same flight path to handle varying image quality.

Chapter 4 addresses the precise geo-referencing of historical images. It explains how the approximate positions of historical images can be converted into exact geo-referenced locations using TP matching

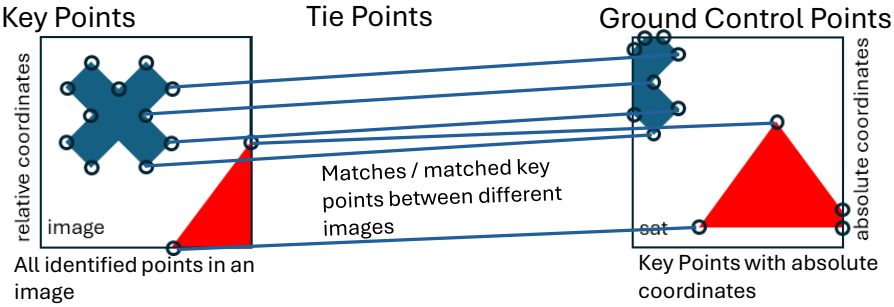
methods and modern satellite imagery, exploring the potential to reduce the labour-intensive process of manual geo-referencing.

Chapter 5 builds upon the results of the previous chapters and looks into the automated creation of 3D models using SfM techniques and its application for the TMA-archive.

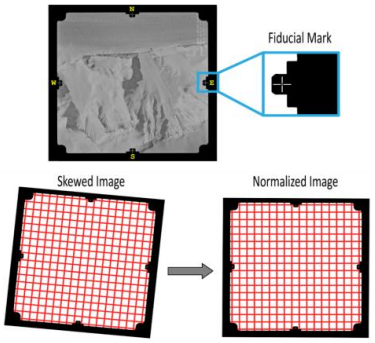
Chapter 6 summarizes and discusses the findings, along with recommendations and implications for future research.

1

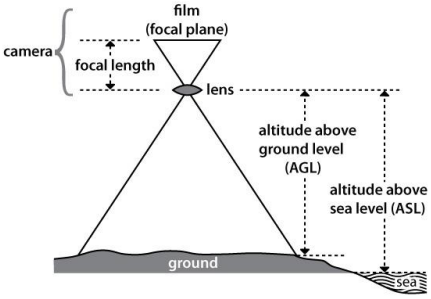
Type of points



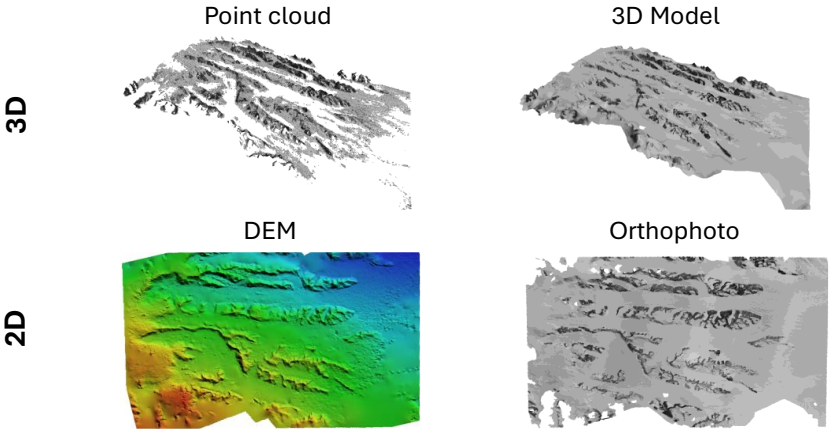
Interior orientation



Exterior orientation



Output from Structure-from-motion



Structure-from-Motion cheat-sheet

2

SEMANTIC SEGMENTATION OF HISTORICAL IMAGES USING U-NETS

The TriMetrogon Aerial (TMA) archive is an archive of historical images of Antarctica taken by the US Navy between 1940 and 2000 with analogue cameras. The analysis of such historic data can give a view of Antarctica's glaciers predating modern satellite imagery and provide unique insights into the long-term impact of changing climate conditions with essential validation data for climate modelling. However, the lack of semantic information for these images presents a challenge for large-scale computer-driven analysis. Such information can be added to the data using semantic segmentation, but traditional algorithms fail on these scanned historical grayscale images, due to varying image quality, lack of colour information and artefacts in the images. To address this, we present a deep-learning-based U-net workflow. Our approach includes creating training data by pre-processing and labelling the raw images. Furthermore, different versions of the U-net are trained to optimize its hyper-parameters and augmentation methods. With the optimal hyper-parameters and augmentation methods, a final model has been trained for a use-case to segment 118 images covering Adelaide Island. We tested our approach by segmenting challenging historical images using a U-net model with just 80 training images, achieving an accuracy of 73% for 20 validation images. While no test data is available for our use case, a visual examination of the

This chapter has been published as *Revisiting the Past: A comparative study for semantic segmentation of historical images of Adelaide Island using U-nets* in the *ISPRS Open Journal of Photogrammetry and Remote Sensing* (Dahle, Lindenbergh, and Wouters 2024b).

segmented images shows that our method performs effectively. The comparison of the hyper-parameters and augmentation methods provides directions for training other U-net-based models so that the presented workflow can be used to segment other archives with historical imagery. Additionally, the labelled training data and the segmented images of the test are publicly available at https://github.com/fdahle/antarctic_segmentation.

2.1. INTRODUCTION

Historical imagery archives provide valuable information about various parts of the world from the pre-satellite era. In recent years, there has been a growing trend of digitising such archives and using these as a data source in geo-sciences (Cowley and Stichelbaut 2012; Heisig and Simmen 2021). However, despite their potential, they remain under-exploited as most images are only available as scans without any metadata. This lack of metadata makes extracting information challenging as it requires a significant amount of manual work to incorporate them into scientific research.

One such historical imagery archive is the Trimetrogon Aerial (TMA) archive, where TMA stands for TriMetrogon Aerial, a system of cameras that takes vertical, left oblique, and right oblique images simultaneously for topographic mapping. The U.S. Navy collected this archive of historical imagery of Antarctica between 1946 and 2000, with a particular focus on the Antarctic Peninsula (AP) (USGS 2018). These photographs were primarily used for topographic mapping and provide a historical snapshot of many parts of Antarctica. Figure 2.1 shows an example image from this dataset.

This data set holds valuable information on historical ice topography and coverage in this area, for which few other data sources are available. In combination with recent observations, the TMA archive provides a unique opportunity to study multi-decadal changes in the state of the Antarctic Ice Sheet (A. J. Cook, Holland, *et al.* 2016; A. J. Cook and Vaughan 2010; Kunz *et al.* 2012). The data set, however, presents several challenges to its use. For example, as can be seen in Figure 2.1, parts of the images can be obstructed by clouds or suffer from degradation due to the vinegar syndrome (decomposing of the film when stored for a longer time (Allen *et al.* 1987), see lower left part of the image). Furthermore, the archive consists of around 330.000 images, without any additional information on the content or quality of the images and with only an approximate geo-localization provided, making it difficult to find specific features in the image archive. Adding semantic information to the TMA image archive would significantly

Figure 2.1: Example of a historical image from the TMA archive (located on the AP in 1969)

increase its usability. For example, this would allow researchers to find images on the boundary of ice and water to study areal changes of ice shelves and marine-terminating glaciers, detect rock outcrops, and provide information about the usability of individual images for specific research purposes, e.g. in terms of cloud cover or variability of the image.

Recent examples of adding semantic information can be found in Heidler *et al.* (2021), Nambiar *et al.* (2022), and Y. Wang *et al.* (2022), where satellite imagery is segmented. In modern satellite imagery, information is available in multiple bands and with high contrast, so that segmentation is easily applicable. However, semantic segmentation of historical imagery is more challenging. The images are only available in grayscale with less contrast, making them less informative for segmentation. As a result, the algorithms developed for satellite imagery segmentation cannot directly be applied to historical imagery. Due to these problems, semantic segmentation of historical imagery is rare but some successful examples have been reported, such as Mboga *et al.* (2020) and Dias *et al.* (2020). In both cases, machine learning algorithms were used to apply semantic segmentation. However, these examples targeted a very diverse environment with very distinct classes, unlike the more monotone scenes in Antarctica.

Another big challenge for the semantic segmentation of historical imagery is the lack of training data: labelled data is often only available for modern data sources and cannot be used for historical images. Thus, all training data must be manually created beforehand, which is time-consuming, resulting in limited availability.

To address these challenges, we propose using a U-net for the semantic segmentation of the TMA archive. A U-net, originally developed for medical purposes by Ronneberger, Fischer, and Brox (2015), is a type of neural network specifically designed for image segmentation with a small amount of training data. Recently, U-nets gained popularity and are also extensively used for semantic segmentation in geo-science(Baumhoer *et al.* 2019; Hartmann *et al.* 2021; Heffels and Vanschoren 2020; Kattenborn, Eichel, and Fassnacht 2019). In a previous paper (Dahle, Tanke, *et al.* 2022), we were able to create a semantic segmentation of part of the TMA historical imagery, even under challenging conditions, with an average accuracy of 74% over six classes using 67 images.

In this contribution, we build upon the use case and establish a fully operational workflow for the semantic segmentation of historical imagery of the cryosphere. To achieve this, we investigate the impact of different model parameters on the quality of the segmentation. It is worth noting that so far, most studies use default parameter settings and standard losses for training. However, adapting these parameters can lead to significantly improved results as shown in Jadon (2020), Kugelman *et al.* (2022), and Solórzano *et al.* (2021). Nevertheless, these parameter comparisons often focus on a single parameter, and a holistic approach considering multiple parameters simultaneously is absent. Moreover, such comparisons are typically conducted on larger datasets with better image quality, making it challenging to extrapolate the findings to historical imagery segmentation.

To demonstrate the performance of our model, successive to the parameter evaluation, we apply the semantic segmentation to a geographical subset of the TMA archive, specifically Adelaide Island (see Figure 2.2). The island is situated in the eastern part of the AP and is an enclosed area with a variety of different classes. As it features multiple flight paths, images are taken by different cameras.

2.2. DATA

As input data, we utilize aerial images from the TMA archive. All pictures within the archives were made in triples, as can be seen in Figure 2.3. Each image is associated with a unique identification number that comprises the flight line, roll, and frame. The roll indicates the direction the camera was facing, while the frame is a unique identifier for each image position on a flight line. For example, the identifier

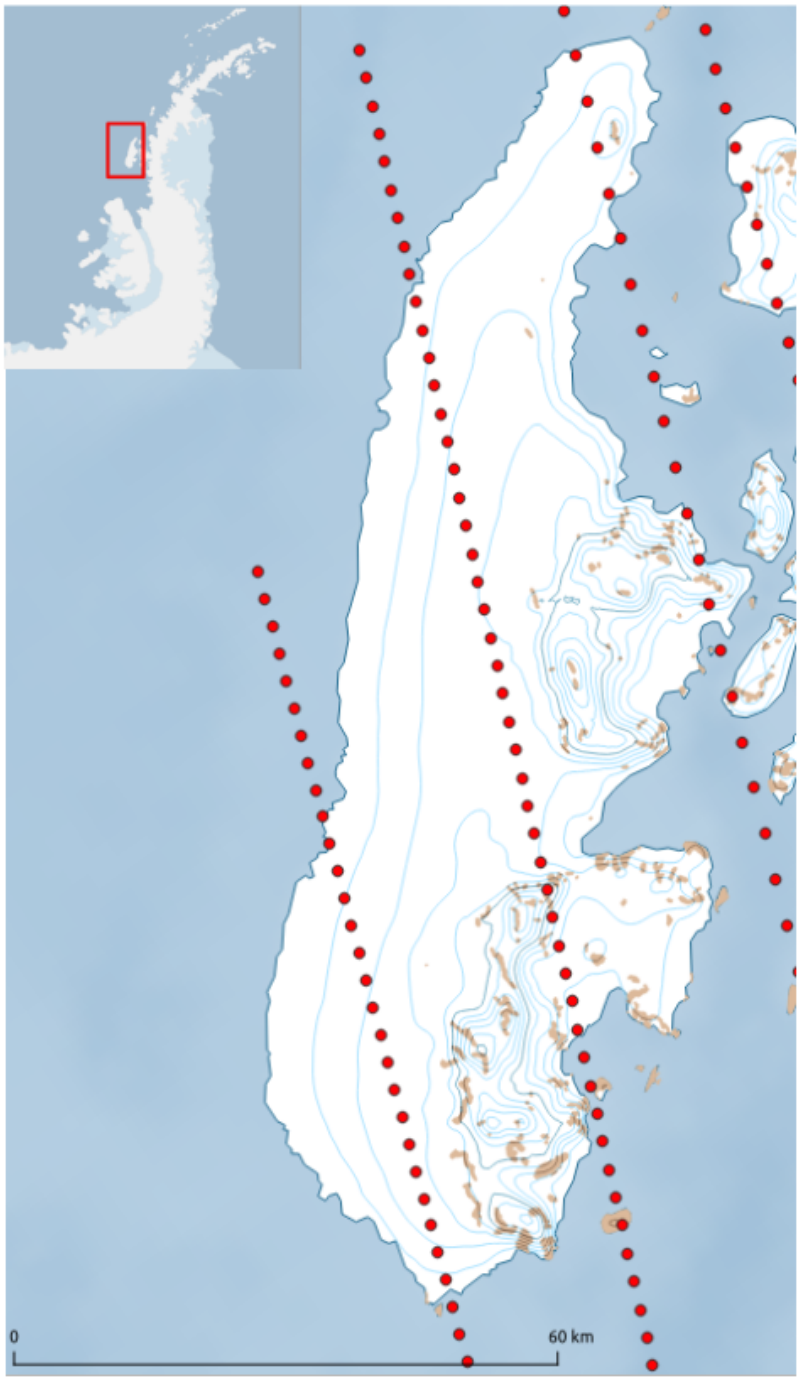


Figure 2.2: Adelaide island with camera positions (red dots)

"CA026433R0058" corresponds to an image captured during flight 0264, using the right roll (33R[1]), and has a frame number of 0058.

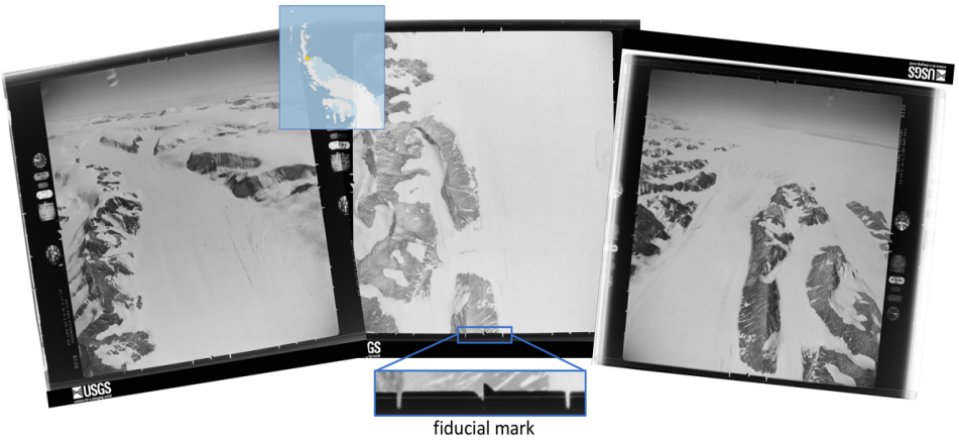


Figure 2.3: Example image triple from Antarctic TMA, consisting of a left-oblique, a vertical and a right-oblique image.

In total, this archive consists of around 330.000 grayscale images, which were digitally scanned in 2009 by the United States Geological Survey (USGS) at a resolution of 25 micron/1000dpi and made publicly available at their website (University of Minnesota 2023) and their FTP-server (Polar Geospatial Center 2023). With this resolution, all features on the historical images are recognizable and even small or objects in the distance can be recognized. However, the scanning process was not flawless, as in many images scanning artifacts (e.g. Newton rings) are introduced.

The images are available as TIFF files with an image size of around 10000×10000 pixels and an 8-bit depth. It is important to note that these images are not geo-referenced, and only have a manually estimated indication of their position. Although the photo centre (position of the camera) is reported for every picture in the Antarctic Polar Stereo-graphic coordinate system (EPSG:3031), a visual inspection has revealed that these coordinates can be inaccurate by several kilometres.

We define six different classes to be segmented in the images, as described in Table 2.1, together with some notes of their influence on the segmentation. Examples of the classes are shown in Figure 2.4.

[1] 33 is a number describing that the camera is looking right, similar to a left (31L) and a vertical (32V) roll

Class	Notes
Ice	Only close to the water
Snow	Most dominant class and can be found on almost every image
Rocks	Small structures scattered in the images, usually easy to segment
Water	Second most dominant class
Clouds	Most difficult to segment; can contain traces of other classes beneath
Sky	Can only be found on oblique imagery
Unknown	Used when a pixel cannot be attributed to another class. Not existing in the final segmentation & no loss will be calculated for this class

Table 2.1: Description of the classes used in the semantic segmentation

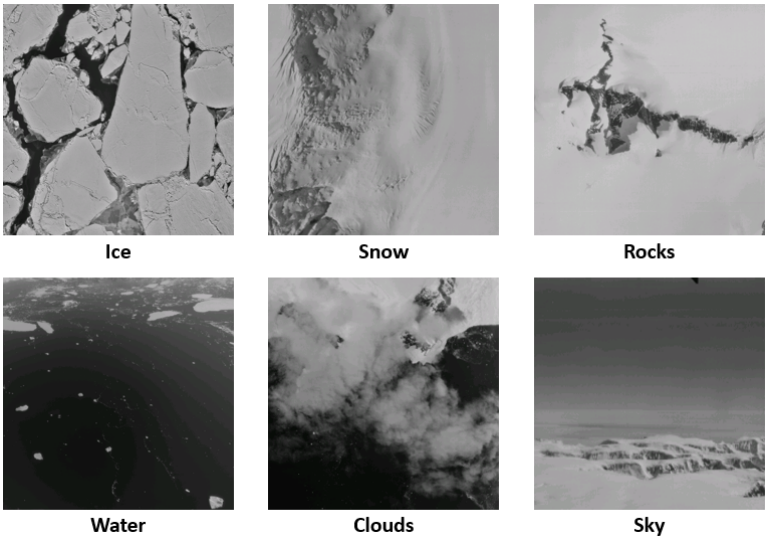


Figure 2.4: Example for the classes with the class being the prominent feature in the image

To train our model, we selected random images from the AP. This region is one of the most varying areas of the Antarctic with a diverse landscape, resulting in the most number of classes during segmentation. However, as there was no labelled training data available, we created the training data ourselves, as will be described in Section 2.3.2.

2.3. METHODOLOGY

In the following, the data pre-processing, the creation of the training data, and different attributes and design decisions of the segmentation process are explained. We compare the training and validation performance of different model parameters, and based on this, choose the best-performing combination of parameters. These are used to train a final segmentation model for more epochs. The parameters of this model can be found in Subsection 2.4.3.

2.3.1. DATA PRE-PROCESSING

During pre-processing, the prevalent borders in the images, as can be seen in Figure 2.3, must be removed. These borders do not contain any semantic information for the scenes and will only limit the efficiency of the model. All images contain fiducial marks that describe the limits of the borders. Using the free library of dlib (King 2009) and computer vision algorithms (e.g. canny edge detection or Hough transform), these fiducial points can automatically be recognized and used to separate the inner part of the images from the borders[2]. Contrast enhancement, like used by McNabb *et al.* (2020) for historical images cannot be used on the data: For some images it improves the segmentation quality, but for other images with scanning errors it decreases the quality of the image & therefore does not improve the general quality of the model.

2.3.2. TRAINING DATA

To generate the training data, we applied an unsupervised neural network for image segmentation to the raw images. This process produced preliminary image segments by identifying and grouping similar regions within the images. Various models of unsupervised segmentation are available, such as those described by Kirillov *et al.* (2023) and Kanezaki (2018). For this study, we adopted the approach

[2]Code is available on Dahle (2023)

outlined by Kanezaki, which follows three main criteria: (1) pixels with comparable features are aggregated under the same label; (2) pixels that are spatially contiguous are also grouped under a single label; and (3) the overall number of unique labels is minimized to simplify the segmentation.

However, the unsupervised segmented images must be further processed to use them for training as they contain misclassified pixels and do not always match the images perfectly. Furthermore, these segments only have consecutive numbers as labels and contain no semantic information. The following steps were applied to improve the unsupervised images: (1) Renumbering segments: The segments created by the unsupervised segmentation could consist of multiple, non-connected parts. These parts are assigned a new number so that every unique segment has its own number as a label. (2) Removing small segments: we remove segments under a threshold size of 20 pixels to simplify the segmentation. Smaller segments are often considered to be noise or irrelevant to the classification task. (3) Filling voids: The removed segments from the previous step are filled with their surrounding pixels by using a watershed algorithm (Kornilov and Safonov 2018). (4) Separating segments: Sometimes, only one segment is created, where in reality two different classes are present. For instance, in Figure 2.5, which displays some examples of self-labelled training data, the classes sky and snow at the top of the left middle image are incorrectly merged in one single segment. These segments are separated manually after visual inspection.

To facilitate efficient labelling, we developed a tool in Python that allows segmenting images using the steps mentioned above. This enables for example relabelling of already labelled images together with adapting segments (e.g. separating) and can be found at Dahle (2024a).

To distinguish the segments and assign the correct classes, the spatial context of the images must be known. Creating labelled images is very time-consuming so the number of images for training and validation was limited to a total of 80 images. Of these, 80% was used as a training set and 20% as a validation set. As the data is highly imbalanced in class occurrences (see Table 2.2), the images cannot be assigned randomly to one of the two sets. This would have resulted in an unequal class distribution in both sets, making the model biased towards certain classes. Instead, the classes are divided using iterative stratification (Sechidis, Tsoumakas, and Vlahavas 2011). This is a technique where a data set is divided into smaller subgroups all containing a similar composition of classes as the whole data set. Using this technique ensures an equal class distribution for both the training and validation set.

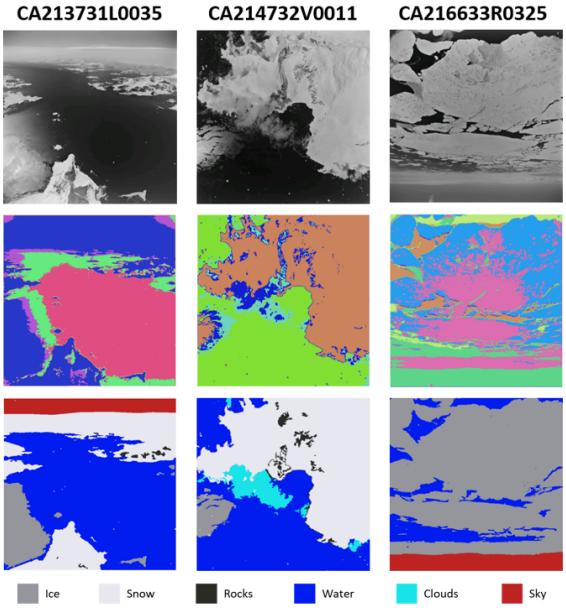


Figure 2.5: Examples of self-labelled images with raw images at the top, the unsupervised segmentation in the middle (colours are assigned randomly) and the final obtained 'ground truth' at the bottom

	Ice	Snow	Rocks	Water	Clouds	Sky	Other
Train	6.99	67.74	2.52	7.6	7.34	6.27	1.55
Validation	2.98	60.9	0.37	10.12	20.62	4.45	0.55
Complete	6.04	66.12	2.01	8.2	10.5	5.84	1.31

Table 2.2: Class composition of train and validation set in percentage

2.3.3. U-NET

In this work, we utilize a U-net (Ronneberger, Fischer, and Brox 2015), a model popular for image segmentation, which was adapted successfully for geospatial tasks (e.g., (Heidler *et al.* 2021), (Baumhoer *et al.* 2019)). In this model, Convolutional Neural Networks (CNNs) and Fully Convolutional Networks (FCNs), two special types of neural networks (Jiang *et al.* 2019), are combined in a U-formed structure. This architecture offers several advantages that are particularly relevant to our study. It can accommodate input images of varying sizes and is specifically designed to perform well even when training data are scarce; a situation we encounter with our set of 80 training images. This extremely limited number of images is also reason for a worse performance of other popular segmentation models like just using FCNs,

Deeplab (L.-C. Chen *et al.* 2018), or SegNet (Badrinarayanan, Kendall, and Cipolla 2017). This is especially noticing for the segmenting of small structures in the images. Despite the computational intensity required by the U-net during the training phase, the resulting model is computationally efficient.

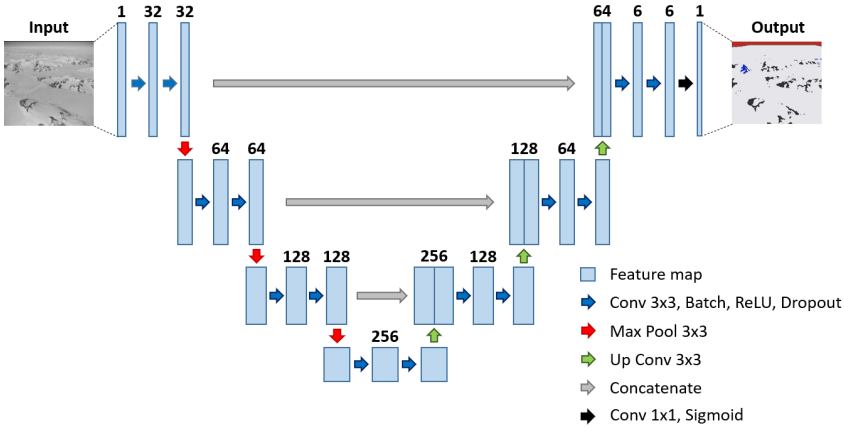


Figure 2.6: The U-net takes a grayscale image as input and returns for each pixel and class a probability, which can be used to create a segmented output image

Our baseline U-net model consists of four layers, as depicted in Figure 2.6. This structure is based on Kattenborn, Eichel, and Fassnacht (2019) and contains multiple encoders (the contracting path) and multiple decoders (the expansive path). The encoders, situated on the left half of the figure, are a classical classification network where convolution blocks are applied followed by a max-pool downsampling to encode the input image into feature representations at multiple different levels. The decoders, located on the right half of the figure, semantically project the discriminative features learnt by the encoder onto the higher resolution pixel space to achieve dense classification at the pixel level. The decoder consists of up-sampling and concatenation followed by regular convolution operations.

Each encoder/decoder in the network is built using the same components with the same attributes. The components include `Conv2D`, `BatchNorm`, `Dropout`, `ReLU`, `MaxPool2D`, and `ConvTranspose2D`. `Conv2D` is a convolutional layer that convolutes to additional feature maps with a kernel size of 3, followed by a stride of 1 to maintain image size. `BatchNorm` normalizes the input batch of 4 images by re-centering and scaling to make the network more stable and to converge faster (Ioffe and Szegedy 2015). `Dropout` temporarily disables 20% of the nodes in the block during training,

making the learning process more challenging, but reducing the chance of overfitting. `ReLU` is the activation function used in our network, which is short for rectified linear units (Goodfellow, Bengio, and Courville 2016). It essentially removes all negative values from the output by setting them to zero. `MaxPool2D` downsizes the image to reduce the computational cost, using a kernel size of 3 with a stride of 2 to halve the image. `ConvTranspose2D` is a transposed convolutional layer that doubles the output image size compared to the input image size, also using a kernel size of 3 and a stride of 2.

The U-net model used in this work reduces the image size while increasing the number of feature maps in each encoding block, except for the first encoder/decoder, effectively reducing the computational cost. After all encoders/decoders are applied, the output image size equals the input image size and consists of 6 channels, one per class, each containing a value between 0 and 1 describing how likely it is for each pixel to belong to that particular class. The segmented image is generated by applying the sigmoid function (Goodfellow, Bengio, and Courville 2016) to the data and selecting the class with the highest probability for each pixel.

The parameters in Table 2.3 are commonly used in neural networks and remain consistent for all combinations of different segmentation models. The Adam-optimizer, a popular optimization algorithm in machine learning, is used to adapt the learning rate during run-time for faster convergence and better performance (Kingma and Ba 2014). As the memory size of the used GPU (NVIDIA Tesla P100 with 16GB RAM) is limited, all data is split up into batches of the maximum possible size and the results of all batches together are averaged.

Type	Value
Optimizer	Adam
Kernel size	3
Batch size	4
Training percentage	80%
Validation percentage	20%

Table 2.3: Consistent model parameters

2.3.4. Tests

Different parameters of the U-net will be tested for their performance in image segmentation. For every parameter, a new model is trained. All models are applied on the same dataset, having the same images in the

training and validation set. The model with the parameters described in Table 2.4 is used as a baseline. For every test category, only the specific parameter of this category is changed. Tests for parameters will be done within six categories: additional layer components, learning rate, losses, number of layers, input size and augmentation, as elaborated below. Every model is trained for exactly 500 epochs with no early stopping.

Parameter	Setting
Nr. of layers	normal (4 layers)
Learning rate	0.001
Loss	Cross entropy
Input-size	1200×1200 (resized)
Augmentation	No Augmentation
Overfitting	Dropout & Batch normalization

Table 2.4: Parameters of the baseline model

Additional layer components
Next to the model components that are required for the model to learn, it is common to add additional components to each layer of the model. These components can help against overfitting as well as improve the quality of the model. For our segmentation, we use dropout and batch normalization, two commonly deployed elements in modern CNN architectures (Garbin, Zhu, and Marques 2020). Dropout layers, as utilized by Baumhoer *et al.* (2019), randomly set a percentage of neuron weights to 0 during training, withholding information from the model. This technique improves the model's robustness against overfitting but may also make it more challenging for the network to interpret the data. Alternatively, Garbin, Zhu, and Marques (2020) recommends using batch normalization, which re-centres and re-scales the inputs to the layers, resulting in improved generalization and faster training. In order to quantify the influence of these components, we are testing both components as a first step, each separately and also using none of these components.

Learning rate
The learning rate is a number between 0 and 1 and influences how quickly a model can adapt to new data and determines how much the weights of a model are changed at every training iteration. A very small learning rate may result in very long training times because the weights only change by a very small amount. Increasing the learning rate will speed up the process, but comes with the danger of learning sub-optimal weights too fast and, therefore, leading to worse model quality. In our tests, we compare learning rates of 0.001, 0.01 and 0.1.

Losses

In machine learning, a loss function is an essential component for the learning process to accomplish its assigned task. The loss function allows describing numerically how well the prediction fits the model. During training, this information is used to change the parameters of the model. The loss is usually expressed as a numeric value starting from 0, where a value of 0 indicates perfect predictions of the data with no deviations, and higher values indicate worse predictions of the model. As the class 'unknown' should not be prevalent in the final classification, no loss is calculated for this class. Numerous loss functions are available for different applications, including image segmentation as documented in Jadon (2020). Here, we compare three common loss functions for semantic segmentation. In the following equations x_i describes the input, y_i is the target and C is the number of classes.

1. Weighted Cross-Entropy loss
   The cross-entropy loss is one of the most commonly used loss functions in machine learning, with the weighted cross-entropy loss being an adaption for imbalanced data sets, as we deal with in our study. Based on a term from information theory, cross-entropy measures the entropy between two different probability class distributions. It is calculated with Equation 2.1. Here w_i is the weight of a class and p_i its probability. The weight of the classes is the inverse probability of each class.

$$Weighted\,Cross\,Entropy\,Loss = \frac{1}{C}\sum_{i=1}^{C} w_i y_i log(p_i) \tag{2.1}$$

2. Focal loss
   The focal loss is another loss especially suited for imbalanced datasets. It was originally designed by Lin *et al.* (2018) for object detection but was used with success for semantic segmentation as well. The loss function is a dynamically scaled cross-entropy loss, where the scaling factor decays to zero as confidence in the correct class increases so that the model is focussing on harder examples. In Equation 2.2, α_i is a weighting factor for each sample, γ is a tunable focusing parameter and p_i is the probability of a class. The value of these parameters depends on the dataset and typically involves setting the weighting factor higher for the minority class and experimenting with different values of gamma to balance the model's ability to learn from hard examples and generalize to new data. Gamma is set to 2, the weight of the classes is again the inverse probability of each class.

$$Focal\,Loss = \frac{1}{C}\sum_{i=1}^{C} -\alpha_i(i-p_i)^{\gamma} log(p_i) \tag{2.2}$$

3. DICE Loss
   This loss is based on the Sørensen–Dice coefficient, which is used to estimate the similarity of two different samples. Like focal loss, it is mainly used to address a class imbalance in images, mainly due to a common imbalance of foreground and background pixels. It is calculated with Equation 2.3. Here N describes the mini-batch size, a further and smaller subdivision of the batches. The choice of mini-batch size depends on factors such as available memory, dataset size, and model complexity, and is typically determined empirically by starting with a moderate size and adjusting based on performance and memory requirements.

$$Dice\ loss = \frac{2\sum_i^N x_i y_i}{\sum_i^C x_i^2 + \sum_i^C y_i^2} \tag{2.3}$$

Model depth
This term describes how 'deep' the model is, so how many encoding and decoding layers the model has. The model size has a direct influence on the number of parameters/weights that are trained. Typically, increasing the number of layers improves the accuracy of a model. However, as more parameters need to be calculated, the training time and model size will increase. Furthermore, it is possible to over-fit the data. On the other hand, it is also important not to have too few layers, as complex scenarios may not be learned, which means the model is underfitting.

To test the impact of model size, four different models with varying numbers of layers will be examined, including 2, 3, 4, and 5 layers.

Input size
Although the U-Net allows for flexibility in input image size, the images were resized prior to training. The images have an original size of around 10000×10000 pixels, which is too large to be used for training due to time and memory constraints. To address this issue, two different options were examined during parameter testing:

1. Resizing images: all images are resized to 1024×1024 pixels. This method is easily applicable, does not change the total number of images and the complete image is still depicted. However, information in the image is lost, especially texture, which is required

for distinguishing classes with similar intensity values, like snow and clouds.

2. Cropping images: instead of using the complete image for training, a crop of 256×256 is taken from the image. This can either be a crop from the same location in every image or a crop based on a random location. In this way, no information on texture is lost. Although this method preserves information on texture, it presents some challenges for the dataset used in this study. Due to the class imbalance, taking a random crop means that the crops may also be imbalanced. In this case, the dominant class would be present in the majority of random crops, whereas other classes would only be present in a few crops. To mitigate this issue, we use the inverted random crop. A weight is assigned to every class, based on the occurrence of the classes in the dataset (see Table 2.2; classes with a smaller occurrence get higher weights), and a random crop is taken with a weighting.

Augmentation

In some cases, there may not be sufficient data to train a model, i.e. due to the lack of training data or the creation of additional training data being too time-consuming. A prime example of these cases is medical imagery, e.g. where some forms of cancer are too rare to have an adequate number of examples (Ayalew, Fante, and Mohammed 2021). For these cases, it is possible to apply data augmentation, a technique that is used to synthetically increase the amount of data, in order to provide a model with more samples and therefore increase the quality of the model. It is possible to either copy and adapt existing data or create completely synthetic data. An example of data augmentation in geo-sciences applications can be found in Feng *et al.* (2022) for hyperspectral image classification. The augmented data are not created before the training nor added to the pool of available images. Instead, whenever the images are required for training an epoch of the model, they are randomly augmented with any of the augmentation methods, each applied with its own probability.

For this paper, we focus on the first option. We test different augmentation methods, used individually and all combined. The following methods are tested: (1) Flipping: The images are randomly (p=0.5) flipped vertically and/or horizontally. (2) Rotation: The images are randomly (p=0.5) rotated 90 degrees for a random number (1-3) of times. (3) Brightness: The pixel values in the images are randomly (p=0.5) increased or decreased by a random number (from 1 to 10). (4) Noise: Gaussian noise is added randomly (p=0.5) to the images. (5) Normalize: The image values are normalized from 0 to 255 to a range

from -1 to 1. Figure 2.7 illustrates a visualization of these augmentation methods (except for normalization, as there is no visual change for this augmentation).

In addition to the benefits conferred by an expanded training dataset, the implementation of image augmentations such as flipping and rotating offers further advantages. These augmentations can simulate common scanning errors, such as images being captured upside-down or rotated by 90 degrees. By introducing these variations during training, the model can be conditioned to effectively process and segment such erroneously scanned images. Similarly, augmentations that adjust brightness and apply Gaussian noise enable the model to better handle images that are underexposed, overexposed, or affected by scanning artifacts. This preparatory step enhances the model's robustness and its ability to generalize from a broader range of input conditions.

Figure 2.7: Four different augmentation methods of the training images together with their probability.

2.3.5. POST-PROCESSING

Another important step to further improve the segmentation quality is post-processing. Our post-processing consists of the following steps: First, the images are resized to 2000×2000 pixels to speed up the post-processing. Subsequently, patches with a size smaller than 50 pixels are removed from the segmentation and filled with the values of surrounding pixels via the watershed algorithm. Next, the images made with cameras facing down vertically are checked for the presence of segments of the class sky. If present, these segments will be replaced with the value of the surrounding pixels via a watershed algorithm, since it is impossible for down-looking images to observe the sky. In non-vertical images, the class sky is enlarged to fill complete rows, when a row exceeds 50% of sky pixels. Furthermore, some combinations of the class sky are physically impossible. Examples would be small patches of the class sky that are located far away from the sky at the top of an image or small patches of the class snow inside the sky. As can be seen in Figure 2.8, these patches are automatically recognized with computer vision methods, are removed and then again filled with

the value of the surrounding pixels. Finally, some logical criteria are applied to handle the confusion of the classes rock and water. Whenever a cluster is smaller than a threshold of 100.000 pixels and does not neighbour any cluster of its class, the class is changed (i.e., a rock cluster gets changed to water and vice versa).

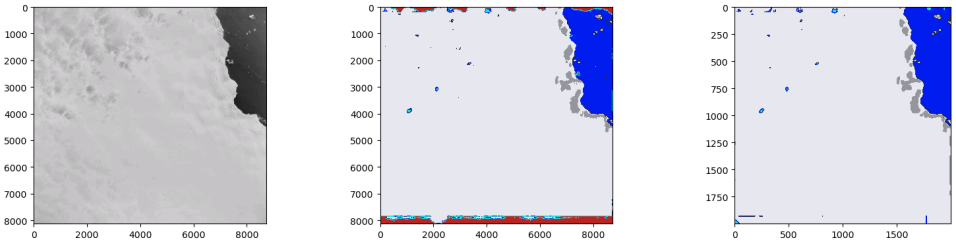


Figure 2.8: For the initial segmented image (middle) post-processing is applied to improve the quality of the final segmentation (right). Here the sky is removed from the borders of the image.

2.3.6. Evaluation metrics

The evaluation of the image segmentation results can be complex as both the accuracy and the correct localization of the segmented images must be considered. Furthermore, an imbalanced distribution of the prominent classes during segmentation can lead to a statistical bias with incorrect high evaluation scores (Müller, Soto-Rey, and Kramer 2022). Finally, not all incorrect segmentation labels are equally wrong. For example, confusion between the classes 'Ice' and 'snow' may have fewer implications than confusion between the classes 'sky' and 'snow'.

Almost all commonly used metrics are based on a computation of a confusion matrix, in which a pixel has one true class and one predicted class, which can result in four different outcomes: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Here, the models will be evaluated with the loss, accuracy and the F1-score. Precision and recall are important metrics required for the F1-score and therefore explained as well: (1) Accuracy: The most classical metric, which gives the percentage of pixels segmented correctly. However, this parameter is skewed by imbalanced datasets. If for example, 90% of an image contains the class 'snow', a model that learns to always classify every pixel to 'snow' will have an accuracy of 90%, even though the model is not useful. (2) Precision: a measure of quality, this parameter shows the number of correctly segmented pixels in relation to all pixels with this class attributed. It shows the ability of the model to segment

a class with only pure results, with as few false segmented pixels as possible. Yet, it does not reflect the ability to capture all pixels of this class category. (3) Recall: a quantitative measure which shows the number of correctly segmented pixels in relation to all pixels with this class in reality. It indicates the ability of the model to segment as many pixels of a certain class correctly but does not take into account whether other classes are wrongly predicted as this class. (4) F1-Score: This parameter is a combination of both precision and recall using a harmonic mean. Only when both values have a high score, does the F1-score as well have a high score. It is generally seen as a more accurate score than accuracy for imbalanced datasets, even though it is less intuitive.

2.4. Results & Discussion

In this section, we present the results of our study on optimizing hyper-parameters for semantic segmentation. We analyze the performance of the model using various hyper-parameters and present our findings. Based on these results, we identify the optimal set of hyper-parameters and train a segmentation model with them. We evaluate the performance of the optimized model on a set of 20 test images and provide a detailed discussion of the results. Additionally, we apply the model in a use-case to a larger set of images of Adelaide Island in Antarctica to demonstrate its effectiveness in real-world scenarios.

2.4.1. Performance of the base-model

Figure 2.9 displays the training and validation performance of the base model with the most basic parameter settings. Even with these most basic settings, the model is observed to be learning, as evidenced by the decreasing training loss and increasing evaluation values.

During training, an interesting pattern emerges for the cross-entropy loss: The validation loss remains constant or even increases, while the training loss is constantly decreasing. This is typically an indication of overfitting, where the model learns to just predict the training data, but not to create generalizations over the data. However, the evaluation parameter values continue to improve for testing and, importantly, for validation as well.

Further experimentation with the base model using more unseen data, however, reveals that this model can segment better than the model

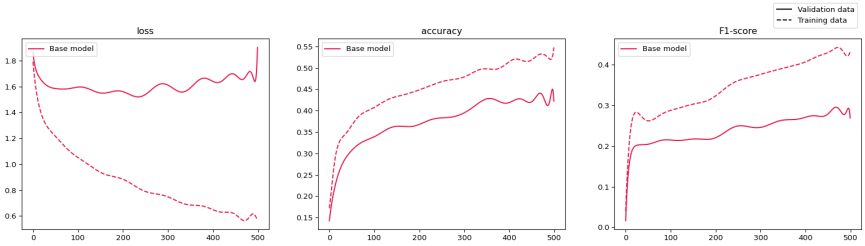


Figure 2.9: Results for base model with loss, accuracy and F1-score for both validation and training data

with fewer iterations. Therefore, classical overfitting does not appear to be the case here. Instead, this phenomenon can be explained by characteristics of the cross-entropy and the input data. Whereas most predictions improve (better scores for the evaluation parameters), other predictions worsen (in this case the prediction of 'ice' as 'snow'). Furthermore, the prediction probabilities for all classes may be increasing, for both incorrect and correct classes, with the right class maintaining the highest probability. This leads to a correct segmentation but also results in a higher loss.

2.4.2. Training & validation performance of individual parameters

Figure 2.10 displays the training and validation performance of the individual parameters. The evaluation performance of the models (accuracy & F1-score) generally increases for each combination, indicating that the models are learning to segment the images. However, depending on the exact parameters, the performance can differ drastically. The average validation values over all different parameter tests are 0.47 for accuracy and 0.48 for the F1-score.

The following can be noticed for the performance of the individual parameters:

Additional components
The results for the test with additional components align with the expectations. Dropout, in which a random percentage of the neurons is disabled so that they do not transmit information while learning, makes it harder for the model to train, leading to higher loss values for both training and validation. Even though the model's performance is decreasing, the chance of overfitting is smaller. On the other hand,

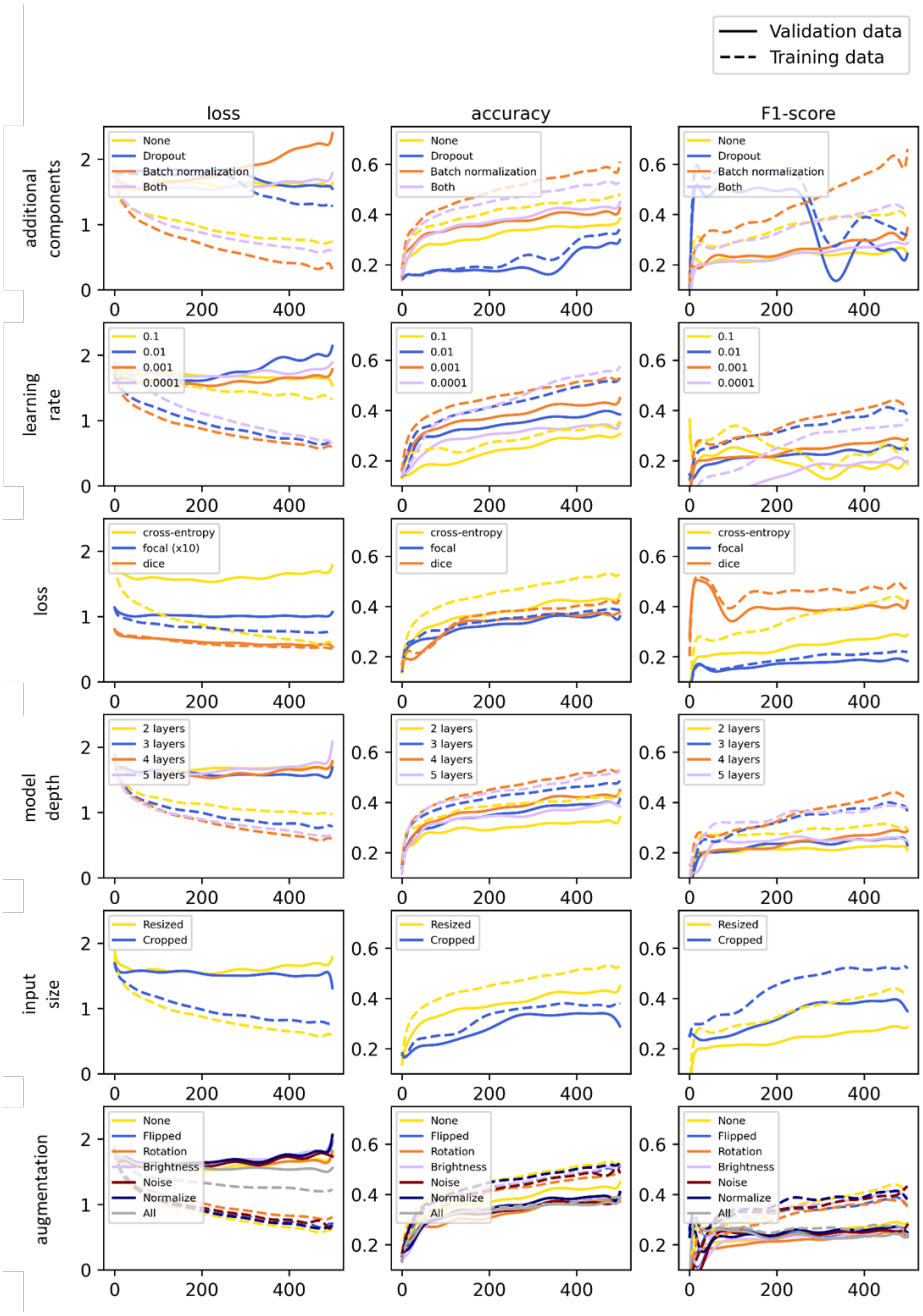


Figure 2.10: Results for the models with the different parameters discussed in Subsection 2.3.4 with each test showing loss, accuracy and F1-score for both validation and training data

batch normalization increases the model's performance, resulting in a smaller loss and increased performance for all evaluation parameters. For these reasons, both components were used in the subsequent tests.

Learning rate
In general, higher learning rates lead to worse model performance. In the extreme case of a learning rate of 0.1, the model is hardly learning anything and the loss remains nearly constant. The model is converging too fast to a sub-optimal solution and gets stuck iterating around it Ketkar and Ketkar (2017). For all other learning rates, the model is improving and learning. With smaller learning rates, the weights are upgraded with smaller steps, allowing the model to converge to more optimal solutions. However, this comes at the cost of increased learning time.

Loss type
All three loss functions allow the model to learn the image segmentation and the evaluation values are improving. For the cross-entropy loss, the training loss is decreasing, but the validation loss remains constant and even increases again at the end, as explained earlier. The focal loss yields minimal loss values in comparison to the other two losses. However, this is not related to a better performance of this loss, instead, the loss values are calculated differently. The evaluation performance of this loss is the lowest of all losses. Whereas the cross entropy loss yields higher accuracy, the dice loss performs better for the F1-score.

Model depth
Regardless of the model depth, the segmentation model is able to improve, but the two-layer model is less capable of learning segmentation, resulting in higher loss and lower evaluation parameters. To a limited extent, this also holds for three and five layers, whereas four layers are the sweet spot with the best loss and evaluation parameters. With more layers, too many parameters are included, so the model is overfitting. With fewer parameters, the model lacks the complexity to capture the relationship between the input and output variables for segmentation (underfitting).

Input size
When comparing the performance of the different input sizes in Figure 2.10, an interesting behaviour similar to the lost type can be noticed. For training and validation, the model with resized input data has a higher accuracy, while cropping the input data yield better F1-scores for both. These conflicting results suggest that neither

input size outperforms the other in terms of overall evaluation metrics. However, considering the better validation loss values and that the F1-score is better suited to evaluate imbalanced data sets, cropped images seem to be a better choice. Comparing the models for the different input sizes visually reveals that the resized model is better at giving the correct classes, whereas the cropped model is more accurate in extracting boundaries between classes.

Augmentation
Including augmentation data in the model did not have the desired effect, as most single methods yield worse evaluation performance than the model without augmentation, even though the loss is smaller. The only exception is the 'noise' augmentation, which delivers better performance at the end of the model training. Interestingly, combining all the augmentation methods leads to the worst results in terms of both loss and evaluation metrics.

Using a single augmentation method only has a small influence, with 'noise' improving the quality the most. However, some augmentation methods, like applying 'rotation' and, to a smaller extent, 'flipping' even decrease the performance of the model. This is in line with the findings of Engstrom *et al.* (2019), who reported that perturbations such as translations and rotations can degrade the performance of a neural network.

2.4.3. Numerical quantification

The insights from the tests for individual parameters are used to train a new model with the combination of the best-performing hyper-parameters. These parameters are described in Table 2.5.

Parameter	Value
Additional components	Dropout & Batch normalization
Learning rate	0.001
Loss	Dice Loss & Cross entropy
Model depth	4 layers
Input size	Cropped & Resized
Augmentation	Noise

Table 2.5: Parameters of the final model used for the numerical quantification

Both dropout and batch normalization are used against overfitting. A learning rate of 0.001 and model depth of 4 are used as the best-performing parameters. For the augmentation parameters, only

noise was included, as it brings the most performance gain. For the loss type and input size, there is no single best-performing hyper-parameter or optimal solution, as it depends on whether accuracy or F1-score is more important.

2

During the training of some initial models, it was noticed through visual inspections that the model using cropped input data is better at detecting the boundaries between classes, whereas the model with resized input data excels at detecting the correct classes. This effect can even be enhanced by combining the different input sizes using different losses: dice loss for cropped input data and cross-entropy loss for resized input data. Therefore, we utilize the strengths of both combinations to further improve the quality of the segmentation by using both models together and merging the results. The segmentation from the cropped image serves as the base segmentation, and some clusters (enclosed segments of the same class values) from the segmentation of the resized images are used as well: The position and class of a cluster from the resized image replace the classes at the same position in the cropped image, as the segmentation performs better for the classes sky, ice, and clouds in the resized model. Figure 2.11 visualizes this combining step: The cropped segmentation is the base image, whereas clouds & ice from the resized segmentation are taken from the resized image.

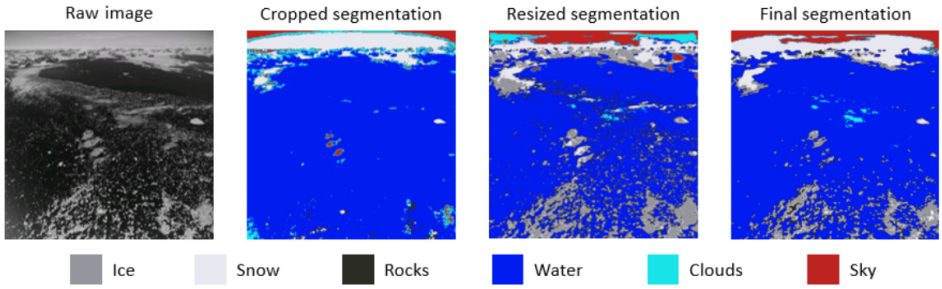


Figure 2.11: Example for combining two models with the images from left to right: raw image, cropped segmentation, resized segmentation and final segmentation after post-processing

The final segmentation is therefore based on a combination of two models with the best-performing parameters for each cropped and resized imagery. The first model utilizes crops of 512×512 pixels and the dice loss, while the second model utilizes resized images of 1200×1200 pixels and the cross-entropy loss. The cropped model was trained for 500 epochs, while the resized model was trained for 360 epochs. The total training time was 74 hours for the cropped model and

45 hours for the resized model.

The model's numerical quantification is based on an evaluation of 20 semi-manually labelled (see Subsection 2.3.2) images as ground truth. These images are randomly selected from the complete archive. Figure 2.12 visualizes the segmentation of all twenty evaluation images with the optimal model. The accuracy, precision, recall, and F1-score are 0.73, 0.84, 0.72, and 0.71, respectively, based on the right confusion matrix of Figure 2.13.

The difference between the confusion matrix of the base model and the optimized model is evident. In the base settings, the model tends to classify pixels as snow with a high probability due to the imbalance of the dataset, where snow is the most dominant class. However, the only two exceptions are the classes water and sky, which have opposite colour values (black instead of white). In contrast, the model with optimized hyper-parameters performs better in terms of correct classifications. Although the accuracy for some classes may have decreased, such as sky from 0.69 to 0.66 or snow from 0.9 to 0.81, this is likely due to the introduction of noise into the data (In return the model performs better on under- or overexposed images) Nonetheless, the model with additional hyper-parameters is more stable and better suited to handle new, unseen data.

When looking at the example images of Figure 2.12, for every image the general semantic meaning can be extracted successfully. However, for some scenarios and images, the model underperforms. Notably, in images T4 and T15, rock structures are accurately recognized, but the model fails to distinguish between snow and sky, particularly when snow covers a significant portion of the ground. This can possibly be attributed to the underexposure of the camera, in which the ground is too dark and resembles the sky of the training data. The augmentation could not help to mitigate this, so more underexposed imagery could be included in the training data to address this limitation, as it would teach the model to differentiate correctly between underexposed snow and sky. Another challenge for the model is distinguishing between water and rocks, as both share similar attributes for the network. However, with post-processing, most of these confusions can be cleared and the correct class can be assigned.

It is worth noting that mismatches between classification and evaluation may be due to errors in the ground truth data. Generating ground truth data is a time-consuming process, and limitations in the accuracy of the data may occur, especially when dealing with small structures. Additionally, there may be limitations in correctly assigning classes. For example, clouds may not completely obstruct the surface beneath them,

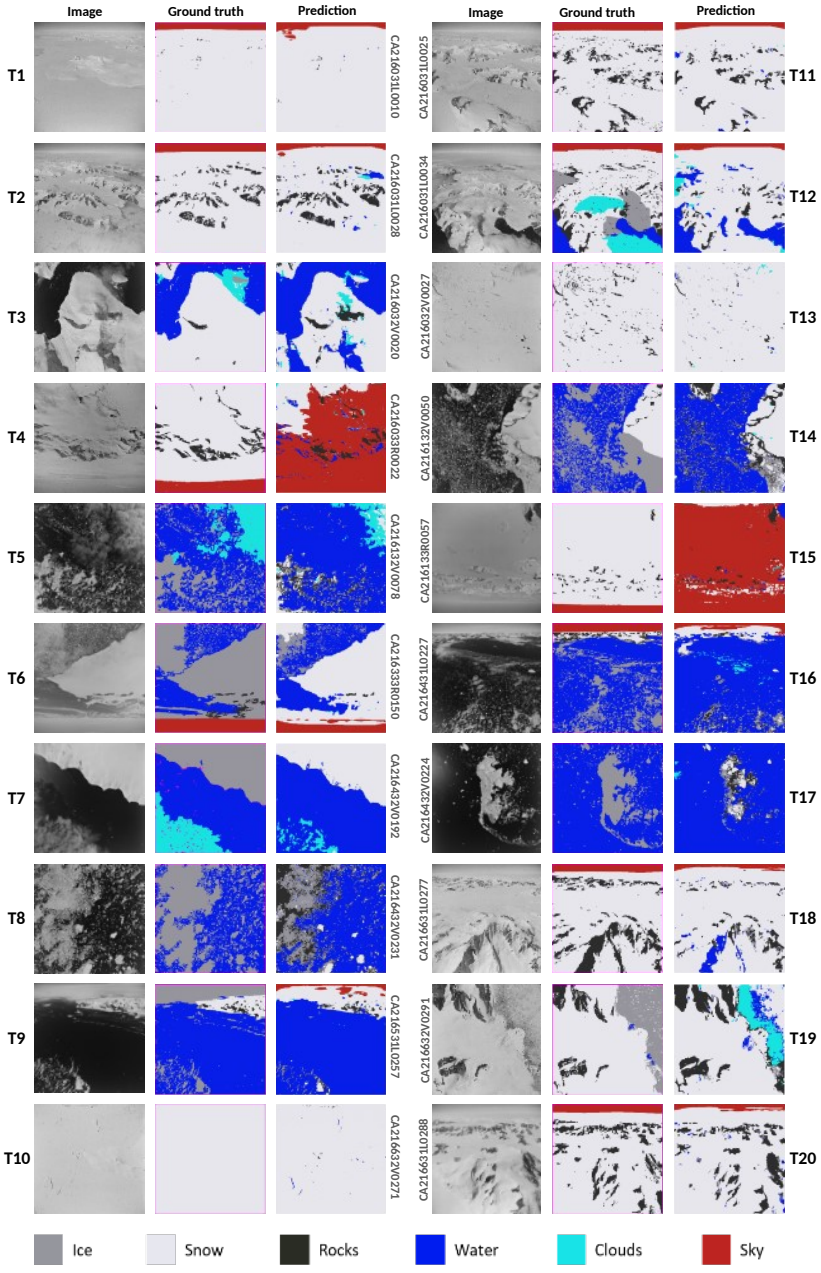


Figure 2.12: Segmentation of twenty randomly selected images (Test image T1 - T20). For each triplet, the raw image is on the left, the manually created ground truth in the middle and the results of the segmentation on the right

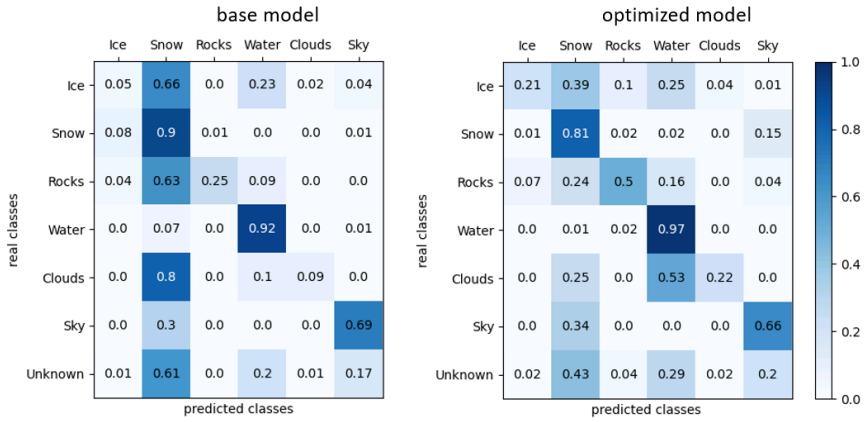


Figure 2.13: Normalized confusion matrix for the base model (left) and the optimized model (right) with real classes on the left and predicted classes on the top

and distinguishing between the clouds and the surface beneath can be challenging. Ice often gets covered with snow, leading to confusion between the two classes. In image T10, the ground truth data identifies the entire image as snow, although the model correctly identifies small rock structures. This leads to lower evaluation scores as it is counted as an error. Nonetheless, it is essential to acknowledge these limitations when interpreting evaluation results.

When compared to the evaluation conducted in Dahle, Tanke, *et al.* (2022), the current assessment, at first sight, does not demonstrate significant improvement in quantitative performance. However, the current study utilized a new set of images that featured more complex scenery, with a higher proportion of cloud cover and greater underexposure. This poses significant challenges for semantic segmentation models as they obscure important details and create inconsistencies in image quality. Despite these more challenging conditions, the new model performed well, maintaining its performance on even more difficult images.

Our model's segmentation approach closely resembles that of human labellers. It performs well in scenes that are easily recognizable by humans while encountering similar difficulties in challenging scenes. However, the model surpasses human capabilities in segmenting smaller structures with greater accuracy (for example T10 or T14 of Figure 2.12). On the other hand, the model struggles with under- or overexposed images. Despite this limitation, the model's speed is a significant advantage over human labellers, taking only 10 seconds on a standard

computer to segment an image compared to 20 minutes on average for a human (based on the creation of the training data for this model).

Overall, while the model shows promising results in adding semantic information to the images, there is still room for improvement, particularly in distinguishing between snow and clouds. Future research could focus on improving the model's performance and incorporating validation data to evaluate its accuracy further.

2.4.4. USE-CASE ADELAIDE ISLAND

To demonstrate the capabilities of our optimized segmentation model beyond the validation set of 20 images, we applied the model to the vertical images of Adelaide Island from 1969 and quantified the results. Figure 2.14 shows the 118 segmented vertical images of Adelaide Island, that were roughly geo-referenced based on the camera positions shown in Figure 2.2. It is important to note that the position is only an approximation based on the (uncertain) reported camera positions, which is evident from the difference in coastline positions between the segmentation and the base map.

Due to the lack of validation data for Adelaide Island, we cannot provide a numerical assessment of the segmentation model's performance in this area. However, a visual inspection of the segmented images reveals that the model successfully adds semantic information to the images, particularly in differentiating between land, water, and ice, as is shown by the segmentations in bounding boxes A and B. Nevertheless, the model has limitations, as can be seen in bounding box C where it struggles to differentiate between snow and clouds, especially in the centre of the clouds, where there is significant visual overlap with snow. However, due to the limited data quality of these images, these areas are also difficult to segment for humans.

2.5. CONCLUSION

In this work, we successfully segmented an archive of historical aerial imagery of Antarctica using a U-net-shaped neural network. We compared and discussed the training and validation performance of different hyper-parameters. Combining the best-performing parameters, we applied a model on a test set and segmented images of Adelaide Island.

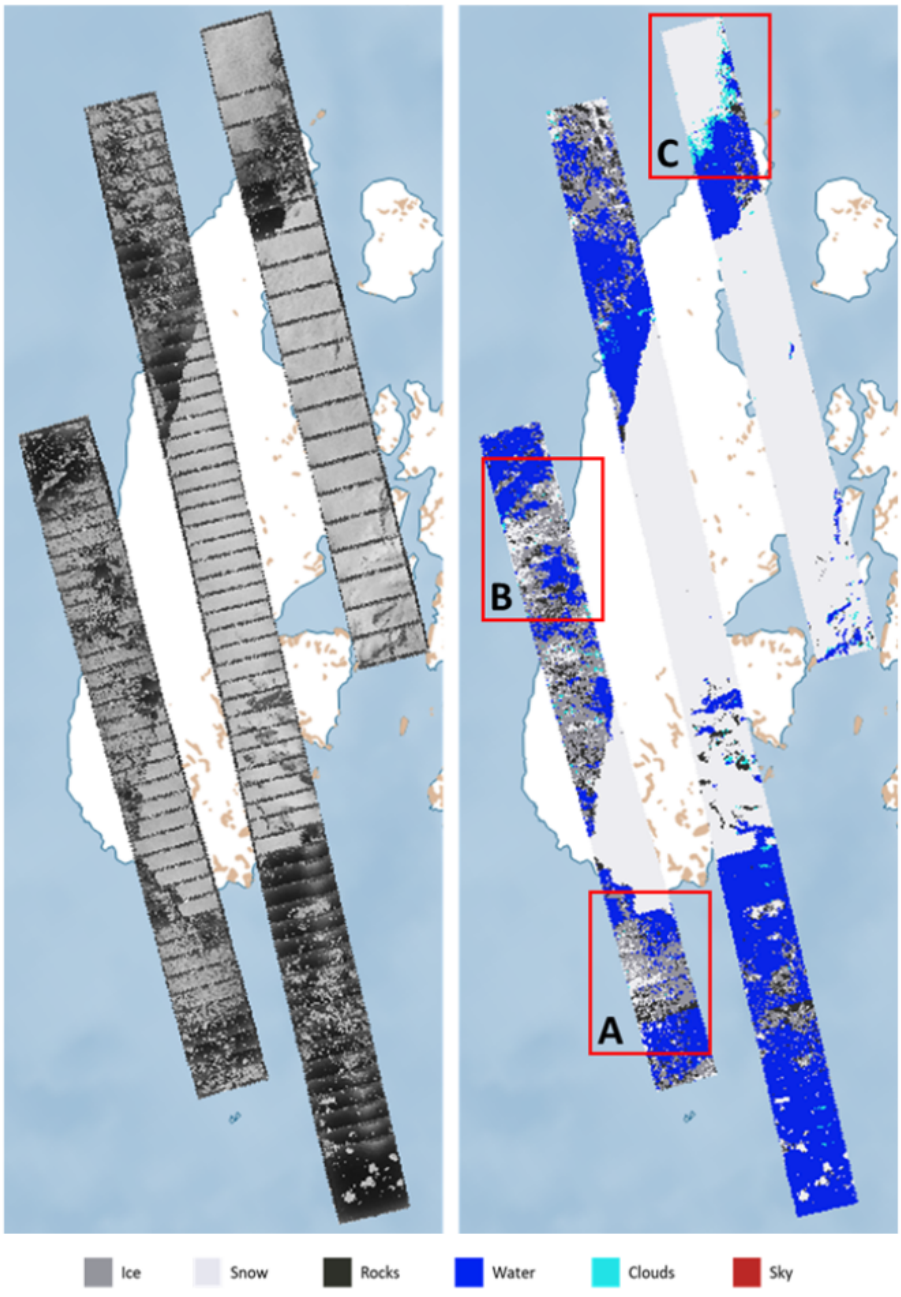


Figure 2.14: 118 raw images (left) and segmented images (right) for Adelaide island

Several other segmentation methods exist, e.g., k-means clustering, random forest approaches, or other deep-learning-based approaches (Lateef and Ruichek 2019). However, these methods seem not to be suited for this dataset, as they require either better quality data or much more labelled training data.

The proposed model is able to get the semantic meaning of a scene for the historical images of the TMA archive, even when using grayscale images with low contrast, conditions for which many other segmentation models would fail. The created model successfully learns to distinguish between most of the different classes with a certain confidence and does not get disrupted by unfavourable conditions, like poor image quality, limited spectral information, difficult semantic classes and only a few training images.

To our knowledge, no other semantic segmentation model exists that can work under these conditions. Based on the added information through segmentation, specific images with certain attributes can be selected, such as images located at the ice-ocean boundary, or images containing sky. This greatly facilitates case-specific inquiries into local conditions within the archive, thereby enabling various applications related to the historical conditions of Antarctica.

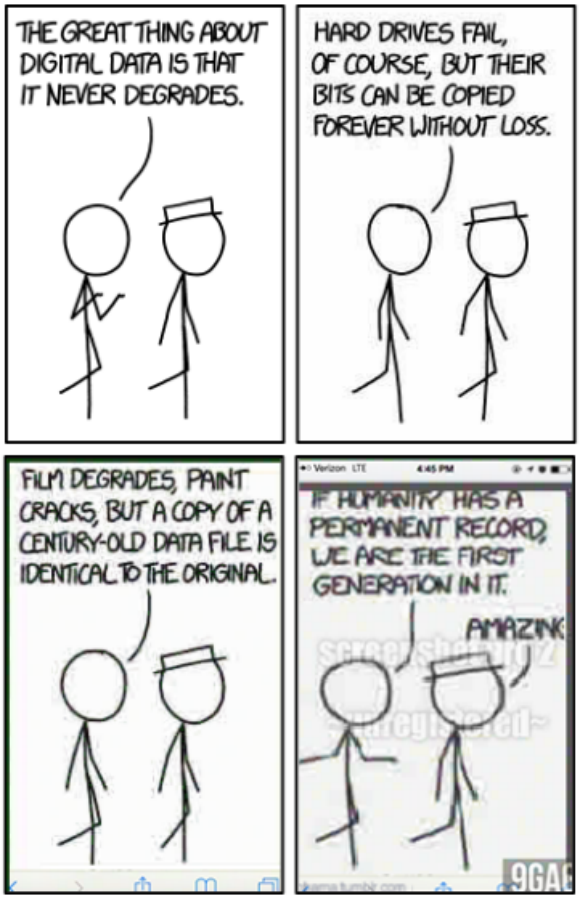
In theory, almost every combination of model parameters allows training a model for semantic segmentation and getting some meaningful output. However, the tests demonstrate that selecting the correct parameters can impact the segmentation's quality and even can account for training with a limited number of images and/or images with sub-optimal quality. The findings can be extended to the segmentation of other historical imagery to enable a better selection of hyper-parameters at the start of the segmentation.

When it comes to parameter selection for semantic segmentation, we can suggest the following guidelines: Parameters that have a direct impact on the model - such as model depth, learning rate, and the choice of loss function have the biggest influence on the quality of the model. While data augmentation is beneficial, particularly when dealing with a limited number of images, its impact is not as strong. However, it is important to keep in mind that there is no universally optimal parameter that can be applied across all scenarios; the choice of parameters must be tailored to the specific characteristics of the dataset and the objectives of the study.

Even though the training time of both models (cropped and resized) takes a significant amount of time, the segmenting of an image itself is fast, with an average time of around 10 seconds per image, including

post-processing. As a next step, we aim to extend the application of the model beyond Adelaide Island and use it to segment the entire TMA archive of 300.000 images, which will allow easy access to the data and encourage its further usage by the community. Besides filtering for specific classes, segmentation allows to correct oblique images which were scanned upside-down, so that the sky is at the top. It also enables the detection of rock outcrops in the images, which can then be used to address the poor geo-location of the imagery by matching with present-day outcrop data sets (Burton-Johnson *et al.* 2016). Furthermore, it can help improve applications, such as Structure-From-Motion (SfM) to create 3D models (Child *et al.* 2020), by excluding irrelevant Tie Points (TPs) (i.e. identical points in overlapping images) in the classes "water" and "sky". TPs detected in these classes are often of worse quality due to movements inside the class. The removal of these points enhances the quality of the 3D models and is a necessary step to automatize the process so that in future efforts the TMA Archive can be used to create historical 3D models at a large scale.

To further improve the quality of the model, we expect that the most effective option would be to increase the size of the training data set. However, even though other historical cryospheric datasets exist, e.g., for Svalbard, Greenland or Alaska (Bjørk *et al.* 2012; Girod *et al.* 2018; Knuth *et al.* 2023), these archives do not currently contain segmented images. Therefore, they cannot be used as training data without further manual work. A viable alternative is to acquire labelled training data from modern satellite images (e.g., Sentinel-2) and artificially degrade them to resemble historical data. Another possible option would be to include metadata of the images as additional data sources, such as whether an image is taken with an oblique- or vertical-facing camera, the flight height or even the date when the picture was taken (to account for seasonal effects). However, these additional sources should be selected carefully, as this would limit the use of this model for other images where these additional metadata are not available.



© xkcd 1683 - Digital Data

3

EXTRACTING METADATA FROM HISTORICAL IMAGES

Historical aerial imagery provides valuable data from regions and periods with limited geospatial information. A common method to utilize this data is through the generation of orthophotos and 3D models using Structure-From-Motion (SfM) techniques. However, many of these images were scanned decades after their acquisition and require geometric calibration, along with internal and external camera parameter estimation, for accurate reconstruction. Manual identification of key features, such as fiducial marks and text annotations, is labour-intensive, while existing automated methods struggle with poor-quality datasets.

This paper presents an automated workflow that combines computer vision and machine learning techniques to detect and extract these key features from historical aerial images. To address challenges related to image quality, we also introduce estimation protocols that compensate for missing or unreliable detections by leveraging redundancy across multiple flight paths. The methodology was evaluated on the Trimetrogon Aerial (TMA) archive, a collection of historical images from the Antarctic Peninsula (AP). Our test dataset comprised over 7,000 images from 20 different flight paths. The workflow demonstrated high success rates in detecting and extracting fiducial marks, image subsets, and textual annotations. Approximately 70% of the images provided usable focal length data, while fiducial mark detection exhibited high accuracy except in cases of severe scanning artifacts. Altitude data extraction from the altimeters proved to be the most challenging, with

This chapter is under review as *From Film to Data: Automating Meta-Feature Extraction in Historical Aerial Imagery* in the *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*.

successful results in only 15% of images due to degraded imagery. Despite these limitations, the automated workflow effectively estimated missing parameters, ensuring a more robust image reconstruction across flight paths. The code for this workflow is open-source and publicly available on GitHub at https://github.com/fdahle/hist_meta_extraction.

3.1. INTRODUCTION

Historical aerial images are a valuable source of information for a wide range of applications, from environmental (Heisig and Simmen 2021), cryospheric (Pope *et al.* 2014) to archaeological sciences (Cowley and Stichelbaut 2012), but still remain under-exploited (Kostrzewa 2024). A particularly prominent application involves generating orthophotos and 3D models through SfM techniques (Farella *et al.* 2022; Mestre-Runge *et al.* 2023; North and Barrows 2024). While SfM can estimate camera parameters from image data alone, its accuracy and reliability are significantly improved when internal parameters (e.g., focal length) and external information (e.g., camera height or position) are known (Stark *et al.* 2022; Zhang, Rupnik, and Pierrot-Deseilligny 2021). Many of these images are scanned with photogrammetric scanners at high resolution (Kostrzewa *et al.* 2024); however, the scanning process itself often introduces subtle distortions, which can affect the overall consistency of the dataset, Schulz, Cramer, and Herbst (2021). Therefore, to perform reliable geospatial analysis, scanned images must be normalized by calibrating their geometry based on fiducial marks.

Fiducial marks are reference points embedded in the original film to correct geometric distortions. While their film-space coordinates are typically provided in camera calibration reports, the corresponding image-space coordinates must be identified on each scanned image. Manual identification of fiducial marks is straightforward but time-consuming and labour-intensive. Unfortunately, in many cases, the original camera calibration reports have been lost over time, leaving only the images themselves. Additionally, scanned images often contain distortions, such as stretching and skewing, introduced by the scanning process, further complicating their use in precise geospatial applications.

Several automated methods have been proposed to address these challenges. Early research by Sun and Wu (2001) employed attribute-based mathematical morphology to segment fiducial marks. Later, Ye *et al.* (2016) introduced a semi-automatic approach based on circle fitting, specifically tailored for ARGON satellite imagery. Girod *et al.* (2018) proposed a shape-based method that identifies bright circular

blobs near the image frame as proxies for fiducial mark locations. Other approaches, such as those by Knuth *et al.* (2023) and Salach (2017), rely on template matching to detect fiducial marks across image sets. These open-source methods vary in robustness and typically require manual correction, especially when applied to degraded or low-quality scans. Commercial software packages, including for example Agisoft Metashape (Agisoft 2025), ArcGis Pro (Esri 2025), or Trimble Inpho (Trimble Inc 2025), offer powerful and largely automated tools for fiducial mark detection that typically perform well, even on challenging imagery. However, these tools are paid and closed-source, which limits accessibility for some users. Their proprietary nature also reduces transparency, as internal algorithms are not openly documented or modifiable. This makes it difficult to adapt or extend their functionality, and complicates integration into fully automated and reproducible processing workflows.

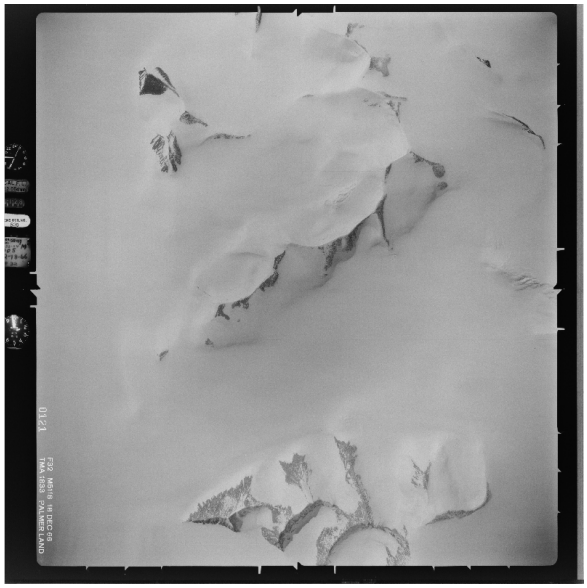


Figure 3.1: Example image CA183332V0121 from the TMA dataset

In this study, we utilize the Trimetrogon Aerial (TMA) archive, a collection of approximately 330,000 black-and-white aerial photographs captured by the U.S. Navy between 1946 and 2000. Around 2012, the imagery has been scanned at a resolution of 25 microns (1000 dpi) by the U.S. Geological Survey (USGS) and the Antarctic Geospatial Information Center (University of Minnesota) and is publicly available at Polar Geospatial Center (2023). An example image from the archive is shown in Figure 3.1. The focal lengths in this dataset range from 151.09 mm to 156.00 mm, while altitudes at the time of image acquisition vary

between 8,000 ft and 24,000 ft, depending on flight line and location. Notably, many images lack accompanying metadata, and approximate values for focal length or altitude are often unavailable.

Our ultimate goal is to extract three-dimensional information from these images using SfM. While in principle, internal and external camera parameters can be estimated during bundle adjustment (Mölg and Bolch 2017), the quality and reliability of the resulting 3D models can benefit from some prior knowledge, especially in the case of archival aerial imagery (AliAkbarpour, Palaniappan, and Seetharaman 2015; Knuth *et al.* 2023; Sevara *et al.* 2018). Historical datasets like the TMA archive often have near-parallel viewing geometries and limited image overlap (around 60%), which can lead to unstable or degenerate configurations with inaccurate approximated camera parameters. This challenge is further increased by strong variations in image quality, caused by improper storage, exposure issues during acquisition, and distortions introduced during scanning (see (Maiwald, Feurer, and Eltner 2023)).

Given the large volume of images, a key objective was to automate the processing as much as possible for all parameters located in the images. Therefore, we have developed a fully automated workflow for detecting fiducial marks directly from the images. Our approach emphasizes robustness, allowing to estimate positions even when detection fails in degraded or incorrect scanned images. In addition to fiducial mark detection, our workflow enhances the geospatial use of the images by extracting height information from three-pointer altimeters embedded in each image. Such altimeters indicate altitude using three separate hands, providing vertical data that can improve the accuracy of 3D reconstructions. Finally, we extract metadata such as focal length from typewritten annotations found on the images.

This paper presents our methodology for the automatic extraction of fiducial marks and height information, with a particular focus on the TMA archive. However, our workflow is designed to be adaptable and can be applied to other historical aerial image collections. To the best of our knowledge, no previous study has developed an automated approach for extracting height data from three-pointer altimeters in historical imagery. By addressing this gap, our work offers an important next step to the challenges of utilizing historical imagery for geospatial analysis and 3D modelling.

3.2. METHODOLOGY

The main objective is recovering metadata from scanned photos, whenever available. This metadata notably includes focal length

and image altitude, which can be estimated through text extraction. Altitude information can also be derived from embedded altimeter readings. Another goal is to identify fiducial marks, which enable precise geometric calibration. These tasks share similar techniques but are designed to function independently. However, the inconsistent quality of the historical images poses challenges, and the methods may not always yield reliable results. To mitigate this, an automatic estimation step follows each extraction, serving as a fallback to ensure robustness and continuity in subsequent processing stages. A detailed description of these metadata recovery steps is presented in the following subsections.

3.2.1. TEXT EXTRACTION

Historical aerial images often contain hand-written or typewritten annotations directly on the photographs, such as focal length or altitude values. These annotations are both valuable and problematic: while they provide crucial metadata, they also introduce noise that can hinder automated processing. To address this, we employ text recognition software to both extract relevant information and mask the regions containing text. Traditional optical character recognition (OCR) methods, such as those based on TensorFlow, were initially considered but yielded poor results in our early trials, particularly when applied to typewritten annotations in historical imagery. While no systematic benchmark was performed, we observed that these general-purpose models, typically trained on modern machine-printed text, struggled with the lower contrast and variability of historical scans. In contrast, the OCR tool PP-OCR (Du *et al.* 2020), originally developed for recognizing Chinese characters, demonstrated significantly more robust performance in our use case. It not only recognized typewritten annotations more reliably but also provided bounding boxes for each detected text region, enabling downstream filtering and spatial reasoning.

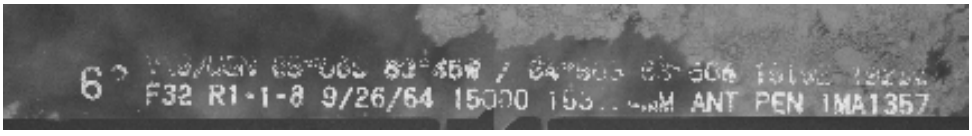


Figure 3.2: Example text from archive image. Note that height (15.000) and focal length (153.xx mm) are included in the description.

Given the large size of the images, applying text recognition to the entire image in one pass is computationally impractical. Instead, we focus on the four border regions (N, E, S, W) where annotations are typically

found. Each border is divided into overlapping sub-regions, and text extraction using PP-OCR is performed on each sub-region individually. To account for annotations that may be written upside down, the process is repeated with a 180-degree image rotation. This approach is sufficient for the current dataset, where text annotations are only present along the top and bottom edges. However, for other datasets where text may also appear along the sides, additional rotations - such as 90 and 270 degrees - can be included with minimal computational overhead. The extracted text and bounding boxes from all sub-regions are then merged, aligning overlapping text based on matching content at the start and end of adjacent regions. Finally, the separate bounding boxes are combined into a single comprehensive bounding box for each text region. This combined text data and bounding box information serves as input for further processing steps, such as masking text regions to reduce noise in subsequent analyses.

Despite successful localization of the text regions, extraction of useful information remains a challenge due to incomplete detections and false positives, as illustrated in Figure 3.2, where parts of the focal length annotation are obscured. To improve reliability, we leverage prior domain knowledge and statistical redundancy across all images from the same flight path and camera orientation. For each image, all detected text elements are concatenated into a single string and stored for analysis. We then apply regular expressions to identify substrings that likely represent focal lengths or altitudes: values containing a decimal point or the unit "mm" for focal lengths, and numerical patterns ending in "000" for altitudes. The identified strings are standardized to a fixed format, such as a 7-character focal length (e.g., XXX.XX) or a numeric altitude value, and recorded per image as a text string in a database. Unrecognized characters are replaced with a placeholder (e.g., "X") to mark uncertainty. To enhance reliability, all extracted values from a given flight path are compared character by character to determine the most frequently occurring digit or symbol at each position. For example, if some images yield 15X.3X mm and others give 154.X4 mm, the inferred result may be 154.34 mm. This voting-based approach helps correcting OCR errors and fill in uncertain or missing characters. Since the OCR does not provide per-character confidence scores, we treat any non-numeric characters in focal length fields as invalid and exclude them from voting. Finally, we incorporate domain-specific constraints, for instance, focal lengths in this archive are always above 150 mm, so any extracted value like for example 54.34 mm is automatically adjusted to 154.34. This structured and knowledge-guided approach enhances metadata extraction by reducing sensitivity to single-image OCR errors and leveraging the redundancy of similar images within a flight line.

3.2.2. FIDUCIAL MARK DETECTION

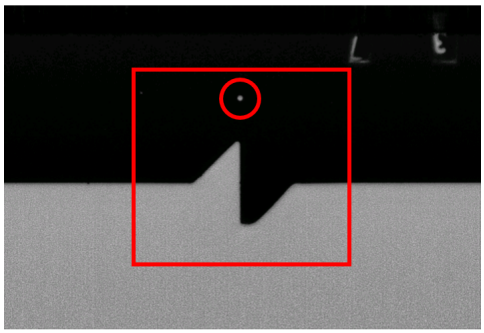


Figure 3.3: Example for a fiducial pattern (red box) together with the fiducial mark (red circle)

Figure 3.3 illustrates how fiducial marks can be identified on historical images. A fiducial mark is a precise reference point, often a small dot, crucial for image orientation and calibration. Surrounding this mark is a distinctive fiducial pattern, which aids in accurately locating the mark. The pattern is designed with lines converging precisely on the fiducial mark to ensure accurate detection.

As shown in Figure 3.4, the fiducial marks are located at eight positions. However, only four of these marks (5–8) can be directly detected, while the remaining four (1–4) must be inferred. Once the positions of the first four fiducial marks are established, the principal point of autocollimation (PPA) (McGlone, Mikhail, and Bethel 2013), can be estimated as the point of intersection of the lines connecting them.

To detect fiducial marks automatically, we employ a structured seven-step approach as shown in Figure 3.5, applied separately for each fiducial mark. For the cardinal directions (North, East, South, West), direct detection is possible, while for the intercardinal directions (Northeast, Southeast, Southwest, Northwest), fiducial marks must be computed based on detected cardinal marks.

Fiducial marks for N, E, S, W

Fiducial marks in the four main directions are detected through a machine-learning and computer-vision-based approach, consisting of the following steps:

1) Initial Cropping: Since fiducial marks are always positioned near the image borders, we extract a cropped region at each respective boundary. This preliminary step enhances processing speed and accuracy by reducing the search space for the fiducial pattern.

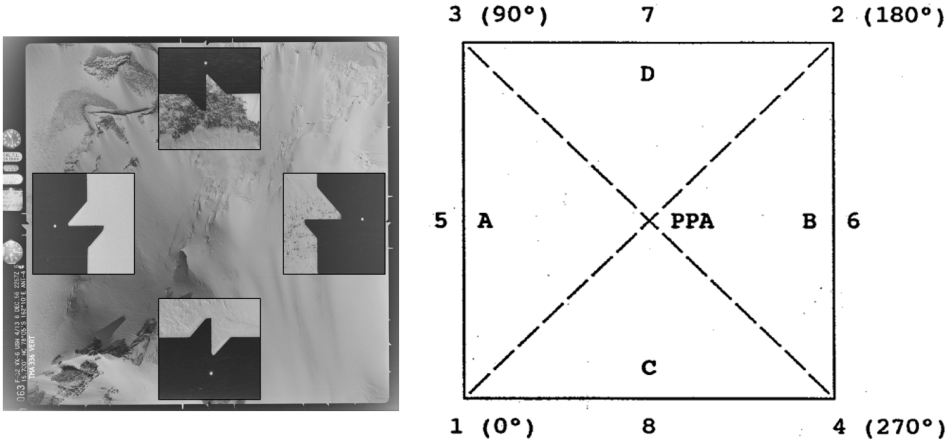


Figure 3.4: Example for fiducial marks in the dataset and the equivalent schematics from a camera calibration report

2) Subset Detection: Once the cropped region has been extracted, we use dlib, a versatile toolkit designed for machine learning and computer vision (King 2009), to recognize the fiducial patterns in that region. We employ the so-called simple object detector to detect fiducial patterns in the images. This efficient object detector is a trainable model based on Histogram of Oriented Gradients (HOG) and a linear Support Vector Machine (SVM) classifier. To enable the detection, we created a training set of 100 fiducial patterns for each direction, enabling the model to learn and generalize effectively. The HOG technique effectively extracts dominant image gradients, which in our case point towards the fiducial marks. Extracted HOG gradients and corresponding fiducial patterns can be seen in Figure 3.6.

At this stage, precise localization is not necessary; rather, we aim to isolate a smaller region as a subset for further refinement by only looking at the border regions of an image. If no suitable subset is found, or if more than one subset is detected, detection for this direction is halted. Figure 3.7 is showing different stages of this extraction in panels A-G.

3) Pre-processing: To simplify subsequent operations, the subset (panel A) is blurred and then binarized using a dynamic threshold (resulting in panel B). Instead of a fixed threshold, we compute the most frequent pixel value (typically the background) and adjust the threshold dynamically by adding a safety margin of 10 px.

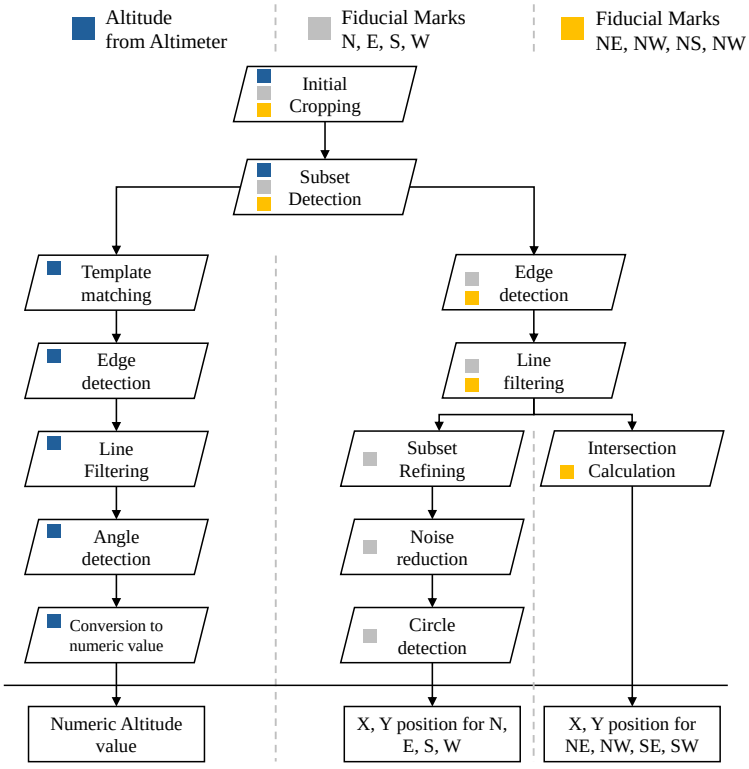


Figure 3.5: Steps required for obtaining altitude, fiducial marks for N, E, S, W and fiducial marks for NE, NW, NS, NW. Each step is colour coded to show which steps are required for each extraction.

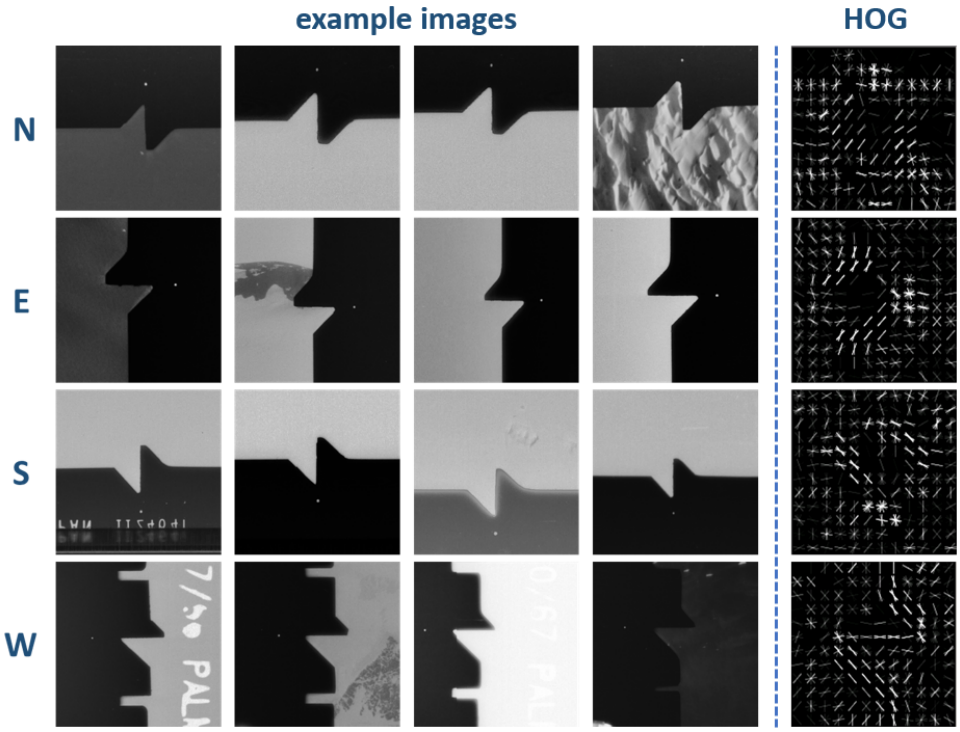


Figure 3.6: Example for fiducial marks from different images and their HOG filters for each cardinal direction N, E, S, W

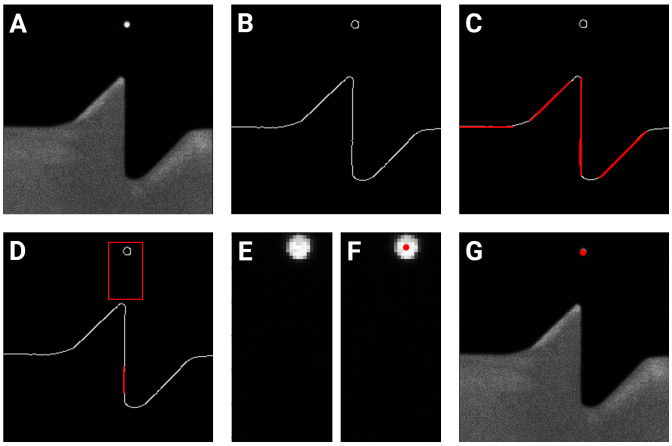


Figure 3.7: Steps for fiducial mark extraction: detected subset (A), binarization (B), Edge detection (C), filtering for centre line (D), blurred subset (E), subset with detected point (F), final absolute coordinates of subset (G)

4) Edge Detection and Line Filtering: Canny edge detection (Xu, Baojie, and Guoxin 2017) is applied to extract straight lines in the subset (panel C), which serve as structural references for fiducial mark localization. Only lines with the correct orientation (vertical for N/S, horizontal for E/W) are retained. Among these, the line closest to the centre is selected (panel D). If multiple lines are detected, they are merged by averaging their endpoints.

5) Refining the Region of Interest: Using the detected line as a guide, we extract a smaller subset (panel E), expanding the region along the fiducial direction while narrowing it perpendicularly.

6) Noise Reduction and Circle Detection: At high zoom levels, image artifacts are common. To mitigate this, we apply the morphological operations erosion and dilation to remove small artifacts while preserving key structures. We then search for circular features within a predefined size range using the circle Hough transform. As a fallback, we detect image moments, which can also identify non-perfect shapes (e.g., blurry or slightly oval ones), but only with pixel-level accuracy. Image moments are weighted averages of image intensities that describe the shape's geometry—such as its centre of mass, orientation, or area—and are useful for locating and characterizing objects when perfect shapes are not present. If multiple circles are detected, the one nearest to the centre is selected (panel F) If no circle is found, the subset is gradually shifted outward until a fiducial mark is detected or the image boundary is reached. If no mark is found, detection for this direction remains

incomplete.

7) Final Localization: Once the position of the fiducial mark is determined within the subset, its absolute coordinates in the full image are calculated (panel G).

Fiducial marks for NE, SE, SW, NW

The remaining four fiducial marks cannot be directly detected, as they lack distinct white dots marking their positions. Instead, their locations are inferred based on the detected cardinal fiducial marks. This process involves the following steps:

1) Resizing Subsets for Edge Detection: Prior to performing Canny edge detection, the relevant subsets from the adjacent cardinal fiducial marks (e.g., North and East for the Northeast fiducial mark) are resized. The subset is elongated in the primary direction of interest while reducing its width in the perpendicular direction.

2) Edge Detection and Line Extraction: Canny edge detection is applied to identify lines perpendicular to the expected fiducial mark location. Only the lines closest to the centre of the subset are retained. If multiple lines are detected due to interruptions in the fiducial pattern, they are averaged to form a single representative line.

3) Intersection Calculation: The detected lines from the two relevant subsets (e.g., North and East for the Northeast fiducial mark) are used to calculate their intersection point. This intersection is assigned as the inferred fiducial mark position.

By combining direct detection for cardinal fiducial marks with computed positions for intercardinal marks, this approach enables robust and accurate fiducial mark localization across historical aerial images. This methodology ensures consistency across the dataset while accounting for variations in image quality, exposure, and potential distortions.

3.2.3. Height Detection

Height detection involves extracting the altitude data from either text annotations or an altimeter display in the images. In some cases, height is clearly visible as text, which can be processed as described in Subsection 3.2.1. However, in many cases, text annotations are absent, and the height must instead be derived from a three-pointer altimeter, as shown in Figure 3.8. This altimeter measures altitude through air pressure and displays it using three clock-like pointers: the longest

pointer indicates increments of 100 feet, the next pointer 1,000 feet, and the shortest pointer 10,000 feet. Similar to fiducial mark detection, height detection required a multi-step approach, as depicted at the left side of the workflow in Figure 3.5.

Figure 3.8: Example for a physical altimeter and it's equivalent from a historical image of the TMA archive. Adapted from UK Airsports (2022).

The first step in this process is to localize the altimeter within the image. Since it is typically found at the bottom-left corner and has a distinct round shape with labelled numbers, its position is easy to approximate. We use again the object detection framework of dlib, trained on 100 manually annotated images with various lighting conditions and image qualities. This small training set suffices due to the standardized design of the altimeter across the dataset.

Next, template matching is applied to refine the position. Perfect scans would allow circle detection through methods like Hough transforms, but the digitisation of these historical images often results in slight distortions. To address this, template matching focuses on the detection of numeric labels such as 3, 5, and 8, which are consistently present on the altimeter face. This approach improves robustness, even in cases where parts of the altimeter are cut off.

Once the position of the altimeter is identified, image enhancement is performed to compensate for inconsistencies in brightness and clarity caused by reflections or poor lighting. We apply histogram equalization to the central area of the altimeter, which improves contrast and prepares the image for binarization. The outer areas, where reflections are more prominent, are set to the median pixel value of the circular region to avoid interference (Step A in Figure 3.9).

After enhancement, edge detection is carried out on the binarized image (Step B) using the Canny method to extract lines within the altimeter

(Step C). Only lines oriented towards the circle's centre are retained, and additional filtering is applied to distinguish each pointer based on its characteristics. The 100 feet pointer is recognized by long, parallel lines (Step D), while the 1,000 feet pointer has intersecting tip lines (Step E).

With the relevant lines detected, we calculate the angle of each pointer relative to the zero position on the altimeter. The angles range from 0° to 360°, where 0° corresponds to position zero and 180° to position five. From these angles, the readings of the 100 and 1,000 feet pointers are determined directly.

Detecting the 10,000 feet pointer presents a unique challenge. Its small size and placement near the bottom of the altimeter often result in poor visibility. However, since flight heights rarely exceed 30,000 feet in the dataset, the pointer's position will always fall within the first few digits (0 to 3). Therefore we calculate three possible positions for the 10,000 feet pointer and search for it within these areas. If detected, this final step confirms the complete altitude reading.

By analysing the positions of all pointers relative to the instrument's centre, the altitude can be converted into a numeric value. The angles of these pointers are measured relative to the north direction and mapped onto a circular scale, where a full rotation corresponds to 10,000 feet for the shortest pointer, 1,000 feet for the middle pointer, and 100 feet for the longest pointer. The final altitude is obtained by summing these values. However, if the calculated altitude falls below 8,000 feet, an offset of 10,000 feet is added, as for the considered dataset no planes flew below this altitude. This correction accounts for cases where the 10,000-feet pointer was not detected correctly. In some cases pointers can overlap, which happens most frequently with the 10,000-feet pointer. In these instances, the algorithm handles the situation similarly to cases when the 10,000-feet pointer could not be detected. Overlaps between the 100-feet and 1,000-feet pointers are very rare and were therefore not explicitly addressed.

3.2.4. Estimation & quality control

Detection of fiducial marks and altimeter readings is not always successful. Low contrast, excessive image artifacts, or incorrect scanning can hinder automated recognition. In some cases, fiducial marks or altimeter displays may have been cut off partially or entirely during the scanning process.

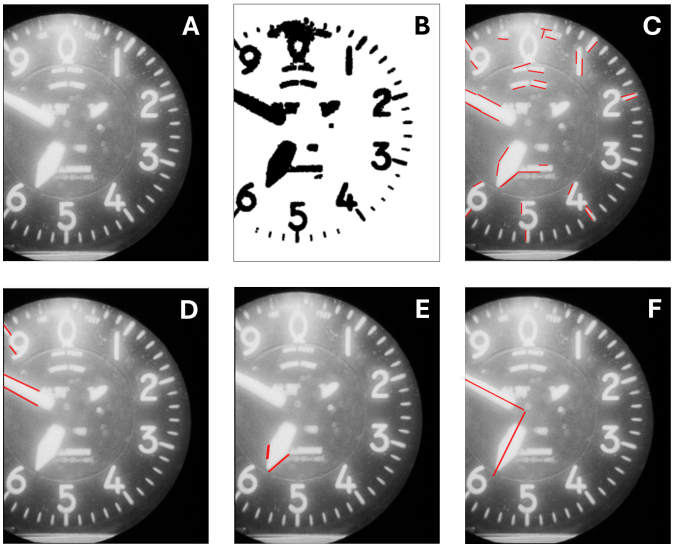


Figure 3.9: Steps for altitude extraction using an altimeter: Original image (A), binarization (B), Edge detection (C), filtering for centre lines (D), filtering for tip lines (E), final lines (F)

To handle these issues, we apply a correction strategy based on redundancy of parameters within the same flight line. Images captured along a flight path typically share the same camera setup, calibration parameters, and flight altitude, which allows for cross-validation between overlapping observations. On average, each flight path contains approximately 54 images, providing in most cases sufficient redundancy for correction. A value is flagged as missing if no meaningful number could be extracted, and it is marked as an outlier if it deviates by more than two standard deviations from the median of all valid values in the same sequence. Only flight paths with at least three valid (non-null) entries are considered for correction.

For the fiducial mark detection, if the initial subset containing the mark is not detected, we first estimate its likely position based on the average coordinates of subsets detected in all other images from the same sequence. Once this position is determined, the fiducial mark detection algorithm is re-applied to the estimated region. If the extraction is still unsuccessful, we calculate the final coordinates of the missing fiducial mark by averaging the coordinates of matching marks from other images of the same flight path.

A similar procedure is applied for altimeter readings. If the altimeter display is not detected in an image, we begin by estimating its expected position. If the reading remains missing, the altitude value is inferred by

analysing the heights recorded in other images within the same flight line. Since the aircraft is assumed to fly along a stable trajectory, any obvious outliers are removed using the interquartile range (IQR) method. A weighted average of the remaining altitude values is then calculated to estimate the missing height. While this estimation is generally less precise than that for fiducial marks due to potential fluctuations in altitude, it provides a reasonable approximation that supports subsequent processing steps.

3.3. RESULTS

To evaluate our methodology, we selected the 20 longest flight paths[1] from the AP section of the TMA archive. This selection of 7719 images provides a diverse range of examples to thoroughly test our approach while still allowing us to apply and validate our estimation techniques across a large dataset. While this facilitates testing the full potential of our method, it may lead to slightly higher estimation success rates compared to shorter or less continuous flight paths found elsewhere in the archive.

Table 3.1 provides an overview of the test results from 20 flight paths. It summarizes the extraction process for each key feature, including subsets used as a preliminary step in fiducial mark detection. The table shows the number of entries that were successfully extracted automatically, those estimated based on extracted features from other photos from the same flight path, and those that remained missing due to failed extraction or estimation. The final column presents the results of a manual extraction of the computer vision based parameters, indicating whether a human observer could identify these patterns on the images. This serves as a baseline for comparison with the automatic extraction, with the exact values per flight path shown in the annex.

Figure 3.10 illustrates examples of successful data extraction. For each identified text, bounding boxes are available to determine the position of the text box. In the case of fiducial marks, precise coordinates are marked at the exact location of each fiducial spot. For height extraction using altimeters, three lines extend from the centre of the circular display to indicate the direction and position of the pointers. The altitude value of these pointers corresponds to 22,800 feet.

[1] flight paths with the following id: 1684, 1813, 1816, 1821, 1822, 1824, 1825, 1826, 1827, 1833, 1846, 2073, 2075, 2136, 2137, 2139, 2140, 2141, 2142, 2143

Field	Extracted	Estimated	Missing	Manual
Text position	7719	0	0	7719
Focal length	2145	3436	2138	-
Subset N	5791	1706	222	-
Subset E	6828	779	112	-
Subset S	6813	794	112	-
Subset W	6859	748	112	-
Fid mark 1	7064	543	112	-
Fid mark 2	6990	508	222	-
Fid mark 3	6950	529	239	-
Fid mark 4	7150	457	112	-
Fid mark 5	6552	1055	112	7711
Fid mark 6	6858	749	112	7714
Fid mark 7	5192	1475	1054	6015
Fid mark 8	6488	1119	112	7662
Height (text)	12	312	7395	-
Height (altimeter)	1279	6440	0	5237

Table 3.1: Summary of extracted, estimated, missing and possible manual values for text extraction, fiducial mark detection and height detection for 7719 photos from 20 different flight paths. (the values with n.e. were not considered in the manual extraction).

Figure 3.10: Examples for successful detections: Fiducial marks (top), text extraction (bottom left) and altitude based on altimeter (bottom right)

3.4. DISCUSSION

The results of the data extraction process demonstrate a high level of effectiveness in several areas. Text extraction performs exceptionally well, with no missing text boxes across the dataset. As can be seen in Table 3.1, for approximately 70% (5581 individual photos) of the images we could successfully extract focal length values. For the remaining 30% of the images, extraction of focal length data failed due to two primary factors. First, the complete absence of focal length information in the text. Second, poor legibility of the written text, as illustrated in Figure 3.2, where the focal length is only partially discernible.

Subset extraction is similarly reliable, with only a small number of images missing critical data. Failures in this area are primarily the result of either portions of the image being cut off during scanning or images being scanned in an inverted orientation, which disrupts the fiducial pattern detection with dlib. This notably affects the left looking direction of flight path 1822, which explains the identical number of missing values for the subsets. Nevertheless, the overall success rate is sufficient to allow substitution of missing subsets with data from successfully processed images.

Fiducial mark extraction performs well overall but faces specific difficulties with fiducial mark 7, located at the top of the image. Investigations revealed that many of these difficulties arise from scanning errors that result in parts of the image, including this fiducial mark, being cut off. Not enough to not find the subset, but so much that the fiducial mark is missing. Figure 3.11, example A, illustrates such a case. In other instances, low contrast between the fiducial pattern and its surroundings complicates detection, as demonstrated by example B in the same figure. In such scenarios, even human observers struggle to pinpoint the fiducial pattern.

When fiducial marks are successfully detected, their positions are generally accurate to within a few pixels, demonstrating strong performance for correctly scanned images. However, errors can occur when a fiducial mark is missing due to scanning issues or when incorrect scanning introduces artifacts - such as additional fiducial marks from the adjacent image - because multiple images were scanned together.

To further evaluate the performance of our fiducial mark extraction, we compared it against the built-in detection algorithm of Agisoft Metashape (see Table 3.2). Metashape successfully recovers nearly all clearly visible markers and slightly outperforms our method in three of the four cardinal directions. In contrast, our approach, augmented with estimation, matches or exceeds the performance of Metashape for

fiducial mark 7 and remains within a few percent for the others. Both methods fail similarly when marks are truncated by scanning errors (e.g., in the case of mark 7). Overall, although Metashape achieves slightly higher detection rates on clean scans, our open-source, fully automated pipeline delivers competitive performance.

Fid Mark	Total	Visible	Our method	Agisoft
Fid Mark 5	7719	7711	7607	7695
Fid Mark 6	7719	7714	7607	7701
Fid Mark 7	7719	6015	6667	6012
Fid Mark 8	7719	7662	7607	7659

Table 3.2: Comparison of fiducial mark detection results between our method and Agisoft Metashape. The table lists the total number of images, the number of images where the fiducial mark is visible, and the number of fid marks by each method. For Fiducial Mark 7, our method detects more instances than the number of visible marks due to its estimation mechanism, which infers likely positions based on consistent placements in neighbouring images.

The extraction of altitude data presents the most significant challenge. Successful extractions based on the altimeter occur in only 15% of all cases (1279 images), often hindered by factors such as inverted images, blurry or illegible altimeter readings, altimeters being cut off, or scanning artifacts affecting the relevant areas. To a lesser extent, these issues also impact manual extraction, which failed in most cases due to altimeters cut off. Figure 3.11, example C highlights one such problematic example. Additionally, some incorrect altitude values are extracted, necessitating further quality control measures. Despite these issues, the successful extractions provide enough data to estimate missing altitudes along flight paths. This marks a considerable improvement over the previous lack of altitude data for most images. In the context of SfM workflows and workflows aiming at geo-referencing historical imagery lacking GNSS data (Craciun and Le Bris 2022; Dahle, Lindenbergh, and Wouters 2024a; Giordano, Le Bris, and Mallet 2018) even approximate altitude information enhances the reconstruction and localization process by providing a valuable starting point.

The extraction of altitude information directly from text, in contrast, is negligible due to the rarity of altitude references within the text boxes. However, in cases where this information is present, the text extraction process typically identifies it correctly.

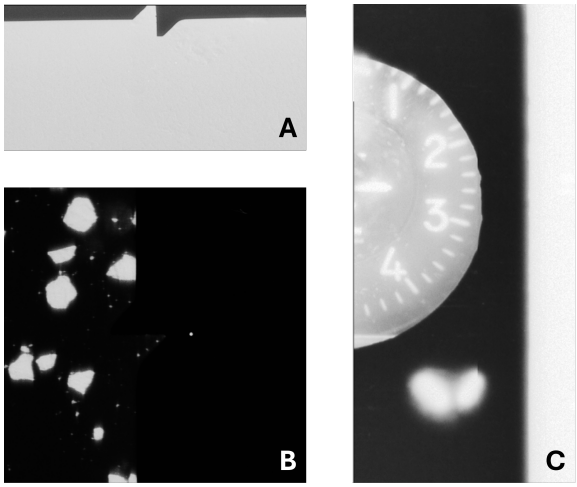


Figure 3.11: Examples for failed extractions: Fiducial mark detection with cut-off or difficult data (A, B) and height detection (C) with missing pointers

3.5. CONCLUSION

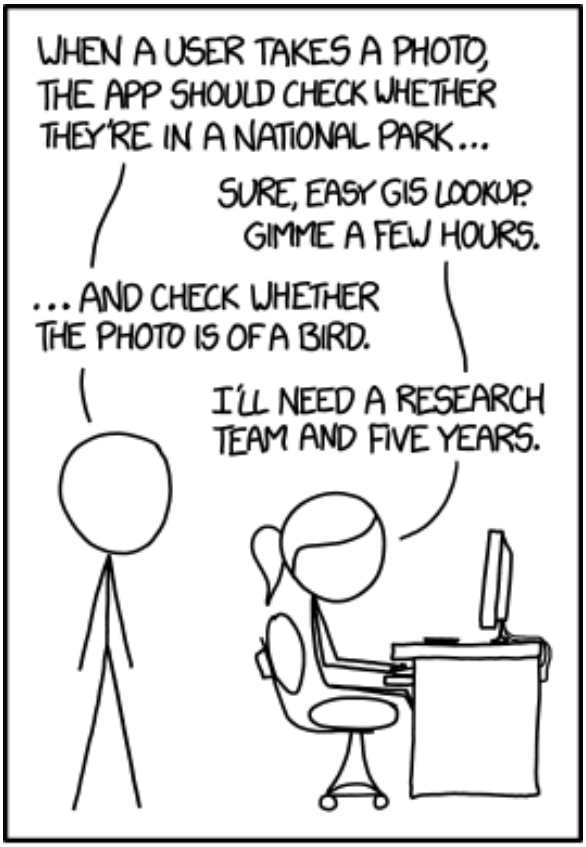
Some challenges remain, particularly in cases involving image degradation, distortions, or non-standard orientations. A notable limitation is the difficulty in reliably detecting fiducial subsets when images are rotated or of poor quality. While our current implementation relies on dlib for object detection, future work could benefit from more advanced methods such as Convolutional Neural Networks (CNNs), which are more robust to variations in image quality and orientation. Integrating a CNN-based approach may further improve detection accuracy and ensure consistent performance across a wider range of image conditions.

Altimeter readings, likewise, remain sensitive to edge detection quality, with factors like noise, lighting artifacts, or obstructions occasionally impairing pointer identification. While altitude text annotations are rare, this reinforces the value of extracting altimeter readings directly from the image. Even though some image collections may include rough altitude metadata, this is not always reliable or consistently available—especially for historical datasets. In these cases, visual extraction of altimeter data offers a valuable complement to existing metadata sources, especially for SfM workflows.

Our current altimeter reading method is based on tailored image processing and known height ranges. While effective, it could be further strengthened by leveraging methods developed for similar

tasks. The challenge of reading multi-pointer altimeters closely parallels the well-studied problem of reading analogue clocks and mechanical gauges. Recent work in these areas has employed both traditional vision techniques and deep learning (e.g., (Chavan, X. Yu, and Saniie 2022; Yang, Xie, and Zisserman 2022)) with promising results. These strategies could serve as a foundation for improving altimeter reading under more variable conditions.

Finally, the methodology could be extended to extract additional visual metadata from historical images, such as analogue capture times or image numbers. Capture time could support sun angle estimation, while image numbers might help link scans to external catalogues. These additions would further enhance metadata completeness and support more powerful, automated processing of historical aerial image collections.



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

4

GEO-REFERENCING OF HISTORICAL IMAGES

The utility of historical image repositories is often limited due to the lack of geo-referencing. A good example is the Trimetrogon Aerial (TMA) archive, a collection of historical aerial images of Antarctica between 1940 and 2000. These images are difficult to use, as their geolocation is only approximate, with location errors in the order of tens of km. This study addresses this challenge by developing an automated geo-referencing workflow that leverages recent advancements in machine-learning-based Tie Point (TP) matching. We use the algorithm LightGlue, to establish TPs between geo-referenced Sentinel-2 satellite imagery and historical non-geo-referenced aerial images. To aid the process, we use already known approximate positions of the historical images. Due to the sub-optimal and inhomogeneous quality of the aerial images, only a portion of the images can be geo-referenced directly by matching. For the remaining images, we employed alternative means of geo-referencing, again based on TP matching. Out of a subset of 4,459 images located inside the research area, 3,393 images could be geo-referenced, a percentage of 76%. Reasons for the geo-referencing failing are insufficient contrast or an approximate position too far away from the real position. The workflow can easily be applied to historical images from other archives, to enhance the usability of historical image repositories for scientific research.

This chapter has been published as *Polar perspectives: a deep dive into geo-referencing historical Antarctic photos* in the *International Journal of Digital Earth* (Dahle, Lindenbergh, and Wouters 2024a).

4.1. INTRODUCTION

Historical aerial imagery offers a window into the past, granting access to valuable information from before the satellite era. Many of these images, initially produced for military purposes on a large scale, were consigned to archives and their contents were inaccessible. However, in recent years these archives have been digitised and made available to the public. Consequently, the potential of these historical image repositories has been recognized across a spectrum of scientific disciplines, with applications ranging from archaeological (Cowley and Stichelbaut 2012; Usmanov *et al.* 2018), environmental (Heisig and Simmen 2021; Nicu and Stoleriu 2019) to cryospheric sciences (Geyman *et al.* 2022; Pope *et al.* 2014).

In cryospheric research, historical aerial images can provide information on the state of the ice sheet before the satellite era for a wide range of subjects: they can be used for evaluating glacier mass balance, shedding light on both historical states and recent changes (Denzinger *et al.* 2021; Korsgaard *et al.* 2016), aid in understanding the structure, dynamics, and flow of ice sheets over time (Z. Yu *et al.* 2022), and provide insight into past ice extents in specific regions (A. J. Cook and Vaughan 2010). Additionally, these images are valuable for monitoring non-ice-related phenomena, such as identifying areas of coastal erosion (J. Wang *et al.* 2022) or observing thermo-erosion gullying processes (Morgenstern *et al.* 2021; Perreault *et al.* 2017).

The use of historical aerial images is not confined to 2D analyses. Advanced techniques like Structure-From-Motion (SfM) enable the generation of historical 3D models from these images (Child *et al.* 2020; Knuth *et al.* 2023; Zhang, Rupnik, and Pierrot-Deseilligny 2021).

However, processing archives of historical images often face a significant challenge that limits their use: many images are not, or incorrectly, geo-referenced. In this context, geo-referencing refers to the process of associating specific points within an image with their corresponding real-world locations. This typically involves identifying TPs, which are points on two distinct images representing the same geographical location. Working with un-referenced images requires researchers to manually identify similar structures and pin-point TPs. This is arguably the primary reason these valuable historical images are underutilized in research.

Efforts have already been made to automate this process (H. Chen and Tseng 2016; Cléri, Pierrot-Deseilligny, and Vallet 2014; Karel *et al.* 2013) by using TP detection and feature matching. Furthermore,

geo-referencing has expanded beyond mere image comparison; photogrammetric methods like SfM have also been employed (Frankl *et al.* 2015; Giordano, Le Bris, and Mallet 2018; Heisig and Simmen 2021). However, these methods typically rely on having scenes with rich textures to provide distinct features for TP matching. Furthermore, in general, images of the same type, such as aerial-to-aerial, are compared. Geo-referencing and SfM are particularly challenging when geo-referencing images of the cryosphere, where many images lack contrast and only have few distinct landmarks. Matching is further hindered, as corresponding modern aerial imagery for these regions is generally not available which implies that historical areal images can only be matched to contemporary satellite imagery.

To address these challenges, we have developed a workflow that applies recent advancements in neural-network-based TP matching techniques to historical aerial images. Our approach aims to identify TPs between historical aerial imagery and contemporary satellite imagery, streamlining the geo-referencing process as much as possible. Recognizing the inherent difficulties in matching aerial and satellite images, we have also devised several algorithms for indirect geo-referencing.

We applied our workflow to images from the TMA archive, a collection of historical photographs of Antarctica, acquired between 1940 and 2000, with the majority of the images taken in the 1960s. The archive has been exploited before to gain insights into the condition of Antarctica prior to the satellite era, in studies focusing on individual or a subset of glaciers, for which manual geo-referencing suffices (Child *et al.* 2020; Ferrigno 2005; Kunz *et al.* 2012; Zhao *et al.* 2017). For the total of around 330,000 images in the archive, such a manual approach is unfeasible. In this study, we present a workflow that can be executed with minimal user interaction and in a reasonable amount of time. We specifically target the Antarctic Peninsula (AP) due to its extensive coverage by overlapping images, which facilitated the evaluation of our geo-referencing accuracy.

While our code was tailored for this specific archive, its design allows for adaptation to other non-referenced datasets. The geo-referencing code and the resultant geo-referenced images are accessible at GitHub (Dahle 2024d). Additionally, we will provide the footprints and transformation matrices for each geo-referenced photograph of the TMA archive. Over time, we aim to geo-reference and publish the entire archive.

4.2. DATA

4.2.1. TMA-ARCHIVE

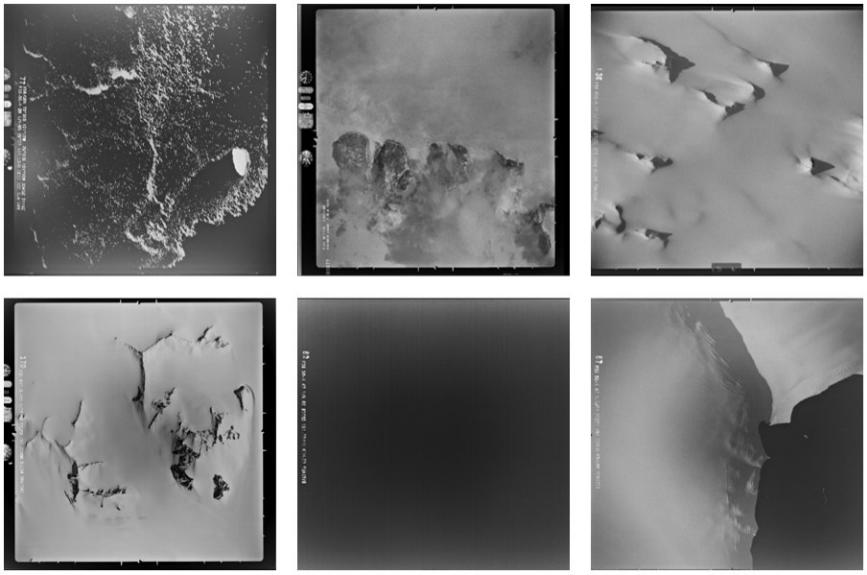


Figure 4.1: Different example images from the TMA archive from different flight-paths

The TMA archive is a collection of photographs retrieved by the U.S. Navy between 1946 and 2000 along the margins of Antarctica at approximately 125,000 locations, consisting of about 330,000 images. This archive consists of black and white trimetrogon (a camera system mounted on airplanes pointing left, right and nadir) aerial photos. Some examples are shown in Figure 4.1. While acquiring the TMA data, airplanes flew along so-called flight paths: approximate straight lines with more or less constant height. About 15,000 images were collected on the AP, most of them collected in the mid-1960s. The imagery has been digitally scanned at high resolution (25 micron/1000 dpi) by the US Geological Survey (USGS) and the Antarctic Geospatial Information Center (University of Minnesota) and is publicly available at University of Minnesota (2023).

To date, the aerial photographs within the collection have not been geo-referenced on a large scale. Nonetheless, as can be seen in Figure 4.2, digitised flight lines exist, which by visual inspection provide approximate locations of the actual flight paths and camera positions. These flight lines are distributed with various attributes such as the altitude of the airplane and the acquisition date. However, these

attributes are not consistently available for every camera position, and their accuracy cannot be assured.

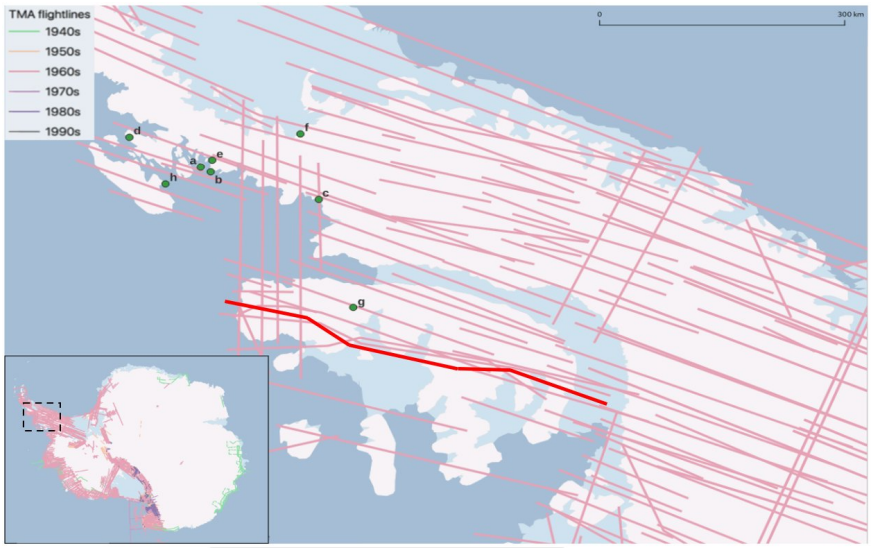


Figure 4.2: Study area with the digitised flight lines for the TMA-archive with their respective recording date. The green dots highlight the position of the use cases discussed in Section 4.4: a) Shoesmith glacier b) McMorrin Glacier c) Fleming Glacier d) Somigliana Glacier e) Unnamed Glacier near Mt. Wilcox f) Sibelius Glacier g) Rothera Research station. The bold red line highlights a non-straight flight path discussed in Section 4.5.

In this work, we focus on the images with nadir view, as these are most straightforward to geo-reference. The camera position is consistent for each triple of left-right-nadir images. Therefore, although the initial goal is to geo-reference the 4,459 available nadir images located in the research area on the AP, a successful procedure will enable a relatively easy consecutive geo-referencing of corresponding left-right images as well.

4.2.2. SENTINEL-2 SATELLITE IMAGERY

For our geo-referencing method, the historical images must be compared with already geo-referenced data. In our study, we employ imagery from Sentinel-2, an Earth observation mission under the Copernicus program of the European Commission. This mission offers freely available optical imagery characterized by a medium spatial resolution. Although our research is based on Sentinel-2 imagery, it is important to mention

that other geo-referenced datasets can also be suitable. The Sentinel-2 data used for geo-referencing spans from 2016 to 2023. Using Google Earth Engine, these images are compiled into a cloud-free composite by masking cloud-cover pixels and computing the median value for each pixel. Given our objective of geo-referencing this data with the historical aerial imagery, we extract bands 2, 3, and 4 to produce an RGB representation. The spatial resolution for these bands is 10 meters. To ensure radiometric consistency with historical images, all images are standardized to an 8-bit format, with RGB values ranging from 0 to 255.

4.3. Methodology

In our geo-referencing approach, we employ TP matching techniques between images, making use of recent technological advancements to design a workflow that can automatically geo-reference historical images.

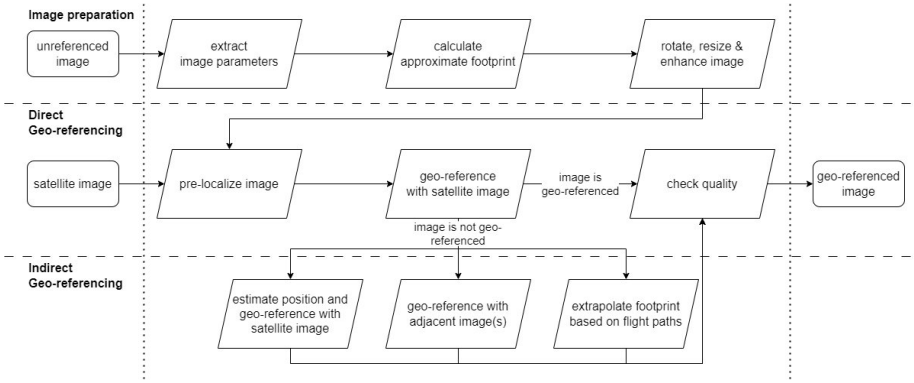


Figure 4.3: Geo-referencing workflow. If direct geo-referencing by matching a prepared historical image to recent Sentinel-2 data fails, indirect strategies will be used.

Figure 4.3 illustrates the workflow of our methodology. The workflow is specified in detail below: first, we prepare the TMA images for matching and geo-referencing by creating approximate footprints (Subsection 4.3.2). Second, we geo-reference part of the TMA images using Sentinel-2 imagery and subsequently perform a quality assessment (Subsection 4.3.3). Due to the mixed quality of the historical images, many of them cannot be geo-referenced directly on the first attempt. Therefore, we apply alternative in-direct strategies to also ensure successful geo-referencing for these challenging images

(Subsection 4.3.4). Due to the importance of geo-referencing in almost every step, our TP matching workflow is explained first in Subsection 4.3.1.

4.3.1. TIE POINT MATCHING

Various TP detection and matching techniques can be incorporated into our geo-referencing workflow. In this research, we use "LightGlue" (Lindenberger, Sarlin, and Pollefeys 2023), a successor to the popular TP matching algorithm "SuperGlue" (Sarlin *et al.* 2020). Both algorithms leverage the TP detection method "SuperPoint" (DeTone, Malisiewicz, and Rabinovich 2018). SuperPoint is a fully convolutional deep neural network, trained to detect interest points (unique points in an image) and compute their corresponding descriptors between pairs of images. Given a collection of these points and descriptors, LightGlue can be used to identify matches between the images.

It is noteworthy, that both LightGlue and SuperGlue were primarily trained on datasets without historical or aerial images. Moreover, the specifics of their training parameters remain undisclosed, rendering retraining for historical images impractical. Nevertheless, they are among the most proficient TP matching techniques currently available that can match images of varying sizes and are already used successfully in many geo-science related applications (Ghuffar *et al.* 2023; Maiwald, Feurer, and Eltner 2023).

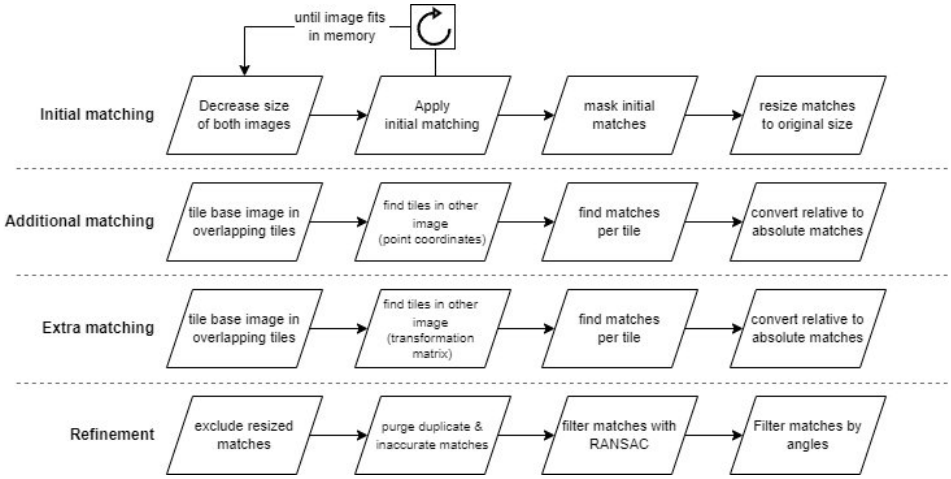


Figure 4.4: TP matching workflow. The first row is mandatory, all other rows can be applied optionally.

However, certain constraints exist: the algorithms are not rotation-invariant and the maximum rotation between two images should not surpass more than 45 degrees, as otherwise fewer matches are found (Tyszkiewicz, Fua, and Trulls 2020). Furthermore, image pairs must fit within the GPU memory, while scanned historical images typically have high resolutions, often surpassing 10,000 pixels in width or height. While downsizing the images is a potential solution to fit the images into the memory, it reduces the information content and consequently the number of detectable TPs. To maintain a dense distribution of TPs, we devised a workflow that optimizes TP detection using LightGlue, specifically for large images. Our TP matching workflow is divided in four different steps, as depicted in Figure 4.4. Only the first step is mandatory, all other steps can be applied optionally.

Initial Tie Point matching

In the initial matching, the selected matching algorithm (in this case 'LightGlue') is applied to resized images. Both images are resized iteratively by 10% until the matching process fits into GPU memory. Parts of the images can be masked to ignore certain areas for TP matching. To account for the rotational dependency of the matching algorithm, the image on the right is matched four times, for four different rotations of 0, 90, 180 and 270 degrees, respectively. The rotation with most matches is fixed for the further workflow. Finally, the TP coordinates are converted to original images size.

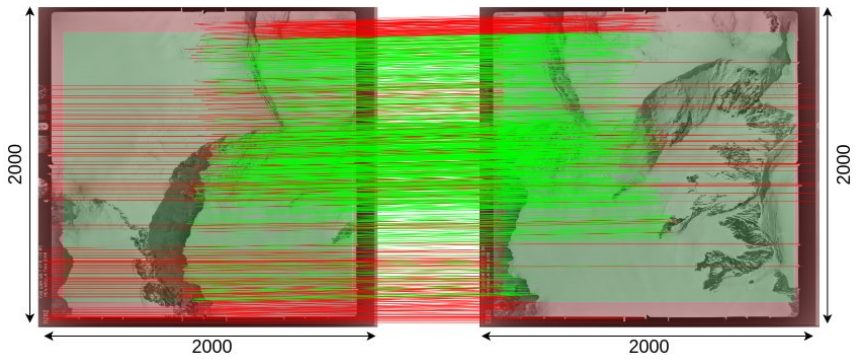


Figure 4.5: Initial matching between two images (IDs CA216632V0282 and CA216632V0283), that are resized to 2000 × 2000 pixels. 881 valid matches (green lines) and 407 masked matches (red lines) are found.

Figure 4.5 shows an initial matching between the two images, which are resized from 10000 × 10000 pixels to ∼ 2000 × 2000 pixels (a reduction of 96 % in image size). The red lines indicate matches that are masked,

while the green lines represent successful matches that will be kept for the further workflow.

Additional Tie Point matching

In the additional matching process, the larger image (or a randomly selected image if both are of the same size) is divided into multiple overlapping tiles, similar to the approach in Zhang, Rupnik, and Pierrot-Deseilligny (2021). In Figure 4.6, the left image is tiled into 30 tiles labelled A to AD. For each tile, all initial TPs within that tile are selected. Corresponding TP coordinates are used to create for each tile a subset of the right image. To account for outliers in the initial matching process, this subset is adjusted in size so that a tile and a subset in both images have identical size. TP matching is then applied between the tile and the subset in the original resolution. This is done iteratively for all tiles containing initial TPs, while tiles without TPs (e.g., the red tiles in Figure 4.6) are excluded.

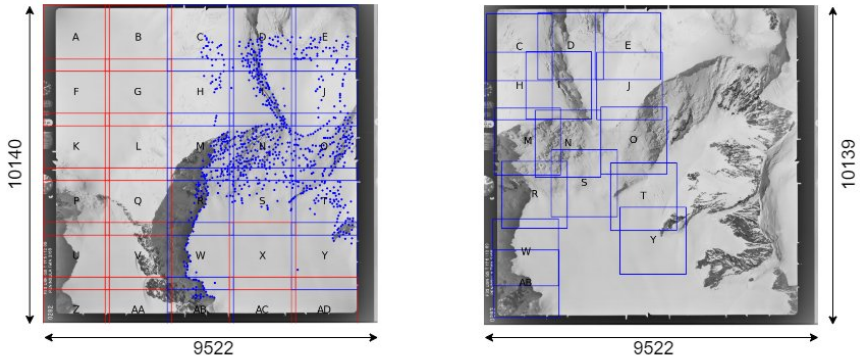


Figure 4.6: Additional matching in original resolution: Only tiles with initial TPs (blue) are matched in original resolution and have a corresponding subset in the right image. Red tiles do not contain any TPs.

Extra Tie Point matching

For now, the matching only focused on parts of the image where in the initial matching step TPs were found. To also use information in other parts of the image, a transformation matrix is derived from the already established TPs to determine the spatial correspondence between the two images (the tiles in Figure 4.7). Again, the left image is tiled in multiple overlapping tiles. For each tile, a corresponding subset on the right image is created by applying the derived transformation matrix on the coordinates of each tile. Like in the previous additional matching step, the matching algorithm is applied at the original resolution between each tile and its correspondent subset. Subsets predicted to

fall outside the image boundaries based on the transformation matrix are not included in this matching step.

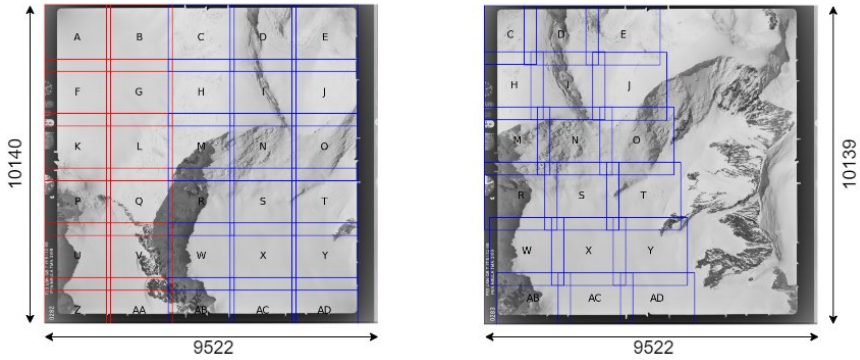


Figure 4.7: Extra matching in original resolution: For each tile on the left image the equivalent subset on the right image is calculated and TP matching is applied in full resolution. Tiles that do not have a subset are highlighted in red.

Refinement

If additional or extra matching is applied, the TPs detected in the initial matching are removed due to their lower accuracy, as these TP were obtained in down-scaled imagery of lower resolution. Subsequently, all remaining matches are combined, and two criteria are applied as detailed in Table 4.1: first, duplicate matches (with the same coordinates in both images) are removed. Second, based on a five-pixel distance, TPs within 5 px distance are averaged, while TPs exceeding this 5 px distance are discarded.

old TPs	x1	y1	x2	y2
T1	100	100	120	120
T2	100	100	120	120
T3	150	150	170	170
T4	150	150	172	172
T5	200	200	210	210
T6	200	200	220	220

new TPs	x1	y1	x2	y2
T1’	100	100	120	120
T2’	150	150	171	171

Table 4.1: Example for the refinement of matches: old TPs T1 and T2 have identical coordinates and are combined to new TP T1’. Old TPs T3 and T4 have different x2/y2 coordinates, which are closer than 5 px. Therefore, they are combined to new TP T2’. Old TPs T5 and T6 are more than 5 px apart and are therefore discarded.

As a final step, TPs are filtered with two methods. The first method employs the classical Random Sample Consensus (RANSAC) algorithm (Fischler and Bolles 1981) for outlier removal. The second method focuses on the angles formed by lines connecting TPs. If a predominant angle is observed among these lines, any TP line deviating more than 45 degrees from this angle is deemed an outlier and discarded. It is crucial, that the images being matched share the same resolution to ensure parallelism of these lines. To achieve this, the resolution of the smaller image is temporarily adjusted to align with that of the larger image.

4.3.2. IMAGE PREPARATION

To prepare the images for the matching, approximate footprints are calculated and the images are enhanced for higher contrast. Both steps are optional, but can speed up and enhance the quality of geo-referencing.

Approximate footprint estimation
Initial positions (x and y in the Antarctic Polar Stereographic coordinate system) and TMA image orientation are available from the Polar Geospatial Center (University of Minnesota 2023). For each image an estimated position is available, manually determined based on historical flight documents. To compute the footprints, the x, y, z position, rotation and the camera's focal length are required. These approximate footprints are neither at the correct position nor have the correct size, but still give a good first indication.

For some images, the focal length and camera height are written on the images themselves, as illustrated in Figure 4.8. To extract this information, we use the Character Recognition (OCR) tool PaddleOCR (Du *et al.* 2020), as it offers the best performance for recognizing such handwritten text . This tool serves two purposes: extracting metadata such as focal length or height (box A) and identifying regions in the images that are excluded from TP searching because they contain text (box B). For images where parameter extraction is not possible, we use average values derived from all other images on the same flight path.

The boundaries of this footprint are determined using the so-called CameraCalculator, adapted from Java to Python (Engel 2020). The determination is initialized with the camera's horizontal and vertical field of view, altitude, roll, pitch, and heading. Four rays are defined to represent the camera's field of view and are rotated based on the roll, pitch, and yaw parameters. The intersection points between these rays and the ground, represented as a flat plane, are computed. These

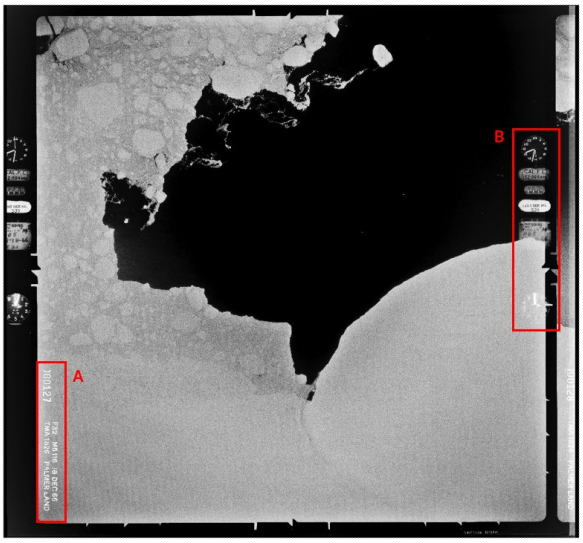


Figure 4.8: Historical image (CA182632V0127, located on Foyn Coast) with its metadata (A: image context information, B: height information) included in the image

intersection coordinates are then aligned with the camera's actual position, providing the approximate footprint's boundaries.

Subsequently, the initial footprint is used to get the average elevation of the encompassed area. The mean elevation is derived from raster elevation data (in this instance, Reference Elevation Model of Antarctica (REMA) elevation data (Howat *et al.* 2019) specific to the footprint's bounding box). The camera's altitude is then adjusted to recalculate its height in relation to the surface's average height instead of sea level. This updated altitude is then used to recalculate the footprint, which is usually smaller than the initial estimate due to the altitude adjustment.

Image enhancement

Before the geo-referencing with the satellite imagery starts, several enhancement steps are applied. These are divided into four primary stages: border masking, resolution adjustment, image rotation, and contrast enhancement.

1) Border masking. Many historical photographs were captured using analogue cameras, often featuring borders with fiducial marks. The high contrast between these borders and the main image content can introduce numerous false errors during TP matching. For the

images in the TMA archive, all borders adhere to a consistent spatial pattern, allowing for straightforward masking during the geo-referencing process.

2) Resolution adjustment. Historical images are usually scanned at 600dpi, resulting in an high resolution, typically measuring $\sim 10,000 \times 10,000$ pixels. Aligning with the approximate footprint shows that the ground sampling distance is in the order of 60 cm. In contrast, the Sentinel-2 data consists of $\sim 1,000 \times 1,000$ pixels, corresponding to a pixel spanning 10m. Conducting TP matching across such disparate resolutions is challenging. To address this, the larger image (the TMA image) is resized based on its approximate footprint to match the spatial resolution of the satellite image.

3) Image rotation. The images are generally not oriented facing North and require a rotation based on the known flight path angle. The necessity of this step depends on the TP matching method in use. For instance, LightGlue is not rotation invariant, and without this adjustment, fewer or no matches may be identified. When the rotation angle is not available, an alternative strategy involves rotating the images in increments of 90 degrees.

4) Contrast enhancement. A substantial portion of the images have sub-optimal contrast or inconsistent brightness levels, further complicating the TP matching process. To account for this, a median filter is first applied to reduce image noise. This is followed by a linear contrast stretch and concluded with a gamma adjustment. The enhancement approach is inspired by the SpyMicMac repository (McNabb *et al.* 2020).

4.3.3. DIRECT GEO-REFERENCING WITH SATELLITE IMAGES

Geo-referencing historical images is based on TP matching between these images and contemporary satellite data. For this method to be effective, certain consistent structures in both the historical and modern imagery need to be available. In most applications, these are human-made structures such as houses or streets. While such structures are absent in Antarctica, rock formations have proven consistent over time, in this instance, spanning nearly 80 years.

Information content
However, not all images from the TMA archive are suitable for geo-referencing due to a lack of distinct structures. To address this, we compute an "information content" value for each image prior to

geo-referencing. The information content is determined based on the count of distinct points in an image, as detected by the aforementioned method SuperPoint. A count of 2500 TPs or more is deemed the highest score with a value of 1. Counts below this are calculated as a percentage of 2500. This benchmark can be adjusted based on specific requirements and the number of TPs deemed necessary. The footprints of images with too low information content or where the geo-referencing with satellite images failed, will be estimated indirectly, as outlined in Subsection 4.3.4.

TMA footprint relocation.
In the case of the TMA archive, the starting boundary of the satellite image (that is used for matching) is defined by the approximate footprint, which however can deviate by several kilometres from the real position. To increase the possibility of a successful TP matching, two refining steps are employed. These adapt the boundary of the satellite to the most likely location for a more successful geo-referencing.

A) Segment based footprint relocation.
Figure 4.9 illustrates the first refinement step of relocating the footprint of the satellite image: the satellite image around the initial approximate footprint is segmented into smaller sections, each characterized by the dimensions of the approximate footprint. An initial TP matching is performed at the position of the approximate footprint (level 1). If the number of recognized TPs falls below a predetermined threshold, a second tier (level 2) is initiated. This phase involves TP matching at eight surrounding positions (cell 2B - cell 2I) encircling the central location of the approximate footprint. Although the positions depicted in Figure 4.9 appear uniformly aligned, this is a simplified representation. In practice, these positions exhibit slight overlaps, ensuring TPs near boundaries are incorporated. In this demonstration, a match is found at cell 2D of the second tier. Without a match in the second tier, another tier of 16 cells can be added resulting in a 25 km search radius. The search concludes if no match is found by this stage. Theoretically, one can endlessly iterate this process, albeit at increasing computational costs.

B) Footprint relocation by TP maximization.
This new approximate footprint is further refined in a second step: the bounds of the footprint are iteratively adjusted to maximize the TP count. This process is shown in Figure 4.10. The historical image is segmented into four quadrants by bisecting its height and width. The TPs within each quadrant are counted and expressed as a fraction of the total number of TPs. For each cardinal direction (N, S, E, W), the quadrant percentages for that direction (e.g. the percentages of

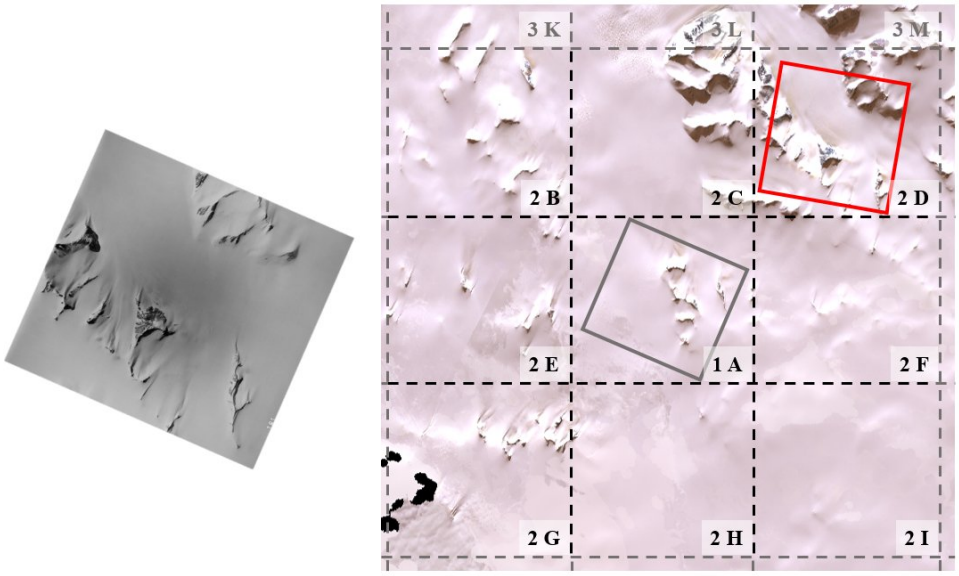


Figure 4.9: Relocating the TMA image (ID CA174632V0181) 20 km from approximate position (grey bounding box) to refined approximate position (red bounding box)

quadrant NE and NW are averaged to a percentage for N) are calculated and then multiplied by a manually predefined step size of 2500 pixels. This yields a measure indicating the image's shift in a given direction. These values are consolidated (with east and south being negative) to determine steps in x and y direction. These adjustments redefine the footprint bounds and a new TP matching is executed. If the number of TPs decreases, the refinement is stopped instantly.

Using the refined footprints, a TP matching between the historical image and the Sentinel-2 image is executed. With a sufficient number of TPs between the satellite and historical images, accurate geo-referencing becomes possible. With the precise bounds of the satellite image known, the relative TP coordinates in the satellite image are converted to absolute coordinates. These serve as the foundation for Ground Control Points (GCPs) that are used for the geo-referencing of images. The Python library Rasterio is used next to derive a transformation matrix from these TPs. This matrix is subsequently be applied to geo-reference the TMA image. The geo-referenced TMA images are then transformed into polygons, enabling the computation of the photo-centre of the TMA-image.

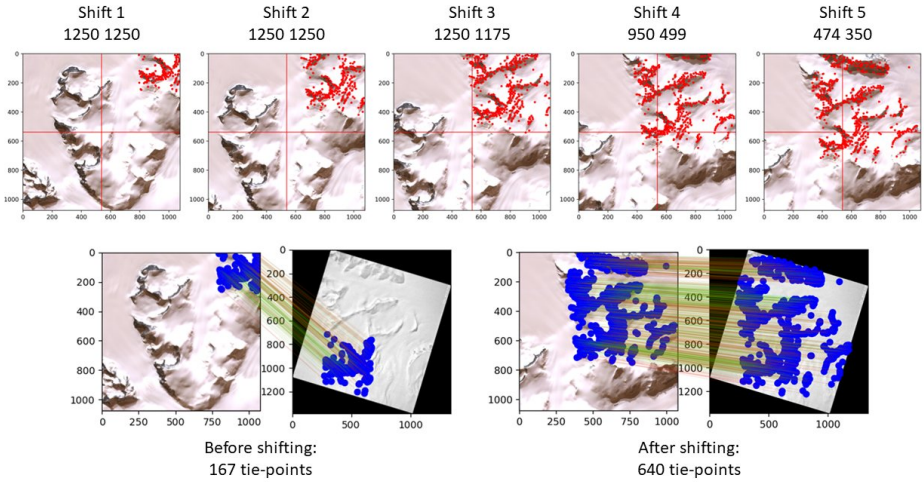


Figure 4.10: Fine shifting of TMA image (ID CA181232V0070) to improve the number of TPs by 473. The images in the top row show the process with the steps in x- and y-direction. The bottom images show the TPs before and after fine shifting.

4.3.4. Indirect Geo-referencing

While geo-referencing with satellite images proves effective for a significant number of images, challenges are encountered in specific scenarios: some images do not yield meaningful TPs, often attributed to factors like cloud cover or an absence of texture. In other situations, the detection of too many incorrect TPs undermines the efficiency of the geo-referencing process. Additionally, there are instances where TPs are primarily concentrated in a corner of an image. Even though geo-referencing might technically succeed under such conditions, the resultant footprints are unfortunately rendered invalid. For such images, we resort to one of three alternative indirect matching methods, proceeding from the first to the last as described below.

First option: flight line based footprint estimation
When for a TMA image direct geo-referencing is not possible, according to above criteria, a new approximate TMA footprint is estimated from already geo-referenced images from the same flight line. Subsequently, these problematic TMA images undergo a second round of geo-referencing with satellite images.

Figure 4.11 illustrates a segment of a flight path, highlighting the

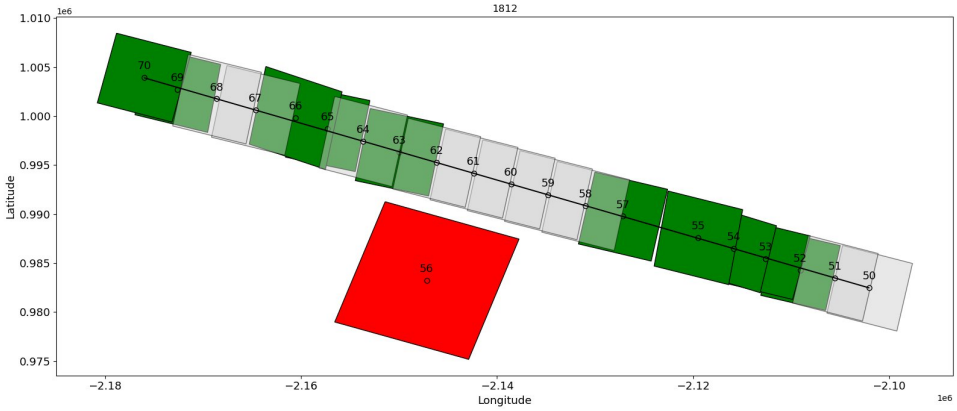


Figure 4.11: Estimation of footprints (grey) using successfully geo-referenced images (green) and ignoring the outlier image with id 56 (red). Ground-projected camera positions are indicated by numbered circles.

ground-projected camera positions of images from this trajectory and their corresponding footprints. The green footprints represent TMA images that were successfully geo-referenced using the direct approach. The coordinates of these TMA images are used to fit the flight path by regression. The singular red footprint highlights a TMA image that, while geo-referenced, is categorized as an outlier due to its position relative to the flight path. For the camera positions associated with the grey footprints, geo-referencing has yet to be successful. The grey footprints are estimated from the established flight path and the intervals between already geo-referenced images. They will subsequently serve as the basis for a renewed attempt at geo-referencing with Sentinel-2 images, since in this scenario the position of the footprint and the actual position are more similar, rendering the geo-referencing usually more successful.

Second option: matching with neighbouring TMA images

Geo-referencing of the TMA images with Sentinel-2 data can sometimes be challenging due to low information content, leading to an insufficient number or unfavourable distribution of TPs across an image. However, the aerial TMA imagery from a single flight path has a substantial overlap of up to 60%. Since these overlapping images are captured within mere seconds of each other, they essentially depict the same scene from slightly different angles. If one image from a sequence is already successfully geo-referenced, it is plausible to geo-reference it by TP matching with its immediate neighbours.

Note that this method is most accurate when limited to TMA images directly geo-referenced with Sentinel-2 data. Extending this approach beyond too many iterations may lead to the accumulation of small geo-referencing errors. As a consequence, TMA images successively geo-referenced using other TMA images can become increasingly distorted with each iteration.

Third option: extrapolating the footprint
Analogous to the footprint estimation in the first option, another polygon estimation is performed here. Between two geo-referenced polygons (from already geo-referenced TMA images from the same flight-path), an averaged polygon, is estimated from the weighted (higher weight for closer polygons) size and shape of adjacent polygons, as the new footprint for the un-referenced image. The four vertices of the estimated polygon serve as GCPs for geo-referencing the image. This modest number of GCPs suffices, given their precise positioning at the four corners.

This method serves as the final solution when all other methods have proven ineffective and is also the least accurate among them. It is only operational when the images follow a discernible order, adhere to a logical flight path, and possess similar footprints. In the case of the TMA archive, this method is viable as the images are sequentially numbered and aligned within a flight path. For images aligned in straight lines, a minimum of two images along a flight path must have been successfully geo-referenced; however, the accuracy of geo-referencing improves with an increase in the number of images available.

4.3.5. QUALITY CHECK AND METRICS

After applying the transformation matrix to a TMA image, images where geo-referencing went wrong must be removed.

Image size. A first quality check involves assessing the image size: images that are either too large or too small (exceeding a difference of 15% compared to other images in the flight path) are excluded, as illustrated by 'A' in Figure 4.12. Rather than solely focusing on the overall size of the image, it is also checked whether the pixel size of the image is consistent by deviating no more than 10% from other images in the flight path.

Footprint shape. A second quality criterion is related to the shape of the footprints. Given that the TMA cameras captured images in a square format, the transformed images should ideally be squares as

well. However, there are instances where images appear as rectangles, with one side longer than the other. In certain cases, the image corners deviate from the expected 90-degree angles, resembling a parallelogram instead. Such images, depicted by 'B' in Figure 4.12 (with the red image representing both a parallelogram and a rectangle), are also excluded if their corner angles fall outside a minimal threshold margin of 10 degrees of a right angle.

Location w.r.t. flight path. In an approach similar to the first step of indirect geo-referencing, when analysing multiple images from a flight path, it is possible to identify TMA images positioned inaccurately, as shown by 'C' in Figure 4.12. The calculated photo-centres of these TMA images are used to fit a regression line among all points, enabling the identification of outliers (points further away then two standard deviations). This approach assumes that the flight path locally follows a straight trajectory.

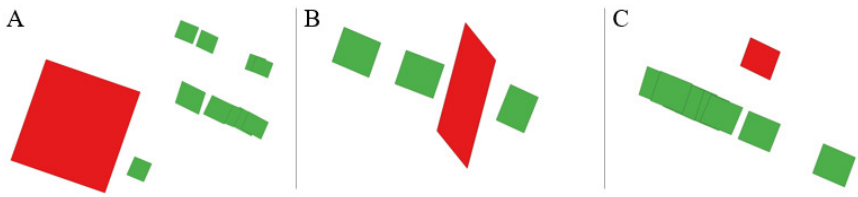


Figure 4.12: Different scenarios for invalid (red) and valid (green) images. Red footprints are rejected because of (A) footprint size, (B) footprint shape, or, (C), footprint position.

Error vector consistency. Recognizing outliers by position is further enhanced by using the original (inaccurate) approximate position of the images. We are calculating a vector between the camera centre of the real position and the approximate position. If this so-called error vector deviates from the other error vectors in the same flight path, we also disregard this image as an outlier.

4.4. RESULTS

In total, 4,459 images were used in the workflow. For the geo-referencing of the images, a desktop computer with a Quadro RTX 5000 (16GB video memory) was used. Approximately 42 hours of calculation time was required, with an average referencing time per image of 30 seconds. First, overall results and results summarized per flight path are presented in Subsection 4.4.1. In Subsection 4.4.2, we present several

case studies illustrating the application of geo-referenced historical images. For several locations on the AP, changes observed between historical photographs and modern Sentinel-2 imagery are discussed.

4.4.1. OVERALL RESULTS

Out of all images, 3,393 (76%) of the images could be geo-referenced with an average length of the error vectors being 28 km. Here, an error vector connects the initial, provided location of a TMA image with the new position of the TMA image, obtained as a result of the geo-referencing procedure. Table 4.2 shows in detail how many images could be geo-referenced by which method. In total, there are 77 flight paths with more than 10 images located in the research area. Table 3 in the appendix gives an overview of the flight paths and some averaged attributes, like information content or the length of the error vector. For 46 flight paths, all images were geo-referenced (over 99% of the images) while 15 flight paths could only be partially geo-referenced. For 13 flights, no images could be successfully geo-referenced.

method	count	information content	TPs	residuals	calculation time (s)
Direct	412	0.32	254	117.26	147
Indirect: option 1	319	0.28	80	93.94	33
Indirect: option 2	181	0.26	1112	58.18	16
Indirect: option 3	2861	0.13	4	3.32	4
not geo-referenced	665	0.12	n.a	n.a	n.a

Table 4.2: Number of geo-referenced images by method and non geo-referenced images together with their averaged parameters.

Figure 4.13 shows the geo-referenced positions of the footprints, colour-coded by the methods they were retrieved by. Each footprint has an error vector, that connects the centre of the resulting geo-referenced footprint to the centre of the initially provided footprint.

Upon closer examination of various flight paths, it is evident that many images are accurately positioned, as confirmed by the alignment with geo-referenced images from different flight paths. This is exemplified in the top two rows of Figure 4.14. However, a detailed visual inspection reveals discrepancies for parts of some flight paths. Although the algorithms successfully compute footprints without distortions or outliers for these images, these images are misplaced and only close to their real positions. Examples of this issue are depicted in the bottom

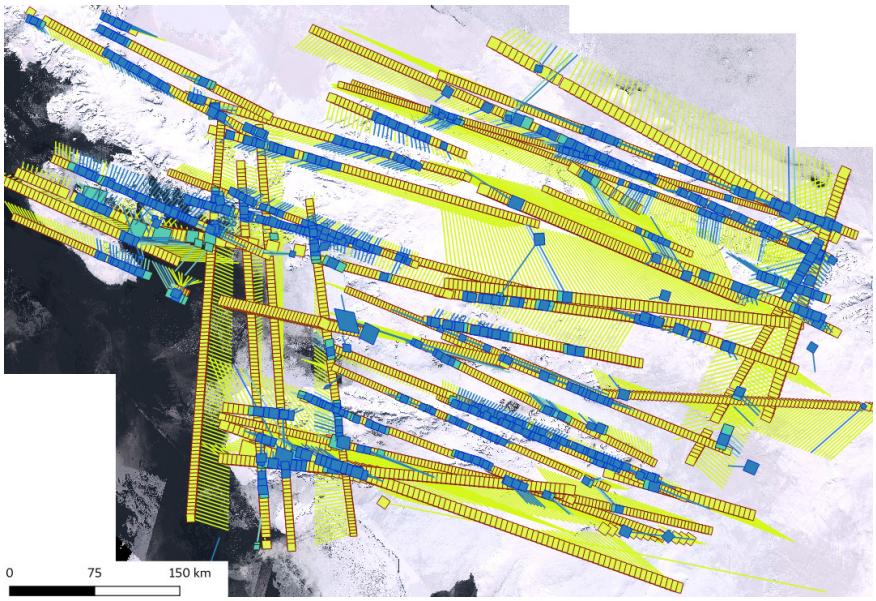


Figure 4.13: Real positions of footprints and error vectors on a Sentinel-2 composite of the AP. Footprints are coloured according to the method used for their estimation: direct geo-referencing and indirect option 1 (blue), indirect option 2 (turquoise) and indirect option 3 (yellow).

row of Figure 4.14. Currently, this problem impacts parts of 6 out of 77 flight paths.

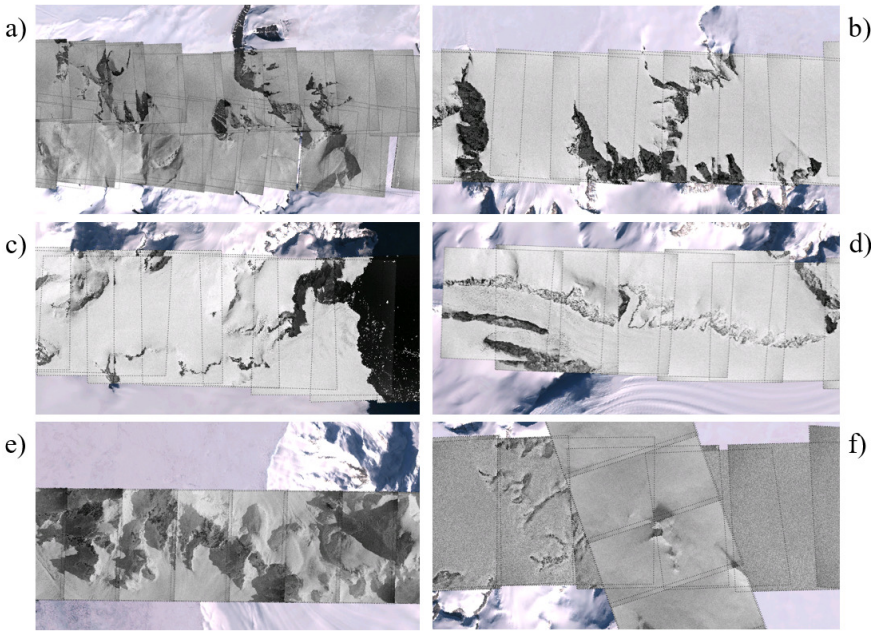


Figure 4.14: Geo-referenced historical images from flight paths 1801/1856 (a), 2142 (b), 2146 (c), 2141 (d), 1824 (e), 1812/1824 (f) overlayed over Sentinel-2 imagery. The dashed line shows the footprint of the images.

Unfortunately, no direct numerical estimation of the geo-referencing quality can be given, as the images have never been geo-referenced before. To have at least some numerical information, we manually placed 20 GCPs distributed over the research area covering different flight paths and calculated the distance to their equivalent geo-referenced position. The results of this comparison can be found in Table 4.3.

4.4.2. Case studies: surface changes revealed by comparing historical and modern imagery

In this section, we demonstrate some practical application of accurately geo-referenced historical imagery, by highlighting its value in environmental change analysis. Our analysis is by no means meant to be comprehensive, but showcases the potential of the accurately geo-referenced TMA imagery for research purposes.

Nr	Id	distance (m)	method
1	CA164432V0071	106	Indirect: option 3
2	CA174632V0171	155	Direct
3	CA182532V0011	108	Indirect: option 2
4	CA183332V0068	33	Indirect: option 1
5	CA184332V0074	99	Direct
6	CA184832V0102	119	Direct
7	CA184832V0120	47	Direct
8	CA184832V0156	14	Direct
9	CA196632V0028	28	Direct
10	CA196632V0056	19	Indirect: option 1
11	CA212332V0016	50	Direct
12	CA212332V0040	83	Direct
13	CA212432V0106	135	Indirect: option 2
14	CA214232V0162	185	Direct
15	CA214232V0193	117	Direct
16	CA216432V0215	12	Indirect: option 1
17	CA216632V0291	161	Indirect: option 1
18	CA216632V0305	67	Direct
	avg	85.44	

Table 4.3: Distance in meter between the geo-referenced position and the real position of sample points for different TMA images, indicated by their TMA Id.

By employing the Structural Similarity Index (SSIM) (Z. Wang *et al.* 2004), we quantified the similarity between historical images and their modern counterparts from Sentinel-2, considering their geo-referenced extents. The SSIM is a metric used to measure the similarity between two images, assessing changes in structural information, luminance, and contrast rather than just pixel differences. Figure 4.15 provides an overview of the similarity scores associated with these geo-referenced images. A blue colour indicates high similarity between images, whereas a red colour indicates that the historical images and their modern counterparts are not similar.

In our analysis, we observed that for most images, particularly those away from coastal regions, there is a high similarity score. This indicates that these areas have not undergone substantial changes over time. Conversely, images near the coast exhibited greater changes, primarily attributed to factors such as changes in sea ice cover. While it is important to note that these variations could also be influenced by other elements such as cloud cover or deviations in image quality, this initial assessment provides a useful baseline for understanding environmental changes in these regions.

With the help of these similarity scores, we identified some locations of interest, i.e. locations in Antarctica, exhibiting significant environmental change over a 60-year period. Figure 4.16 and 4.17 present these comparative examples: on the left, the geo-referenced historical imagery, on the right modern co-located Sentinel-2 imagery with enhanced natural colour visualization. The dates of the historical images are given in the figure, the dates of Sentinel-2 satellite images are from the corresponding months in 2023. As for subfigure d) no month is known, the month of January was chosen. The positions of these examples are indicated in Figure 4.2.

In Figure 4.16 and Figure 4.17, a series of subfigures (a to e) illustrate glacier retreats in various highlighted areas. Focusing on subfigure a), the Shoesmith Glacier on Horseshoe Island is featured, characterized by two distinct branches: one leads into Lystad Bay to the west and the other into Gaul Cove to the northeast. From a direct image comparison, we estimate a retreat of 700m at the western terminus, contrasted with a 300m retreat at the northeastern terminus. Rodrigo *et al.* (2021) previously reported a retreat of approximately 200m in the western part from 1999 to 2012, yet no data were available for the northeastern part. Coincidentally, a British gravity survey was conducted in this region in 1969, aligning with the time of the photograph (Smith 1973). The McMorrin Glacier, showcased in subfigure b), is located on the Fallières Coast, extending westwards from Mount Metcalfe to Marguerite Bay. Based on the partial visibility in the historical imagery, a retreat

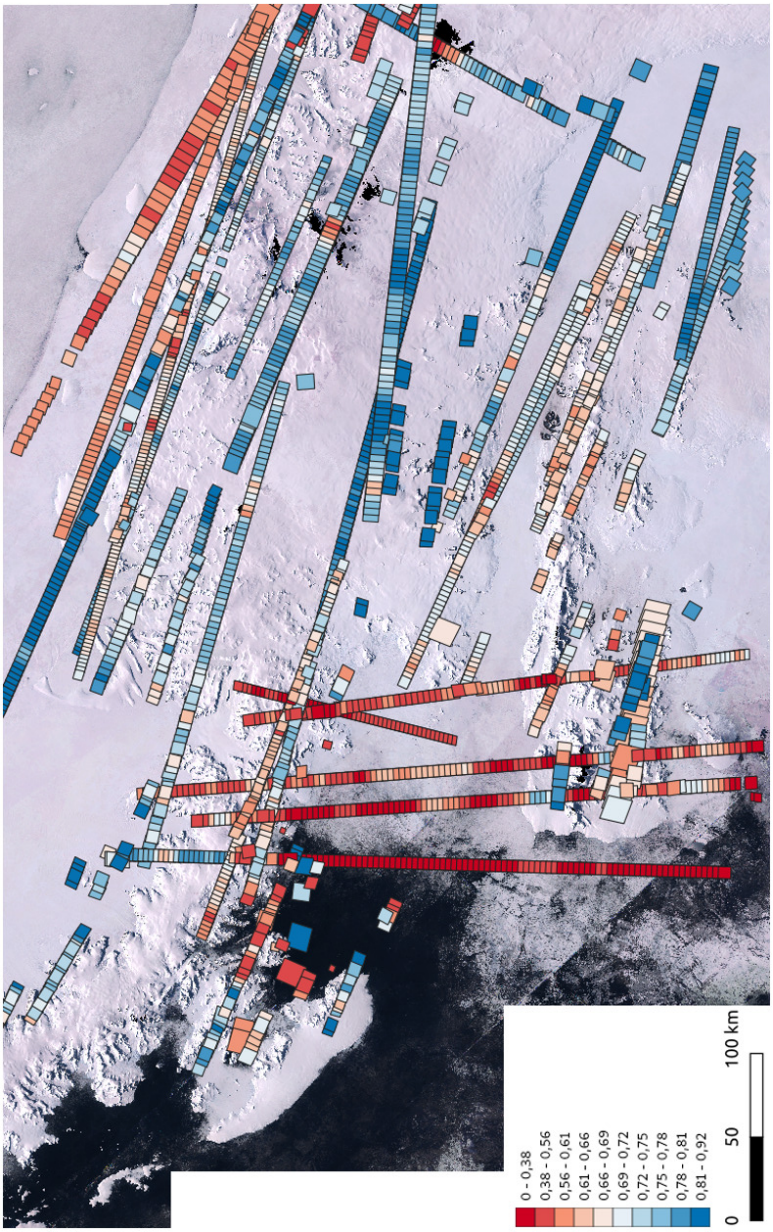
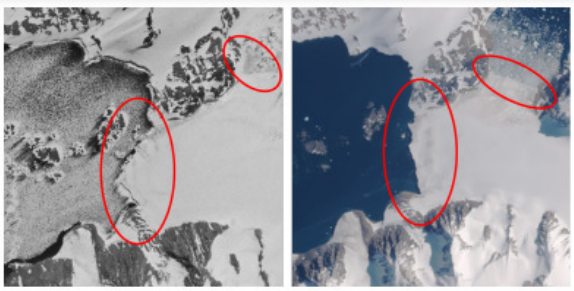
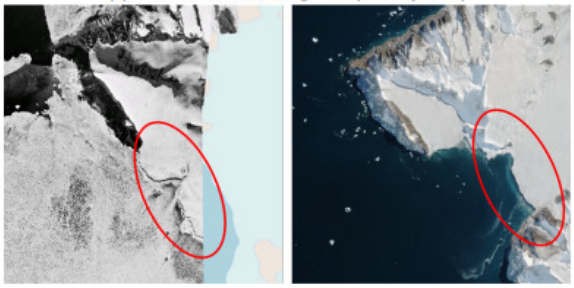


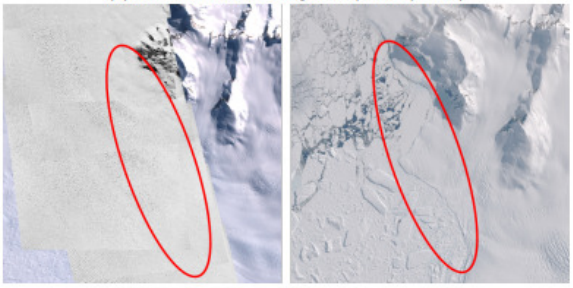
Figure 4.15: Similarity between historical images and co-located modern Sentinel-2 imagery. Red indicates low similarity, while blue indicates high similarity.



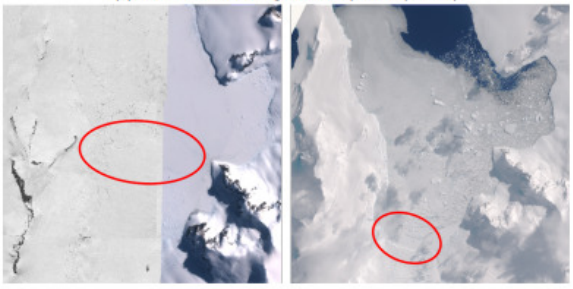
(a) Retreat of Shoesmith glacier (January 1969)



(b) Retreat of McMorrin glacier (January 1969)

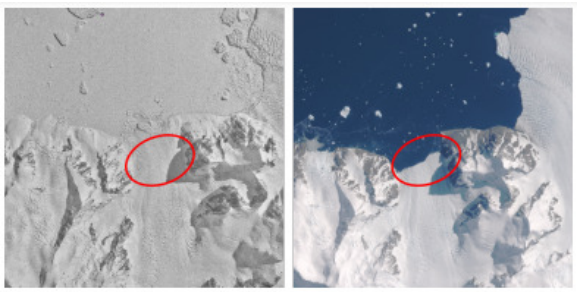


(c) Retreat of Fleming Glacier (January 1969)

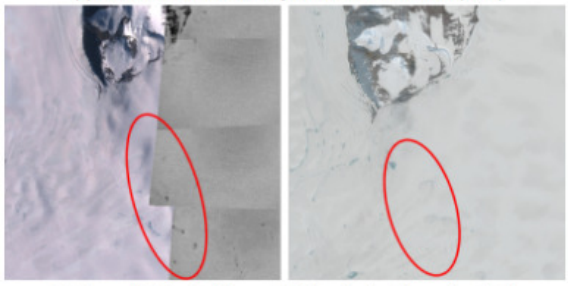


(d) Retreat of Somigliana Glacier (1956)

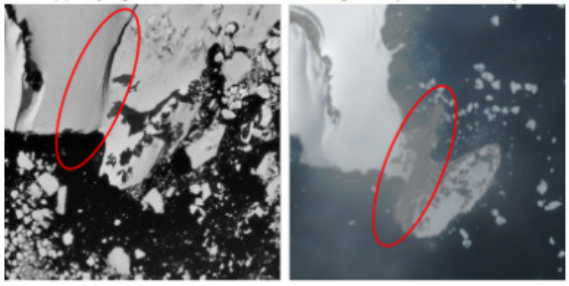
Figure 4.16: Examples for changes between historical images (left) and modern Sentinel-2 images (right). Notable changes are marked with red.



(e) Retreat of an unnamed glacier near Mt Wilcox (1966)



(f) Supraglacial melt lakes on Sibelius glacier (December 1966)



(g) Rothera research station (January 1969)

Figure 4.17: [Continued] Examples for changes between historical images (left) and modern Sentinel-2 images (right). Notable changes are marked with red.

of 1500m is estimated, leading to the isolation of a small northern sub-glacier. In subfigure c), Fleming Glacier is depicted, located along the Fallières Coast and terminating at the Forster Ice Piedmont, to the east of the Wordie Ice Shelf. Zhao *et al.* (2017) observed a significant decrease in elevation at the glacier's front, noting a reduction of 60 meters from 2002 to 2014. Similarly, Friedl *et al.* (2018) reported a retreat of the grounding line by 6-9 km from 1996 to 2011. These findings are consistent with the historical imagery covering the northern part of the glacier, which reveals a clear change in the glacier front's position. In the modern image, the glacier front is distinctly visible, whereas in the historical image, the glacier front cannot be discerned at the marked position, nor further seawards. subfigure d) presents the Somigliana Glacier, located on the northern Arrowsmith Peninsula and extending towards Langmuir Cove. Analysis reveals a retreat of 2,000m from 1956 to 2023. Similarly, subfigure e) examines a glacier near Mt. Wilcox, displaying a 600m retreat from 1966 to 2023. In A. Cook *et al.* (2014) and A. J. Cook, Holland, *et al.* (2016), a surface area reduction of 15% (1957 to 2009) and 5% (1940 to 2001) was reported for these glaciers, respectively. In subfigure f), we observe supraglacial melt lakes on the Sibelius Glacier in central Alexander Island. Sentinel-2 imagery shows annual lake formations from 2014 to 2023, a phenomenon dating back to at least 1966 as per TMA images (Corr *et al.* 2022; Moussavi *et al.* 2020). Lastly, subfigure g) captures Rothera Point on Adelaide Island, where the British Rothera Research Station is situated. These historical images predate the construction of the research base. In the satellite imagery, the station's landing strip is discernible (marked with a red square).

With the images now correctly oriented vertically, we can automatically obtain the accurate positioning of the oblique images, too, which facilitates additional comparisons. As a first example, Figure 4.18 presents a comparative oblique view of the Fleming Glacier (at position c) in Figure 4.2), juxtaposing an image from January 1969 with a Sentinel-2 image from January 2023. The latter is artificially rendered in 3D while approximating the same viewpoint.

4.5. DISCUSSION

A large fraction (around 80%) of the images from the TMA archive in our study area have been successfully geo-referenced. While in the past only an approximate position was available, which deviated by several kilometres and lacked exact image footprints, these images are now situated at their actual position, scaled to their true size, and aligned with their genuine orientation. As the cameras for the vertical and

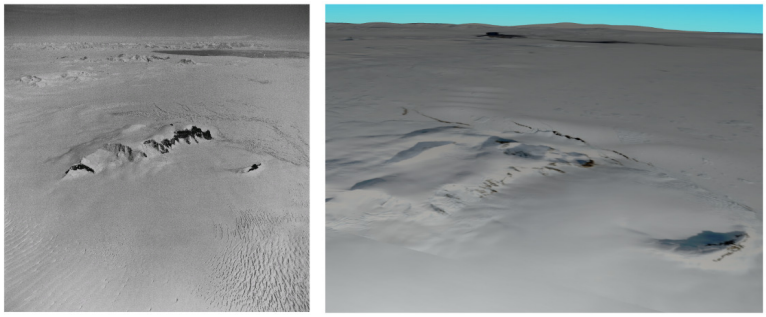


Figure 4.18: Oblique view on Fleming Glacier from TMA imagery in January 1969 (left) and a Sentinel-2 3D rendering from January 2023 (right).

both oblique directions are located at the same position, the vertical geo-referencing can easily be extended to the oblique images.

Given the inherent limitations of the historical data and the lower geometric resolution of the Sentinel-2 satellite imagery, geo-referencing accuracy is often confined to several meters. Moreover, the quality of some image scans is compromised due to the poor image quality (degradation of the images & scanning errors), posing challenges in accurate geo-referencing. This level of precision, while not comparable to that of modern images, represents a substantial advancement compared to the approximate camera positions.

Even though only 20% of the historical images can be geo-referenced directly with satellite images, these images are the key to further successful geo-referencing. They play a pivotal role in the broader geo-referencing process and serve as references for geo-referencing adjacent images. Furthermore, they prove instrumental as reference points when computing the footprints of the remaining images. These direct geo-referenced key images are typically characterized by distinct rock structures, which have proven to be remarkably stable for many decades. For most of these key images successful geo-referencing is possible; however, there are three limitations: first, for some images, the changes between historical and recent images are too substantial to establish meaningful TPs (top part of Figure 4.19). This issue is particularly prevalent in coastal areas, where over time the extent of glaciers and ice shelves has drastically changed. Furthermore, long-term and interannual variations in sea ice cover further frustrate the geo-referencing process when images contain a large fraction of the ocean. Second, TPs must be dispersed across various parts of the image. When distinct structures are concentrated in just one corner of the image (middle part of Figure 4.19), geo-referencing tends to

be unsuccessful, as in such cases, while the image might be correctly positioned, its footprint becomes distorted due to the imbalanced distribution of reference points. Lastly, even though the information content computed in Subsection 4.3.3 can serve as a valuable metric to identify these key images, it is not without limitations. In some cases, images may lack sufficient meaningful content (right part of Figure 4.19) or the TPs are randomly distributed, often due to cloud coverage, thus hindering the identification of meaningful TPs. On the opposite, we found that prevalent shadows in both the Sentinel-2 imagery and the historical images are only marginally influencing the TP matching and geo-referencing.

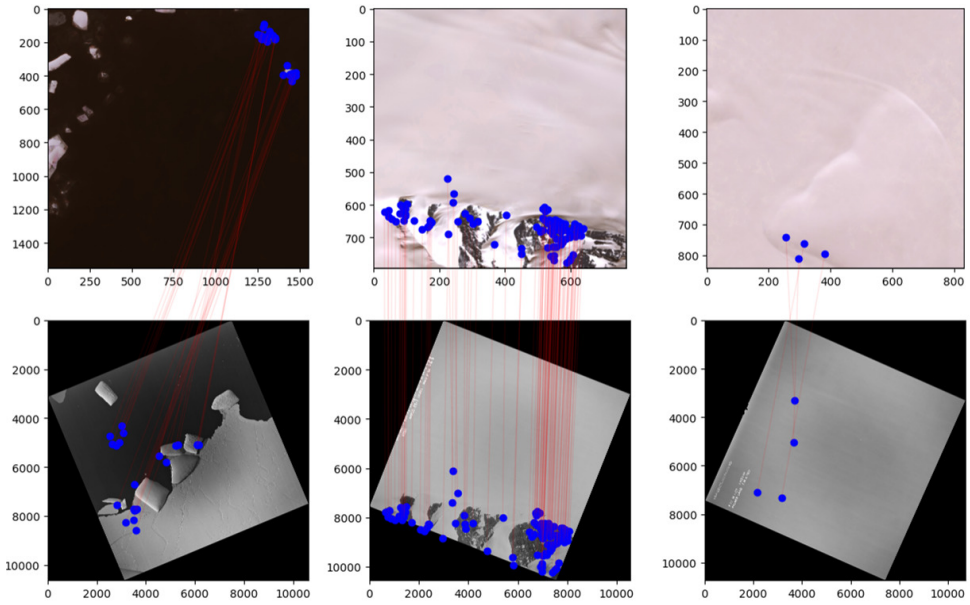


Figure 4.19: Examples for images where the geo-referencing failed due to changed landscape (left), unequal distribution of TPs (middle) and too few TPs (right). Modern Sentinel-2 imagery is on the top, historical images are on the bottom.

The main reason that for some flight paths no images can be geo-referenced, is the low information content of all their images. The average information for these flight paths is significantly lower (average 0.07) compared to those where geo-referencing is successful (average: 0.18). For instance, the images of some flight paths predominantly feature snowfields, leading to monochromatic and uniformly white frames. These conditions render even manual inspection methods for geo-referencing ineffective, underscoring again the importance of diverse landscape features in images for accurate geo-referencing.

Conversely, a high information content can also hinder the geo-referencing. This often indicates dynamic landscapes, such as coastal regions, where significant changes in the landscape can occur. Images of this flight path feature coastal areas that have undergone substantial alterations, complicating the geo-referencing process. Additionally, images with extensive cloud coverage present a high information content as well. However, the clouds obscure essential landscape features, rendering the further geo-referencing task impossible.

Despite these challenges, when at least 2 images of a flight path can be geo-referenced successfully, the workflow can usually successfully extrapolate the position of the other images. During the geo-referencing, it was noticed that some images were scanned with an incorrect rotation, leading to accurate footprint extents but misaligned orientations. Furthermore, for some images, there were gross errors in the position of the provided approximate footprint: initially, we assumed that positional deviations in images would be limited to 30 km. However, our recent findings indicate discrepancies as large as 70 km. Increasing the search radius of the TMA footprint relocation can detect these deviations, but they impose a significant computational burden, increasing processing times to approximately 15 minutes per image (from the majority of images having a calculation time of 46 seconds).

As described in the results, for some flight paths, the images are technically successfully geo-referenced but misaligned from their true positions. This issue predominantly arises for images geo-referenced with the third option (calculating the position of images), especially those situated further from the directly geo-referenced images within the same flight path. Minor discrepancies in angle and distance between the directly geo-referenced images accumulate, leading to the misplacement of the images with calculated positions. However, they are generally located in the vicinity to their real positions, allowing for the possibility of manual geo-referencing. Additionally, contrary to the initial hypotheses that all flight paths are straight lines, it has been observed that some flight paths deviate significantly from linear trajectories, incorporating sharp turns (see bold red flight path in Figure 4.2). This leads to greater deviations of the images from their actual positions, challenging the traditional assumption of uniformly straight and predictable flight paths. To address this, it is necessary to manually segment the flight path into different sections to guarantee accurate geo-referencing.

In conclusion, our observations underscore the multifaceted challenges in geo-referencing aerial images. From issues related to low information content and dynamic landscapes to errors in image rotation and underestimations of deviations, each factor significantly impacts the

accuracy of geo-referenced data. As we continue to refine our methodologies and deepen our understanding of these complexities, we aim to enhance the precision and reliability of geo-referencing in aerial imaging.

4.6. CONCLUSION

In this study, we outlined the methodology for geo-referencing historical aerial photographs. We employed advanced neural network-based TP matching techniques, creating a unique workflow specifically optimized for high-resolution, large historical imagery. Geo-referencing is achieved by matching TPs between the historical photos and contemporary satellite imagery, guided by preliminary footprints. If initial attempts are unsuccessful, we employ a tiered approach: first, re-attempting with refined footprints from satellite data; second, utilizing geo-referenced neighbouring images; and finally, computing the image's footprint directly. Rigorous quality control measures are then implemented to guarantee the precision of the geo-referenced outputs. The efficacy of our method was showcased in a practical application, in which we successfully geo-referenced 76% of the images of an extensive collection of aerial photographs from the AP and highlighted differences between the historical and modern images for 7 scenarios.

While our geo-referencing technique proves successful for a majority of the images in the subset and these images are geo-referenced on a large scale for the first time, certain challenges persist. Many images, due to their sparse informational content, cannot be geo-referenced through TP matching alone, necessitating the deployment of our tiered strategy. Due to the geometric resolution of our modern data and the limited quality of the historical data, our method remains constrained to meter-level precision. For the majority of the flight paths in our research area, we were able to geo-reference at least some images from the flight path using satellite data, ensuring the following geo-referencing via extrapolations. However, should a flight path entirely lack satellite-based geo-referencing, estimating the location of its images becomes unattainable.

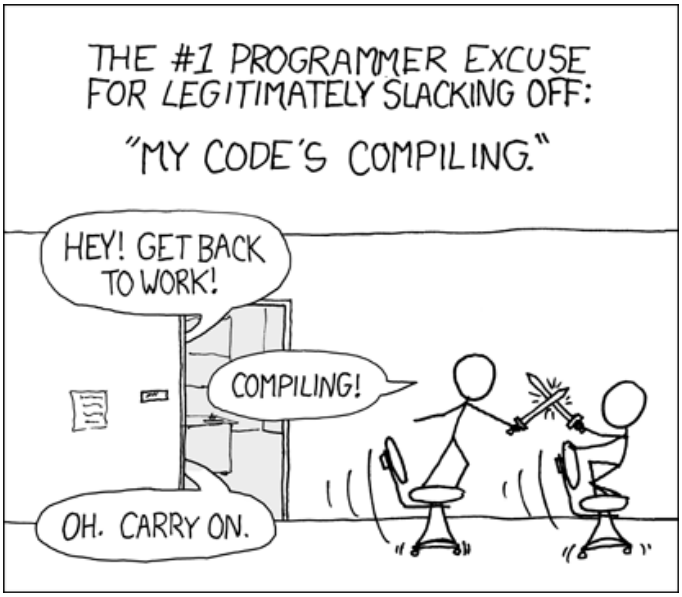
For the future, our primary focus is on increasing the number of images that can be directly geo-referenced with modern data. This is crucial as it significantly influences the quality of our method. Enhancements can be achieved either by refining the TP matching process (e.g. LoFTR (Yuan *et al.* 2022) or DISK (Tyszkiewicz, Fua, and Trulls 2020)), retraining the neural network for TP matching with historical images, or by further optimizing the historical imagery to ease the identification of TPs. In

the current approach, an initial approximation of the image's location is required, which should be relatively close to its actual location. The step of refining the approximate position is notably time-consuming. We plan to employ machine learning, specifically U-nets, for this initial approximation refining, allowing for the faster identification of a historical image's location within a larger area. To enhance the functionality of the geo-referencing workflow, non-linear flight paths should be automatically recognized in the future and be taken into account during the line-fitting. Finally, it would be beneficial to evaluate this workflow's adaptability to other regions and contexts, including historical aerial imagery from places like Greenland or Svalbard, or entirely different settings such as urbanized regions.

Geo-referencing the images from these historical archives offers many opportunities for subsequent research. Potential applications include monitoring changes in glaciers, ice shelves, and other cryospheric features over time, thereby furnishing tangible proof of Antarctica's response to global warming. Given that these images overlap, they can be used to construct 3D models using SfM. These models can then be compared with contemporary elevation data, like satellite stereo-photogrammetry and altimetry (e.g. ICESat-2), to procure an understanding of elevation and mass variations.

Data Availability
The data for this submission is available at Dahle (2024c). An online visualization of the footprints of the historical data & more recent geo-referenced images can be found at https://polar-archive.citg.tudelft.nl/.

Code Availability
The code for this submission is available on https://github.com/fdahle/hist_georef.



© xkcd 303 - Compiling

5

3D RECONSTRUCTION OF HISTORICAL GLACIERS

Historical aerial imagery is an underutilized resource for studying glacier evolution over multi-decadal timescales due to challenges such as inconsistent metadata, degraded image quality, and the complexity of processing oblique imagery. Here, we introduce a fully automated Structure-From-Motion (SfM) workflow specifically designed for digitised analogue images from the Trimetrogon Aerial (TMA) archive of the Antarctic Peninsula (AP), dating between 1940 and 1990, with the majority of the images taken around 1960. Our pipeline automates critical steps including Tie Point (TP) matching, image pre-processing, geo-referencing via modern satellite data, and Ground Control Point (GCP) generation using stable terrain features. Crucially, our method does not depend on known camera positions and effectively integrates vertical and oblique images. We demonstrate the workflow by successfully reconstructing historical Digital Elevation Models (DEMs) for 49 glaciers across the AP without any manual intervention. Comparison with contemporary datasets reveals median elevation differences of approximately 90 meters across entire glaciers and 76 meters in stable areas. While these differences reflect both reconstruction uncertainty and real glacier change over several decades, the results are sufficiently accurate to support regional-scale assessments of long-term elevation and mass balance change. This automated approach significantly enhances the feasibility of large-scale glacier reconstructions, offering valuable new opportunities for quantifying long-term cryospheric changes at a time interval of 50 years. The reconstruction workflow is openly available on GitHub and can be easily adapted to other historical or modern datasets.

This chapter is intended for later publication as a journal article.

5.1. INTRODUCTION

Long-term observational data are essential for understanding glacier evolution and quantifying mass balance changes over decadal timescales. Consistent monitoring of Antarctic glaciers began with satellite altimetry in the early 1990s and was complemented by satellite gravimetry (GRACE) from 2003 onward, enabling comprehensive assessments of ice sheet mass balance (Shepherd, Ivins, van den Broeke, *et al.* 2019). This relatively short observational record leaves substantial uncertainties regarding the long-term response of Antarctic glaciers to climate variability. Historical aerial imagery offers a unique opportunity to extend the observational record and reconstruct past glacier states, providing valuable insights into long-term glacier behaviour prior to the satellite era. Among the most significant historical data sources is archival aerial photography, especially from mid-20th century survey campaigns. Using photogrammetric techniques, these analogue images allow for the creation of DEMs and orthophotos that represent glacier conditions at the time of acquisition. Advancements in SfM photogrammetry (for example Brynte *et al.* (2024) and Elflein *et al.* (2025)), combined with developments in machine learning-based TP matching (Lindenberger, Sarlin, and Pollefeys 2023), have improved the feasibility and accuracy of such reconstructions.

Several studies have demonstrated the effectiveness of SfM techniques on archival imagery across various geographic settings. In New Zealand, Vargo *et al.* (2017) used historical photographs to reconstruct equilibrium-line altitudes and length changes of the Brewster Glacier from 1981 to 2017. Mölg and Bolch (2017) compared SfM and conventional photogrammetry on Swiss glaciers and emphasized the importance of accurate GCPs (geo-referenced locations used to align imagery) and quality control. In the Arctic, Holmlund (2021) reconstructed Svalbard's Aldegondabreen using oblique photographs from the early 20th century, and Eskandari *et al.* (2024) applied SfM techniques to historical imagery of the Mont Blanc Massif in the European Alps. Similarly, Mertes *et al.* (2017) applied SfM to multiple European datasets dating back as far as 1896. Girod *et al.* (2018) presented an optimized workflow for processing oblique aerial imagery acquired in 1936, demonstrating that modern SfM methods can deliver high-quality digital elevation models with sufficient accuracy to reliably quantify terrain changes over decadal scale. In Svalbard, Geyman *et al.* (2022) reconstructed historical glacier elevations for over 1,500 glaciers using archival aerial imagery from 1936–1938, quantifying long-term glacier thinning rates and their sensitivity to temperature changes.

A recurring challenge across these studies is the high degree of manual labour required for image preparation, alignment, and GCP

placement. Additionally, most workflows rely exclusively on vertical aerial images, often excluding oblique views due to the complexity of their geometry and the difficulty of manual TP matching. As a result, many studies are restricted to single-glacier case studies and achieve only limited automation. This limits the scalability of such approaches and constrains their applicability to large, heterogeneous archives like those of Antarctica. To fully harness the potential of historical aerial imagery for regional-scale and long-term glacier monitoring, the development of robust and automated workflows is essential.

In our study, we make use of the TMA archive to reconstruct historical surface elevations for glaciers on the AP. This dataset contains over 300,000 high-resolution black-and-white photographs collected by the U.S. military between the 1940s and 1990s. Despite its potential, relatively few studies have fully leveraged this archive for deriving elevations. Notably, Child *et al.* (2020) used TMA images from 1978–1979 to reconstruct glacier surface elevation over Byrd Glacier in East Antarctica, demonstrating surface stability over four decades. More recently, North and Barrows (2024) created 3D models of glaciers in the Larsen B region using TMA imagery, though their workflow was restricted in spatial coverage and required significant manual input. Zhao *et al.* (2017) reconstructed historical surface elevations over the Wordie Ice Shelf region by generating a DEM from TMA imagery acquired in 1966, which they compared to a 2008 SPOT-derived DEM to quantify long-term glacier mass loss.

Recent approaches have aimed to increase the level of automation through the use of machine learning and advanced image-matching algorithms. For example, Maiwald, Feurer, and Eltner (2023) applied new machine learning based methods for improved TP detection, resulting in 3D-reconstruction in higher quality. Knuth *et al.* (2023) developed a pipeline to automatically process archival imagery from North America into consistent time series of high-resolution (0.5–2 m) DEMs and orthomosaics without requiring manual GCPs. However, these methods benefit from high-quality metadata and datasets with known camera positions and depicting heterogeneous terrain, which simplifies the reconstruction process. To date, no fully automated approach has been demonstrated that operates reliably on datasets lacking this metadata and covering the complex terrain of the AP.

Building on previous efforts, we present a fully automated workflow for reconstructing historical glacier elevations across the AP, using imagery from the TMA archive as a case study. The workflow is implemented in Python and leverages Agisoft Metashape, a photogrammetric software widely used for 3D reconstruction. Our pipeline addresses several limitations of prior work by incorporating both vertical and oblique

images and automating critical steps such as TP extraction, model alignment, and elevation correction. In contrast to studies relying on precise metadata or surveyed camera positions, our workflow is designed to operate without reliable camera location data, making the automation task substantially more difficult. The AP poses additional challenges due to its rugged terrain, frequent cloud cover, and limited availability of contemporary support datasets. Nevertheless, its high density of archived images and scientific importance make it an ideal region for testing the scalability and robustness of an automated historical SfM pipeline.

By reconstructing DEMs for numerous glaciers across the Antarctic Peninsula, this study aims to expand the availability of long-term glaciological data and offer new insights into regional glacier dynamics. Our work demonstrates the feasibility of automating SfM-based historical reconstructions under difficult conditions, providing a valuable tool for cryospheric science and climate change research. The accuracy of the reconstructed DEMs was assessed through comparisons with the REMA reference DEM and qualitative validation.

5.2. METHODOLOGY

5.2.1. DATA

This study relies on two main types of data: historical aerial imagery and modern geospatial datasets. The historical images form the basis for 3D reconstruction using SfM-techniques. However, to fully automate the SfM workflow and enable accurate geo-referencing, modern data are also incorporated. While the historical data are used to generate the final models, the modern datasets serve only as external references to support alignment and calibration; they are not included in the final outputs.

Historical Data

The historical aerial images used in this study were captured by the U.S. military for mapping purposes between 1940 and 1990 in different campaigns. This dataset comprises approximately 300,000 images and is part of the TMA photography archive. Pictures were taken in combination of three images (as shown in Figure 5.1): one vertical and two oblique images taken at 30-degree angles to the left and right of the aircraft. The images were collected during multiple flight paths (see Figure 5.2), primarily covering the AP and the coastlines of

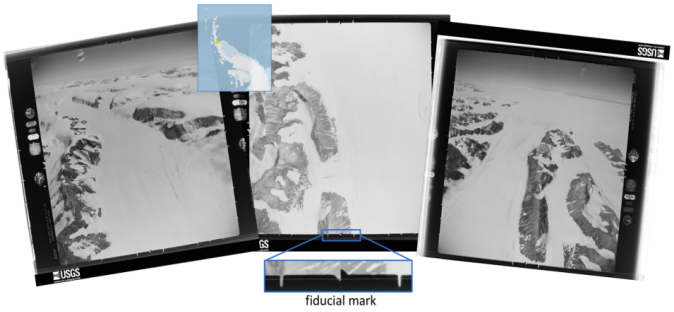


Figure 5.1: Example of a TMA image triplet, consisting of one vertical and two oblique photographs.

West Antarctica. With an overlap of approximately 60%, the images are well-suited for SfM workflows. Despite their age, the resolution of these images is remarkably high, typically around 1m depending on flight height, capturing fine details that provide a unique glimpse into the state of glaciers before the satellite era, a period with very limited alternative data sources.

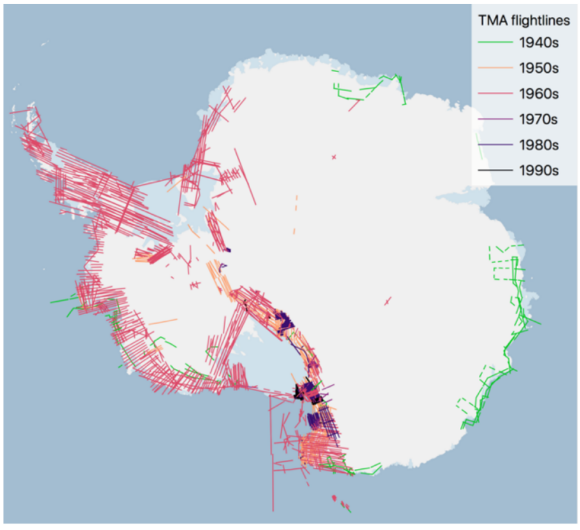


Figure 5.2: Historical flight paths from the TMA archive over the AP, colour-coded by approximate acquisition dates.

The images were captured using black-and-white analogue film cameras and were later digitised in 2012 by the Polar Geospatial Center [1]. The scanning was performed at a high resolution of 0.025 mm

[1]The images are freely available at Polar Geospatial Center (2023)

per pixel, preserving the fine details critical for photogrammetric analysis. However, the scanning process introduced several challenges, including the presence of Newton rings, upside-down scans, and occasional artifacts. Additionally, while some metadata exists, it is often incomplete or unreliable, as the original records of the images are not available. For instance, internal camera parameters (focal length, lens distortion) and external camera parameters (position and orientation) are not consistently available and, in some cases, are erroneous. The rough geo-referencing of the images further complicates processing, with deviations sometimes reaching several dozen kilometers. Furthermore, not all images in the dataset are usable. Certain images lack discernible features, are entirely obscured by clouds, or suffer from severe underexposure or overexposure. These unusable images are excluded from the SfM workflow to maintain data quality and ensure reliable outputs.

In our previous work, we developed methods to address some of the inherent challenges in processing this dataset. First, we introduced an image segmentation approach (Dahle, Lindenbergh, and Wouters 2024b) to automatically mask unusable regions of the images, such as those obscured by clouds or containing only ocean. Second, we presented a workflow for automated metadata extraction from historical aerial imagery (Dahle, Liu, *et al.* 2025), which combines computer vision and machine learning techniques to detect fiducial marks and extract metadata from the historical images. Third, we implemented a geo-referencing methodology (Dahle, Lindenbergh, and Wouters 2024a) that aligns the historical images with modern geospatial datasets, enabling the creation of absolute elevation models. These two pre-processing steps are integral to automating the SfM workflow in Agisoft Metashape and ensuring the quality of the resulting 3D models.

Modern Data

To enable the automatic reconstruction process and especially convert relative measurements to absolute values, we incorporate modern geospatial data from three key sources.

First, Sentinel-2 data, a multi-spectral satellite mission developed by the European Space Agency, provides medium-resolution optical imagery. Sentinel-2 was selected because it is freely available, easy to access, and provides complete coverage of Antarctica; however, other satellite datasets with comparable resolution and availability could also be used within this workflow (e.g Landsat-8 and 9 or WorldView). For this study, we utilize only the Sentinel-2 optical bands (B02-B04) with a 10-meter resolution, as they offer a good balance between detail and computational efficiency. To ensure consistent and reliable coverage, a cloud-free composite of all images from 2024 is generated.

Second, the Reference Elevation Model of Antarctica (REMA) (Howat *et al.* 2019), developed by the Polar Geospatial Center, is used to extract the z-component of the GCPs. REMA is a high-resolution DEM derived from stereoscopic satellite imagery, with multiple resolutions available. For our purposes, we use the 10-meter resolution, which balances data quality with storage efficiency, making it ideal for large-scale geospatial analysis.

Finally, the Quantarctica GIS package (Matsuoka *et al.* 2021), a freely available collection of geospatial datasets tailored for Antarctic research, provides a wide range of geospatial data layers, including a 30m resolution rock mask derived from Landsat8 imagery (Burton-Johnson *et al.* 2016). In this study, it is used to further refine the GCPs by identifying stable, non-dynamic areas such as exposed rock outcrops. This mask is crucial for restricting GCPs to stable features such as rock outcrops, thereby enhancing the accuracy of the reconstructed model. While the Quantarctica rock mask is a valuable supplementary dataset, its inclusion in this workflow is optional.

5.2.2. PREPARATION OF THE PICTURES

While the workflow can use the input images directly, applying a series of pre-processing steps to each image increases the chances of a successful reconstruction and improves the overall quality of the result. The complete workflow for image preparation is shown in Figure 5.3.

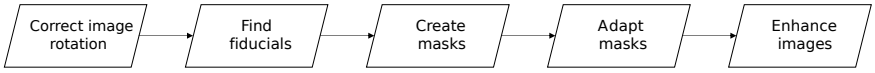


Figure 5.3: Workflow illustrating the sequential pre-processing steps applied to historical images before initiating the SfM reconstruction pipeline.

Correct image rotation
Many of the original images, both vertical and oblique, are incorrectly rotated. In some cases, images were scanned upside down, leading to an incorrect placement of the image borders that contain the fiducial markers (for example the right image in Figure 5.1). This misalignment prevents the automatic extraction of internal camera parameters and hinders camera calibration, as the fiducials are also inverted. In other cases, especially in oblique imagery, the camera itself was mounted upside down during acquisition, resulting in inverted images with correctly placed borders. These rotation issues complicate feature matching, as many algorithms are not fully rotation-invariant and may fail to find sufficient TPs between images.

To address these issues, all images must be checked and, if necessary, rotated. For the first case, where borders are misplaced due to upside-down scanning, we use a subset of the images that contain either the left or right side of the border. For each subset, we calculate a complexity score using the homogeneity analysis by Haralick, Shanmugam, and Dinstein (1975). The correct side (typically the left) contains elements such as the altimeter, which results in a higher complexity score. If the right-side subset has the higher score, the image is rotated accordingly.

In the second case, where the camera was mounted upside down, placing the sky at the bottom of the image, we applied the image segmentation method described in Dahle, Lindenbergh, and Wouters (2024b). In oblique images, the segmentation often includes the class 'sky', which is expected to appear at the top of the image. Based on this, images can be correctly rotated. This method does not apply to vertical images or oblique images without visible sky, where no reliable class information is available. In those cases, we apply a fallback strategy during the TP matching step: if no TPs are found using the original image rotation, we attempt matching with a 180-degree rotated version. However, correcting rotation in advance is preferred, as it improves matching efficiency and reduces processing time.

Find fiducials

Fiducial marks are typically used to define the internal geometry of a film camera and establish a link between the scanned image coordinates (in pixels) and the calibrated camera model. Agisoft Metashape provides built-in tools for detecting these fiducial marks to support parameter extraction. However, the automatic detection is not always reliable, especially in degraded or non-standard imagery. In such cases, we use our own parameter extraction workflow, which combines computer vision techniques and estimation methods, as described in Chapter 3.

Create & adapt masks

Because these images were not originally acquired with SfM applications in mind, many standard photogrammetric practices were not followed during capture. As a result, the images often include non-relevant elements such as borders, fiducial marks, and annotations. These elements do not represent real-world features but can still produce artificially high matching scores during alignment, which may distort the reconstructed geometry. To avoid such errors, it is essential to mask these elements before processing.

Agisoft Metashape includes an automatic masking tool for detecting and masking image borders. However, its effectiveness is sometimes

limited, particularly with the grayscale historical images used in this study. The masks are often placed too close to the image edges, leading to incorrect matches near the borders. Furthermore, in some cases, valid image content such as dark rock outcrops near the edges are mistakenly masked. To address these limitations, we apply a two-step masking strategy.

First, we expand the default masks by adding a buffer zone around the image edges. This reduces the likelihood of incorrect feature matches near the borders. Second, we apply our own masking method by automatically masking the detected annotation text boxes and estimating the image borders based on the fiducial mark positions, as identified through our parameter extraction workflow (see Chapter 3). These additional masks are combined with the default Agisoft masks to ensure more complete and consistent masking across the dataset.

Enhance images
Photo enhancement is applied to improve feature detection, particularly in low-contrast images dominated by snow-covered scenes. By highlighting subtle features such as small snowdrifts, this step improves TP detection during the image-matching process and enhances the visual quality of the resulting DEMs. The enhancement process involves automatic adjustments to brightness and contrast values, following the methodology described in McNabb *et al.* (2020). This optimization ensures that subtle surface features are more easily detected, contributing to improved feature matching and better overall model quality.

5.2.3. STRUCTURE-FROM-MOTION WORKFLOW

Our project is based on the conventional SfM workflow, which typically involves feature detection, TP matching, camera pose estimation, and dense point cloud generation. However, the unique challenges associated with historical aerial imagery, such as limited image quality, incomplete metadata, and geometric distortions, combined with the objective of automation, requires specific adaptations to this standard pipeline.

Notably, the alignment of cameras and subsequent product generation are performed in two stages. The first stage involves a relative reconstruction to produce intermediate outputs required for the generation of GCPs. In the second stage, these GCPs are integrated to obtain an absolute, geo-referenced model. Several steps in the workflow have been adapted or newly introduced to address issues inherent in the

dataset. These customized components are marked with an asterisk in the workflow diagram (Figure 5.4) and are described in detail in the following sections.

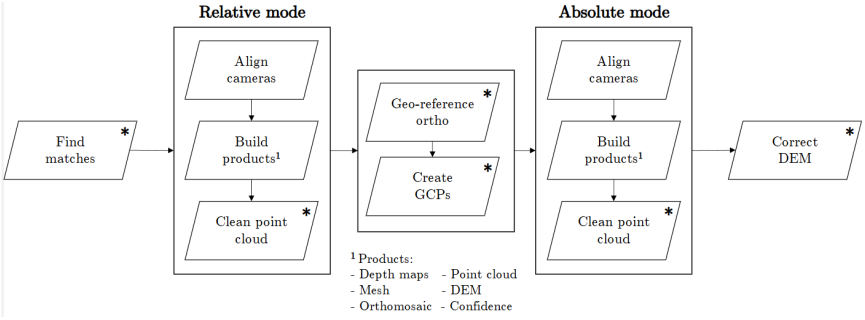


Figure 5.4: SfM workflow diagram highlighting key processing steps and automated adaptations specific to historical aerial imagery (adapted steps marked with an asterisk).

5

Match photos

While the default matching algorithm of Agisoft is effective in many standard scenarios, it struggles with historical aerial imagery. This is largely due to its reliance on traditional feature detection methods such as SIFT, which are not always robust to the challenges posed by degraded image quality, low contrast, and geometric inconsistencies common in archival data. As a result, the number and quality of TPs can be reduced, potentially leading to poor camera alignment and lower-quality reconstructions. To overcome these limitations, we developed a custom matching methodology and integrated the resulting matches into the Agisoft workflow.

Our approach begins with pairing images based on one of three specific criteria: exhaustive pairing, pairing by image IDs, or pairing guided by overlap data when available. For the matching process itself, we adapt a methodology developed in Dahle, Lindenbergh, and Wouters (2024a), originally designed for geo-referencing historical images. This method is based on the *LightGlue* deep neural network TP matching algorithm of Lindenberger, Sarlin, and Pollefeys (2023) and has a significantly higher match density compared to the default approach of Agisoft. As illustrated in Figure 5.5, our method enhances the quality and detail of matches, providing a robust foundation for subsequent analyses while minimizing the occurrence of invalid matches.

Since Agisoft does not natively support the direct incorporation of external matches, we introduced an intermediate step to integrate our

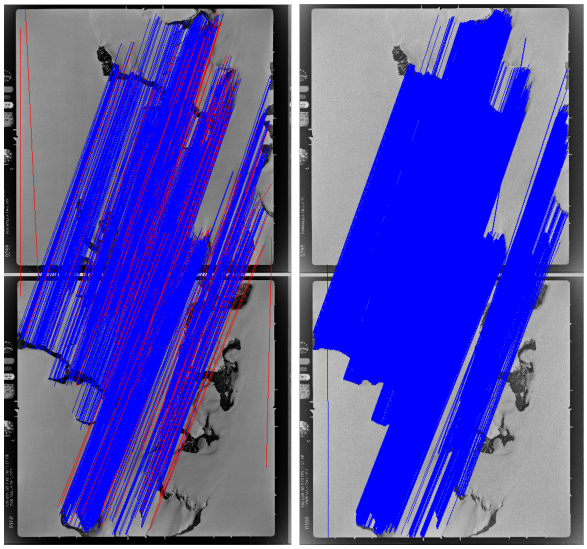


Figure 5.5: Comparison of image matching results: Agisoft's default matching algorithm (left; medium quality, 1248 valid and 136 invalid TPs) versus our custom matching method (right; 6829 valid TPs and no invalid TPs), demonstrated on two example images (CA215132V0299 and CA215132V0300).

custom matches. The matched TPs are first converted into a Bundler-format file (Snavely, Seitz, and Szeliski 2006), a format originally developed for scenes with already pre-aligned cameras. Within this file, default values for camera positions and orientations are assigned, enabling seamless import into Agisoft. Once the Bundler file is imported, the camera alignment is reset and re-performed, ensuring that the externally derived TPs are fully utilized in the alignment process. This integration not only enhances the alignment accuracy but also provides a scalable solution for processing large datasets with minimal manual intervention.

Clean Point cloud

The point cloud serves as primary data source for generating the DEM. However, due to the quality of the input data and imperfections in the reconstruction process, the raw point cloud may contain artifacts such as "floating islands" or points located significantly below the true surface. These anomalies can distort the resulting DEM and compromise therefore the accuracy of subsequent processing steps.

To mitigate this issue, we apply a Principal Component Analysis (PCA)-based filtering method to the point cloud. PCA is a widely used technique

for identifying the main orientation of local point clusters and filtering outliers that deviate significantly from this dominant surface direction In our implementation, all three principal components are retained, and the third component (corresponding to the smallest eigenvalue) is interpreted as the local surface normal. By computing the distance of each point to the fitted local plane and removing those that exceed a threshold of 3 standard deviations from the mean plane distance, this method effectively removes outliers while preserving the underlying surface structure. The result is a cleaner, more consistent point cloud that forms a robust basis for generating accurate elevation models.

Geo-reference Orthophoto

As a prerequisite for the automated identification of GCPs, the relative orthophoto must first be geo-referenced. Our geo-referencing approach uses the methodology presented in Dahle, Lindenbergh, and Wouters (2024a), which involves matching TPs between historical imagery and modern, geo-referenced datasets. Although the individual input images have already been roughly geo-referenced based on the digitised flight charts, we re-apply geo-referencing to the orthophoto mosaic.

We observed that this additional geo-referencing step significantly enhances the spatial accuracy of the final alignment. By leveraging the spatial continuity and contextual information inherent in the orthophoto mosaic, the matching process becomes more robust and yields more reliable results than matching based solely on individual frames. This is particularly beneficial in complex or repetitive terrain, where isolated features may be ambiguous in single images but more clearly defined in the stitched mosaic.

Since the relative orthophoto may be arbitrarily rotated due to image alignment processes and our matching is not rotation-invariant, we perform TP-matching against the modern geo-referenced image by generating rotated versions of the orthophoto (in 5-degree increments) over a full 360-degree range. This increases the likelihood of successful TP detection regardless of the initial orientation.

Create ground control points

GCPs are points with known absolute coordinates in x, y, and z, which are also identified in the relative 3D model. They play a crucial role in positioning the model within an absolute reference frame and improving its spatial accuracy. Traditionally, GCPs are placed manually by identifying distinctive features in both the relative model and a geo-referenced dataset, and then manually placing corresponding markers. However, to enable full automation of the SfM workflow,

we implemented an automated GCP generation method based on TP matching with modern geospatial datasets.

In a preceding step, the relative orthophoto was geo-referenced. This serves as the basis for further TP matching. At this stage, a second round of matching is conducted between the already geo-referenced orthophoto and the Sentinel-2 composite. Unlike the initial coarse matching, this step is performed at the full resolution of the orthophoto to capture finer details. As the spatial extent is too large for global matching in a single step, both the orthophoto and the satellite composite are divided into tiles. Since the historic image has already been geo-referenced, each tile covers the same scenery in both datasets, allowing localized and accurate matching.

Within each tile, TPs are identified. Due to the increased resolution, more TPs are generally detected compared to the initial geo-referencing step. These provide the x and y components of the GCPs. To obtain the elevation (z) component, the absolute height is extracted from the REMA at the TP's location. The relative elevation is retrieved from the historic DEM, which was generated using the same bounds and pixel size as the orthophoto. As a result, the transformation applied to the orthophoto can be directly transferred to the DEM.

To further improve the quality of the GCPs, TPs are filtered using a composite mask. This mask restricts selection to (a) rock outcrops as delineated in the Quantarctica dataset, (b) TPs with a confidence score above a defined threshold, and (c) areas with moderate slope. The slope criterion is used to avoid steep terrain, where small horizontal errors can lead to large vertical discrepancies. To prioritize quality, an initial maximum slope of 20° is used, as these areas generally yield more accurate and stable GCPs. However, in cases where this strict threshold results in too few candidates (<10), the slope limit is incrementally increased in 5° steps, up to a maximum of 60°, to ensure sufficient spatial coverage even in challenging terrain. These threshold values were determined through manual testing as a suitable compromise; sufficiently restrictive to minimize vertical error while still allowing for adequate GCP selection.

To ensure a uniform spatial distribution of GCPs and avoid redundancy, a ranking system based on TP confidence is applied. The highest-confidence TPs are selected first, and a circular buffer of 150 meters is applied around each selected point to exclude neighbouring candidates. This iterative selection strategy prevents spatial clustering and ensures that each GCP contributes unique, independent information to the model.

After identifying and filtering the GCPs, they are incorporated into the photogrammetric project, and the image alignment is recomputed. The pixel and meter residuals of each GCP are evaluated, and GCPs with errors exceeding predefined thresholds are iteratively removed. This refinement step is repeated until all remaining GCPs fall within acceptable error margins. While this automated geo-referencing approach may not reach the precision of manually surveyed GCPs, it offers the major advantage of full automation, making it well-suited for large-scale or archive-based studies. Its success depends on the availability of a modern reference DEM and the presence of stable surface features, such as exposed rock outcrops, within the study area.

DEM co-registration

To enable meaningful comparison between the modern and historical DEMs, we perform a co-registration procedure. For this, we use the framework introduced in xDEM contributors (2024), following an approach similar to that of North and Barrows (2024), which is based on standard techniques commonly used in large-scale glacier mass balance assessment. However, we adapted this framework to operate automatically and to better accommodate the challenging terrain in our study area, as outlined below.

The co-registration is performed iteratively to ensure accurate alignment between historical and modern DEMs. In the first step, systematic elevation biases in the historical DEM are corrected through polynomial surface fitting (de-ramping), which removes large-scale planar artifacts such as tilts or vertical trends that commonly arise from sensor distortions or processing errors. Subsequently, horizontal and vertical misalignment is addressed using a feature-based co-registration method following the approach of Nuth and Kääb (2011), with stable terrain defined by exposed rock outcrops. To minimize elevation differences over stable terrain, the method examines how the differences depend on slope and aspect. This relationship is then used to estimate and correct the horizontal offset between the DEMs. If the initial co-registration attempt fails due to an insufficient number of stable pixels (usually at least 100-200 pixels are required as an absolute minimum), the slope threshold used to define stable ground is incrementally relaxed from 20° in 5° steps. Similar to the previous step relaxation, this dynamic adjustment increases the number of valid reference points and improves the likelihood of successful convergence, even under challenging conditions. In some cases, co-registration remains unsuccessful, when stable features are spatially clustered, leading to localized alignment that causes distortions elsewhere in the image. In this case a fallback approach is applied. This backup method involves subtracting the

median elevation difference computed solely from the stable areas, thereby achieving a first-order correction without introducing large-scale distortions.

5.2.4. EVALUATION

The primary output products of this study are the DEM and the orthophoto. Evaluating the accuracy and quality of these products involves a combination of qualitative and quantitative approaches. One method involves measuring the residual pixel and meter errors associated with the GCPs. While this provides an indication of alignment accuracy, it is important to consider that GCPs inherently carry positional uncertainties, which may influence the results.

In addition to residual error analysis, a manual inspection is conducted by visually comparing the orthophoto to modern geospatial datasets. This evaluation focuses on assessing the spatial consistency of key structural features, such as the position of rock outcrops and identifying major misalignments and systematic distortions.

Quantitative evaluation of the DEMs is also performed by comparing them to modern datasets using statistical metrics. For each glacier, both across the full extent and restricted to stable areas, the mean and standard deviation of absolute and relative elevation differences are calculated, along with additional metrics such as Root Mean Square Difference (RMSD) and the Median Absolute Deviation (MAD). Since the modern datasets are not true ground truth, these values reflect elevation differences rather than absolute errors. Nonetheless, these numerical comparisons provide a detailed assessment of the relative accuracy and consistency of the derived elevation models.

However, two significant factors complicate the evaluation process. First, many areas, particularly glacier-covered regions, have undergone substantial changes over time, such as glacier retreat and thinning due to climate change. These changes introduce large discrepancies between historical and modern datasets, which do not reflect errors in the methodology but rather real-world environmental changes. Second, the inherent nature of SfM processes affects the quality of the outputs. Areas with abundant TPs and GCPs, such as stable rock outcrops, yield higher accuracy. In contrast, glacier-covered regions, which often lack sufficient TPs, rely heavily on interpolation, introducing additional uncertainty.

To address these challenges, the evaluation is divided in three scenarios. The first considers the entire study area, the second focuses exclusively

on the regions covered by the mask used for filtering, and the third evaluates only the areas outside the mask. This approach allows for a refined assessment of accuracy across different regions. The highest accuracy is expected within the masked areas, where the density of TPs and GCPs is greatest, representing the upper limit of the workflow's performance. In contrast, glacier-covered regions, where interpolation is more prominent, are expected to exhibit lower accuracy.

5.3. RESULTS

The workflow is applied to all glaciers located on the Antarctic Peninsula, as defined by the GLIMS glacier database (GLIMS and NSIDC 2005). Glacier outlines from this dataset are used to define the spatial boundaries for processing. Images were selected for each glacier based on their overlap with the outlines, specifically if at least 20% of an image footprint intersected a glacier or if the image covered at least 30% of the glacier area, thereby ensuring adequate coverage even for smaller glaciers. Both vertical and oblique images were considered, and image footprints were refined using updated positions from the geo-referencing step rather than relying on original metadata. A minimum of three overlapping images was required to initiate the SfM workflow.

Out of the 296 glaciers in the dataset, 143 (48.3%) were excluded due to insufficient image overlap. An additional 104 (35.1%) glaciers failed during the workflow, predominantly in the geo-referencing and GCP generation steps. Notably, in 90 of these cases, relative models and DEMs were still successfully created, and only 14 (4.7%) glaciers failed entirely. Ultimately, 49 (15.5%) glacier reconstructions were completed with full automation. An overview for the spatial coverage and performance of the glaciers can be seen in Figure 5.6. The processing time for each glacier differs, depending mostly on the number of input images but also on the complexity of the images. The maximum processing time was observed for Crane Glacier, which required approximately 7 hours on a standard desktop computer[2], whereas most glaciers were processed in about 1 hour on average.

Each successfully completed project produced a dense point cloud, an orthophoto, a confidence array, and a DEM at 2-meter resolution. Elevation values not supported by the point cloud were interpolated using Agisoft Metashape. A confidence array was created by averaging Metashape's per-point confidence values per cell, indicating the local reliability of the elevation estimates. To evaluate reconstruction

[2]Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz, 64 GB RAM, Quadro RTX 5000 16GB GPU

accuracy, historical DEMs were compared with the REMA using both qualitative and quantitative approaches.

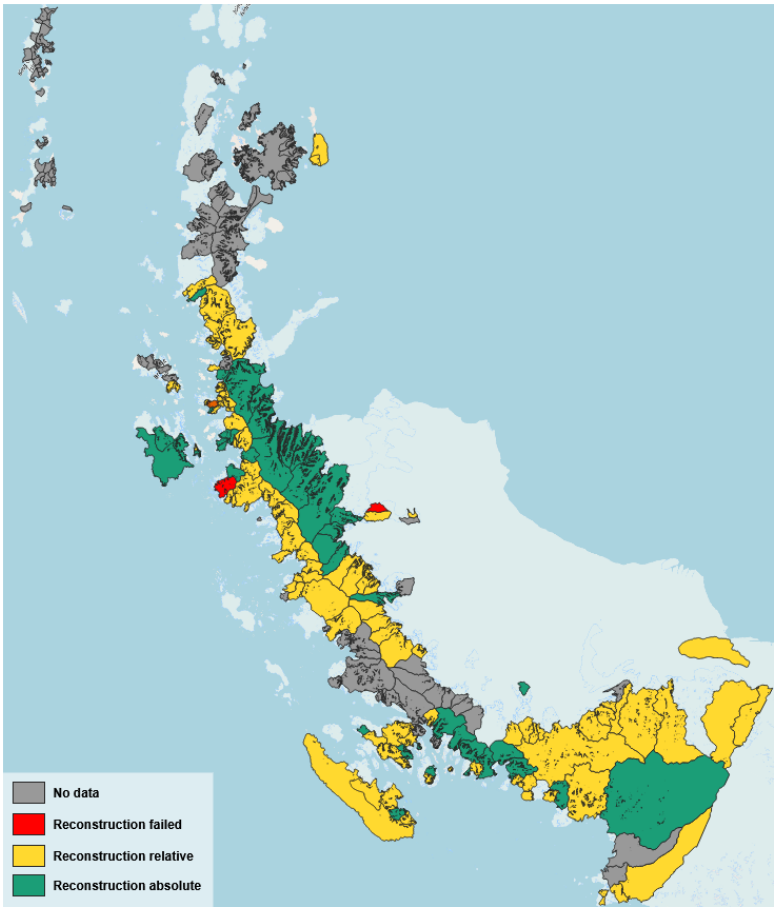


Figure 5.6: Spatial overview of reconstruction outcomes for named glaciers on the Antarctic Peninsula, based on the GLIMS dataset. Gray indicates glaciers without TMA coverage, red indicates failed reconstructions, yellow denotes successful relative reconstructions without absolute referencing, and green marks glaciers with completed absolute reconstructions.

Quantitative evaluation was conducted for the 49 completed reconstructions using statistical comparison between historical DEMs and REMA. For each glacier, the median difference, RMSD, and MAD were calculated across two regions: the entire DEM and only rock outcrop areas. Table 5.1 summarizes these values (visual examples for four glaciers, two with good and two with poor results, can be found in the Annex at Figure 1). The results show substantial variation in elevation

differences, with higher discrepancies typically observed when including glacierized areas ('all') due to interpolation and limited TP coverage there. In contrast, rock outcrop areas ('masked') yielded more consistent and accurate results, though some outliers persisted due to limitations in GCP accuracy and model drift.

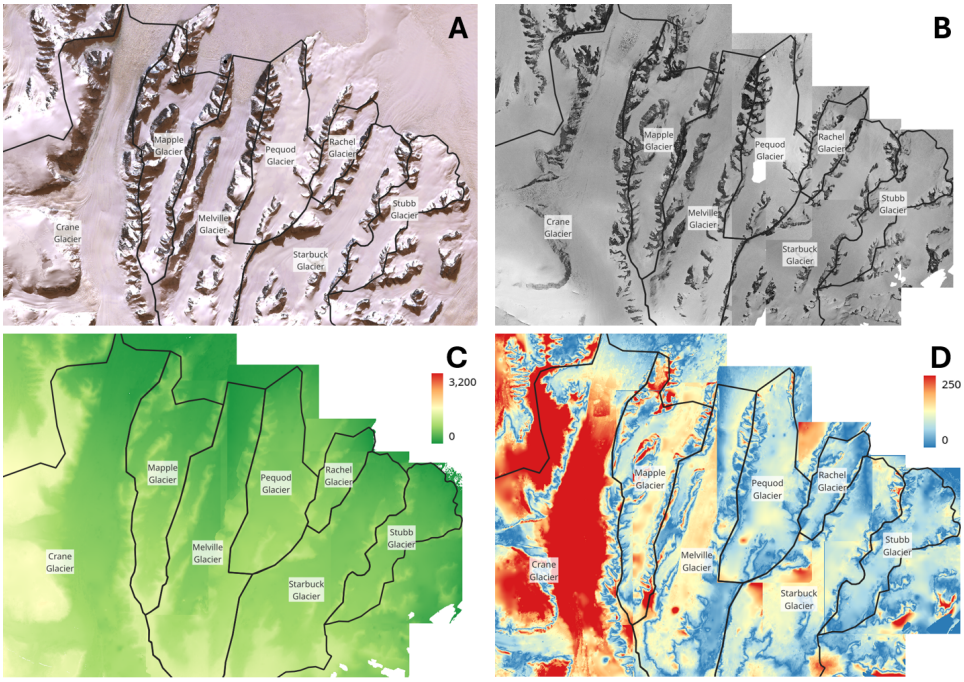


Figure 5.7: Example results showing modern Sentinel-2 satellite imagery (A) with derived historical orthophotos (B), derived historical DEMs (C) and the difference between the historical DEMs and REMA elevation data (D) in a composite encompassing reconstructions of Crane, Melville, Pequod, Rachel, Starbuck, and Stubb glaciers.

Of the 49 successfully reconstructed glaciers, two were excluded from the subsequent quantitative assessment. Although their geo-referencing was successful and their spatial positioning was broadly correct, the distribution of GCPs was suboptimal, resulting in high marker errors and significant elevation differences compared to modern datasets, even after co-registration. The remaining outputs, assessed through visual comparison with contemporary reference data, showed generally good spatial agreement. As illustrated in Figure 5.7, representative results for selected glaciers exhibit generally a good orthophoto alignment and elevation structure, although local inaccuracies are apparent near the borders of the glaciers. This is expected due to the lower number of tie points and ground control points available for Structure-from-Motion in

Project	Area (km²)	nr img	marker err(px)	marker err(m)	mdn diff (all)	mdn diff (msk)	RMSD (all)	RMSD (msk)	MAD (all)	MAD (msk)
Aagaard Glacier	232	28	0.23	23	53	52	112	111	50	51
Archer Glacier	255	20	1.96	3163	536	715	1493	806	502	472
Attlee Glacier	909	37	0.18	57	72	78	800	2263	68	76
Bevin Glacier	910	29	0.19	13	58	62	119	100	53	62
Bolton Glacier	111	13	0.29	58	431	220	3849	1098	415	227
Centurian Glacier	14	6	0.43	32	102	32	1217	1135	102	37
Clarke Glacier	523	28	0.33	33	332	181	6117	4890	278	169
Crane Glacier	2954	207	0.52	48	139	95	268	159	139	95
Eden Glacier	542	42	0.25	33	61	86	1409	126	60	74
Evans Glacier	653	42	0.34	59	130	69	2785	1243	102	69
Flask Glacier	2592	173	0.63	24	90	73	495	115	92	71
Forbes Glacier	885	25	0.22	40	84	101	145	154	74	97
Getman Ice Piedmont	143	10	0.56	27	26	48	986	66	26	48
Hektoria Glacier	2502	156	0.71	48	193	96	253	163	150	96
Hooper Glacier	58	7	0.43	108	41	52	730	84	41	41
Hurley Glacier	238	13	0.14	20	57	50	2194	1715	53	50
Iliad Glacier	507	22	0.18	16	149	131	1319	227	139	129
Jones Is	34	4	0.24	29	110	96	4537	1599	108	84
Jorum Glacier	1022	121	0.51	35	125	75	179	123	123	75
Leonardo Glacier	0	17	0.19	15	56	50	1421	1572	55	50
Leppard Glacier	3694	199	0.71	41	85	89	1433	456	85	83
Mamut Glacier	12	6	0.44	19	32	32	60	44	28	29
Mapple Glacier	320	49	0.13	37	91	64	106	97	53	64
Mcmorrin Glacier	65	10	0.72	189	74	76	910	229	64	71
Melville Glacier	629	76	0.37	12	99	60	112	99	68	59
Miethe Glacier	93	9	0.29	21	67	150	613	230	65	128
Mitterling Glacier	242	9	0.25	59	55	76	516	148	55	76
Montiel Glacier	2996	58	0.90	75	173	152	1117	218	132	152
Morrison Glacier	436	42	0.63	118	97	82	1501	1724	93	70
Nemo Glacier	114	6	0.03	10	58	63	1094	852	45	63
Neny Glacier	741	41	0.30	46	96	90	633	205	92	84
Niepce Glacier	328	17	0.42	33	167	172	364	381	170	162
Northeast Glacier	553	24	0.26	24	131	139	4135	3677	102	98
Orel Ice Fringe	40	6	0.82	86	86	70	3767	2021	81	74
Pequod Glacier	296	55	0.26	16	73	44	90	69	31	43
Perutz Glacier	719	19	0.60	28	129	158	234	216	124	93
Petzval Glacier	110	14	0.51	88	330	142	5056	3341	302	110
Rachel Glacier	127	14	0.19	11	63	50	1183	73	37	48
Reid Glacier	240	16	0.14	27	77	63	879	92	65	61
Seller Glacier	9923	212	0.42	12	91	93	4220	3252	105	87
Sikorsky Glacier	214	7	0.11	29	202	189	3046	1742	126	74
Snowshoe Glacier	629	54	0.59	41	54	81	1667	576	54	70
Somigliana Glacier	113	5	0.38	49	341	282	337	287	29	47
Starbuck Glacier	880	76	0.20	31	78	50	102	83	57	49
Stubb Glacier	324	39	0.60	25	71	49	1254	88	30	50
Swithinbank Glacier	617	34	0.27	16	96	70	723	129	53	64
Thunder Glacier	51	7	0.57	37	82	48	2183	2409	73	47
Vogel Glacier	96	4	2.30	621	730	485	6744	6667	971	489
William Glacier	187	10	0.21	32	217	84	2556	552	179	88
mean	841	45	0.38	41	118	93	1464	856	96	80
median	324	24	0.33	32	90	76	986	227	73	71
std	1601	56	0.21	33	87	52	1516	1143	75	38

Table 5.1: Summary of evaluation metrics for the 49 successfully reconstructed glaciers. Glaciers highlighted in gray were excluded from the final quantitative assessment due to high marker errors and are not included in the aggregated statistics (mean, median, and st. dev.) presented at the bottom of the table. Key parameters include glacier area, number of input images, marker error (in pixels and meters), and elevation difference metrics (median difference, RMSD, and MAD) computed across the full DEM and over stable, masked regions. glaciers names in bold are illustrated in Figure 5.7 as representative examples.

those areas

5.4. DISCUSSION

This study demonstrates that a fully automated SfM pipeline is capable of reconstructing historical glacier surfaces using mid-20th century aerial photographs from the TMA archive. For nearly all glaciers with sufficient imagery, the workflow was able to generate relative models. However, the transition from relative to absolute reconstructions posed the greatest challenge, particularly in the geo-referencing and GCP generation stages.

Of the 90 projects that failed to produce absolute DEMs, multiple factors were responsible. In approximately 60% of these cases, the relative orthophoto could not be geo-referenced. This typically resulted from insufficient TP matches due to poor image quality, low image overlap, or a lack of stable terrain features. In some cases, the scenery had changed significantly, such as glacier termini now bordering open water, making geo-referencing difficult. In roughly 25% of the failures, the orthophoto was successfully geo-referenced, but no suitable GCPs could be found, largely due to the absence of distinguishable, stable surface features. The remaining failures were caused by technical issues such as unreadable image files or missing metadata.

Manual geo-referencing could likely resolve many of these cases, especially where only a small number of TPs or GCPs were missing. However, our goal was to assess the potential of a fully automated workflow; thus, we deliberately avoided any manual intervention.

Despite these limitations, the pipeline successfully produced absolute DEMs for 49 glaciers without any manual input. Visual inspection confirmed that for all but two reconstructed glaciers the orthophotos aligned well with modern imagery, and DEMs displayed realistic topographic structures. Minor shifts in image alignment or elevation were observed in a few cases, likely due to insufficient GCP coverage or drift in steep terrain.

Quantitative comparisons with the REMA reveal that elevation differences are generally larger in glacierized areas than in stable, rocky terrain. This pattern is clearly visible in the six glaciers shown in Figure 5.7, where the smallest discrepancies between historical and modern DEMs are concentrated near rock outcrops. This observation is supported by the numerical values in Table 5.1, which show lower differences, RMSD, and MAD for the masked areas corresponding to these stable

zones. However, these discrepancies are expected and stem from two main factors. First, many glaciers have undergone significant surface lowering over the 60-year period between the historical and modern data, primarily due to climate-driven retreat and thinning (For example, Fleming Glacier, located on the AP as well, experienced a surface lowering of approximately 60 meters between 1966 and 2008 (Zhao *et al.* 2017)). This is evident when comparing elevation differences across all pixels versus only in masked, stable areas in which the elevation differences are smaller. Second, some residual inaccuracies arise from the 10-meter resolution of Sentinel-2 imagery used for GCP matching and from the inherent limitations of the rock mask and automated GCP placement. In rugged areas, even small horizontal errors can lead to substantial vertical deviations. In addition, high RMSD values observed for some glaciers can be attributed to the co-registration step. As discussed in Section 5.2.3, distortions may occur when the number of valid pixels used for alignment is low and spatially clustered, often leading to edge effects at the DEM borders. While these localized errors can elevate statistical error metrics, they are visually prominent and easily identifiable, and they do not compromise the quality of the remaining DEM.

These technical limitations are also reflected in comparisons with other recent studies. For example, when comparing our DEMs to those of North and Barrows (2024), which used the same TMA imagery but a partially manual workflow, we found a RMSD of approximately 135 meters and a MAD of 84 meters. The median elevation difference between the DEMs was -25 meters. Despite these elevation differences, a visual inspection as seen in Figure 5.8 revealed similar orthophoto alignment and structural features, suggesting that both reconstructions capture the glacier geometry and flow characteristics. The observed discrepancies are due to a combination of factors: differences in the underlying datasets used for validation (REMA vs ASTER), a substantially lower number of GCPs (five versus twenty), and the inherently lower precision of automatically generated GCPs compared to manually placed ones.

Crucially, however, our pipeline achieves these results in a fully automated manner, without any manual geo-referencing or GCP placement, demonstrating its potential for large-scale, archive-based glacier reconstructions with minimal human effort. In traditional workflows, geo-referencing and manual GCP placement are among the most time-consuming steps. Additionally, depending on the image quality and content, manually masking irrelevant image areas (e.g., clouds, borders, or fiducial frames) can further slow down processing. By contrast, our approach requires no user interaction once launched, allowing the entire workflow to run in the background without

supervision. If higher precision is required, manual refinement, such as adding high-precision GCP remains possible.

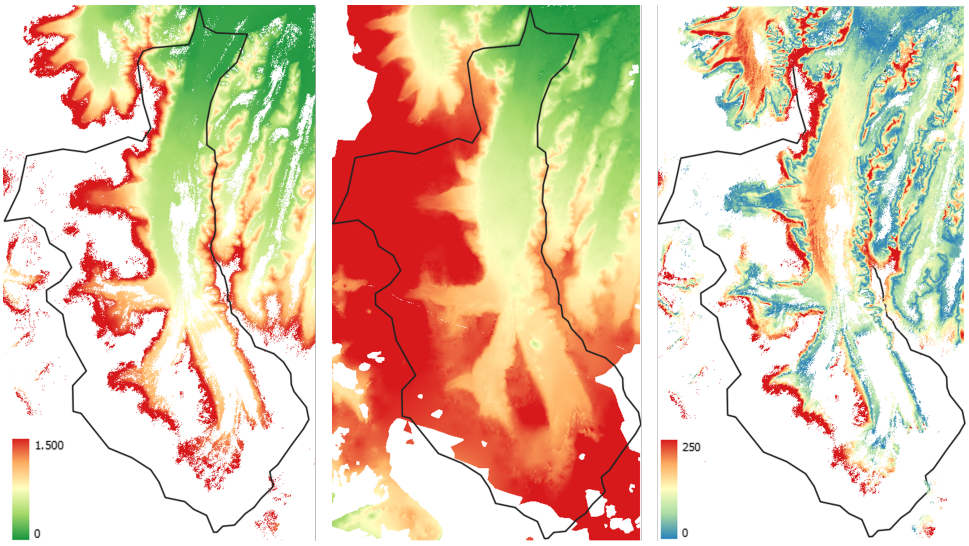


Figure 5.8: Difference (right) between the manually-assisted DEM by North and Barrows (2024) (left) and our automated DEM reconstruction (middle) for Crane Glacier in December 1968, demonstrating prevailing differences, but also showing the similarities in glacier geometry and structural features.

The workflow is adaptable and can be transferred to other datasets beyond the TMA archive. While it is optimized for digitised aerial photographs, it can also be applied to more recent platforms such as satellite or drone imagery, provided suitable reference data are available. If precise camera metadata exist, the need for automated GCP generation may be reduced or even eliminated, simplifying the reconstruction process. Furthermore, the integration of segmentation-based image masks helps exclude problematic regions such as open water or heavily shadowed areas, improving both the accuracy and robustness of the results.

In summary, this study marks an important step toward scaling up historical photogrammetry through automation. By demonstrating that archival aerial imagery can be processed with minimal manual labour, we open the door to large-scale reconstructions in other glacierized regions, and ultimately to improved understanding of long-term cryospheric change.

5.4.1. Scientific applications

The ability to automatically reconstruct historical glacier surfaces at scale opens new possibilities for glaciological research, particularly in remote and data-sparse regions such as the AP. By transforming large volumes of archival imagery into geo-referenced orthophotos and DEMs, this workflow enables a wide range of scientific applications.

One of the most immediate benefits is the ability to quantify elevation change over multi-decadal timescales. With reconstructions now available for a considerably larger number of glaciers, the long-term response of the AP to climate change can be assessed with improved spatial detail and coverage. However, these results must be interpreted in the context of their associated uncertainties. In some areas, elevation differences carry standard deviations of several tens of meters, which limits the detection of subtle elevation changes but still allows robust identification of major trends, especially when aggregated over larger spatial or temporal scales. These reconstructions also provide valuable input for validating climate and ice flow models, many of which are limited by the scarcity of pre-satellite observations.

From the generated elevation products, various glaciological indicators can be derived. Surface elevation change can be directly used to estimate glacier mass balance, assuming reasonable density estimates. Although this workflow does not explicitly identify calving fronts, previous studies have demonstrated that front-line positions can be extracted from the DEMs using automated techniques (Dong *et al.* 2022; Vacek *et al.* 2024). The availability of geo-referenced imagery from approximately 60 years ago offers a valuable reference point for assessing terminus change. While earlier studies, such as those by A. J. Cook, Holland, *et al.* (2016) and A. J. Cook and Vaughan (2010), have used archival data to document ice shelf and glacier front retreat across the Antarctic Peninsula, our workflow enables the extraction of such changes from fully geo-referenced orthophotos and elevation models at higher spatial resolution and with minimal manual effort.

More broadly, the flexibility of the workflow represents a significant advantage. Although it is currently optimized for historical imagery, the same methodology can be applied to modern datasets, including drone or satellite-based imagery. This generalization makes it possible to process heterogeneous datasets within a consistent framework, facilitating the integration of historical and contemporary observations. As many historical aerial archives remain largely untapped due to the complexity of processing, the presented automated pipeline provides a practical solution for unlocking the scientific potential of these datasets and extending our understanding of long-term glacier change.

5.5. CONCLUSION

In this study, we developed and evaluated a fully automated SfM pipeline to reconstruct historical glacier surfaces from mid-20th century aerial imagery from the AP using the TMA archive. The workflow was successfully applied to 47 glaciers, providing absolute DEMs without manual intervention. Quantitative evaluation against modern datasets indicated median elevation differences of approximately 90 m across entire glaciers and 76 m in stable rocky areas, demonstrating the method's practical viability for large-scale reconstruction tasks.

However, the primary limitation to complete automation was observed during the geo-referencing and automated GCP generation steps. Small inaccuracies in automatically derived GCPs often propagated into larger spatial errors, especially in steep and featureless glacierized terrain. Additionally, challenges related to poor image quality, incomplete metadata, and significant terrain changes since the original acquisitions frequently complicated the automated workflow, causing approximately one-third of glaciers with adequate imagery to fail at the geo-referencing stage.

To overcome the practical limitations, future research should focus on using higher-resolution reference imagery to significantly enhance the positional accuracy of automatically generated GCPs. Additionally, developing more robust segmentation techniques tailored explicitly for historical aerial imagery could reduce errors during TP matching, particularly in problematic regions such as glacier termini, open water interfaces, or shadowed areas. Further improvements could be achieved by creating adaptive strategies for the SfM workflow, allowing dynamic parameter adjustments that account for variations in image quality, terrain complexity, and data availability on a per-glacier basis. Finally, integrating methods for systematic uncertainty quantification would provide valuable insights into error propagation, allowing targeted refinements within the pipeline and supporting more confident interpretations of glacier change analyses derived from historical datasets.

This study significantly advances the scalability and practicality of automated historical SfM, demonstrating the potential to efficiently convert archival imagery into scientifically valuable data for glacier change analysis. By explicitly addressing the identified limitations, the presented workflow can become a powerful, broadly applicable tool for reconstructing long-term cryospheric dynamics at large scales.

Beyond these refinements of the methodology, future plans include extending this automated workflow to additional regions and datasets,

such as Svalbard and other parts of the TMA archive outside the AP. Given that the workflow requires minimal manual intervention after initial setup, its application to new datasets is straightforward and primarily limited by data availability.

6

CONCLUSIONS & RECOMMENDATIONS

This dissertation presents a fully automated approach to processing historical aerial imagery with the aim of addressing the following research question.

How can historical aerial imagery of the Antarctic Peninsula be processed in an automated manner using Structure-from-Motion to generate geospatial datasets suitable for long-term glacier change analysis?

To address this question, I developed an end-to-end workflow for processing historical aerial imagery of the Antarctic Peninsula (AP) using Structure-From-Motion (SfM). While the underlying principles follow established methods, historical imagery introduces a range of challenges that must be addressed, such as missing metadata, degraded image quality, and acquisition not originally intended for SfM. To handle these challenges, the workflow is divided into four dedicated components, each targeting a specific aspect, as detailed in the respective chapters. Although full automation is not feasible for every flight path, particularly those with extremely poor image quality or insufficient metadata, the approach proves effective under the right conditions and for many glaciers of the AP. Therefore, the workflow enables the large-scale generation of geospatial datasets suitable for long-term glacier change analysis with minimal manual input.

6.1. CONCLUSIONS

More specifically, the following conclusions were found answering the different sub-questions:

(Q1) *How effectively can semantic segmentation be applied to historical aerial imagery of the cryosphere?* **(Chapter 2)**

Semantic segmentation is needed to identify and mask regions unsuitable for 3D reconstruction. A custom U-Net model was trained on manually labelled images from the archive to classify each pixel into six semantic classes (ice, water, clouds, snow, rocks and sky). Training was done with an extremely limited training dataset of 80 images. These results show semantic segmentation can be effectively applied to historical aerial imagery of the cryosphere, though some caveats must be acknowledged. In our current implementation, the segmentation model reaches an overall accuracy of 73% and an F1-score of 71%. The performance, however, varies significantly between classes. Distinct classes such as rock or water perform particularly well, while others like sky perform moderately. Classes with high visual ambiguity, most notably snow and sea ice, remain challenging due to their semantic similarity and the limited visual cues in the historical imagery. In some cases, even human annotators struggle to distinguish between them.

Obvious cloud formations can also be detected, though this depends on their contrast and visibility. Post-processing the segmented images is essential to improve segmentation quality, especially for oblique imagery, where the sky class in particular benefits from this refinement (see Section 2.3.5). Despite some limitations, the segmentation results are sufficient for our intended purposes. The model enables filtering based on semantic content, such as identifying images that contain ocean. It can also infer the image orientation—for instance, distinguishing between vertical and oblique imagery—even in the absence of metadata. In the context of SfM, segmentation also facilitates pre-processing by masking out large areas unsuitable for Tie Point (TP) matching, such as ocean and sky, while highlighting relevant terrain features like rocks. While the segmentation is not reliable for identifying fine details, due to both the limited training data and inherent image quality, it still provides significant practical value and lays the groundwork for more targeted downstream processing. Overall, semantic segmentation proves to be a valuable pre-processing step for working with historical cryospheric imagery, even under constrained conditions.

(Q2) *How reliable can camera calibration metadata be extracted automatically from degraded historical aerial imagery?* **(Chapter 3)**

Accurate camera parameters are essential for reliable 3D reconstruction, yet they are often missing or difficult to access in historical imagery. To address this issue, a method was developed to extract key metadata such as focal length, altitude, and fiducial mark positions directly from the image borders. The approach combines deep learning, Optical Character Recognition (OCR), and classical computer vision techniques. The workflow was applied to hundreds of images and successfully retrieved metadata in most cases. When specific information was unavailable, estimates from neighbouring images within the same flight path were used to fill in the gaps.

Camera calibration metadata can be partially extracted from degraded historical aerial imagery, although the reliability of extraction varies depending on the type of metadata. Text recognition proves to be particularly effective when a tiling strategy is combined with the PP-OCR pipeline, resulting in a complete detection rate of text boxes in the test set. From these boxes, key parameters such as focal length can be accurately derived, provided that the text remains legible. Grouping images by flight line further improves the robustness of the extraction, as recurring information can reinforce correct values. However, altitude values are less frequently present in the text and are more challenging to recover automatically.

Fiducial marks can also be detected reliably in most cases. Exceptions occur primarily when images are strongly rotated or when the fiducial areas are partially degraded or cropped. In contrast, the extraction of altitude values from analogue altimeters remains particularly difficult. These areas often suffer from blur, scanning artifacts, or cropping, which limits the success of both edge detection and digit recognition. In many cases, even manual interpretation becomes infeasible. Despite these challenges, the current implementation demonstrates that valuable metadata can still be recovered from these often-overlooked regions, offering a promising direction for further refinement.

Estimation techniques play an important role in improving metadata reliability. They not only help compensate for missing values by inferring plausible alternatives, but also act as a validation step by identifying and filtering out incorrect detections. Although fully automated extraction of all metadata types is not yet possible, the approach already enables substantial progress in retrieving useful information from historical imagery. This significantly enhances the potential of automated reconstruction workflows.

(Q3) *To what extent is it possible to automatize geo-referencing of historical aerial imagery?* **(Chapter 4)**

Geo-referencing is essential to place reconstructed models into a modern coordinate system and to enable meaningful comparisons with contemporary datasets. To address this need, a fully automated method was developed using LightGlue for TP detection between historical aerial imagery and modern Sentinel-2 data. This approach generates Ground Control Points (GCPs) with positional accuracies within a few meters, representing a substantial improvement over earlier estimates, which were often off by several kilometres. Validation through visual overlays and numerical metrics confirms the effectiveness of the method. Nevertheless, applying geo-referencing in glacier interiors remains challenging, as these regions often lack stable and distinct reference features.

The success of this automated approach is largely enabled by advances in modern TP matching techniques. Unlike traditional methods that rely on pixel-level similarity, these techniques identify correspondences based on structural and textural features, making them more robust to differences in imaging conditions and sensor characteristics. This flexibility is particularly valuable when working with historical images, which often exhibit considerable variation in quality and appearance.

However, certain conditions must still be met for geo-referencing to succeed. A key requirement is the availability of a second, geo-referenced image source that overlaps with the area of interest. In addition, an approximate location must be known to guide the matching process. Reliable TP detection also depends on the presence of visually consistent and identifiable features, such as exposed rock outcrops. These features must remain sufficiently stable over time to serve as effective reference points. Automated workflows, in particular, require a high degree of visual consistency and clarity to perform reliably.

In situations where historical images follow a regular acquisition pattern, such as a linear flight path with consistent spacing and altitude, it becomes possible to geo-reference images that lack clear visual anchors. This is achieved by estimating their position based on the spatial relationships with neighbouring geo-referenced images. Such interpolation methods complement direct matching and expand the coverage of the automated workflow.

Applied to the AP, this combined approach enabled the geo-referencing of approximately 76 percent of the historical image set. About one quarter of the images were matched directly through automated TP detection, while the remainder were positioned through interpolation

from adjacent frames. Comparison with manually placed GCPs revealed an average positional deviation of 85 meters, which represents a major improvement over traditional methods, where errors of several kilometres were common.

Despite these advances, the accuracy of the method is inherently limited by the resolution and quality of the reference data. In this study, Sentinel-2 imagery was used due to its consistent coverage across the AP and its spatial resolution of 10 meters. While sub-pixel matching can help mitigate some of the resolution-related limitations, the reference dataset ultimately sets a ceiling for achievable precision.

Importantly, the presented workflow is not restricted to Sentinel-2 imagery. Any geo-referenced dataset with suitable spatial resolution and coverage can be used, allowing for further improvements in regions where higher-resolution data are available. Nonetheless, complex terrain, cast shadows, and uneven distribution of GCPs continue to present challenges. Even so, the automated method represents a significant advance over earlier geo-referencing efforts, which often relied on hand-drawn maps and were subject to substantial spatial uncertainty.

(Q4) *How feasible is the 3D reconstruction from historical aerial imagery using Structure-from-Motion?* **(Chapter 5)**

SfM forms the final step in generating 3D models of glacier surfaces from prepared historical imagery. The automated pipeline integrates segmentation masks, extracted metadata, and GCPs to produce dense point clouds and elevation models using the photogrammetric software Agisoft Metashape. Reconstructions were completed for 49 glaciers and evaluated by calculating elevation differences relative to the Reference Elevation Model of Antarctica (REMA) reference Digital Elevation Model (DEM). While REMA is not a true ground truth, it provides the best available large-scale baseline for comparison.

Median elevation differences were approximately 93 meters across all glacier surfaces and 76 meters within stable, rocky areas. These magnitudes exceed the level of uncertainty typically expected from both datasets, suggesting that the reconstructions capture substantial glacier surface change over the past several decades. Although individual changes on the order of a few tens of meters remain below the detectable threshold given current uncertainties, the results are nonetheless valuable for identifying large-scale elevation trends, spatial patterns of glacier change, and regions of pronounced surface lowering. Remaining limitations are primarily linked to snow-covered areas and

image sets with poor texture or limited overlap, where reconstruction quality tends to degrade.

This study demonstrates that automatic 3D reconstruction from historical aerial imagery is feasible, even under the challenging conditions typical of archival data. With a carefully adapted workflow, digitised images from the Trimetrogon Aerial (TMA) archive of the AP were successfully converted into digital elevation models with minimal manual input.

The developed pipeline was able to reconstruct relative 3D models for nearly all glaciers where image overlap was sufficient. For 49 out of 78 glaciers, absolute 3D products, including orthophotos and DEMs, were generated fully automatically. Key workflow components such as modern TP matching, automatic camera alignment, and iterative DEM co-registration proved essential for overcoming issues like degraded image quality, missing metadata, and substantial terrain change between acquisition dates.

Despite these promising results, certain limitations persist. Transitioning from relative to absolute models proved to be the most error-prone step. Around one third of glaciers with otherwise sufficient input data failed during geo-referencing or GCP placement. These failures often stemmed from low image contrast, a lack of stable reference features, or significant environmental changes since the time of acquisition. In such cases, reconstructions were affected by image drift or geometric distortions.

Quantitative validation confirmed median elevation differences with respect to REMA of roughly 93 meters across glacier surfaces and 76 meters in stable terrain, reflecting the cumulative impact of small inaccuracies throughout the processing chain. Where GCPs could be reliably detected or manually added, accuracy improved substantially. These results highlight the continued importance of expert input, as manual workflows still offer higher precision, particularly in areas with poor texture, heavy shadowing, or insufficient visual contrast.

Nevertheless, the strength of the automated workflow lies in its efficiency and consistency. The ability to process large volumes of data quickly enables coverage at a spatial scale that would be unmanageable through manual methods alone. While the current level of accuracy does not yet match that of expert-guided reconstructions, the workflow establishes a strong foundation for scalable and reproducible processing of historical imagery. With future enhancements in GCP detection and the integration of auxiliary datasets such as modern DEMs, the quality gap can be narrowed while preserving the advantages of automation. This positions the method as a valuable tool for reconstructing glacier

surfaces at scale and for advancing long-term change analyses based on historical aerial data.

While this study focused on establishing a robust and scalable workflow for processing historical imagery, the generated DEMs already enable the reconstruction of glacier-specific elevation change time series. This temporal analysis, however, requires additional steps such as inter-epoch co-registration, uncertainty propagation, and glacier-wide aggregation, and is therefore beyond the scope of this thesis. A dedicated follow-up study will address this aspect in more detail, building directly on the methods and datasets presented here.

6.2. MAIN CONTRIBUTIONS

This PhD thesis advances the integration of historical aerial imagery into modern cryospheric research by developing novel, largely automated workflows for data extraction, segmentation, geo-referencing and SfM. These contributions collectively reduce the dependency on manual processing and open up the TMA archive for broader scientific use. The contributions fall into two main categories: data-driven outputs and methodological innovations.

On the data side, this work generated new datasets for each of the four core workflow components: semantic segmentation, metadata extraction, geo-referencing, and 3D reconstruction. These include derived metadata, pixel-level semantic masks, geo-referenced image positions, and partial 3D models. All data products are hosted on a dedicated online platform[1], which provides accessible entry points to both the historical imagery and its derived data. In addition, all source code and processing pipelines are published on GitHub ([2]) to ensure transparency, reproducibility, and future reuse within the research community.

Methodologically, several novel contributions were made. First, a custom U-Net model was developed for semantic segmentation of historical Antarctic imagery, capable of delivering meaningful results even with a training set of just 100 labelled images. A systematic comparison of hyper-parameter configurations provided insights into which combinations work best for such limited and degraded datasets. The trained model was successfully applied to a large portion of the AP archive, and the resulting segmentation masks are now publicly available for use by other researchers.

[1] https://polar-archive.citg.tudelft.nl/
[2] https://github.com/fdahle/Antarctic_TMA

Second, in the area of metadata extraction, this thesis applies computer vision-based methods to support the automatic recovery of key camera parameters. This includes the use of OCR techniques to extract focal length values from printed text on degraded image borders, algorithms for detecting and measuring fiducial marks, and to our knowledge for the first time a workflow for estimating image altitude from analogue altimeter dials. This final approach represents an important first step toward recovering missing acquisition metadata, although its accuracy remains strongly dependent on image quality.

Third, a novel geo-referencing workflow was developed by matching TPs between historical and modern imagery using state-of-the-art feature matchers. This pipeline includes a custom adaptation of LightGlue, optimised for large-format and low-quality historical images by implementing a progressive tiling and matching strategy. This approach not only overcomes memory limitations but also improves robustness for degraded visuals. The geo-referencing pipeline is suitable for other historical datasets and other types of modern reference imagery.

Fourth, this thesis advances the application of SfM techniques to historical aerial imagery by adapting several key steps of the workflow to address challenges such as missing metadata and complex terrain. The approach supports both vertical and oblique photographs and is implemented in Python using Agisoft Metashape. The entire workflow has been automated to require minimal user interaction. While it was developed for the TMA archive, the workflow is designed to be flexible and can be readily adapted for other historical datasets and geographic regions.

Taken together, the developed workflows form a coherent system for the automated processing of historical aerial imagery. Each component was designed for interoperability, enabling them to function both independently and as part of a larger pipeline. This modular design ensures flexibility for different data conditions and research needs. The overarching contribution of this thesis is not a single algorithm or model, but a replicable system for unlocking the scientific value of historical aerial archives. The full system was demonstrated through a case study on the AP, and its modular components were designed with adaptability in mind, making them potentially transferable to other regions and applications.

6.3. RECOMMENDATIONS

This research identifies several important opportunities to enhance the automated processing of historical aerial imagery. A key limitation encountered throughout the workflow is the frequent absence and often irretrievable loss of essential auxiliary information such as camera calibration parameters and precise acquisition positions. Since these data are typically not recoverable from the imagery itself, future work should focus on the integration of supplementary datasets, such as early satellite imagery or modern remote sensing products, to fill spatial gaps and enhance the reliability of reconstructions. Stable terrain features, such as exposed rock outcrops, play a critical role by providing consistent reference points across time, which support accurate alignment and validation of image orientation. However, in glacierized regions where the surface has often changed dramatically historical imagery alone may be insufficient to reconstruct past conditions. Here, modern datasets can help infer plausible glacier extents or surface elevations based on recent morphology, aiding the interpolation of missing or ambiguous areas. This integration of stable reference points and modern proxies is particularly valuable in regions affected by cloud cover, cast shadows, or extensive glacier retreat.

In the context of semantic segmentation, machine learning-based approaches have demonstrated considerable potential, yet their effectiveness remains limited by the availability and diversity of training data. Expanding the training dataset to include a broader range of imagery, particularly examples that contain rare or visually ambiguous classes such as clouds or thin snow, is essential. Furthermore, a standardised classification protocol should be adopted to resolve common ambiguities. This includes for example decisions on the appropriate level of detail for class boundaries (e.g., whether to label isolated small rock outcrops or only larger contiguous areas), the introduction of a 'mixed' or 'ambiguous' class for uncertain regions, or how to treat clouds that are only partially visible or where the underlying terrain is still discernible. Finally, training data should ideally include annotations from multiple contributors to reduce individual biases and increase the generalisability of the model. The addition of auxiliary data sources may further enhance classification performance, provided that the images are geo-referenced to ensure spatial alignment. This is particularly beneficial for distinguishing between visually similar features such as rock and water. While historical DEMs would be ideal for this purpose, even modern elevation data can offer valuable contextual information.

In terms of metadata extraction, traditional computer vision techniques have enabled meaningful progress, particularly in detecting fiducial marks and printed camera parameters. However, these methods often

fail when confronted with severely degraded or cropped imagery. To improve robustness, future efforts should focus on developing dedicated deep learning models for metadata recognition. Such models could be trained on synthetic imagery derived from historical camera data sheets, allowing them to learn plausible parameter ranges and recognise distorted or partially obscured text. Estimation and redundancy across flight lines should remain an integral part of the workflow, both to fill in missing values and to validate automatically extracted information.

Geo-referencing presents one of the most significant challenges in historical image processing. The current approach relies heavily on identifying corresponding features in modern imagery, which becomes difficult when environmental changes have altered the visual landscape or almost no outstanding features are present. To address this, future research should incorporate semantic information into the matching process, enabling systems to recognise that two regions may correspond even if their visual characteristics differ. This includes the ability to distinguish between permanent and transient features, such as bedrock and snow cover. Ongoing developments in machine learning-based TP matching offer promising solutions, particularly those that integrate semantic cues or make use of high-level scene understanding (compare for example Pix4D (2024)).

Finally, while SfM has proven effective for reconstructing relative 3D models from historical imagery, transitioning to absolute models remains a bottleneck, much like the geo-referencing step itself. This is especially challenging on glacier surfaces, where reliable GCPs are often unavailable due to dynamic changes and a lack of stable reference features. One potential workaround is to use modern DEMs to derive approximate elevations for features visible in historical images, even on the glacier itself. These estimates can serve as pseudo-GCPs, temporary reference points that help to initially constrain and stabilize the model during geo-referencing. While their accuracy may be limited due to elevation changes over time, they can still guide the alignment process. Once a better fit is achieved or more reliable constraints are identified, these pseudo-GCPs can be refined or removed.

Overall, future research should prioritise the integration of semantic understanding, additional data sources, and learning-based methods to overcome the inherent limitations of historical imagery and to further automate and generalise the processing pipeline developed in this work.

6.4. PERSONAL STATEMENT

Pursuing this PhD has taught me as much about the research process as about the subject matter itself. One of the most important lessons I've learned is the value of adaptability. In the early stages of the project, I invested nearly a year into developing a workflow using MicMac. Despite the time and effort involved, it became clear that this approach would not deliver the required results. Changing direction was a difficult decision, but it reminded me of the importance of not holding on to past efforts when more suitable alternatives are available. Even after switching tools, it took longer than expected to get the automated workflow fully operational. While the basic structure came together quickly, each successful case revealed new exceptions. Developing a reliable pipeline meant identifying and resolving a wide range of edge cases. This experience showed me that progress in science often comes not through a single breakthrough, but through many small and incremental improvements. I also came to appreciate the collaborative nature of scientific research. My work benefited greatly from methodologies developed by others, such as the TP matching approach introduced by Lindenberger, Sarlin, and Pollefeys (2023). Integrating these advances into my own work underscored the importance of a strong, open research community where tools, data, and ideas are freely shared.

At the same time, I've gained a better understanding of the strengths and limitations of automation. While my workflow can now generate 3D models and orthophotos with minimal manual input, the results still don't reach the quality of expert-guided reconstructions. Yet the outputs are visually consistent and structurally accurate in many cases, showing that this approach is moving in the right direction. Automating the process makes it possible to handle large datasets efficiently, which is essential for studying long-term change across wide areas. Looking ahead, I believe that artificial intelligence will become increasingly important in geospatial research. It offers promising ways to fill in gaps in historical or incomplete datasets, though it is essential to ensure that the results remain grounded in real-world conditions. In my postdoctoral work, I plan to build on this foundation by incorporating higher-resolution data and more flexible algorithms that can better handle variability. My goal is to develop methods that are not only efficient, but also robust and interpretable, bridging the gap between automation and expert insight.

BIBLIOGRAPHY

Agisoft (June 21, 2025). *Agisoft Metashape software*. Version 2.2.0. URL: https://www.agisoft.com/downloads/installer/.

AliAkbarpour, H., Palaniappan, K., and Seetharaman, G. (2015). "Fast Structure from Motion for Sequential and Wide Area Motion Imagery". In: *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*. DOI: 10.1109/ICCVW.2015.142.

Allen, N., Edge, M., Appleyard, J., Jewitt, T., Horie, C., and Francis, D. (1987). "Degradation of historic cellulose triacetate cinematographic film: The vinegar syndrome". In: *Polymer Degradation and Stability* 19. DOI: 10.1016/0141-3910(87)90038-3.

Ayalew, Y. A., Fante, K. A., and Mohammed, M. A. (2021). "Modified U-Net for Liver Cancer Segmentation from Computed Tomography Images with a New Class Balancing Method". In: *BMC Biomedical Engineering* 3. DOI: 10.1186/s42490-021-00050-y.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39. DOI: 10.1109/TPAMI.2016.2644615.

Baumhoer, C. A., Dietz, A. J., Kneisel, C., and Kuenzer, C. (2019). "Automated Extraction of Antarctic Glacier and Ice Shelf Fronts from Sentinel-1 Imagery Using Deep Learning". In: *Remote Sensing* 11. DOI: 10.3390/rs11212529.

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). "SURF: speeded up robust features". In: *Proceedings of the 9th European Conference on Computer Vision - Volume Part I*. Springer-Verlag. DOI: 10.1007/11744023_32.

Bjørk, A. A., Kjær, K. H., Korsgaard, N. J., Khan, S. A., Kjeldsen, K. K., Andresen, C. S., Box, J. E., Larsen, N. K., and Funder, S. (2012). "An Aerial View of 80 Years of Climate-Related Glacier Fluctuations in Southeast Greenland". In: *Nature Geoscience* 5. DOI: 10.1038/ngeo1481.

Bracegirdle, T., Caton Harrison, T., Holmes, C., Lu, H., Martineau, P., and Phillips, T. (2024). "Antarctic extreme seasons under 20th and 21st century climate change". In: *npj Climate and Atmospheric Science* 7. DOI: 10.1038/s41612-024-00822-y.

Brynte, L., Iglesias, J. P., Olsson, C., and Kahl, F. (2024). "Learning Structure-from-Motion with Graph Attention Networks". In: *Pre-print arXiv*. DOI: 10.48550/arXiv.2308.15984.

Burton-Johnson, A., Black, M., Fretwell, P. T., and Kaluza-Gilbert, J. (2016). "An Automated Methodology for Differentiating Rock from Snow, Clouds and Seain Antarctica from Landsat 8 Imagery: A New Rock Outcrop Map and

Areaestimation for the Entire Antarctic Continent". In: *The Cryosphere* 10. DOI: 10.5194/tc-10-1665-2016.

Chavan, S., Yu, X., and Saniie, J. (2022). "High Precision Analog Gauge Reader Using Optical Flow and Computer Vision". In: *2022 IEEE International Conference on Electro Information Technology (eIT)*. DOI: 10.1109/eIT53891.2022.9813954.

Chen, H. and Tseng, Y. (2016). "Study of automatic image rectification and registration of scanned historical aerial photographs". In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B8. DOI: 10.5194/isprs-archives-XLI-B8-1229-2016.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2018). "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence*. DOI: 10.1109/TPAMI.2017.2699184.

Child, S. F., Stearns, L. A., Girod, L., and Brecher, H. H. (2020). "Structure-From-Motion Photogrammetry of Antarctic Historical Aerial Photographs in Conjunction with Ground Control Derived from Satellite Data". In: *Remote Sensing* 13. DOI: 10.3390/rs13010021.

Cléri, I., Pierrot-Deseilligny, M., and Vallet, B. (2014). "Automatic Georeferencing of a Heritage of old analog aerial Photographs". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II–3. DOI: 10.5194/isprsannals-II-3-33-2014.

Cook, A. J., Holland, P. R., Meredith, M. P., Murray, T., Luckman, A., and Vaughan, D. G. (2016). "Ocean Forcing of Glacier Retreat in the Western Antarctic Peninsula". In: *Science* 353. DOI: 10.1126/science.aae0017.

Cook, A. J. and Vaughan, D. G. (2010). "Overview of Areal Changes of the Ice Shelves on the Antarctic Peninsula over the Past 50 Years". In: *The Cryosphere* 4. DOI: 10.5194/tc-4-77-2010.

Cook, A., Vaughan, D., Luckman, A., and Murray, T. (2014). "A new Antarctic Peninsula glacier basin inventory and observed area changes since the 1940s". In: *Antarctic Science* 26. DOI: 10.1017/S0954102014000200.

Corr, D., Leeson, A., McMillan, M., Zhang, C., and Barnes, T. (2022). "An inventory of supraglacial lakes and channels across the West Antarctic Ice Sheet". In: *Earth System Science Data* 14. DOI: 10.5194/essd-14-209-2022.

Cowley, D. C. and Stichelbaut, B. B. (2012). "Historic Aerial Photographic Archives for European Archaeology". In: *European Journal of Archaeology* 15. DOI: 10.1179/1461957112Y.0000000010.

Craciun, D. and Le Bris, A. (2022). "Automatic algorithm for georeferencing historical-to-nowadays aerial iamges acquired in natural environments". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 43. DOI: 10.5194/isprs-archives-XLIII-B2-2022-21-2022.

Dahle, F. (2023). *antarctic_segmentation*. https://github.com/fdahle/antarctic_segmentation. Version 1.0.0. Accessed 11-07-2025.

– (2024a). *Antarctic_TMA*. https://github.com/fdahle/Antarctic_TMA. Accessed 11-07-2025.

– (2024b). *Data underlying the publication: Revisiting the Past: A comparative study for semantic segmentation of historical images of Adelaide Island using U-nets*. DOI: 10.4121/9A08DC50-8957-4E43-8238-3A84386CAFBD.V1.

– (2024c). *Geo-referenced image positions of the historical images of the TMA archive*. DOI: 10.4121/CD746A00-15BB-479B-A669-37B5AEEE44B5.V1.

– (2024d). *Historical Geo-referencing*. https://github.com/fdahle/hist_georef. Accessed 11-07-2025.

– (2024e). *Training data semantic segmentation historical Images*. DOI: 10.4121/052FA1DC-CF10-4BEF-8B64-AEDA99FBE3A4.V1.

Dahle, F., Lindenbergh, R., and Wouters, B. (2024a). "Polar perspectives: a deep dive into geo-referencing historical Antarctic photos". In: *International Journal of Digital Earth* 17. DOI: 10.1080/17538947.2024.2406384.

– (2024b). "Revisiting the Past: A comparative study for semantic segmentation of historical images of Adelaide Island using U-nets". In: *ISPRS Open Journal of Photogrammetry and Remote Sensing* 11. DOI: https://doi.org/10.1016/j.ophoto.2023.100056.

Dahle, F., Liu, Y., Lindenbergh, R., and Wouters, B. (2025). "From Film to Data: Automating Meta-Feature Extraction in Historical Aerial Imagery". In: *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*. DOI: 10.1007/s41064-025-00357-8.

Dahle, F., Tanke, J., Wouters, B., and Lindenbergh, R. (2022). "Semantic segmentation of historical photographs of the Antarctic Peninsula". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-2-2022. DOI: 10.5194/isprs-annals-V-2-2022-237-2022.

Denzinger, F., Machguth, H., Barandun, M., Berthier, E., Girod, L., Kronenberg, M., Usubaliev, R., and Hoelzle, M. (2021). "Geodetic mass balance of Abramov Glacier from 1975 to 2015". In: *Journal of Glaciology*. DOI: 10.1017/jog.2020.108.

DeTone, D., Malisiewicz, T., and Rabinovich, A. (2018). "SuperPoint: Self-Supervised Interest Point Detection and Description". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE. DOI: 10.1109/CVPRW.2018.00060.

Dias, M., Monteiro, J., Estima, J., Silva, J., and Martins, B. (2020). "Semantic Segmentation and Colorization of Grayscale Aerial Imagery with W-Net Models". In: *Expert Systems* 37. DOI: 10.1111/exsy.12622.

Dong, Y., Zhao, J., Floricioiu, D., and Krieger, L. (2022). "Automatic calving front extraction from digital elevation model-derived data". In: *Remote Sensing of Environment* 270. DOI: 10.1016/j.rse.2021.112854.

Du, Y., Li, C., Guo, R., Yin, X., Liu, W., Zhou, J., Bai, Y., Yu, Z., Yang, Y., Dang, Q., and Wang, H. (2020). "PP-OCR: A Practical Ultra Lightweight OCR System". In: *Pre-print arXiv*. DOI: 10.48550/ARXIV.2009.09941.

Elflein, S., Zhou, Q., Agostinho, S., and Leal-Taixé, L. (2025). "Light3R-SfM: Towards Feed-forward Structure-from-Motion". In: *Pre-print arXiv*. DOI: 10.48550/arXiv.2501.14914.

Eltner, A., Kaiser, A., Castillo, C., Rock, G., Neugirg, F., and Abellán, A. (2016). "Image-based surface reconstruction in geomorphometry – merits, limits and developments". In: *Earth Surface Dynamics* 4. DOI: 10.5194/esurf-4-359-2016.

Engel, F. (2020). *camera-footprint-calculator*. https://github.com/frank-engel-usgs/camera-footprint-calculator. Accessed 05-09-2023.

Engstrom, L., Tran, B., Tsipras, D., Schmidt, L., and Madry, A. (2019). "Exploring the Landscape of Spatial Robustness". In: *Proceedings of the 36th International Conference on Machine Learning*. PMLR. DOI: 10.48550/arXiv.1712.02779.

Eskandari, R., Genzano, N., Fugazza, D., and Scaioni, M. (2024). "Investigation on Miage/Brenva Glaciers in The Alps From 50s to-date Based on Remote-Sensing Data". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-3-2024. DOI: 10.5194/isprs-archives-XLVIII-3-2024-147-2024.

Esri (June 21, 2025). *ArcGIS Pro software*. URL: https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview.

Farella, E. M., Morelli, L., Remondino, F., Mills, J. P., Haala, N., and Crompvoets, J. (2022). "The EuroSDR TIME benchmark for historical aerial images". In: *International archives of the photogrammetry, remote sensing and spatial information sciences* 43. DOI: 10.5194/isprs-archives-XLIII-B2-2022-1175-2022.

Feng, F., Zhang, Y., Zhang, J., and Liu, B. (2022). "Small Sample Hyperspectral Image Classification Based on Cascade Fusion of Mixed Spatial-Spectral Features and Second-Order Pooling". In: *Remote Sensing* 14. DOI: 10.3390/rs14030505.

Ferrigno, A. J. C. A. J. F. D. G. V. J. G. (2005). "Retreating Glacier Fronts on the Antarctic Peninsula over the Past Half-Century". In: *Science* 308. DOI: 10.1126/science.1104235.

Fischler, M. A. and Bolles, R. C. (1981). "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". In: *Communications of the ACM* 24. DOI: 10.1145/358669.358692.

Frankl, A., Seghers, V., Stal, C., De Maeyer, P., Petrie, G., and Nyssen, J. (2015). "Using image-based modelling (SfM–MVS) to produce a 1935 ortho-mosaic of the Ethiopian highlands". In: *International Journal of Digital Earth* 8. DOI: 10.1080/17538947.2014.942715.

Friedl, P., Seehaus, T. C., Wendt, A., Braun, M. H., and Höppner, K. (2018). "Recent dynamic changes on Fleming Glacier after the disintegration of Wordie Ice Shelf, Antarctic Peninsula". In: *The Cryosphere* 12. DOI: 10.5194/tc-12-1347-2018.

Garbin, C., Zhu, X., and Marques, O. (2020). "Dropout vs. Batch Normalization: An Empirical Study of Their Impact to Deep Learning". In: *Multimedia Tools and Applications* 79. DOI: 10.1007/s11042-019-08453-9.

Geyman, E., Pelt, W., Maloof, A., Aas, H., and Kohler, J. (2022). "Historical glacier change on Svalbard predicts doubling of mass loss by 2100". In: *Nature* 601. DOI: 10.1038/s41586-021-04314-4.

Ghuffar, S., King, O., Guillet, G., Rupnik, E., and Bolch, T. (2023). “Brief communication: Glacier mapping and change estimation using very high-resolution declassified Hexagon KH-9 panoramic stereo imagery (1971–1984)”. In: *The Cryosphere* 17. DOI: 10.5194/tc-17-1299-2023.

Giordano, S., Le Bris, A., and Mallet, C. (2018). “Toward automatic georeferencing of archival aerial photogrammetric surveys”. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2. DOI: 10.5194/isprs-annals-IV-2-105-2018.

Girod, L., Nielsen, N. I., Couderette, F., Nuth, C., and Kääb, A. (2018). “Precise DEM Extraction from Svalbard Using 1936 High Oblique Imagery”. In: *Geoscientific Instrumentation, Methods and Data Systems* 7. DOI: 10.5194/gi-7-277-2018.

GLIMS and NSIDC (2005). *Global Land Ice Measurements from Space (GLIMS) Glacier Database*. Updated 2018. DOI: 10.7265/N5V98602.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press.

Haralick, R., Shanmugam, K., and Dinstein, I. (1975). “Texture Features for Image Classification”. In: *Systems, Man and Cybernetics, IEEE Transactions on* 3.

Harris, C. G. and Stephens, M. J. (1988). “A Combined Corner and Edge Detector”. In: *Alvey Vision Conference*. URL: https://api.semanticscholar.org/CorpusID:1694378.

Hartmann, A., Davari, A., Seehaus, T., Braun, M., Maier, A., and Christlein, V. (2021). “Bayesian U-Net for Segmenting Glaciers in Sar Imagery”. In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE. DOI: 10.1109/IGARSS47720.2021.9554292.

Heffels, M. and Vanschoren, J. (2020). “Aerial Imagery Pixel-level Segmentation”. In: *Pre-print arXiv*. DOI: 10.48550/arXiv.2012.02024.

Heidler, K., Mou, L., Baumhoer, C., Dietz, A., and Zhu, X. X. (2021). “HED-UNet: Combined Segmentation and Edge Detection for Monitoring the Antarctic Coastline”. In: *IEEE Transactions on Geoscience and Remote Sensing*. DOI: 10.1109/TGRS.2021.3064606.

Heisig, H. and Simmen, J.-L. (2021). “Re-Engineering the Past: Countrywide Georeferencing of Archival Aerial Imagery”. In: *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 89. DOI: 10.1007/s41064-021-00162-z.

Holmlund, E. S. (2021). “Aldegondabreen glacier change since 1910 from structure-from-motion photogrammetry of archived terrestrial and aerial photographs: utility of a historic archive to obtain century-scale Svalbard glacier mass losses”. In: *Journal of Glaciology* 67. DOI: 10.1017/jog.2020.89.

Howat, I. M., Porter, C., Smith, B. E., Noh, M.-J., and Morin, P. (2019). “The Reference Elevation Model of Antarctica”. In: *The Cryosphere* 13. DOI: 10.5194/tc-13-665-2019.

Ioffe, S. and Szegedy, C. (2015). “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *Proceedings of the*

32nd International Conference on Machine Learning. DOI: 10.48550/ARXIV.1502.03167.

Jadon, S. (2020). "A Survey of Loss Functions for Semantic Segmentation". In: *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE. DOI: 10.1109/CIBCB48159.2020.9277638.

James, M. R. and Robson, S. (2012). "Straightforward reconstruction of 3D surfaces and topography with a camera: Accuracy and geoscience application". In: *Journal of Geophysical Research: Earth Surface* 117. DOI: 10.1029/2011JF002289.

Jiang, X., Wang, Y., Liu, W., Li, S., and Liu, J. (2019). "CapsNet, CNN, FCN: Comparative Performance Evaluation for Image Classification". In: *International Journal of Machine Learning and Computing* 9. DOI: 10.18178/ijmlc.2019.9.6.881.

Kanezaki, A. (2018). "Unsupervised Image Segmentation by Backpropagation". In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. DOI: 10.1109/ICASSP.2018.8462533.

Karel, W., Doneus, M., Verhoeven, G., Briese, C., Ressl, C., and Pfeifer, N. (2013). "Oriental: automatic geo-referencing and ortho-rectification of archaeological aerial photographs". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. II-5/W1. DOI: 10.5194/isprsannals-II-5-W1-175-2013.

Kattenborn, T., Eichel, J., and Fassnacht, F. E. (2019). "Convolutional Neural Networks Enable Efficient, Accurate and Fine-Grained Segmentation of Plant Species and Communities from High-Resolution UAV Imagery". In: *Scientific Reports* 9. DOI: 10.1038/s41598-019-53797-9.

Ketkar, N. and Ketkar, N. (2017). "Stochastic gradient descent". In: *Deep learning with Python: A hands-on introduction*.

King, D. E. (2009). "Dlib-ml: A Machine Learning Toolkit". In: *Journal of Machine Learning Research* 10.

Kingma, D. P. and Ba, J. (2014). "Adam: A Method for Stochastic Optimization". In: *Pre-print arXiv*. DOI: 10.48550/ARXIV.1412.6980.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R. (2023). "Segment Anything". In: *Pre-print arXiv*. DOI: 10.48550/arXiv.2304.02643.

Knuth, F., Shean, D., Bhushan, S., Schwat, E., Alexandrov, O., McNeil, C., Dehecq, A., Florentine, C., and O'Neel, S. (2023). "Historical Structure from Motion (HSfM): Automated Processing of Historical Aerial Photographs for Long-Term Topographic Change Analysis". In: *Remote Sensing of Environment* 285. DOI: 10.1016/j.rse.2022.113379.

Kornilov, A. S. and Safonov, I. V. (2018). "An Overview of Watershed Algorithm Implementations in Open Source Libraries". In: *Journal of Imaging* 4. DOI: 10.3390/jimaging4100123.

Korsgaard, N. J., Nuth, C., Khan, S. A., Kjeldsen, K. K., Bjørk, A. A., Schomacker, A., and Kjær, K. H. (2016). "Digital elevation model and orthophotographs of

Greenland based on aerial photographs from 1978–1987". In: *Scientific Data* 3. DOI: 10.1038/sdata.2016.32.

Kostrzewa, A. (2024). "Geoprocessing of archival aerial photos and their scientific applications: A review". In: *Reports on Geodesy and Geoinformatics* 118. DOI: 10.2478/rgg-2024-0010.

Kostrzewa, A., Farella, E. M., Morelli, L., Ostrowski, W., Remondino, F., and Bakuła, K. (2024). "Digitizing Historical Aerial Images: Evaluation of the Effects of Scanning Quality on Aerial Triangulation and Dense Image Matching". In: *Applied Sciences* 14. DOI: 10.3390/app14093635.

Kugelman, J., Allman, J., Read, S. A., Vincent, S. J., Tong, J., Kalloniatis, M., Chen, F. K., Collins, M. J., and Alonso-Caneiro, D. (2022). "A Comparison of Deep Learning U-Net Architectures for Posterior Segment OCT Retinal Layer Segmentation". In: *Scientific Reports* 12. DOI: 10.1038/s41598-022-18646-2.

Kunz, M., King, M. A., Mills, J. P., Miller, P. E., Fox, A. J., Vaughan, D. G., and Marsh, S. H. (2012). "Multi-Decadal Glacier Surface Lowering in the Antarctic Peninsula: Peninsula Glacier Surface Lowering". In: *Geophysical Research Letters* 39. DOI: 10.1029/2012GL052823.

Lateef, F. and Ruichek, Y. (2019). "Survey on semantic segmentation using deep learning techniques". In: *Neurocomputing* 338. DOI: 10.1016/j.neucom.2019.02.003.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2018). "Focal Loss for Dense Object Detection". In: *Pre-print arXiv*. DOI: 10.48550/arXiv.1708.02002.

Lindenberger, P., Sarlin, P.-E., and Pollefeys, M. (2023). "LightGlue: Local Feature Matching at Light Speed". In: *Pre-print arXiv*. DOI: 10.48550/arXiv.2306.13643.

Lowe, D. G. (2004). "Distinctive Image Features from Scale-Invariant Keypoints". In: *International Journal of Computer Vision* 60. DOI: 10.1023/B:VISI.0000029664.99615.94.

Maiwald, F., Feurer, D., and Eltner, A. (2023). "Solving photogrammetric cold cases using AI-based image matching: New potential for monitoring the past with historical aerial images". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 206. DOI: 10.1016/j.isprsjprs.2023.11.008.

Matsuoka, K., Skoglund, A., Roth, G., de Pomereu, J., Griffiths, H., Headland, R., Herried, B., Katsumata, K., Le Brocq, A., Licht, K., Morgan, F., Neff, P. D., Ritz, C., Scheinert, M., Tamura, T., Van de Putte, A., van den Broeke, M., von Deschwanden, A., Deschamps-Berger, C., Van Liefferinge, B., Tronstad, S., and Melvær, Y. (2021). "Quantarctica, an integrated mapping environment for Antarctica, the Southern Ocean, and sub-Antarctic islands". In: *Environmental Modelling & Software* 140. DOI: 10.1016/j.envsoft.2021.105015.

Mboga, N., Grippa, T., Georganos, S., Vanhuysse, S., Smets, B., Dewitte, O., Wolff, E., and Lennert, M. (2020). "Fully Convolutional Networks for Land Cover Classification from Historical Panchromatic Aerial Photographs". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 167. DOI: 10.1016/j.isprsjprs.2020.07.005.

McGlone, J., E. Mikhail, and J. S. Bethel, eds. (2013). *Manual of photogrammetry*. 5th. ASPRS.

McNabb, R., Girod, L., Nuth, C., and Kääb, A. (2020). *An open-source toolset for automated processing of historic spy photos: sPyMicMac*. DOI: 10.5194/egusphere-egu2020-11150.

Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., Melbourne-Thomas, J., Muelbert, M., Ottersen, G., Pritchard, H., and A.G., S. (Sept. 2019). "Chapter 3: Polar Regions. IPCC Special Report on the Ocean and Cryosphere in a Changing Climate". In.

Mertes, J. R., Gulley, J. D., Benn, D. I., Thompson, S. S., and Nicholson, L. I. (2017). "Using structure-from-motion to create glacier DEMs and orthoimagery from historical terrestrial and oblique aerial imagery". In: *Earth Surface Processes and Landforms* 42. DOI: 10.1002/esp.4188.

Mestre-Runge, C., Lorenzo-Lacruz, J., Ortega-Mclear, A., and Garcia, C. (2023). "An Optimized Workflow for Digital Surface Model Series Generation Based on Historical Aerial Images: Testing and Quality Assessment in the Beach-Dune System of Sa Ràpita-Es Trenc (Mallorca, Spain)". In: *Remote Sensing* 15. DOI: 10.3390/rs15082044.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D. (2022). "Image Segmentation Using Deep Learning: A Survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44. DOI: 10.1109/TPAMI.2021.3059968.

Mölg, N. and Bolch, T. (2017). "Structure-from-Motion Using Historical Aerial Images to Analyse Changes in Glacier Surface Elevation". In: *Remote Sensing* 9. DOI: 10.3390/rs9101021.

Morgenstern, A., Overduin, P. P., Günther, F., Stettner, S., Ramage, J., Schirrmeister, L., Grigoriev, M. N., and Grosse, G. (2021). "Thermo-erosional valleys in Siberian ice-rich permafrost". In: *Permafrost and Periglacial Processes* 32. DOI: 10.1002/ppp.2087.

Moussavi, M., Pope, A., Halberstadt, A. R. W., Trusel, L. D., Cioffi, L., and Abdalati, W. (2020). "Antarctic Supraglacial Lake Detection Using Landsat 8 and Sentinel-2 Imagery: Towards Continental Generation of Lake Volumes". In: *Remote Sensing* 12. DOI: 10.3390/rs12010134.

Müller, D., Soto-Rey, I., and Kramer, F. (2022). "Towards a Guideline for Evaluation Metrics in Medical Image Segmentation". In: *BMC Research Notes* 15. DOI: 10.1186/s13104-022-06096-y.

Nambiar, K. G., Morgenshtern, V. I., Hochreuther, P., Seehaus, T., and Braun, M. H. (2022). "A Self-Trained Model for Cloud, Shadow and Snow Detection in Sentinel-2 Images of Snow- and Ice-Covered Regions". In: *Remote Sensing* 14. DOI: 10.3390/rs14081825.

Nicu, I. C. and Stoleriu, C. C. (2019). "Land use changes and dynamics over the last century around churches of Moldavia, Bukovina, Northern Romania – Challenges and future perspectives". In: *Habitat International* 88. DOI: 10.1016/j.habitatint.2019.04.006.

Nilsson, J., Gardner, A. S., and Paolo, F. S. (2022). "Elevation change of the Antarctic Ice Sheet: 1985 to 2020". In: *Earth System Science Data* 14. DOI: 10.5194/essd-14-3573-2022.

North, R. and Barrows, T. (2024). “High-resolution elevation models of Larsen B glaciers extracted from 1960s imagery”. In: *Scientific Reports* 14. DOI: 10.1038/s41598-024-65081-6.

Nuth, C. and Kääb, A. (2011). “Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change”. In: *The Cryosphere* 5. DOI: 10.5194/tc-5-271-2011.

Otosaka, I. N., Shepherd, A., Ivins, E. R., Schlegel, N.-J., Amory, C., van den Broeke, M. R., Horwath, M., Joughin, I., King, M. D., Krinner, G., Nowicki, S., Payne, A. J., Rignot, E., Scambos, T., Simon, K. M., Smith, B. E., Sørensen, L. S., Velicogna, I., Whitehouse, P. L., A, G., Agosta, C., Ahlstrøm, A. P., Blazquez, A., Colgan, W., Engdahl, M. E., Fettweis, X., Forsberg, R., Gallée, H., Gardner, A., Gilbert, L., Gourmelen, N., Groh, A., Gunter, B. C., Harig, C., Helm, V., Khan, S. A., Kittel, C., Konrad, H., Langen, P. L., Lecavalier, B. S., Liang, C.-C., Loomis, B. D., McMillan, M., Melini, D., Mernild, S. H., Mottram, R., Mouginot, J., Nilsson, J., Noël, B., Pattle, M. E., Peltier, W. R., Pie, N., Roca, M., Sasgen, I., Save, H. V., Seo, K.-W., Scheuchl, B., Schrama, E. J. O., Schröder, L., Simonsen, S. B., Slater, T., Spada, G., Sutterley, T. C., Vishwakarma, B. D., van Wessem, J. M., Wiese, D., van der Wal, W., and Wouters, B. (2023). “Mass balance of the Greenland and Antarctic ice sheets from 1992 to 2020”. In: *Earth System Science Data* 15. DOI: 10.5194/essd-15-1597-2023.

Perreault, N., Lévesque, E., Fortier, D., Gratton, D., and Lamarque, L. J. (2017). “Remote sensing evaluation of High Arctic wetland depletion following permafrost disturbance by thermo-erosion gullying processes”. In: *Arctic Science* 3. DOI: 10.1139/as-2016-0047.

Pix4D (2024). *Semantic intelligence meets photogrammetry in PIX4Dmatic*. https://www.pix4d.com/labs/semantic-intelligence-photogrammetry-pix4dmatic/. Accessed 11-07-2025.

Polar Geospatial Center (2023). *Public HTTP Data Repository*. https://data.pgc.umn.edu/aerial/usgs/tma/photos/. Accessed 03-01-2023.

Pope, A., Rees, W., Fox, A., and Fleming, A. (2014). “Open Access Data in Polar and Cryospheric Remote Sensing”. In: *Remote Sensing* 6. DOI: 10.3390/rs6076183.

Rodrigo, C., Varas-Gomez, A., Bustamante-Maino, A., and Mena-Hodges, E. (2021). “High-concentration sediment plumes, Horseshoe Island, western Antarctic Peninsula”. In: *Antarctic Science* 33. DOI: 10.1017/S0954102021000055.

Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Vol. 9351. Springer International Publishing. DOI: 10.1007/978-3-319-24574-4_28.

Salach, A. (2017). “SAPC – Application for adapting scanned analouge photographs to use them in Structure from Motion technology”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-1/W1. DOI: 10.5194/isprs-archives-XLII-1-W1-197-2017.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., and Rabinovich, A. (2020). “SuperGlue: Learning Feature Matching With Graph Neural Networks”. In: *2020 IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. DOI: 10.1109/CVPR42600.2020.00499.

Scambos, T., Bohlander, J., Shuman, C., and Skvarca, P. (2004). "Glacier Acceleration and Thinning after Ice Shelf Collapse in the Larsen B Embayment, Antarctica". In: *Geophysical Research Letters* 31. DOI: 10.1029/2004GL020670.

Schönberger, J. L. and Frahm, J.-M. (2016). "Structure-from-Motion Revisited". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR.2016.445.

Schulz, J., Cramer, M., and Herbst, T. (2021). "Evaluation of Phase One Scan Station for Analogue Aerial Image Digitisation". In: *PFG – Journal of Photogrammetry Remote Sensing and Geoinformation Science* 89. DOI: 10.1007/s41064-021-00174-9.

Sechidis, K., Tsoumakas, G., and Vlahavas, I. (2011). "On the Stratification of Multi-label Data". In: *Machine Learning and Knowledge Discovery in Databases*. Springer Berlin Heidelberg.

Sevara, C., Verhoeven, G., Doneus, M., and Draganits, E. (2018). "Surfaces from the Visual Past: Recovering High-Resolution Terrain Data from Historic Aerial Imagery for Multitemporal Landscape Analysis". In: *Journal of Archaeological Method and Theory* 25. DOI: 10.1007/s10816-017-9348-9.

Shepherd, A., Ivins, E., Rignot, E., Smith, B., Van den Broeke, M., Whitehouse, P., Briggs, K., Joughin, I., Krinner, G., Nowicki, S., Payne, A., Scambos, T., Schlegel, N.-J., Aa, G., Agosta, C., Ahlstrøm, A., Babonis, G., Barletta, V., Blazquez, A., and Wouters, B. (2018). "Mass balance of the Antarctic Ice Sheet from 1992 to 2017". In: *Nature* 558. DOI: 10.1038/s41586-018-0179-y.

Shepherd, A., Ivins, E., van den Broeke, M., Schrama, E., van der Wal, W., Wouters, B., Rietbroek, R., McMillan, M., Wilton, D., and More Authors (2019). "Mass balance of the Greenland Ice Sheet from 1992 to 2018". In: *Nature* 579. DOI: 10.1038/s41586-019-1855-2.

Smith, I. F. (1973). "Gravity survey on Shoesmith Glacier, Horseshoe Island, Graham Land". In: *British Antarctic Survey Bulletin* 33.

Snavely, N., Seitz, S. M., and Szeliski, R. (2006). "Photo tourism: exploring photo collections in 3D". In: *ACM Trans. Graph.* 25. DOI: 10.1145/1141911.1141964.

Solórzano, J. V., Mas, J. F., Gao, Y., and Gallardo-Cruz, J. A. (2021). "Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery". In: *Remote Sensing* 13. DOI: 10.3390/rs13183600.

Stankiewicz, O., Lafruit, G., and Domański, M. (2018). "Multiview video: Acquisition, processing, compression, and virtual view rendering". In: *Academic Press Library in Signal Processing: Image and Video Processing and Analysis and Computer Vision*. DOI: 10.1016/B978-0-12-811889-4.00001-4.

Stark, M., Rom, J., Haas, F., Piermattei, L., Fleischer, F., Altmann, M., and Becht, M. (2022). "Long-term assessment of terrain changes and calculation of erosion rates in an alpine catchment based on SfM-MVS processing of historical aerial images. How camera information and processing strategy affect quantitative analysis". In: *Journal of Geomorphology* 1. DOI: 10.1127/jgeomorphology/2022/0755.

Sun, C. and Wu, X. (2001). “Automatic segmentation of fiducial marks using attribute-based mathematical morphology”. In: *Journal of Electronic Imaging* 10. DOI: 10.1117/1.1344186.

Trimble Inc (June 21, 2025). *Trimble Inpho software*. URL: https://geospatial.trimble.com/en/products/software/trimble-inpho.

Turner, J., Lu, H., White, I., King, J., Phillips, T., Hosking, S., Bracegirdle, T., Marshall, G., Mulvaney, R., and Deb, P. (2016). “Absence of 21st century warming on Antarctic Peninsula consistent with natural variability”. In: *Nature* 535. DOI: 10.1038/nature18645.

Tyszkiewicz, M., Fua, P., and Trulls, E. (2020). “DISK: Learning local features with policy gradient”. In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc.

University of Minnesota (2023). *Aerial Photography - Antarctic Single Frames (1946-2000)*. https://www.pgc.umn.edu/data/aerial/. Accessed 03-01-2023.

USGS (2018). *USGS EROS Archive - Aerial Photography - Antarctic Single Frame Records*. DOI: 10.5066/F7MW2FDP.

Usmanov, B., Nicu, I. C., Gainullin, I., and Khomyakov, P. (2018). “Monitoring and assessing the destruction of archaeological sites from Kuibyshev reservoir coastline, Tatarstan Republic, Russian Federation. A case study”. In: *Journal of Coastal Conservation* 22. DOI: 10.1007/s11852-017-0590-9.

Vacek, F., Deutsch, C., Kuttenkeuler, J., and Kirchner, N. (2024). “Short-term calving front dynamics and mass loss at Sálajiegna glacier, northern Sweden, assessed by uncrewed surface and aerial vehicles”. In: *Journal of Glaciology* 70. DOI: 10.1017/jog.2024.34.

Vargo, L., Anderson, B., Horgan, H., Mackintosh, A., Lorrey, A., and Thornton, M. (2017). “Using structure from motion photogrammetry to measure past glacier changes from historic aerial photographs”. In: *Journal of Glaciology* 63. DOI: 10.1017/jog.2017.79.

Wang, J., Li, D., Cao, W., Lou, X., Shi, A., and Zhang, H. (2022). “Remote Sensing Analysis of Erosion in Arctic Coastal Areas of Alaska and Eastern Siberia”. In: *Remote Sensing* 14. DOI: 10.3390/rs14030589.

Wang, L., Chen, X., Hu, L., and Li, H. (2020). “Overview of Image Semantic Segmentation Technology”. In: *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*. Vol. 9. DOI: 10.1109/ITAIC49862.2020.9338770.

Wang, Y., Su, J., Zhai, X., Meng, F., and Liu, C. (2022). “Snow Coverage Mapping by Learning from Sentinel-2 Satellite Multispectral Images via Machine Learning Algorithms”. In: *Remote Sensing* 14. DOI: 10.3390/rs14030782.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). “Image quality assessment: from error visibility to structural similarity”. In: *IEEE Transactions on Image Processing* 13. DOI: 10.1109/TIP.2003.819861.

Westoby, M., Brasington, J., Glasser, N., Hambrey, M., and Reynolds, J. (2012). “‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications”. In: *Geomorphology* 179. DOI: 10.1016/j.geomorph.2012.08.021.

XDEM contributors (Dec. 2024). *xDEM*. Version v0.1.0. DOI: 10.5281/zenodo.11492983. URL: https://doi.org/10.5281/zenodo.11492983.

Xu, Z., Baojie, X., and Guoxin, W. (2017). "Canny edge detection based on Open CV". In: *2017 13th IEEE international conference on electronic measurement & instruments (ICEMI)*. IEEE.

Yang, C., Xie, W., and Zisserman, A. (2022). *It's About Time: Analog Clock Reading in the Wild*. URL: https://arxiv.org/abs/2111.09162.

Ye, W., Qiao, G., Kong, F., Guo, S., Ma, X., Tong, X., and Li, R. (2016). "Rigorous geometric modelling of 1960s Argon satellite images for Antarctic ice sheet stereo mapping". In: *ISPRS Annals of the Photogrammetry*. URL: https://api.semanticscholar.org/CorpusID:56390992.

Yu, Z., Cao, Z., Yu, C., Qiao, G., and Li, R. (2022). "Ice flow velocity mapping in Greenland using historical images from 1960s to 1980s: scheme design". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B3-2022. DOI: 10.5194/isprs-archives-XLIII-B3-2022-799-2022.

Yuan, X., Yuan, X., Chen, J., and Wang, X. (2022). "Large Aerial Image Tie Point Matching in Real and Difficult Survey Areas via Deep Learning Method". In: *Remote Sensing* 14. DOI: 10.3390/rs14163907.

Zhang, L., Rupnik, E., and Pierrot-Deseilligny, M. (2021). "Feature matching for multi-epoch historical aerial images". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 182. DOI: 10.1016/j.isprsjprs.2021.10.008.

Zhao, C., King, M., Watson, C., Barletta, V., Bordoni, A., Dell, M., and Whitehouse, P. (2017). "Rapid ice unloading in the Fleming Glacier region, southern Antarctic Peninsula, and its effect on bedrock uplift rates". In: *Earth and Planetary Science Letters* 473. DOI: 10.1016/j.epsl.2017.06.002.

IMAGE SOURCES

550 Cord Software (2024). *AERIAL PHOTOGRAPHS*. URL: https://550cord.com/land-navigation-training-fm-3-25-26/map-reading-land-navigation-ch8/#fig8-7.

Leventhal Map & Education Center (2022). *Georeferencing*. URL: https://cartinal.leventhalmap.org/guides/georeference.html#georeferencing-on-the-web.

Li, F.-F., Johnson, J., and Yeung, S. (2017). *Detection and Segmentation*. URL: https://cs231n.stanford.edu/slides/2017/cs231n%5C%5F2017%5C%5Flecture11.pdf.

Sweeney, C. (2016). *Theia Vision Library*. URL: http://theia-sfm.org/sfm.html.

UK Airsports (2022). *Altimeter Settings & Calibration*. URL: https://www.ukairsports.com/news/how-to-set-up-your-instrument-altimeter-settings/.

APPENDIX

1. SEMANTIC SEGMENTATION OF HISTORICAL IMAGES

Table 1 displays all images that were used during the training (t) and validation (v) of the models and the respective class composition per image in percent.

Image	Sea ice	Snow	Rocks	Water	Clouds	Sky	Other	set
135431L0337	4.6	20.8	0.0	6.8	36.6	12.8	18.4	t
135431L0343	0.0	0.0	0.0	0.0	72.2	10.9	16.9	t
135433R0343	0.5	0.0	0.0	19.2	65.9	9.6	4.9	t
135631L0036	29.8	0.0	0.0	8.7	59.6	0.0	1.9	t
135632V0032	23.3	0.0	0.0	42.9	33.4	0.0	0.4	t
135633R0037	0.0	2.8	0.0	1.1	83.0	11.1	2.2	t
139132V0154	0.0	98.0	2.0	0.0	0.0	0.0	0.0	t
179231L0038	70.8	16.4	3.5	0.8	0.0	8.6	0.0	t
180031L0060	0.0	89.0	2.2	0.0	0.0	8.8	0.0	t
181331L0123	0.0	86.5	0.8	0.0	0.0	10.0	2.7	t
181332V0125	0.0	97.5	0.0	0.0	0.0	0.0	2.5	t
181333R0125	0.0	97.3	0.1	0.0	0.0	2.7	0.0	t
182433R0047	0.0	90.4	0.2	0.0	0.0	8.1	1.3	t
182433R0050	0.0	85.3	0.4	0.0	0.0	12.3	2.0	t
182933R0037	0.0	88.4	0.0	0.0	0.0	11.6	0.0	t
183431L0012	0.0	99.2	0.0	0.0	0.0	0.0	0.8	t
183432V0034	0.0	100.0	0.0	0.0	0.0	0.0	0.0	t
183432V0045	0.0	95.0	3.7	0.0	0.0	0.0	1.3	t
183532V0067	0.0	98.9	0.0	0.0	0.0	0.0	1.1	t
183533R0058	0.0	86.2	1.7	0.0	0.0	12.1	0.1	t
184333R0078	0.0	89.2	0.0	0.0	0.0	9.8	1.0	t
184431L0143	0.0	84.5	4.6	0.0	0.0	9.1	1.8	t
184432V0094	0.0	99.7	0.0	0.0	0.0	0.0	0.3	t
184432V0113	0.0	100.0	0.0	0.0	0.0	0.0	0.0	t
184432V0115	0.0	100.0	0.0	0.0	0.0	0.0	0.0	t
184432V0154	0.0	99.6	0.0	0.0	0.0	0.0	0.4	t
184531L0226	0.0	83.1	1.4	0.0	0.0	15.1	0.4	t
184532V0199	0.0	99.6	0.0	0.0	0.0	0.0	0.4	t
184532V0201	0.0	100.0	0.0	0.0	0.0	0.0	0.0	t
184532V0219	0.0	100.0	0.0	0.0	0.0	0.0	0.0	t

184532V0229	0.0	96.5	3.5	0.0	0.0	0.0	0.0	t
184532V0231	0.0	87.1	12.0	0.0	0.0	0.0	0.9	t
184533R0206	0.0	88.9	1.8	0.0	0.0	9.3	0.0	t
184533R0229	0.0	83.4	4.1	0.0	0.0	12.6	0.0	t
184533R0238	0.0	84.5	2.5	0.0	0.0	11.8	1.1	t
184733R0095	0.0	94.0	0.0	0.0	0.0	6.0	0.0	t
212333R0050	0.0	83.6	0.4	0.0	0.0	14.3	1.7	t
213731L0035	8.9	26.0	0.5	55.3	0.0	7.7	1.6	t
213731L0038	0.0	20.9	1.6	53.8	8.6	5.9	9.3	t
213733R0050	3.5	27.3	0.1	52.2	0.0	14.8	2.1	t
214732V0011	3.9	49.2	2.2	37.4	6.0	0.0	1.3	t
214831L0099	0.1	23.2	1.9	48.9	13.5	10.2	2.1	t
214832V0090	0.0	97.1	2.0	0.0	0.0	0.0	0.9	t
214833R0100	0.1	55.0	3.5	25.5	0.7	12.3	2.9	t
214932V0146	0.0	89.9	9.6	0.0	0.0	0.0	0.4	t
215032V0257	0.0	78.8	20.6	0.0	0.0	0.0	0.6	t
215131L0274	0.0	76.6	10.2	0.0	0.0	11.8	1.3	t
215131L0288	0.0	73.4	16.3	0.0	0.0	10.2	0.1	t
215132V0275	0.0	37.3	7.8	0.0	54.2	0.0	0.6	t
215331L0411	0.0	74.2	11.6	0.0	0.0	11.9	2.3	t
215333R0402	0.0	83.2	1.5	0.0	0.0	12.9	2.4	t
215731L0063	0.0	85.4	6.7	0.0	0.0	7.7	0.2	t
216631L0328	27.2	55.6	4.3	5.9	0.0	6.1	0.8	t
216632V0331	85.6	0.0	0.0	13.9	0.0	0.0	0.4	t
216633R0325	78.4	0.1	0.1	14.1	0.0	7.4	0.0	t
216633R0332	88.6	0.0	0.0	4.1	0.0	7.3	0.0	t
216731L0333	0.0	82.7	7.3	0.0	0.0	9.9	0.0	t
216733R0338	0.0	89.4	0.3	1.3	0.0	9.0	0.0	t
216733R0346	0.0	90.9	0.1	0.8	0.0	8.2	0.0	t
216733R0367	0.0	90.5	0.5	0.0	0.0	9.0	0.0	t
512933R0013	1.1	0.0	0.0	70.6	14.2	13.5	0.6	t
035131L0077	45.8	0.0	0.0	40.2	0.0	14.0	0.0	v
135431L0352	3.2	8.2	0.2	3.9	83.8	0.0	0.8	v
135432V0337	0.0	0.0	0.0	48.6	50.3	0.0	1.1	v
135433R0350	5.1	0.0	0.0	32.0	49.9	9.1	3.9	v
135632V0031	26.0	0.0	0.0	24.5	72.0	0.0	0.8	v
168431L0207	0.0	0.0	0.0	43.1	45.5	11.3	0.0	v
172032V0190	0.0	100.0	0.0	0.0	0.0	0.0	0.0	v
172733R0183	0.0	92.2	1.0	0.0	0.0	6.8	0.0	v
180031L0079	0.0	0.0	0.0	0.0	90.3	9.6	0.1	v
182033R0051	0.0	87.7	1.8	0.0	0.0	10.5	0.0	v
182431L0059	0.0	93.0	2.0	0.0	0.0	5.0	0.0	v
183032V0009	0.0	99.5	0.0	0.0	0.0	0.0	0.5	v
183432V0005	0.0	99.6	0.0	0.0	0.0	0.0	0.4	v
183432V0041	0.0	99.8	0.2	0.0	0.0	0.0	0.0	v
183433R0044	0.0	92.2	0.0	0.0	0.0	7.3	0.5	v
183531L0087	0.0	86.8	0.4	0.0	0.0	11	1.8	v
183532V0060	0.0	99.7	0.0	0.0	0.0	0.0	0.3	v
184332V0060	0.0	98.4	1.5	0.0	0.0	0.0	0.2	v

184432V0105	0.0	100.0	0.0	0.0	0.0	0.0	0.0	v

Table 1: Class composition per image

2. EXTRACTING METADATA FROM HISTORICAL IMAGES

Table 2 presents an overview of 20 distinct flight paths, indicating for each direction whether it is feasible to derive elevation information using the onboard altimeter and the recorded positions of the fiducial marks.

Flight Path	Direction	Nr. of images	Altimeter	Fid N	Fid E	Fid S	Fid N	
	L		110	1	110	110	110	110
1684	R		110	102	0	110	110	110
	V		110	77	110	110	110	110
	L		123	123	123	123	123	123
1813	R		123	121	0	123	123	123
	V		123	123	123	123	123	123
	L		182	118	182	182	182	182
1816	R		182	179	182	182	182	182
	V		182	161	159	182	169	182
	L		191	186	191	191	191	191
1821	R		191	190	191	191	191	191
	V		191	185	191	191	191	191
	L		114	112	114	114	114	114
1822	R		114	103	114	114	114	114
	V		114	1	1	114	114	114
	L		109	10	105	105	106	105
1824	R		109	40	109	109	109	109
	V		109	2	0	109	109	109
	L		142	30	142	142	142	142
1825	R		142	142	0	142	142	142
	V		142	142	142	142	142	142
	L		129	129	129	129	129	129
1826	R		129	129	0	129	129	129
	V		129	129	123	129	129	129
	L		111	0	109	111	111	111
1827	R		111	78	0	111	111	111
	V		111	90	111	107	111	110
	L		156	149	156	156	156	156
1833	R		156	156	156	156	156	156
	V		156	156	156	156	156	156
	L		107	72	107	107	107	107
1846	R		107	107	107	107	107	107
	V		107	107	107	107	107	107

	L	115	115	115	115	115	115
2073	R	115	115	115	115	115	115
	V	115	69	37	115	91	115
	L	107	73	107	107	107	107
2075	R	107	21	107	107	107	107
	V	107	1	18	107	90	107
	L	141	16	141	141	141	141
2136	R	141	141	0	141	141	141
	V	141	140	141	141	141	141
	L	145	46	145	145	145	145
2137	R	145	144	145	145	145	145
	V	145	70	86	145	145	145
	L	108	3	108	108	108	108
2139	R	108	38	108	108	108	108
	V	108	30	10	108	108	108
	L	116	7	116	116	116	116
2140	R	116	37	116	116	116	116
	V	116	47	116	116	116	116

Table 2: Overview of parameters that may be manually derived

3. GEO-REFERENCING OF HISTORICAL IMAGES

Table 3 displays the geo-referenced flight paths together with their respective average values.

Flight path	Nr. of images	Nr. of georef. images	information content	length error vector (m)
1644	67	66	0.10	26,939
1741	18	0	0.07	n.a.
1742	36	0	0.04	n.a.
1743	28	28	0.02	51,102
1744	10	10	0.07	28,265
1745	16	15	0.03	19,384
1746	19	18	0.12	18,507
1747	15	14	0.08	10,705
1800	76	76	0.12	20,709
1801	85	85	0.18	11,751
1803	14	0	0.52	n.a.
1804	20	2	0.18	92,024
1805	41	41	0.05	24,839
1807	41	41	0.07	27,766
1809	36	0	0.02	n.a.
1810	61	0	0.03	n.a.
1811	21	0	0.03	n.a.
1812	80	80	0.08	22,013

1813	123	123	0.15	18,837
1814	45	45	0.04	23,785
1815	76	76	0.16	26,304
1816	182	182	0.06	20,490
1817	35	35	0.11	14,912
1818	70	70	0.09	15,613
1821	188	188	0.07	7,203
1822	114	114	0.13	18,893
1823	103	103	0.19	12,645
1824	109	109	0.30	28,286
1825	115	114	0.09	59,068
1826	87	0	0.05	n.a.
1827	62	10	0.24	57,809
1829	45	2	0.06	40,189
1830	54	0	0.05	n.a.
1831	78	78	0.13	79,129
1832	81	81	0.08	57,454
1833	156	156	0.17	17,072
1834	50	50	0.46	32,106
1835	93	93	0.34	17,655
1843	89	89	0.12	13,001
1844	97	97	0.19	56,554
1845	86	86	0.20	91,998
1846	108	107	0.20	9,441
1847	95	95	0.19	13,301
1848	88	88	0.24	5,082
1849	101	101	0.16	160,741
1962	53	53	0.19	66,112
1963	11	0	0.08	n.a.
1965	20	20	0.21	44,150
1966	41	41	0.09	11,386
1967	46	1	0.08	13,889
1968	36	36	0.10	20,710
1969	34	0	0.07	n.a.
1993	26	1	0.14	20,039
2073	89	89	0.02	36,656
2074	73	73	0.02	41,598
2075	91	0	0.00	n.a.
2121	76	75	0.09	36,459
2123	54	54	0.07	5,665
2124	66	66	0.04	6,243
2133	15	1	0.38	55,343
2134	13	13	0.37	6,242
2135	25	0	0.35	n.a.
2141	53	53	0.12	15,077
2142	75	75	0.24	13,084
2160	28	28	0.50	16,189
2161	38	38	0.53	35,423
2163	36	0	0.94	n.a.

2164	48	47	0.52	18,065
2165	26	7	0.76	27,797
2166	67	67	0.60	20,396
2167	37	37	0.23	3,338

Table 3: Overview of all flight paths located inside the research area

4. 3D RECONSTRUCTION OF HISTORICAL GLACIERS

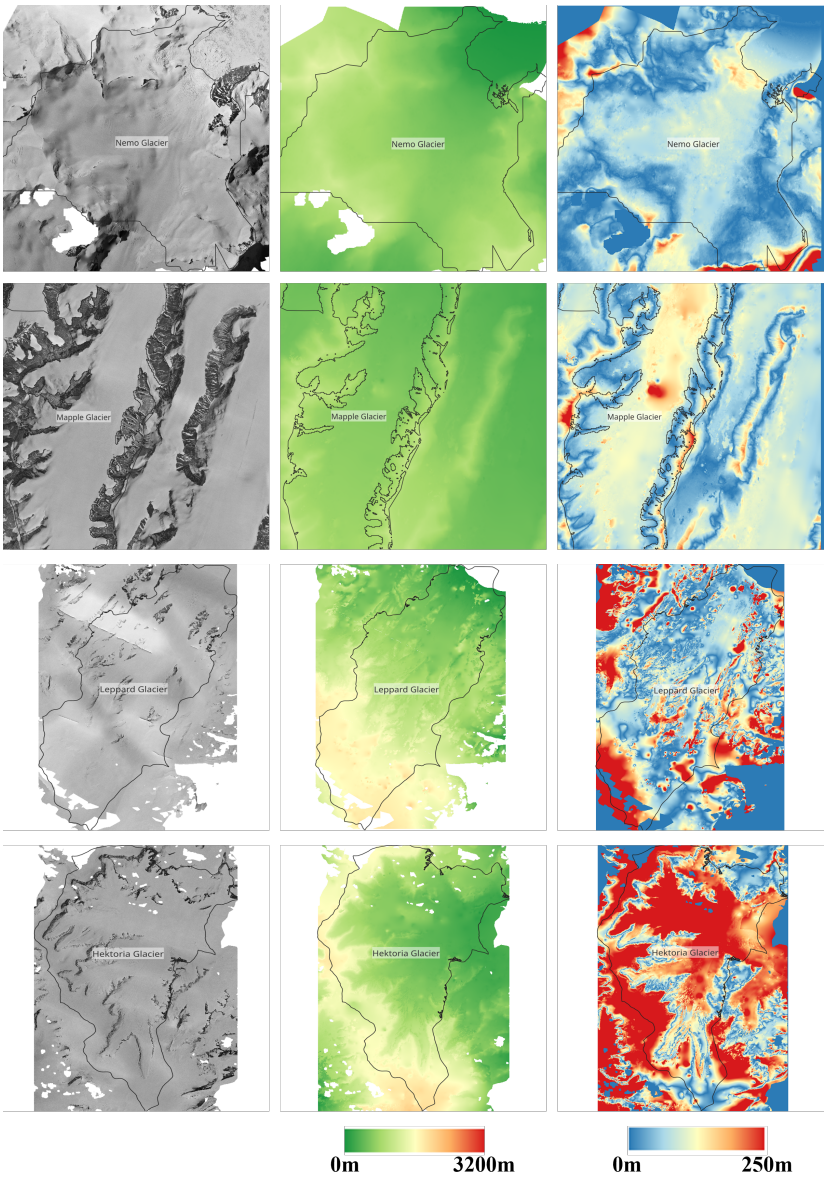


Figure 1: Orthophoto, Elevation (0-3200m) and Digital Elevation Model (DEM) (0-250m) for Nemo Glacier, Mapple Glacier, Leppard Glacier & Hektoria Glacier. The first two rows are examples for good reconstructions, the bottom two rows are examples for poor reconstructions

5. BEST PRACTICES IN PYTHON FOR SCIENTIFIC RESEARCH

Throughout this research, Python has been instrumental in processing large datasets, automating workflows, and ensuring reproducibility. Several practical approaches proved especially useful for keeping the code organized, reducing errors, and making the workflow easier to extend or adapt:

By applying these best practices, it is possible to significantly improve code readability, efficiency, and robustness, making scientific workflows more scalable and reproducible. These insights may benefit other researchers integrating Python into their computational research.

Factorization
A well-structured codebase is crucial for maintainability and reproducibility. Frequently used operations, such as loading images or connecting to a database, should be encapsulated in reusable functions. A good rule of thumb is that if a piece of code is used more than twice, it should be turned into a function. Each function should perform a single, well-defined task and have a meaningful name.

Factorization Example

```python
import cv2
import numpy as np

"""Load an image as a NumPy array."""
def load_image(image_path: str) -> np.ndarray:
    return cv2.imread(image_path, cv2.IMREAD_GRAYSCALE)
```

Type-hinting
While Python is dynamically typed, type hints improve code clarity and catch potential errors early, especially in large projects. Type annotations provide useful information to IDEs and static analysis tools, improving maintainability.

Type-hinting Example

```python
"""Compute the mean of a list of values."""
def compute_mean(values: list[float | int]) -> float:
    return sum(values) / len(values)
```

Code Performance
For large datasets, vectorization significantly improves performance compared to traditional for loops. NumPy's built-in operations execute much faster than explicit loops.

Code Performance Example

```python
import numpy as np

# Inefficient loop-based approach
squared_values = [x**2 for x in range(1000000)]

# Efficient vectorized approach
squared_values_np = np.arange(1000000) ** 2
```

Data types
When handling large datasets, selecting the correct data type is crucial for memory efficiency. By default, NumPy uses int64 or float64, which may be unnecessary. Historical images often contain values in the range [0, 255], which can be stored as int8, while binary masks can be stored as boolean. The memory savings for instance for a $50,000 \times 50,000$ array are significant: using float64 or int64 requires approximately 19,073 MB, whereas int8 or boolean only requires around 2,384 MB, resulting in an 87% reduction in memory usage.

Data Types Example

```python
# Default memory-intensive array with np.float64
large_array = np.ones((50000, 50000))

# Optimized memory-efficient array
optimized_array = np.ones((50000, 50000),
  dtype=np.uint8)
```

Code hardening
Scientific workflows often involve long-running computations. If an error occurs in one dataset, it should not halt the entire workflow. Wrapping processes in a try block ensures that errors are logged, and execution continues. Additionally, intermediate results should be cached to avoid reprocessing the entire workflow in case of a crash.

Code Hardening Example

```python
import os
import numpy as np

def process_glacier(glacier_id: str):
    # Define path to cached data
    cache_file = f"cache_{glacier_id}.npy"

    # Use existing data
    if os.path.exists(cache_file):
        return np.load(cache_file)

    try:
        # Load and work with data
        data = load_glacier_data(glacier_id)
        result = run_analysis(data)

        # Save intermediate results
        save_results(result, glacier_id)
    except Exception as e:
        print(f"Error processing {glacier_id}: {e}")
```

Efficient storage
Efficient data storage reduces disk usage and speeds up file access. TIFF files, for instance, support lossless compression using the LZW-algorithm, reducing file sizes without losing data quality. Similarly, point clouds stored in the LAZ format instead of LAS can significantly reduce file sizes while preserving precision.

Efficient Data Storage Example

```python
import rasterio

"""Save a raster with LZW compression."""
def save_lzw_tif(data: np.ndarray, filename: str):
    with rasterio.open(
        filename, 'w', driver='GTiff',
        dtype=data.dtype, compress='lzw',
        height=data.shape[0], width=data.shape[1],
        count=1) as dst:
            dst.write(data, 1)
```

ACKNOWLEDGEMENTS

The time for a PhD went by in an instant, and looking back, I am surprised to find myself already writing these words. They say this is the part of the dissertation most people will read—and for some, perhaps the only part—which makes it all the more difficult to find the right words to express my gratitude.

First and foremost, I want to thank my **supervisors**. Both of you constantly supported me and made me feel like a valued researcher and colleague rather than someone just following orders. Our weekly meetings were a highlight, providing a space where I gave updates of my research, but could also share private news. Your incredibly quick feedback was invaluable, especially as deadlines approached.

Bert, thank you for offering me the chance to pursue a PhD, fulfilling a goal I've had since I started my studies. Thank you also for the postdoc opportunity and for bringing me into the world of the cryosphere; I never expected to work in this field, but I find it endlessly fascinating. I truly appreciate the trust you placed in me and your constant support for my ideas, even those times when I managed to get myself a little lost in the middle of them.

Roderik, thank you for always bringing new ideas and interesting opportunities to the table, including the excursions to France and Colombia. Whenever I felt an idea wouldn't work or my motivation dipped, you were there to build my spirits back up and get me back on track.

To my family, who have been my foundation throughout all these years: **Anja**, thank you for the original spark that started the whole journey — the suggestion to study outside of Germany. You have supported me every step of the way, listening to my endless talk about almost any topic and helping me constantly to become a better version of myself. Thank you for encouraging me to reach new heights and try things I never thought possible. **Leon**, even though you have only been in my life for a short time, you have changed it in so many ways. Finishing this thesis while you were sleeping on my lap is a memory I will always cherish.

To my **parents**, you are the constant in my life. Thank you for always being there. No problem is too small to bring to you, and I am comforted knowing I always have a place to call home. To my sister **Carolina**, you are my mental sparring partner and another person I know I can turn to whenever I need help. Seeing you start (and, I must admit, finish!) your own PhD was a great source of external motivation for me to reach the finish line. I also want to thanks my cousins **Julian** and **Simon** for our occasional round of Age of Empires II, that alway have been (and still are) a lot of fun!

My life in the Netherlands would not have been the same without the roots back at home: To my old friends in Germany: **Mark**, **Tobi**, **Dani**, **Pia**, **Kristopher** and **Sabine** — every time I return to Nuremberg or Würzburg, it feels as though no time has passed at all.

A huge role in my life here was played by **Slopend**. This association gave me a reason to stay in the Netherlands and introduced me to survival running, as well as a world of outdoor recreation—hiking, camping, bouldering, mountain biking, trail running, adventure racing, canoeing, and climbing. These activities were my mental reset button. The yearly expeditions, often without mobile connection, forced me to take the breaks I sometimes find difficult to take myself.

Many of my Dutch friends originated from this community: **Lydia**, **Daniel**, **Max**, **Ruben**, **Kristel**, **Maikel**, **Anna**, **Schelte**, **Rik**, **Carlijn**, **Mayke**, **Mirte**, **Samuel**, **Jorgis**, **Wouter**, **Willemijn**, **Margot**, **Pim**, **Fay**, **Elise**, and **Lara**. Thank you for making me feel completely integrated into Dutch culture and giving me a second home here. A special thanks to my **hardloop & zwem group**; our weekly "craziness" helped me keep a cool head, sometimes quite literally in the winter.

This sense of community and support extended beyond my free time and back at the university: I want to thank the secretaries: **Debbie**, **Natascha**, and **Cindy**. You helped me navigate the complexities of university bureaucracy with ease, but more importantly, your doors were always open for a chat. As well I want to mention **Adriaan**, who supported me at the begin of the PhD and was always available to help me with any question about the university life.

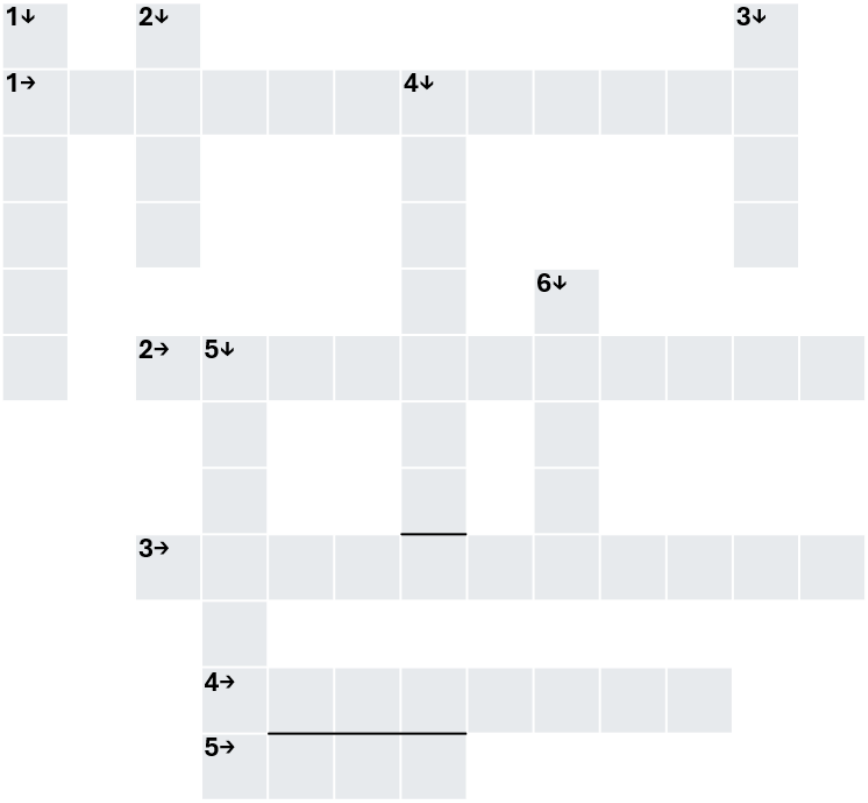
Finally, to my colleagues, many of whom became close friends. I truly enjoyed our coffee breaks, trips to PSOR, dinners, and pub quizzes. Whether you were there from the start or joined later, you made the office a better place: **Anna**, **Max**, **Paco**, **Ben**, **Maaike**, **Sophie**, **Bob**, **Carla**, **Daan**, **Frithjof**, **Marleen**, **Annelot**, **Kevin**, **Phil**, **Shashwat**, **Steven**, **Weilun**, **Wietske**, **Zach**, **Arturo**, **Ann-Sofie**, **Koen**, **Lotfi**, **Mahmoud**, **Pouriya**, **Serge**, **Shafi**, and **Yi**. (the random order of

which I credit to a quickly coded Python program, just to stay in the mood). A special thank you to those who shared the writing retreats with me. Those trips are among my favourite memories: from extended breakfasts with French baguettes and cheese to the fun of coding while beer tasting. We survived getting lost in Belgian forests, followed the US elections together, and enjoyed the occasional sauna. It was the perfect change of pace from work.

To everyone mentioned here, and those I may have missed — thank you all for being part of this journey.

CROSSWORD CHALLENGE

The following crossword covers some terminology and locations discussed in this thesis. The solutions can be found on the last page of the thesis.



Across

1. A reference symbol on the image border used for camera calibration (two words).
2. Scanning artifacts that appear as concentric interference patterns on images (two words).
3. The distance from the optical center of the lens to the sensor (two words).
4. The geometric constraint used to limit the search for matching points between two images to a single line.
5. A standard statistical measure used to quantify reprojection or prediction error.

Down

1. A geometric transformation that preserves lines and parallelism.
2. The optimization algorithm used for training the neural networks.
3. The parameter describing the angle between the x and y pixel axes.
4. The photogrammetry software utilized in this work.
5. The section of the U-Net architecture that "contracts" the image to extract features.
6. The glacier that required the most extensive processing time.

CURRICULUM VITæ

Felix Dahle

2021–2025 **PhD. in Geoscience & Remote sensing**
TU Delft, Netherlands

Thesis: *Antarctic Time Machine: 3D Reconstruction of Glaciers in the Antarctic Peninsula using Historical Structure-from-Motion*

2018–2020 **M. Sc. in Geomatics**
TU Delft, Netherlands

Thesis: *Automatic change detection in digital maps using aerial images and point clouds*

2015–2018 **B. Sc. in Geography**
Julius-Maximilians-Universität Würzburg, Germany

Thesis: *A socioeconomic comparison of regional units in Europe based on administrative and remote sensing regions in cooperation*

2012–2015 **B. Sc. in IT-Engineering & Business Consulting**
Steinbeis Hochschule Berlin, Germany

Thesis: *Remote Services to reduce the reaction time for suppressing of customer's security technology*

2001–2011 **Secondary School (Gymnasium)**
Adam-Kraft-Gymnasium Schwabach, Germany

08-05-1992 **Born in Nuremberg, Germany**

Public & Academic Profiles

Website:
www.fdahle.de

GitHub:
github.com/fdahle

ResearchGate:
Felix-Dahle

ORCID:
0000-0001-6706-7099

Google Scholar:
Felix Dahle

LIST OF PUBLICATIONS

Journal Articles

1. F. Dahle, Y. Liu, R. Lindenbergh, and B. Wouters (2025). "From Film to Data: Automating Meta-Feature Extraction in Historical Aerial Imagery". In: *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*. DOI: 10.1007/s41064-025-00357-8

2. F. Dahle, R. Lindenbergh, and B. Wouters (2024b). "Revisiting the Past: A comparative study for semantic segmentation of historical images of Adelaide Island using U-nets". In: *ISPRS Open Journal of Photogrammetry and Remote Sensing* 11. DOI: https://doi.org/10.1016/j.ophoto.2023.100056

3. F. Dahle, R. Lindenbergh, and B. Wouters (2024a). "Polar perspectives: a deep dive into geo-referencing historical Antarctic photos". In: *International Journal of Digital Earth* 17. DOI: 10.1080/17538947.2024.2406384

Publications in peer reviewed conference proceedings

1. F. Dahle, J. Tanke, B. Wouters, and R. Lindenbergh (2022). "Semantic segmentation of historical photographs of the Antarctic Peninsula". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-2-2022. DOI: 10.5194/isprs-annals-V-2-2022-237-2022

Selected conference talks

1. F. Dahle, R. Lindenbergh, J. Tanke, and B. Wouters (May 2022). "Semantic segmentation of historical images in Antarctica with neural networks". In: *EGU General Assembly*. Presented at the EGU General Assembly 2022, 23–27 May 2022. Vienna, Austria, EGU22–10597. DOI: 10.5194/egusphere-egu22-10597

2. F. Dahle, R. Lindenbergh, and B. Wouters (Sept. 2023). "Unlocking the Potential of Historical Aerial Imagery for the Antarctic Peninsula: Automating 3D Reconstruction using Python and MicMac". In: *5th Virtual Geoscience Conference*. Presented at the 5th Virtual Geoscience Conference 21-22 September 2023. Dresden, Germany

3. F. Dahle, R. Lindenbergh, and B. Wouters (Apr. 2024c). “Utilizing historical aerial imagery for change detection in Antarctica”. In: *EGU General Assembly*. Presented at the EGU General Assembly 2024, 14–19 April 2024. Vienna, Austria, EGU24–15439. DOI: 10.5194/egusphere-egu24-15439

Data publications

1. F. Dahle (2024e). *Training data semantic segmentation historical Images*. DOI: 10.4121/052FA1DC-CF10-4BEF-8B64-AEDA99FBE3A4.V1

2. F. Dahle (2024b). *Data underlying the publication: Revisiting the Past: A comparative study for semantic segmentation of historical images of Adelaide Island using U-nets*. DOI: 10.4121/9A08DC50-8957-4E43-8238-3A84386CAFBD.V1

3. F. Dahle (2024c). *Geo-referenced image positions of the historical images of the TMA archive*. DOI: 10.4121/CD746A00-15BB-479B-A669-37B5AEEE44B5.V1

Crossword Solutions

Across: 1. Fiducial Mark, 2. Newton Rings, 3. Focal Length, 4. Epipolar, 5. RMSE

Down: 1. Affine, 2. Adam, 3. Skew, 4. Agisoft, 5. Encoder, 6. Crane

TU Delft