Master's thesis

Semantic Cloud Segmentation for All Sky Images using Deep Learning and LWIR Data

Jakob Brunen

28.06.2026

Supervision:

PD Dr. Gudrun Thäter
Prof. Dr. Rudolph Triebel
Eduardo Saez Martinez

Karlsruhe Institute of Technology
Institute for Applied and Numerical Mathematics

German Aerospace Center
Institute of Solar Research

Erklärung

Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde, sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Karlsruhe, den 28.06.2026

Acknowledgements

This master thesis was conducted at the Institute of Solar Research of the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt e.V., DLR) in Almería, Spain.

First of all, I would like to thank my supervisor PD Dr. Gudrun Thäter for the opportunity to write this thesis and the support in professional matters.
I highly appreciate her feedback and confidence in my research.

I would also like to thank my second supervisor Prof. Dr. Rudolph Triebel for the interdisciplinary exchange and his valuable academic insight.

Special thanks go to my advisor, Eduardo Saez Martinez, for his mentorship, professional advice, and encouragement, as well as for his support during my time in Almería.

I would also like to thank Dr. Bijan Nouri and Yann Fabel for the discussions and feedback which were essential for the successful completion of this thesis.

Finally, I wish to thank the students and colleagues I encountered during my stay in Spain. Their advice and collegiality made this an exceptional professional and personal experience.

Abstract

The transition toward sustainable and renewable energy sources represents a significant challenge. Among different alternatives, solar power is the most abundant, however, its integration into the global energy mix is complicated, due to the variability of solar irradiance. While seasonal and diurnal changes are predictable and well-researched, intra-hour fluctuations, which primarily come from cloud cover, introduce instability into the power generation grid. The development of reliable short-term *Nowcasting* systems is important to maintain grid stability, prevent damages through power ramps in photovoltaic parks and optimize the thermal efficiency of concentrated solar power plants. A state-of-the-art approach to Nowcasting utilizes ground-based all-sky imagers (ASI) to capture hemispherical sky images. Clouds are detected and tracked in real-time to generate future irradiance maps. Furthermore, the utility of ASI systems extends to optical satellite communications, where a clear line of sight is mandatory. By leveraging thermal infrared sensors, which remain operational during night, cloud-free windows can be detected to optimize transmission schedules.

The fidelity of these nowcasts is highly dependent on the accuracy of the cloud detection phase. In recent years, deep learning methods have emerged as dominant, outperforming conventional detection techniques. However, those models introduce the critical dependency on large amounts of high-quality manually labeled ground truth data. Since manual pixel-wise annotation is time-intensive and costly obtaining such datasets remains a primary bottleneck in the field.

This thesis aims to remove that bottleneck by using the recently acquired long-wave infrared data to generate an automatically labeled ground truth height map dataset. A combination of infrared imagery and reanalysis data is used to generate height maps as pixel-wise ground truth to train a convolutional neural network for cloud detection and pixel-wise height estimation. The current state-of-the-art cloud detection model is used to generate binary ground truth masks for cloud detection in infrared images.

Evaluation on a predefined test set containing different cloud formations, shows accurate cloud detection with a mean IoU of 0.777. The cloud base height is predicted with a mean absolute error of 1046.09 m and a bias of -399.76 m from a single visible light input image. The conversion capability from the obtained height predictions to cloud height layer maps is limited with a mean IoU over all 4 classes of 0.372. The cloud detection in infrared images performs strongly on the test set with a mean IoU of 0.835 and F1 score of 0.908.

Kurzfassung

Der Übergang zu nachhaltigen und erneuerbaren Energiequellen stellt eine erhebliche Herausforderung dar. Unter den verschiedenen Alternativen ist die Solarenergie am reichlichsten vorhanden; ihre Integration in den globalen Energiemix wird jedoch durch die Variabilität der Sonneneinstrahlung erschwert. Während saisonale und tageszeitliche Schwankungen vorhersagbar und gut erforscht sind, führen kurzfristige Fluktuationen innerhalb einer Stunde – primär verursacht durch Bewölkung – zu Instabilitäten im Stromnetz. Die Entwicklung zuverlässiger Nowcasting-Systeme ist daher essenziell, um die Netzstabilität zu gewährleisten, Schäden durch Leistungsgradienten (Power Ramps) in Photovoltaikanlagen zu vermeiden und den thermischen Wirkungsgrad von konzentrierenden Solarkraftwerken (CSP) zu optimieren.

Ein aktueller Ansatz für das Nowcasting nutzt bodengestützte All-Sky-Imager (ASI), um hemisphärische Himmelsbilder aufzunehmen. Wolken werden dabei in Echtzeit erkannt und verfolgt, um zukünftige Einstrahlungskarten zu generieren. Darüber hinaus finden ASI-Systeme Anwendung in der optischen Satellitenkommunikation, bei der eine freie Sichtverbindung zwingend erforderlich ist. Durch den Einsatz von thermischen Infrarotsensoren, die auch nachts funktionsfähig sind, können wolkenfreie Zeitfenster identifiziert werden, um Übertragungspläne zu optimieren.

Die Genauigkeit dieser Nowcasts hängt maßgeblich von der Präzision der Wolkenerkennungsphase ab. In den letzten Jahren haben sich Deep-Learning-Methoden als dominant erwiesen und übertreffen konventionelle Erkennungstechniken. Diese Modelle weisen jedoch eine kritische Abhängigkeit von großen Mengen hochwertiger, manuell annotierter Referenzdaten (Ground Truth) auf. Da die manuelle pixelweise Annotation zeitintensiv und kostspielig ist, bleibt die Beschaffung solcher Datensätze ein primärer Engpass in diesem Forschungsfeld.

Das Ziel dieser Arbeit ist es, diesen Engpass durch die Nutzung seit kurzem verfügbarer langwelliger Infrarotdaten zu überwinden und einen automatisch annotierten Referenzdatensatz von Höhenkarten zu erstellen. Eine Kombination aus Infrarotbildern und reanalysis- Daten wird verwendet, um Höhenkarten als pixelweise Ground Truth zu generieren, mit denen ein Convolutional Neural Network (CNN) für die Wolkenerkennung und die pixelweise Höhenschätzung trainiert wird. Zur Generierung binärer Masken für die Wolkenerkennung in Infrarotbildern wird das aktuelle State-of-the-Art Modell verwendet.

Die Evaluierung auf einem vordefinierten Testset mit verschiedenen Wolkenformationen zeigt eine präzise Wolkenerkennung mit einem mean IoU (mIoU) von 0,777. Die Höhe der Wolkenbasis wird basierend auf einem einzelnen Bild im sichtbaren Licht mit einem mittleren absoluten Fehler (MAE) von 1046,09 m und einem Bias von -399,76 m vorhergesagt. Die Fähigkeit zur Konvertierung der gewonnenen Höhenvorhersagen in Wolkenhöhen-Layer-Karten ist mit einem mIoU über alle vier Klassen von 0,372 begrenzt. Die Wolkenerkennung in Infrarotbildern erzielt auf dem Testset starke Ergebnisse mit einer mIoU von 0,835 und einem F1-Score von 0,908.

Contents

List of Figures

List of Tables

Nomenclature

ASI All-Sky Imager

BCE Binary Cross Entropy

CBH Cloud Base Height

CIEMAT Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

CNN Convolutional Neural Network

CSL Clear Sky Library

CSP Concentrated Solar Power

DL Deep Learning

DNI Direct Normal Irradiance

HDR Hyper-Dynamic Range

HTF Heat Transfer Fluid

IoU Intersection over Union

ISA International Standard Atmosphere

LWIR Long-Wave Infrared

MAE Mean Absolute Error

ML Machine Learning

NWP Numerical Weather Prediction

PSA Plataforma Solar de Amería

PV Photovoltaics

PWV Precipitable Water Vapor

ReLU Rectified Linear Unit

RMSE Root Mean Square Error

SSIM Structural Similarity

WMO World Meteorological Organization

1 Introduction

1.1 Motivation

The automatic detection and classification of clouds is essential across several scientific and industrial fields, ranging from meteorology [1] and aviation [2] to satellite communications [3, 4] and renewable energies [5]. Clouds have always been studied however the shift towards automation is driven by the need for consistent, high-resolution data that removes human subjectivity factors.

The world shifts away from fossil fuels toward renewable energy including solar power generation [6]. A critical challenge is the variability of solar irradiance, which is highly influenced by geographical location and weather. While large scale phenomena, like seasonal or even daily changes are predictable, *intra-hour* fluctuations, which are mostly caused by passing clouds [5], are far more difficult to anticipate. Two main technologies are affected by this: Photovoltaics (PV) and concentrated solar power (CSP). As photovoltaics become more and more present in electrical power generation, growing each year, it is responsible for about 10% of all electricity generation [6]. One downside of the technology are the sharp drops in electricity caused by short-time weather phenomena like clouds. This can lead to a destabilization of power grids and induce high financial penalties [7]. CSP on the other hand relies on direct normal irradiance (DNI) from the sun to heat heat transfer fluids (HTF). The DNI changes with cloud cover variation. If this variation is too large the plant can not be operated effectively, not reaching predefined market agreements again inducing financial penalties [8].

To manage these risks, so-called *Nowcasting* systems are installed to provide short-term cloud or irradiation forecasts [9, 10, 5]. While satellite data and numerical weather prediction (NWP) models can supply broad large scale data, they lack fine grid spatial and temporal resolution crucial for real-time plant operation [11]. Ground-based All-Sky Imagers (ASI) aim to fill this gap by providing high-resolution spatial and temporal data for the region of interest. This enables the projection of cloud shadows to generate irradiation maps on the ground and forecast DNI variation [5]. To widen the coverage over larger areas up to several square kilometers multiple ASI systems can be combined to imaging networks using stereographic approaches[5, 12], enabling large scale Nowcasting.

Beyond the energy sector, cloud detection is vital for optical satellite communication [3, 4]. Laser- based satellite communication requires a clear line of sight to not loose information in the atmosphere [4]. Hence the prediction of cloud-free windows allows the optimization of data transfer times between ground station and satellite. While solar applied Nowcasting approaches are only relevant during daytime and therefore mostly use visible light cameras, satellite communication also needs to be available during the night, motivating the use of thermal infrared cameras, which do not rely on visible light. Other critical applications include aviation safety and climate science. Accurate real-time data on cloud height and coverage helps pilots avoid transitions into instrument meteorological conditions when flying under visual flight rules [2]. In climate science ground-based observations can be used to validate satellite-based measurements.

1.2 Objectives and challenges

The work is guided by the following three objectives:

- To explore the possibility of using long-wave infrared (LWIR) data to obtain cloud base height (CBH).
- To establish a method to generate automatically labeled ground truth data using LWIR all-sky images.
- To develop a real-time applicable cloud detection system for infrared cameras.

Several technical challenges complicate the fulfillment of those goals. First, as the LWIR all-sky cameras were just recently installed there is a very limited amount of historical data. The acquisition, processing and enhancing of the newly captured data is therefore of high importance for future research.

Although there has been a lot of research on cloud detection in infrared images [13, 14, 15], the accurate detection in real-time remains challenging. Certain cloud types, especially high cirrus clouds are difficult to distinguish from clear sky [16]. Additionally hazy atmospheric conditions can cause unexpected camera responses. In all-sky infrared imagery there is the additional challenge of view distortion from the fish-eye lens, yielding different responses dependent on the elevation angle [17]. Those near horizon regions pose challenges for different reasons. The different atmospheric path lengths from zenith to horizon and the accumulation of particles towards the horizon causes enlarged atmospheric radiance, influencing the output image. Furthermore clouds near the horizon appear very small even though physically they are of large scale.

A major bottleneck for deep learning (DL) approaches in the topic of cloud detection and segmentation remains the need for large amounts of high quality ground truth data. In past research this was typically done manually, requiring meteorological experts and large amounts of time [18]. This work will try tackle this issue by creating an automatically labeled dataset with the use of LWIR data.

1.3 Physics principles

To fully understand the work, this section introduces how clouds form and what types exist, some basic principles about radiation, atmospheric behaviors and how this influences infrared imagery. To close of the introduction, the different possible camera observations are discussed.

Cloud formations

Cloud formation is primarily driven by the process of adiabatic cooling. This occurs when moist air rises in the atmosphere and expands due to lower ambient pressure. This additionally causes the air parcels to cool down and the air becomes saturated with water vapor, reaching its dew point. At this point the vapor starts to condense into liquid droplets or directly form ice crystals at the surface of microscopic particles [19].

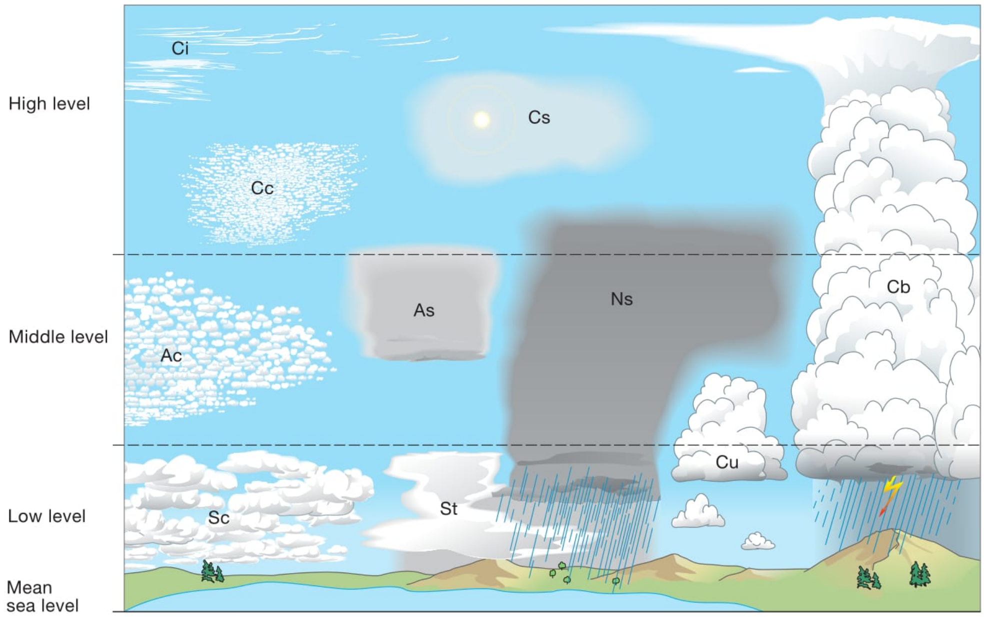

Figure 1: Illustration of the cloud genera defined by the WMO [20]. The cloud genera: Cumulus (Cu), Stratus (St), Stratocumulus (Sc), Cumulonimbus (Cb), Altocumulus (Ac), Altostratus (As), Nimbostratus (Ns), Cirrus (Ci), Cirrocumulus (CC), Cirrostratus (Cs)

The composition of a cloud is mainly governed by the ambient temperature. In lower, and therefore warmer layers of the atmosphere, clouds consist of liquid water droplets. As altitude increases the temperature decreases drastically, and even though supercooled liquid droplets may persist, clouds eventually consist almost exclusively of ice crystals [19].

Those thermodynamic conditions are almost entirely limited to the troposphere because it contains the majority of water vapor and temperature decreases with an almost constant gradient. The overlying stratosphere is characterized by extreme dryness and thermal inversion, preventing vertical air motion and the formation of clouds [20]. Given those atmospherical constraints a visible cloud can be described and classified based on its dimensions, shape, structure, luminance and color resulting in ten cloud genera [20].

A schematic of the different cloud genera as defined by the World Meteorological Organization (WMO) is shown in Figure 1.

Now that the general types of clouds have been introduced the focus is shifted towards atmospheric behavior.

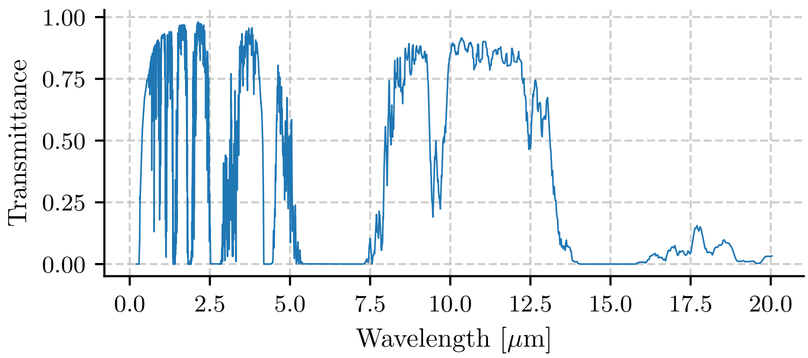

Figure 2: Atmospheric transmittance over wavelength through 1976 US Standard Atmosphere model, simulated using MODTRAN6 [22]

Atmospheric radiation

All particles and molecules emit thermal radiation due to their vibrations and rotations. Kirchhoff's radiation law states that the absorptivity must equal the emissivity for a body to be in thermal equilibrium [21]. This principle also applies to the atmosphere. Consequently the radiation absorbed by the atmosphere is emitted at corresponding wavelengths [16].

Figure 2 shows an atmospheric transmittance- wavelength diagram calculated using MODTRAN6 [22] for a zenith path from sea level to space.

Regions with high transmittance are called *atmospheric windows* [16]. The focus in this work is set to the long-wave infrared window between $8 - 14\mu m$. This window is disrupted at $9.6\mu m$ by ozone absorption mainly present in high altitudes. Otherwise the LWIR transmittance is highly vulnerable to the water vapor contained inside the atmosphere [16]. The combination of high atmospheric transmittance and it being highly effected by water vapor content, makes the window ideal for cloud detection. This itself motivates the use of long-wave infrared cameras for cloud detection purposes.

Infrared cloud imagery

Infrared imagery is excellent to detect clouds in different conditions, since it does not depend on lighting conditions and is not affected by sun scattering. Especially low and mid layer clouds are easily distinguishable. However clouds like cirrus clouds that are more transparent and yield a different spectrum can be challenging to distinguish from clear sky [16]. If they have visible structure they often can be detected, however if the sky is covered the detection is almost impossible.

The clear sky atmosphere itself also emits radiation in the infrared spectrum. The so called clear sky radiance depends highly on precipitable water vapor (PWV), hence the water vapor content inside the atmosphere. This water vapor content can be used to calculate the clear sky radiance inside the desired wavelength spectrum.

As seen above, the atmosphere has a high transmission window between the LWIR wavelenghts $8 - 14\mu m$. Inside this window opaque clouds emit thermal radiation with emissivity close to 1 and a near cloud temperature, similar to a black body. Therefore Plancks radiation law (1.1) can be used to obtain cloud temperatures from the camera response.

Planck's radiation gives a relationship between emitted spectral radiance of a black body L_{bb}, wavelength λ and temperature T. It states

$$L_{bb}(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k T}} - 1}, \tag{1.1}$$

where $h = 6.626 \cdot 10^{-34}$ Js is the Planck constant, $k = 1.381 \cdot 10^{-23} \frac{\mathrm{J}}{\mathrm{K}}$ the Boltzmann constant and $c = 2.998 \cdot 10^{8} \frac{\mathrm{m}}{\mathrm{s}}$. This relation can be used to obtain temperature values from the sensor response of the infrared cameras, by assuming black body emissivity $\varepsilon = 1$ or measuring cloud emissivity in another way. The temperature obtained from this relation by transforming the spectral radiance from the camera response to temperature is less than or equal to the physical temperature of the clouds depending on the emissivity ε of the cloud and is often times referred to as brightness temperature [16].

Atmospheric temperature

As described previously there is a direct relation between height and temperature in the troposphere, which can be used to approximate a clouds base altitude using the obtained image information. The temperature values typically range from 15 °C to -56.5 °C as defined in the international standard atmosphere (ISA)[23]. There are multiple options to obtain a temperature height relation. For the most exact information one can use data from a nearby radiosonde, which is rarely available.

The relationship can also be described using formulas. Since the temperature decreases almost linearly in the troposphere, where most clouds are present, a first approach can be to simply assume a linear temperature height relation. The ISA suggests the use of an environmental lapse rate of $\delta_{LR} = -6.5 \frac{^\circ\mathrm{C}}{\mathrm{km}}$.

Other approximations vary with humidity and pressure ranging from dry lapse rates $\tau_{LR,dry} \approx -9.8C/km$ to moist lapse rates $\tau_{LR,moist} \approx -4.5C/km$. Note that the moist lapse rate depends highly on location specific thermodynamics and can be lower [24]. Those approximations can yield acceptable results, however a more accurate approach is to use reanalysis data. Reanalysis data combines measurements from weather stations and simulation results from NWP models to provide accurate climate information which is physically sound over a large scale area [25].

To detect the clouds correctly and accurately estimate a clouds temperature careful and accurate radiometric calibration of the LWIR cameras is necessary. In the scope of this work this is done by the company Miratlas. All the available calibration information is given in section 3.

1.4 Camera detections

This part briefly introduces all the objects and conditions that are observed in the visible and infrared cameras. What the cameras show depends highly on the weather conditions. Rain and snow can block the sensors and result in unusable images. Depending on the infrared camera type those weather conditions can produce unwanted artifacts that are visible even hours after the rainfall. Furthermore birds, animals and humans can enter the cameras FOV and produce unwanted artifacts inside the captured images. Such images are fairly rare but nevertheless present. More present artifacts include the consequences of sand storms or hazy conditions. Those can produce soiling on top of the cameras which only disappears after manual camera cleaning. The final artifact in that sense comes from camera errors and false responses, leading to unwanted stripes or different artifacts. How they are treated is discussed in section 3.

Now that the artifacts have been mentioned, the actual monitoring possibilities are described.

Visible light images can detect all kinds of clouds and cloud conditions. For a human it is fairly easy to distinguish between aerosols, clouds and clear sky. Additionally it is easy to detect contrails. Depending on the conditions and camera exposure time the sun can yield very high scattering effects, making parts of the image hard to differentiate. However to really distinguish all the conditions, the sky colors are important, and therefore the imagery is only suitable during day hours. For night time the exposure time can be adjusted, so clouds can be detected, this however yields only a grayscale image, which makes differentiating between different conditions a lot harder.

This is where one of the advantages of infrared imagery comes into play. Low and mid level clouds are easily detected with very clear borders. However the sensors have difficulties distinguishing highly transmissive clouds from clear sky atmospheric radiation. Artifacts can appear in the same way as they do in visible light images, without the presence of sun scattering effects. The cameras are not sensitive to flaring or halos. They are however more vulnerable to humidity and when the captured scene is foggy, the sensors yield obscured responses. Since the atmosphere itself emits infrared radiation and the cameras have a high FOV where the atmospheric path is longer near the horizon, there exists an elevation dependent radiation offset. The elevation dependence is often described using airmass instead, which aims to describe the relative ratio from the longer atmospheric path compared to the shortest one at zenith. This visual effect is stronger the wetter the air is and needs to be considered when processing the images.

2 State of the art

Now that the usage and necessity of a cloud Nowcasting system was introduced and motivated, this chapter will turn to some previous works, that relate to the problem at hand. A few basic terms and practices in the field will be introduced and previous works done on the topic of ground based cloud observation, infrared cloud imaging and Nowcasting systems will be discussed. Special attention will be given to the approaches previously developed and examined at DLR that this work builds upon.

2.1 Ground based cloud classification

In the past there have been a lot of different approaches to classifying clouds in images obtained from ground based cameras [26, 27, 28, 29, 30]. Since the problem is very complex, it is often times split into multiple subtasks. One of the most important tasks is binary cloud segmentation and cloud cover estimation. A lot of works have been published on different criteria which can be used to distinguish clouds from the clear sky. Often times they use image statistics based on RGB channels. They range from basic color criteria like the RB- criterion to more complex algorithms (OTSU, HYTA ... [27, 31]) and since recent years the application of different machine learning techniques (e.g. [32, 33, 34, 18, 35]).

Algorithmic techniques

The algorithmic techniques make use of different statistical image parameters. Those statistics can be used to apply thresholds to distinguish between clear sky and clouds. One can use the fact that for clear sky conditions without aerosols the atmosphere scatters more blue light than red. From this different color ratio criteria can be defined [29]. One of the most basic approaches is to take the ratio or difference between the red and blue channel to define a threshold [29, 26]. Those fixed thresholding techniques yield good results in single layer conditions with sufficient lighting, however often times multilayer cases are not considered and the ratios do not perform as well. More advanced techniques use histogram thresholding combined with, for example, standard deviation to better distinguish the clouds [27]. Hasenbalg et al. [31] benchmarked different thresholding algorithms and improved upon the original HYTA technique.

A very general limitation using thresholding techniques is that they work well for low aerosol conditions but struggle as soon as the conditions vary from the ideal and the atmosphere is hazy. To tackle this issue clear sky Libraries (CSL) [36] are used for reference for similar sky conditions. They consist of historical data for various conditions and help a lot in segmenting hazy conditions, especially in near sun pixels and near the horizon where more total aerosols are present due to a larger atmospheric path lengths.

Other techniques classify using graph cuts [28]. They work by considering the pixels as connected nodes and weigh the edges between them using a cost function. If two pixels have similar value separating them has high cost and vice versa. The image is then separated into two groups by finding the *cheapest* graph cut.

Deep learning techniques

Since the binary detection of clouds from ground based images is an image classification problem many Machine learning (ML) techniques can be applied. One of the first approaches by Taravat et al. [37] already indicated improvements over traditional thresholding methods. Later convolutional neural networks (CNN) became the primary used binary cloud detection model, due to their capability of extracting different cloud features [34].

Other researches went on to distinguish the WMO defined cloud types 1 inside visible light images [33].

Fabel et al. [18] aimed to classify clouds into three separate height classes.

A big problem however remains the need for good ground truth data. There are always new ways being developed in the topic of self-, semi- and unsupervised learning [13, 18, 38] or other ways to artificially enhance already available datasets [39]. However up until now it was almost always unavoidable to label at least some of the data by hand, which is not only time consuming but also expensive and prone to human errors.

2.2 Cloud height estimation

Another important cloud parameter is the cloud base height. It is used to validate numerical weather prediction models [40] and climate models [41]. It is also of interest for air traffic controllers in aviation [42]. For solar irradiance Nowcasting applications cloud height is needed to accurately calculate the cloud shadow projection [43].

The most accurate cloud base height retrieval technique are Lidar systems also called ceilometers. They work with a light backscattering technique and yield the cloud base height for a single point accuracy of typically up to 30m. However for higher clouds this error can get bigger to up to +-1050 m [44].

Other techniques include cloud radars, stereo camera systems, numerical weather predictions or approximations from satellite imagery [45, 46]. Another approach is to obtain cloud height from infrared imagery. This has mainly been done using satellite imagery and cloud top height [47] but with the rise of ground based infrared imaging systems this can also be applied to approximate cloud base height [48, 49, 50].

2.3 Infrared sky imagery

Infrared imagery has been used in the past for cloud detection purposes and different researchers have considered using infrared images to obtain cloud base height information [49, 51, 14]. Using infrared imagery for cloud detection is promising as motivated in 1, but a clear distinction from clouds and clear sky is challenging. Fixed thresholds often fail due to hazy conditions, fogs or aerosols and make detecting thin cirrus clouds hard [52]. Other researchers opted to an adaptive thresholding strategy which is triggered using cloud lidar responses, which can detect high cirrus and aerosol levels [53]. This is an option for processing the data, where ceilometers are available, however this is not always the case.

Most other authors are using fixed empirical thresholds to detect cloud presence in infrared images [51]. Another approach is to use atmospheric radiative transfer models to calculate clear sky thresholds individually per image [54]. Similar to visible light cameras wind information can also be used in infrared imagery to distinguish different cloud layers [14]. With the rise of Deep learning methods they have also been applied with success to infrared imagery [15, 50]. Sommer et al. [15] have used a U-Net architecture to segment clouds in non-all-sky infrared images with high accuracy.

Just like in visible light images, most works used manually labeled data for training.

2.4 Previous works at DLR

The topic of Solar Nowcasting has been continuously researched at the Institute of Solar research at DLR in recent years. There have been multiple different approaches making use of all sky imaging systems in different ways.

The start of the series of works which this thesis builds upon was done by Yann Fabel in his Master thesis [55] which was later published [18]. In his work a semantic segmentation model based on the U-Net architecture was trained to classify images pixel-wise into four cloud height classes. Those layer classes were defined based on the WMO definitions [20] and Sassen and Wang [56] and slightly adapted. Table 1 gives an overview over those cloud height classes.

Within his work a hand labeled dataset containing 770 all sky images with pixel-wise cloud layer annotations was created which is used up till now to validate new approaches. The trained models achieve around 95% accuracy in binary cloud detection and a mean intersection over union (IoU) of 74.6% for cloud layer segmentation on the test set. Two enhanced pretraining methods were investigated which improved the mean IoU to over 80%. Some core issues in the topic have been recognized and addressed as the distinction of adjacent cloud layers due to their visual similarity. Oftentimes this occurred between mid and high layer classes, where both clouds contain a mix of water and ice particles.

Afterwards this was improved upon by David Magiera during his Bachelor's thesis [38, 35] who developed a semi-supervised learning approach in which ceilometer cloud height measurements are introduced into the training of a layer detection model. He created a dataset of over 47000 weakly annotated all sky images. The approach proved to be successful and yielded mean IoU of 80.52 %.

Raphael Gut then took it further by using time series information to propagate the original masks to enhance the dataset further [39]. The masks were propagated using video segmentation models to artificially enlarge the hand labeled dataset by a factor of 20. He showed that increasing the dataset like this can also increase model performance compared to the imageNet pretrained cloud layer segmentation model. The model achieved 79.4% mean IoU on the test set. Since images taken just a few seconds apart often hold almost the same information, this approach needs careful consideration

Class	Height Level	Cloud Genera	Characteristics
High-Layer	> 6 km	Cirrus	Detached, white, fibrous filaments, narrow bands
		Cirrocumulus	Thin, white, very small elements without shading
		Cirrostratus	Transparent, whitish cloud veil, fibrous or smooth
Mid-Layer	$1.8 - 8$ km	Altocumulus	White and/or gray, small elements with shading
		Altostratus	Grayish/Blueish layer, fibrous or uniform
Low-Layer	$0 - 2.4$ km	Cumulus	Sharp outlines, detached, developing vertically, dark bases
		Stratus	Fairly uniform gray base, no clear outlines
		Stratocumulus	Gray and/or whitish, visible outlines
		Cumulonimbus	Heavy dense, large vertical extent (over all layers)
		Nimbostratus	Gray (often darken) layer, can extend to high-layer

Table 1: Classification of Clouds by Height Level and Genera [55]

Input	Camera distance	MAE	RMSE	Bias
Metas-Kontas (Training)	879 m	**605.7 m**	**979.7 m**	**-150.7 m**
Metas-PVot	833 m	1648.2 m	2383.2 m	-469.8 m
Mono-cam PSA	-	1276.2 m	1677.5 m	-774.1 m
Mono-cam Oldenburg	-	3221.4 m	3635.2 m	-1881.8 m
Cross-Correlation	879 m	1370.8 m	2231.1 m	+186.8 m

Table 2: Error metrics for cloud base height estimation from [58]

which data is used for training and validation to not unwillingly take influence on the evaluation metrics.

Felix Häusler then implemented a fold split, optimizing the training-validation split during an internship. This concludes the recent works on the topic of semantic cloud segmentation.

The estimation of cloud base height is also researched extensively [57, 44]. Previously cloud base height was calculated using stereoscopic Cross-Correlation. Geometric shifts between different image sections and the knowledge of camera positions are utilized to calculate the base height.

The most recent thesis by Franka Weiler [58] had a look at CNNs with multiple input images, compared to a triangulation approach [57] and validated with ceilometer data. The trained CNN uses two camera RGB images as input and approximates the ceilometer measurement. Across all cloud heights the model reaches a mean absolute error (MAE) of 605.7 m and a root mean square error (RMSE) of 979.7 m with bias of -150.7 m. Generally speaking the model performs best in low-layer conditions with decreasing accuracy for higher clouds. The evaluation was also performed for different input images of which the results are summarized in Table 2.

Although this work improved upon the previous approach for a given site a good cloud base height estimation which is applicable to different conditions and different geographical locations is still to be developed. This is the goal of multiple ongoing projects including this one.

3 Dataset

This chapter gives an overview about the available data and the hardware used to acquire it. Afterwards it is explained how the data is processed and additionally filtered to obtain a high quality dataset. In current state of the art deep learning methods the accuracy of the model is highly dependent on the data quality. Therefore a lot of effort was put into preparing the dataset.

3.1 Data acquisition

The hardware and sensor equipment to acquire the data used for analysis and training the machine learning model is spread over two different locations. One site is the *Plataforma Solar de Amería* (PSA) located in Tabernas, south of Spain. It is one of the largest concentrating solar power research centers in the world and run by the *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas* (CIEMAT). Since it is located near the Tabernas desert it has very unique conditions with almost no precipitation. The other site is located in Trauen, Germany. It is part of the responsive space cluster competence center of DLR. Here the weather conditions are very different to the ones near Tabernas. This can be estimated by looking at the yearly precipitation for the two sites, being 772 mm for Trauen and 305 mm for Tabernas [59] [60]. Another difference is that Tabernas is located near the sea where specific large scale weather phenomena occur, which do not happen in Trauen. Additionally near Tabernas there are often air conditions with high levels of aerosols and other small particles present.

The main reasons for using such different sites is the availability of infrared cameras in combination with the goal of a resulting model that generalizes well and is applicable in different geographic locations.

The cameras used for this work include All-Sky-imaging systems for visible light and infrared radiation. The visible light cameras are *standard* security surveilance cameras from Mobotix, which fulfill the boundary condition of creating a cheap sky observation system. At the PSA the Mobotix Q25 is mounted, while in Trauen it is the Q71 model. Such a camera can be seen in Figure 3.

The long wave infrared cameras in use are managed and calibrated by Miratlas. They provide the images already converted to temperature. This lack of control proved challenging during the work and will be addressed in the next subsection and will also be part of the discussion in the end.

Table 3 shows different camera parameters for the infrared cameras.

Site	Camera type	Bandwidth	Precision	Radiometric calibration range
Tabernas	External	$8 - 14$ μm	$+/-2$ °C	-40 to 600°C
Trauen	Internal	$8 - 12$ μm	$+/-5$ °C	-5 to 60°C

Table 3: Camera specifications for the infrared cameras.

As already mentioned in 1, the temperature of the sky in the troposphere varies

Figure 3: Image of a Mobotix Q71 camera.

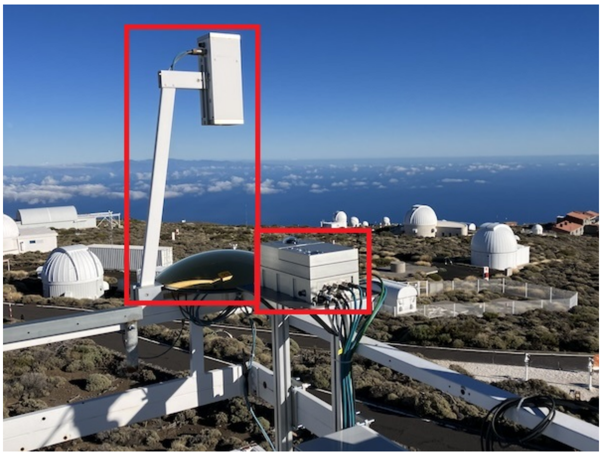

Figure 4: Foto of the external (left) and internal (right) infrared camera types. Image taken from [61].

between roughly -56 °C and 15 °C. Therefore the calibration range is not ideal for our application. Additionally the error in precision can produce errors of more than thousand meters when transferring temperature to height, depending on current weather conditions. A rough estimate of the influence on the final result is given in 5.

The camera type refers to the build version of the camera. External refers to a system where a camera is pointed down onto a gold coated mirror creating the effect of an effective LWIR ASI system with a field of view (FOV) of 180 degrees. This leads to a fairly big region in the center of the image in which no information about the sky is available due to the camera own reflection. The internal camera type has an fisheye lens built in, leading to a direct availability of a 180 degree FOV. Figure 4 shows the two cameras that have been used.

To validate the cloud height approximation a ceilometer is used. A ceilometer is a cloud height measuring device, that enables detecting different particles in the air using the laser distance measurement technique LiDAR. The available ceilometer CHM15k is located at the PSA and can measure the cloud base height of up to 3 different cloud layers and up to 15000 m height. It is based on a single wavelength backscattering technique. Additionally to measuring CBH it can also create aerosol backscatter profiles, measure precipitation, cloud cover and distinguish clouds from fog [62]. It is accurate up to 15 meters. The measurement is still to be considered with caution because the errors can increase with altitude as mentioned in 2. However since it is the most accurate measurement available and is commonly used as the *true* cloud base height, its response value will be assumed to be true in this work. This distinction into clouds and aerosols will also be part of the analysis later, and can provide challenging conditions.

3.1.1 Camera calibration

The visible light cameras are setup, such that images are taken in a frequency of 30 second time frames. They take images with different exposure times, to make detection of clouds possible at different times of the day and for diverse weather conditions. The available exposure times range from $80\mu s$ up to $1280\mu s$. In the past the works done on this topic focused on one fixed exposure time. In this work the option of combining multiple exposure times to one single pseudo hyperdynamic range (HDR) image is explored as well and will be discussed further later in this chapter.

The cameras also take images at night with higher exposure times, however those images were excluded from this work. Since the main application of the model in development is meant to be for solar forecasting applications, night time images do not provide additional information, and make the processing even more time intensive, due to the different image format.

Unfortunately the infrared cameras are setup to take images only every five minutes, and even then not perfectly regular, making it impossible to match some images with the ones captured from the visible light cameras. The low sampling rate also prevents the use of cloud movement information between images. The option of using those cloud shifts in between images to detect and segment clouds in normal and infrared imagery to distinguish between different clouds is to be explored in the future, when a higher sampling rate is available. Other authors partly used similar information successfully, when higher capturing frequency was available [50] [14].

The temporal *unalingment* is different for the two sites of interest. In Trauen the LWIR camera takes pictures with an exact frequency of one image every five minutes, but the timestamp is not aligned with the visible light camera. It it offset by about ten seconds.

At the PSA the frequency varies and is not set to exactly one image every five minutes, leading to an offset of up to 15 seconds.

Additionally the PSA infrared camera is placed about 250 meters apart from the visible light camera, making additional transformations necessary to create valid ground truth masks.

The sensor of both infrared cameras is an uncooled microbolometer. The cameras are then calibrated on 4 different temperatures on a hemispherical blackbody. A shutter is used for non-uniformity correction.

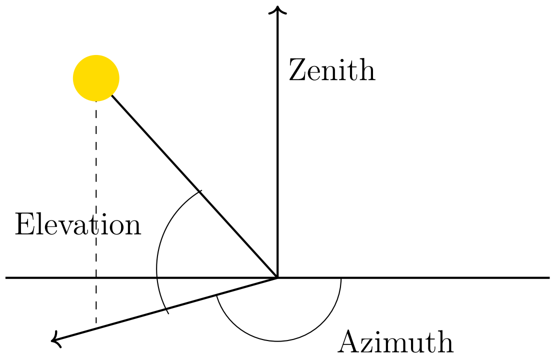

Figure 5: Angle definitions

The target emissivity used to obtain temperature values is $\varepsilon \in [0.95, 1]$. The exact value and how it is obtained is not shared with the author. Emissivity can then be used with the Planck Radiation law which has been introduced in 1. This results in the parameters mentioned in Table 3.

Since the calibration is done by Miratlas and the hardware is originally from a french defense subcontractor, there is no further information available on the calibration. Unfortunately this absence of information limits some processing and validation options.

In ASI systems the images are strongly distorted due to the high FOV. Hence to work with the images, there is also a software-sided calibration required.

This is done using a code created at DLR, which makes use of the position of the sun and moon inside the images to calculate the azimuth and elevation matrices of high FOV all sky images. Those matrices provide pixel-wise information, to which position in the sky a pixel belongs. Figure 5 shows the definition of those angles.

In the past this code has only been used on visible light images, which work with multiple color channels. The calibration parameters for the visible light cameras in use are already available, since they have been calibrated in the past.

To apply it on the infrared images, multiple challenges have to be overcome. First of all, the moon is not visible in infrared imagery because it does not emit or reflect enough radiation in this spectrum. Additionally the sun appears smaller in infrared images and does not produce scattering effects, making it hard to detect it sometimes.

Finally, the infrared cameras where only mounted end of 2025, the one in Trauen in October, the one at PSA produces correctly calibrated images since December. Therefore only low elevation angles are represented in the data and sun elevation from spring and summer times are missing, which complicates the process further.

Figure 6 shows a plot of the detected and theoretical sun positions after the calibration has been successfully applied on the camera at PSA.

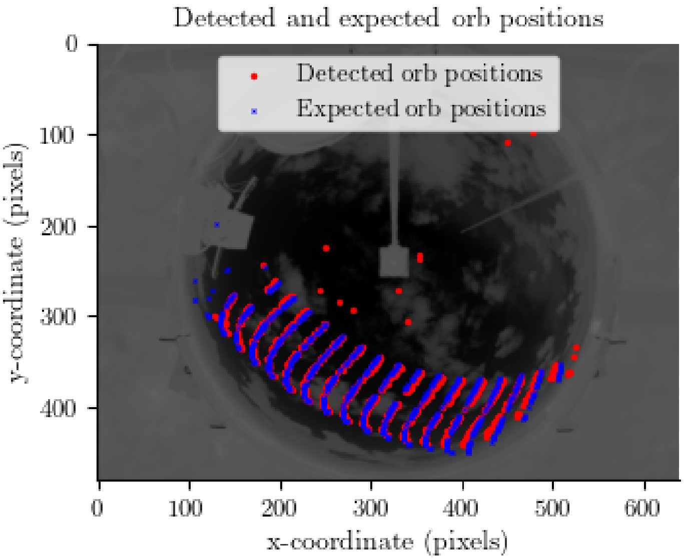

Figure 6: Detected and theoretical sun positions for camera calibration between October 2025 and March 2026 at PSA

3.2 Preprocessing regression

Now that the *pixel to sky* relation is defined for the images, further processing can be applied. This includes the handling of clear sky radiation, the conversion from temperature values to height, the transformation of camera frames and image filtering based on boundary conditions.

To start of, obstacle camera masks are defined. These masks are used to remove regions from the image that are not of interest. For the external infrared camera at PSA this mask includes a big part in the center of the image, due to the above mentioned camera structure. Additionally a zenith cropping angle $\theta_{crop} = 80°$ is defined. The zenith angle resolution is very high for high zenith angles [63]. For those regions there applies very strong view distortion. Another reason for the usage of angle cropping is that the clear sky subtraction in the next chapter works as a function of airmass which increases drastically for high zenith angles $\theta > 80°$. This makes this region prone to higher errors.

Figure 7 shows a raw image and the cropped and masked result.

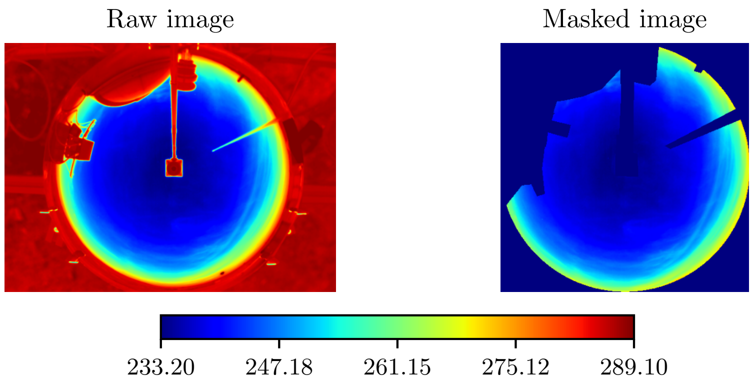

Figure 7: Raw image and masked image

From this point onwards the masked image can be referred to as the raw image.

3.2.1 Clear sky subtraction

In section 1 it was introduced that the atmosphere itself emits radiation in the long wave infrared spectrum. In section 2 different state of the art ways for calculating or approximating the clear sky radiation were mentioned.

To tackle the elevation dependance of the clear sky radiation and enhance the image for further processing Klebe et al. [17] ran multiple MODTRAN simulations to obtain the clear-sky radiance as a function of PWV. They then showed that this clear sky radiation is almost identical to the lower envelope of an infrared sky image in normalized radiance units for standard atmospheric conditions. More accurately, the lower envelope of the normalized radiance data per airmass coincides really well with their function of PWV and airmass. They state that their clear-sky subtracted images are independent of the conducted simulations, by just using the lower envelope of the normalized radiance data. Additionally they introduced an offset term which depends on the radiation environment on the imaging system.

In this work the approach is adapted and slightly modified.

Since the provided images are already given in physical temperature values, it is not possible to recover the radiance response of the camera without knowing the exact calibration parameters. A normalized radiance image can only be obtained by assuming the emissivity $\varepsilon = 1$ and using Planck's radiation law (1.1) to obtain a spectral emitted radiance, which can then be normalized using black body radiation for a reference temperature T_{ref}. After the conversion, a function can be fitted to the lower envelope of the data. This function is then subtracted from the normalized radiance and converted back to Temperature.

However since those calibration parameters are unknown their approach is slightly changed while keeping the core idea the same. The images are already calibrated in temperature values and since the clear sky radiance can be obtained from the image envelope, it is not necessary to transfer it to normalized radiance.

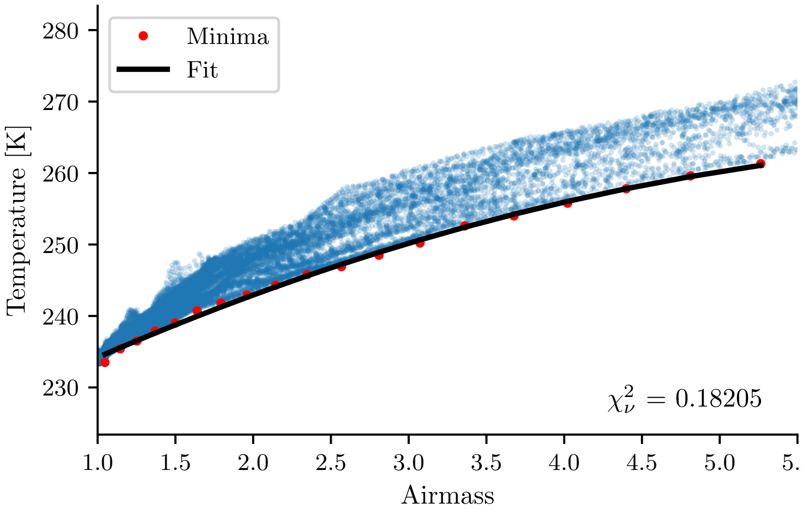

Figure 8: Pixel temperature vs. airmass data for December 09, 2025 16:13 UTC. Black line is 2nd-order polynomial fit to lower envelope of the minima (red)

If the envelope fitting is applied to the temperature image directly it is a similar representation, just in the form of temperature.

The only downside to this is that it is hard to embed a physical meaning to it. Since the images are provided in physical temperature, the lower envelope is in the order of roughly 250 K. A direct subtraction does therefore not work and makes the temperature values incorrect. Instead the fitted second order polynomial $p_{fit}(AM(\theta))$ is moved such that $p_{fit}(AM(\theta) = 1) = 0$ K and the temperature at zenith is not modified. This is important to be consistent with the calibration of the camera which provides physical temperature at zenith. Hence only the change of temperature with airmass $AM(\theta)$ is of interest and can be correctly modeled using the described approach. The airmass was modeled using the formula by Kasten and Young [64]

$$AM(\theta) = \frac{1}{\cos\theta + a(b + 90^\circ - \theta)^{-c}}, \tag{3.1}$$

where θ is the zenith angle, and $a = 0.50572$, $b = 6.07995^\circ$ and $c = 1.6364$ are empirically fit parameters. Figure 8 shows the fitted function to the lower envelope of $T(AM(\theta))$. Figure 9 shows an example image before and after the clear sky is subtracted. It also shows why a clear sky subtraction is necessary and how big of an impact it has on the incoming radiation. The impact or ratio of the clear sky radiation changes highly in between days, sometimes even in a few hours, depending also highly on the weather conditions themselves.

The clear sky subtraction proves to be most challenging for very cloudy days, because, the lower envelope approach assumes that for each airmass bin, there is a clear sky pixel available. If the sky is covered in clouds this is not the case. Klebe et al. proposed the usage of a χ^2 goodness of fit criterion (3.2) [65] to ensure a well fit polynomial, where $p_{fit}(x_i)$ is the function to be tested at discrete points i, ν is the degree of the polynomial

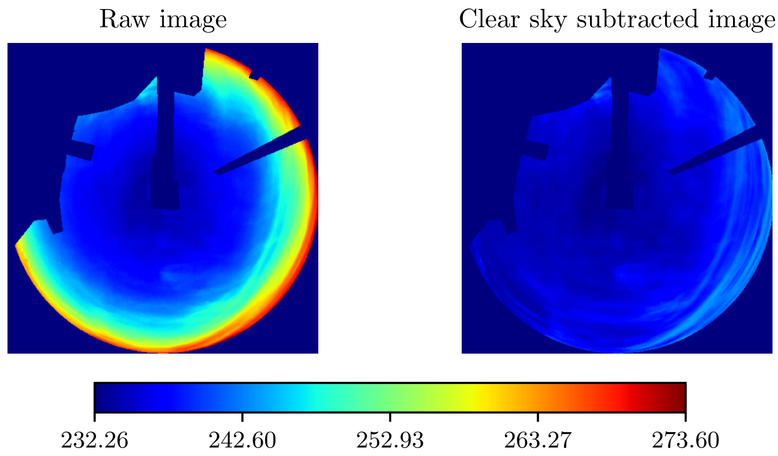

Figure 9: Raw temperature image vs clear sky subtracted image for December 09, 2025 16:13 UTC. The camera mask is set to the image minima for visualization purposes.

and σ is the standard deviation.

$$\chi^2_\nu = \frac{1}{\nu} \sum_i \frac{(y_i - p_{fit}(x_i))^2}{\sigma_i^2} \overset{!}{\leq} 0.05 \tag{3.2}$$

If this criterion is under a predefined threshold the fitted polynomial is subtracted, if not the previous polynomial equation is used. This is also incorporated in this work. Additionally to that criterion another criterion is used which discards an image if the previous fit is also not fitting well. This reduces the size of the dataset slightly, but prevents nonphysical temperature values. Since in between days, there are no RGB images available, this processing is for now only applied to daytime images. This leaves a data-gap of a few hours each night, hence it does not make sense to start a new day using a polynomial criterion from the previous image which was taken multiple hours before. Therefore a third criterion is added where a weaker fit is allowed if no better option is available. In that case a polynomial is fitted to only the bins for the higher airmasses, because they have a higher probability of containing clear-sky pixels. The fitted function is then evaluated on the whole lower envelope using a similar χ^2 goodness of fit criterion. This prevents the discarding of images in early-day hours.

Now that the elevation dependent clear sky radiation has been removed successfully from the images, the pixel values correspond to actual sky temperature everywhere. In the following it will be discussed how we can use that temperature information to obtain knowledge about the cloud base height.

3.2.2 Temperature to height

As mentioned in section 1 the best option to acquire a relationship between temperature and height is the use of a radiosonde. This is unfortunately not available at the locations where the cameras are located. Hence primarily, different linear lapse rates have been investigated for simplicity. Other authors have used formulas for a moist adiabatic lapse rate based on atmospheric pressure, temperature and the dew point [50] to improve accuracy. Since basic ground based meteo data is available, this approach has also been investigated. However since the PSA is located between mountains and near the sea in a very hot and dry region there exists very complex winds and weather phenomena, that also influence the temperature height relations.

Therefore for this work ERA5 reanalysis data on pressure levels is used [25, 66] for a best possible approximation of the true relation. The Climate Data Store from Copernicus Climate Change Service provides many meteorological variables hourly from 1940 until today. The data is gridded on a regular latitude-longitude grid and has a resolution of $0.25° \times 0.25°$ which is about 30km $\times$ 30km. The vertical resolution ranges from 1000 hPa to 1 hPa in 37 different pressure levels.

The provided variables include fluid dynamics parameters like wind velocity vector components, vorticity and divergence, and meteorological variables like geopotential, humidity, temperature and more.

In this work only the geopotential Φ and temperature T data is used. Those variables are close to the true values and have been validated for the eastern Mediterranean region by Hassan et al. [67]. The values are assumed to be accurate to a similar degree for southern Spain and Germany. The geopotential data is used to obtain a value for the geopotential height z above sea level $z = \frac{\Phi}{g}$ where $g = 9.80665 \frac{\mathrm{m}}{\mathrm{s}^2}$.

The data is interpolated bilinearly in between grid points to match the sites locations. Additionally the data is also interpolated in time to match the exact timestamps of the images.

Figure 10 shows the height plotted over temperature for the environmental lapse rate and the ERA5 reanalysis data for an example day in December.

For the chosen timestamp the assumption of a constant lapse rate is reasonable. However this is not always the case.

Furthermore there exists the phenomenon of temperature inversion in the troposphere. While temperature typically decreases with increasing altitude, an inversion occurs when a layer of warmer air overlies a layer of cooler air. In such a special case, the Temperature to height mapping is not well-defined anymore, because a single temperature can yield multiple heights, hence the mapping is not injective. However those conditions are very rare at the observation sites. Additionally they occur mostly in the highest layers of the troposphere, where infrared imagery is fairly limited anyways. If such conditions appear the lower of the two heights is used for the mapping.

Now that a temperature height relation is defined, the image can be transformed into height space.

Figure 11 shows the previously obtained clear sky subtracted image and the same image converted into height values for each pixel.

Two important remarks to the data have to be added. First, the presented height conversion assumes that the measured IR temperature corresponds to cloud base height. This is generally not true in two conditions. As previously mentioned the infrared image assigns the temperature where the cloud becomes optically thick, whilst the ceilometer

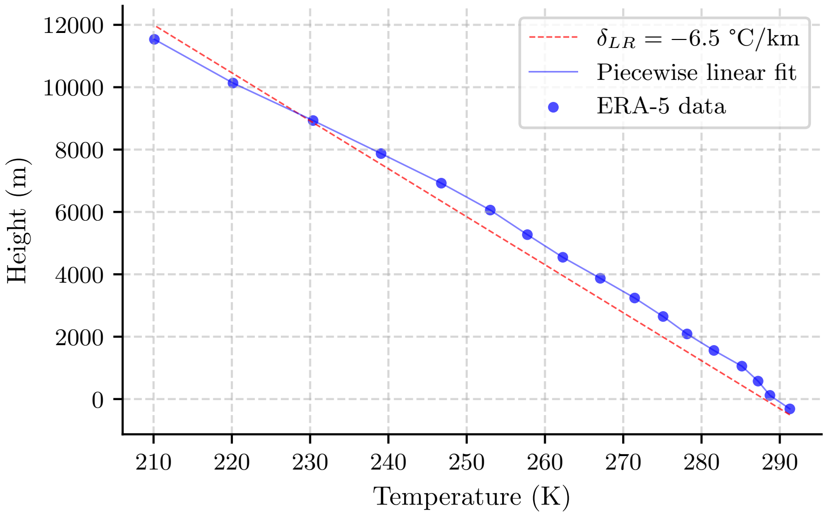

Figure 10: Height vs. temperature for December 09, 2025 16:13 UTC. For the environmental lapse rate a reference temperature of 288K is used.

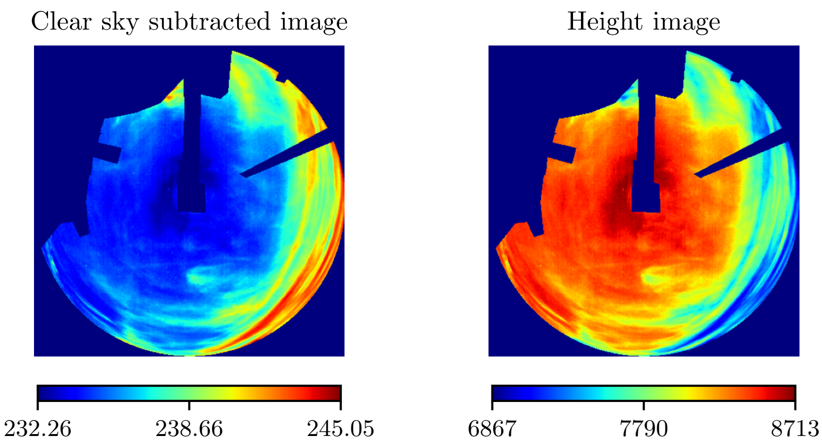

Figure 11: Clear sky subtracted image (left) and height image (right) for December 09, 2025 16:13 UTC.

measures the first point of contact. Additionally the clouds which are present at high elevation angles far from zenith are actually photographed from the side. Meaning the temperature is more representative for cloud side than it is for cloud base. Hence the assigned temperature is not really cloud base height. For example a big cumulus cloud photographed from the side ranges through multiple temperature layers. It is therefore possible that multiple temperatures get measured for a single cloud. As a consequence the conversion to height then does not get the cloud base height, but pixel-wise cloud height. This will be observed, especially for low layer clouds, which often range into the mid-layer.

Second, the way the infrared cameras are calibrated, they always return a temperature value hence the conversion to height always yields a height value for each pixel. The distinction of cloud and clear sky therefore becomes more complicated than in normal RGB imagery, where the in section 2 introduced criteria can be used. The chosen approach is explained after the camera frame transformation.

Height- validation Since the ceilometer measures the first contact points with water particles, the measurements are sensitive to sudden changes, like very small holes inside the clouds or single particles in the atmosphere on clear condition days. The lower spatial resolution of the infrared imagery cannot detect small artifacts like this, which is why the data is filtered for one layer conditions before comparing to image data. To filter the ceilometer data, a ten minute window is defined. If two measurements inside this time window vary by more than 1% with respect to the previous measurement, the complete time window is discarded. This ensures that no small holes or random lower clouds or haze are detected and makes the measurements comparable to the height maps obtained by the infrared images. A mean of the ceilometer measurement inside this time window is taken. An algorithm then searches for the closest infrared image timestamp to the middle of that time window.

For the five available months of data, 1169 timestamps satisfy this condition and are used for the comparison. Those days still include clear sky days which were excluded for the height validation. To obtain a single comparison value from the infrared image the ceilometer measurement is mapped into the infrared camera frame using its location and the respective height measurement. Unfortunately except for very low cloud conditions the sky partition measured by the ceilometer lies behind the obstacle camera mask. This makes a direct comparison impossible. To circumvent this issue the closest pixel to the mapped ceilometer coordinate is evaluated. This is not ideal, but since the timestamps are filtered for one layer conditions, this approximation is valid and should not vary much from the actual value.

Figure 12 shows the ceilometer measurements on the x- axis and the heights for the closest non-masked pixel obtained from infrared image on the y-axis.

It is visible that the data does not fit well to the reference line. To improve the models performance the data is calibrated to better fit the ceilometer measurements. The layer-wise linear polynomial fits in Figure 12 overlap nicely with a 2nd order polynomial fit $g(x)$ through all the data-points. Hence this polynomial is chosen to calibrate the system.

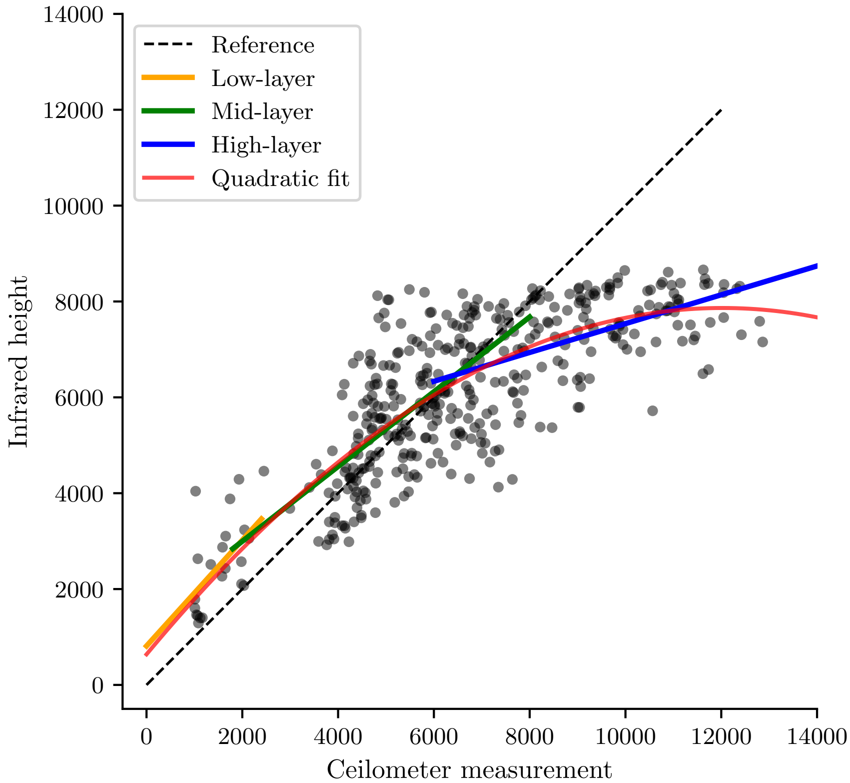

Figure 12: Calibration curve: Ceilometer measurements vs Infrared image height. Black line is the reference, the colored lines show a linear fit for the different layers, Red line is a quadratic polynomial fit

	MAE	RMSE	Bias
Before	1298.97	1716.76	-554.51
After	1213.54	1587.69	-38.87

Table 4: MAE and bias before and after calibration.

The data-points are transformed using the inverse of the fitted polynomial $y\prime = g^{-1}(x)$ to reduce the bias and MAE. The values before and after the calibration are shown in Table 4 However this transformation can induce negative height values which is non-physical. Therefore a correction is implemented, that maps all negative values from its negative minimum $g(0)$ up to a predefined ramp up threshold in this case arbitrarily chosen to be the end of the low-layer height $h = 2400$ to the interval $[0, 2400]$.

This prevents negative heights and still corrects most of the data to fit a better relation. The applied function is then applied to all the image data.

Visualizing the data like this also raises a few concerns. For once the mean absolute error is 1298.97 meters before the correction and 1213.54 meters afterwards and a quite big spread can be observed in the mid layer range. This comes from error accumulation up to this point. Starting with measurement sensor errors of $+/-2°$C which in mid-layer region often translates to a height error of roughly 500 meters. This overshoot in mid-layer might also be explained, by the fact that infrared imagery assigns the temperature to the cloud at the point where it is optically thick while the ceilometer measures the first point of contact. Low layer clouds are often filled with water and therefore there is rarely a big difference between first contact point and optical opaque regions. This changes for mid and high layer clouds which statistically speaking transmit more light. However from ceilometer measures of 6000 meters and up most of the infrared image heights are underestimations. Part of this might come from the nature of high layer clouds and that they are hardly detected in infrared imagery in general. The graph also shows that the infrared height rarely reaches values above 8000 meters. This might be a calibration issue because clouds at that height reach temperatures below $-40°$C to which the camera is not calibrated. Another explanation is the aggregation of dust and aerosols inside the atmosphere up to those heights. Since infrared imagery is vulnerable to hazy conditions, the accumulation of molecules might also cause this limit.

If a ceilometer is available, it might be the best approach to feed the ceilometer height as additional input into the machine learning model, to not assume some polynomial relations. The model could then learn a nonlinear relation itself. Nonetheless the large spread indicates, that there will always be a quite big error range through sensor and assumption errors.

Although the non-linear data correlation raises concerns, ML models are known to generalize remarkably well. Since the errors are not simply additive white noise and are instead linked to the intensity following a non-linear trend, traditional bias correction alone is insufficient [68]. A ML model however can effectively map the noisy inputs to the underlying physical trend [69]. Furthermore the inherent measurement uncertainty may act as a natural regularizer, preventing model overfitting and encouraging the learning

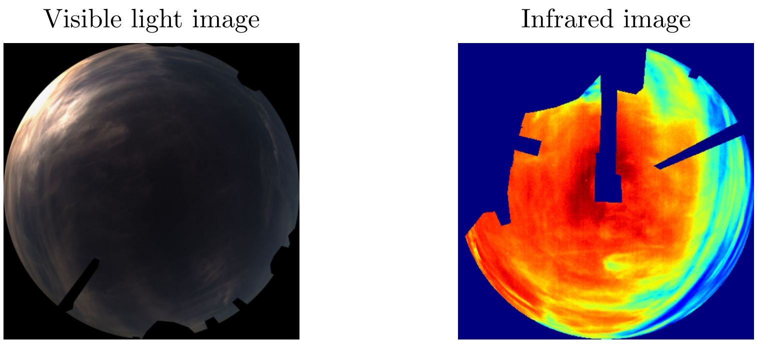

Figure 13: Visible light camera image (left) and from the infrared image acquired height map (right) for December 09, 2025 16:13 UTC.

of more general trends.

In addition the layers defined in Table 1 have a high range of multiple thousands of meters, suggesting that the model output can be transformed to useful layer maps.

It has to be added, that this validation/ calibration is suboptimal, because there is always only a single pixel value compared to a single measurement for a total of only 1169 images. Therefore the process is very vulnerable to faulty images and uncertainties which for now can not be ignored, due to limited data. This process should be repeated once more data has been gathered and additional filtering can be applied.

In the next step a camera transformation is needed.

3.2.3 Camera transformation

To use the images obtained from the infrared cameras as ground truth for a ML model, the cloud structures need to match well inside the images. This is not the case without any transformations. Figure 13 shows a raw image at the same time stamp for the infrared camera and the visible light camera at PSA.

Without any proper adjustments it is hard to tell which structure corresponds to which, due to different camera domains. At the PSA the cameras are located roughly 260 meters apart from each other, making it necessary to account for parallax.

Therefore to match the structures inside the images of an infrared camera and a visible light camera a set of rotations and translations have to be applied.

To achieve this a mapping $\mathcal{F}$ from the visible light camera domain Ω_{vis} to the infrared camera domain Ω_{IR} is needed. Pixel coordinates in a visible light image are given by the pixel coordinates in azimuth $\psi_{vis} \in [0, 2\pi)$ and elevation angle $\phi_{vis} \in (0, \frac{\pi}{2}]$. To start of, the camera coordinates are converted from latitude and longitude to the geocentric coordinate system. The cameras position is then given as $\mathbf{P}_{vis}(x, y, z)$. In the next step the azimuth and elevation matrices from the calibration procedure are used to obtain a unit directional vector $\mathbf{v}_{vis}(\phi_{vis}, \psi_{vis})$, that points to the information that is present in a certain pixel from the cameras position. This vector is given by

$$\mathbf{v}_{vis}(\phi_{vis}, \psi_{vis}) = \begin{pmatrix} \cos\phi_{vis}\sin\psi_{vis} \\ \cos\phi_{vis}\cos\psi_{vis} \\ \sin\phi_{vis} \end{pmatrix}. \tag{3.3}$$

This is then normed by dividing with the Frobenius norm $\|\mathbf{v}(\phi_{vis}, \psi_{vis})\|$ to obtain a unit vector. To transform the actual pixel values, we use the height information given by the infrared image conversion in the previous step. This is essential such that the mapping is deterministic. Since the height information $h(\phi_{vis}, \psi_{vis})$ is only dependent on azimuth and elevation angles with respect to where the cloud lies in the image, but not the actual height information, it is only preserved correctly if $h(\phi_{vis}, \psi_{vis})$ is divided by the z-component of $\mathbf{v}_{vis}(\phi_{vis}, \psi_{vis})$. The whole information is then preserved in the following ray equation

$$\mathbf{X}_{vis} = \mathbf{P}_{vis} + \frac{h}{v_z}\frac{\mathbf{v}_{vis}}{\|\mathbf{v}_{vis}\|}. \tag{3.4}$$

Arguments are omitted for visibility. The vector $\mathbf{X}_{vis}(\phi_{vis}, \psi_{vis})$ now describes the cloud base heights obtained from the infrared image in world coordinates. Afterwards the position is adjusted to the infrared cameras location and transformed into the azimuth elevation space of that camera. This happens by adding the translation vector and applying a *world* to *camera* space transformation. This transformation looks as follows

$$\mathbf{r}_{IR} = \mathbf{X}_{vis} - \mathbf{P}_{IR}\ , \tag{3.5}$$

$$\mathbf{v}_{IR} = \frac{\mathbf{r}_{IR}}{\|\mathbf{r}_{IR}\|}\ , \tag{3.6}$$

$$\psi_{IR} = \arctan 2\left(\frac{v_{IR,x}}{v_{IR,y}}\right)\ , \tag{3.7}$$

$$\phi_{IR} = \arcsin\big(v_{IR,z}\big). \tag{3.8}$$

This defines the geometric projection $\mathcal{F} : (\psi_{vis}, \phi_{vis}) \to (\psi_{IR,target}, \phi_{IR,target})$, which provides the ideal angular coordinates in the infrared camera domain Ω_{IR} that correspond to a specific pixel in the visible camera domain Ω_{vis}.

However, because the infrared image is composed of a discrete grid of pixels with non-linear spacing due to lens distortion, $\mathcal{F}$ cannot be used for direct pixel indexing. Instead, a set $\mathcal{S}_{IR}$ is defined containing the angular coordinates (ψ_p, ϕ_p) of all valid pixels p in the infrared image domain Ω_{IR}:

$$\mathcal{S}_{IR} = \{(\psi_p, \phi_p) \mid p \in \text{Pixels}_{IR}, \text{mask}(p) = \text{True}\} \tag{3.9}$$

To retrieve the actual height value h_{IR}, a nearest-neighbor optimization problem is solved for every projected point. For each pixel in Ω_{vis}, we identify the pixel index p^* that minimizes the Euclidean distance in angular space:

$$p^* = \arg\min_{p\in\mathcal{S}_{IR}} \sqrt{(\psi_{IR,target} - \psi_p)^2 + (\phi_{IR,target} - \phi_p)^2} \tag{3.10}$$

To ensure the validity of the match and prevent the mapping of points that fall outside the infrared sensor field of view, a distance threshold ϵ is applied. The final mapped height value h_{vis} in the visible domain is thus constructed as:

$$h_{vis}(\psi_{vis}, \phi_{vis}) = \begin{cases} h_{IR}(p^*) & \text{if } \min_{p \in \mathcal{S}_{IR}} \|(\psi_{IR,target}, \phi_{IR,target}) - (\psi_p, \phi_p)\|_2 < \epsilon, \\ \text{NaN} & \text{otherwise} \end{cases} \tag{3.11}$$

This approach ensures that the parallax transformation is robust to lens distortion and non-linear pixel mapping, as it relies on angular proximity in the physical domain rather than linear interpolation in the pixel coordinate space.

To smoothen the result pixel defects are in-painted and binary erosion is applied to the transformed camera object mask. Afterwards the mask is combined with the obstacle mask of the visible light camera. Figure 14 shows the image before and after the successful transformation.

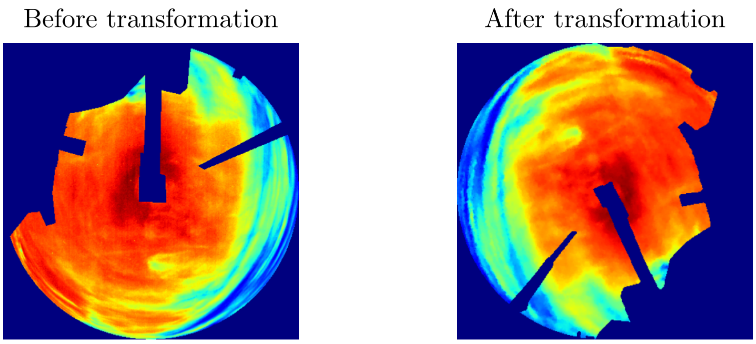

Figure 14: Image in the original LWIR camera space (left) and the same image transformed into the visible light camera space (right) for December 09, 2025 16:13 UTC

Figure 15 shows a flow chart of the important transformation steps.

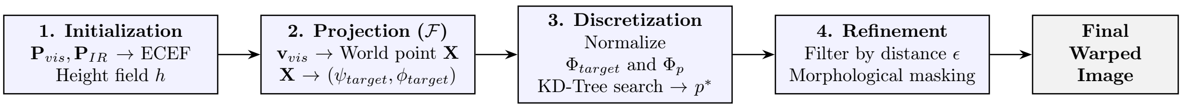

Figure 15: Flowchart of the image transformation

Now that both images have a representation in the same domain, the structures of the clouds can be matched nicely.

Clear sky distinction To differentiate between clouds and sky a distinction method is needed. Many different techniques were analyzed and evaluated ranging from fixed thresholds, over dynamic histogram thresholding and many more. Since Miratlas also provides some meteorological quantities in there Monitoring systems, a more physical approach is investigated, which uses the PWV to calculate a threshold. PWV is directly related to downwelling infrared radiance as described in section 1. Liu et al. [53, 70] describe the relationship mathematically and suggest its use as a clear sky threshold. Miratlas already provides downwelling infrared radiance as a parameter for each timestep. The downwelling radiance can be transformed into an approximated threshold temperature using Stefan-Boltzmann law $T = (\frac{L}{\sigma})^{1/4}$ which can then again be transformed into a corresponding height value. Everything below that height is identified as cloud, while every pixel value above it is identified as cloudless. This approach gives reasonable results for some days, however for the majority of days this thresholding strategy did not provide satisfactory thresholds. Oftentimes the temperature obtained from the conversion is too high and more detail about the provided variables would be needed. However those details are not shared from Miratlas in the scope of this work. This distinction is not accurate enough to train a ML model.

The most accurate solution found for this challenging distinction is to use the current best segmentation model on the visible light images. The binary segmentation already yields very good results with an accuracy of 0.949% [39]. However to not inherit too many errors from the old model, the results are used to calculate a clear sky threshold for the infrared obtained height maps, instead of using the binary map directly. To do this the clear sky regions from the binary segmentation are evaluated on the infrared image. Depending on the size of the region, the lower percentiles of this region are used as a clear sky threshold for the height map. This percentile approach is necessary to avoid clouds that might be inside those regions due to the offset in the infrared data to impact the threshold negatively. This thresholding approach turns out to work really well and also helps detect high layer cloud one layer conditions, that the pure infrared image does not distinguish from clear sky on its own. Figure 16 shows an example image of the thresholding technique.

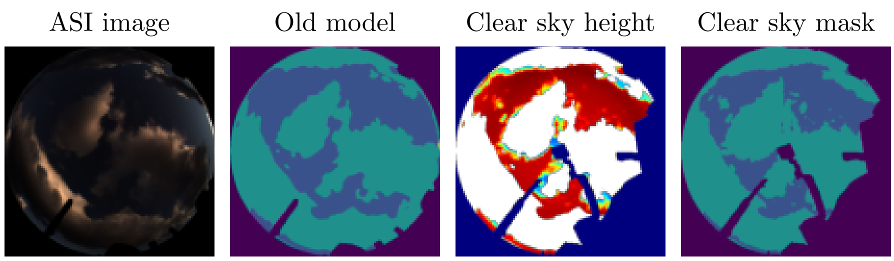

Figure 16: Demonstration of the clear sky masking using the old model. From left to right: Input ASI image, segmentation result from current RGB model, height in clear sky masked region, obtained clear sky mask using percentile thresholding.

One drawback of this approach is the need for a segmentable RGB image. Currently only RGB images with enough lighting can be accurately segmented making a clear sky distinction during night impossible. However this motivates the need for an IR segmentation model even further.

Not all images provide good enough information to identify the same clouds in RGB and infrared accurately. Why this is the case and how the data is filtered and adjusted accordingly is explained in the next chapter.

3.3 Data distribution and filters

The performance of a ML model is directly dependent on the quality of the dataset that is used to train the model. Before the details of the ML model are discussed an overview of the dataset is given.

Multiple models were trained using either images for just one of the cameras or both of them combined. The height-distribution of the different site datasets is displayed in Figure 17.

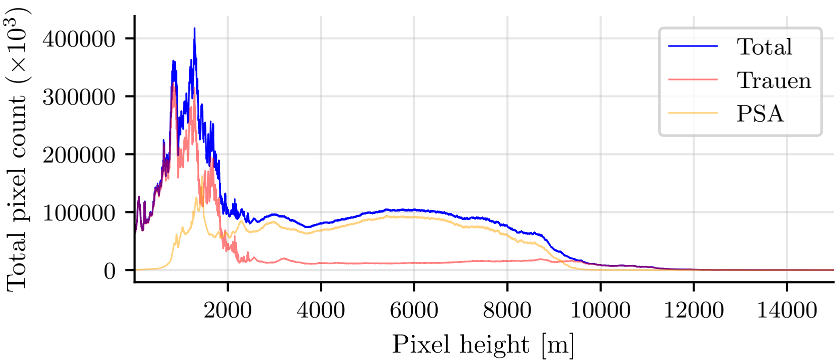

Figure 17: Pixel-wise height distribution of the dataset without brightness filter. All images combined (blue), at Trauen (red) and at PSA (orange).

It can be seen that the low layer peak up to 2000 m is almost solely from the images of the Trauen camera in Germany. This influences the performance of the different model runs when it comes to detecting different cloud layers which will be seen in the evaluation.

Ideally the dataset would be more evenly distributed. Visual observation of the dataset shows, that a lot of the images are very dark, due to a combination of low sun elevation and low cloud conditions in Germany. Those images do not provide a lot of useful information, hence they can be filtered from the dataset. This is done using a simple mean intensity threshold on the visible light camera images. However this brightness filtering removes a high percentage of the Trauen images, indicating that more data from summer months is needed to represent the Trauen site meaningfully. As a consequence getting useful predictions for sites further North is challenging in winter months.

A possible solution to this is described in the upcoming section.

To enhance the dataset further *bad* images are extracted and removed by multiple filtering operations. The first filter is already mentioned in 3 which filters out data where the elevation dependence of the clear sky irradiation can not be estimated accurately.

A big problem inherent in how the data is captured from Miratlas is that the timestamps do not align with the ones of the visible light cameras. Naturally clouds move in the atmosphere with different speeds that are mostly unknown. This bad synchronization makes a lot of images have non matching cloud borders, because the closest visible light image timestamp is up to 15 seconds apart from the infrared image.

Figure 18 shows an example of such a timestamp. For visualization purposes a third image is added where a height threshold is applied.

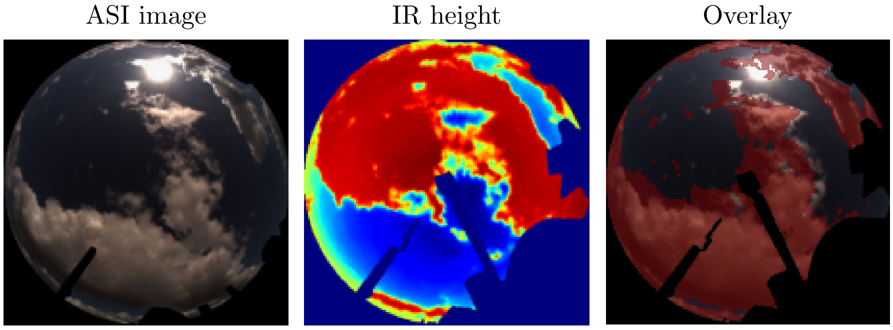

Figure 18: (Visible light image (left), normalized height map obtained from infrared image (middle) and a threshold overlay to visualize the difference for December 18, 2025 12:02 UTC

As an attempt to improve the data quality multiple models will be trained, some with the following filters applied. One approach for filtering is to simply only consider images for which the timestamps align well and the difference is under a predefined threshold in this case chosen to be 3 seconds: $|\mathrm{t}_{IR} - \mathrm{t}_{vis}| \leq 3\mathrm{s}$. This is only applicable to the images obtained at PSA, since the images in Trauen have an almost constant offset of $\Delta \mathrm{t} = 10\mathrm{s}$. Applying this reduces the number of usable timestamps from PSA drastically from 7616 to only 1765. A different approach to filter these images is to imply wind based conditions and use the available ceilometer data. The Miratlas sky observation system provides ground measurements for wind speed. Those are used with fixed thresholds to throw away images in which ground wind speed and the time difference between the images are high. In reality the wind speed at the ground is not necessarily comparable to the one at a clouds height, but it is the best approximation available. Additionally those thresholds are adjusted if the cloud height measured by the ceilometer is low, because the movement of low layer clouds is more visible in the image than that of clouds that are located higher. Since there is no ceilometer available in Trauen and the timestamps are offset by almost constantly 10 seconds the filtering is applied relying on the ground measured windspeed only, not incorporating cloud height or temporal offset.

Because the images use different channels it is in general challenging to find suitable metrics to decide if a shift is present or not. Some approaches to reduce shifts (e.g. Optical flow applied to cloud edges) have been investigated, but did not always improve the positions as desired. Hence the evaluation of the thresholds is done visually by manually checking examples of the dataset.

A combination of brightness and wind filtering reduces the dataset size from 3429 available infrared images to 609 usable images for Trauen and from 7616 to 3792 for the site at PSA.

Additionally a few images do not have good quality due to different anomalies captured by the cameras as introduced in 1. Those range from waterdrops on the lense in rainy conditions, snow or ice blocking the lense in winter, over soiling through dirt or sand up to people, animals or insects blocking part of the viewing field. Further there are camera errors that happen fairly frequently and make the affected images unsuitable. This is problematic because those objects are not present on both cameras, hence the *ground truth* obtained from the infrared image is not comparable to the input image from the visible light camera.

Figure 19 shows examples for present anomalies and errors in the infrared camera.

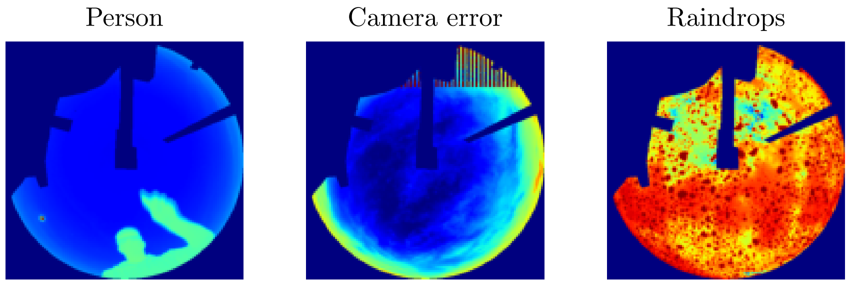

Figure 19: Example images of captured anomalies in infrared cameras, a person (left), an error (middle) and raindrops (right).

The detection of such anomalies is in general challenging as well, especially in one channel images where anomalies can look very similar to clouds. However in recent years good object detection systems and models have been developed, that can detect predefined objects with high confidence [71, 72]. In this work a YOLO11 [73] nano model is trained on close to 400 manually labeled images per camera. It is trained to detect covered lenses, false images with camera errors and a spot class is added which prevents misclassifications due to camera spots that are not weather dependent. The YOLO model performs reasonably well and detects 67 % of covered lenses correctly sometimes confusing them with the background. Camera errors are detected 80 % of the time. It is assumed that this filtering improves the model quality a lot more than the additional few misclassified images would since a lot of unusable faulty images are excluded.

Note that there are some additional enhancements to be made for the dataset in the future once there is more data available. For example the data should represent different sun elevation angles spread all over the year. Furthermore some weather phenomena are maybe only present in specific times of the year which are now not present or overrepresented in the data.

3.4 Input images

The model input images are the visible light All sky camera images. In previous works a fixed exposure time was used for the input images. This was set to perform better than varying the exposure times between images [55]. However if different sites should be used for the training of the model a fixed exposure time is not ideal. Due to the higher latitude in Trauen the sun elevation angle is lower, leading to generally darker images. In addition the local weather conditions favor low layer clouds as seen in Figure 17, which block even more light if a high cloud cover is present. This reduces the usable images from Trauen drastically.

To tackle this issue one could use location dependent exposure times or combine the exposure times to obtain a pseudo HDR image that combines different lightning to one image.

In this work the same exposure time as for the previous best model is used which is 160 μs. This is needed to compare its performance to the old model which only works with this exposure time. Additionally the 5 different available exposure times are used to create such a high quality image using Mertens' *exposure fusion* method [74].

It works by taking the series of input images with different exposure times, then computes for every pixel and every image how *useful* that pixel is. Pixel quality weight is defined based on saturation and contrast information to surrounding pixels in addition to a term that addresses *well-exposedness*. This information is gathered in weights which are added up across images for each pixel and normalized. The images are then decomposed into a Laplacian pyramid, while the weight maps are decomposed into a Gaussian pyramid. At each frequency level, the images are blended using the corresponding weights. Finally, the pyramid levels are recombined to produce the output image.

This produces an image with well-exposed shadows and highlights where sharp details are preserved.

Figure 20 shows the 5 images captured with different exposure times and the resulting improved quality image using the described method for an example timestamp in December.

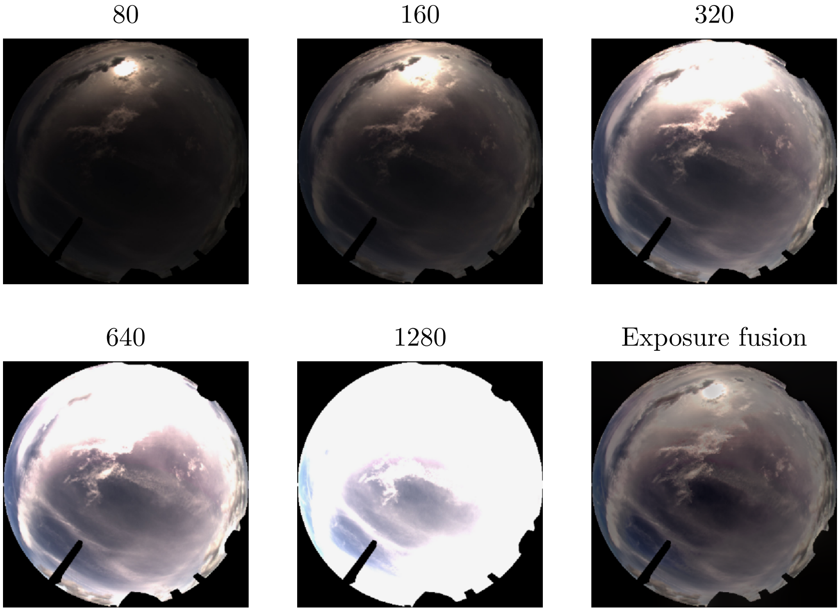

Figure 20: Images captured with different exposure times and the high quality image using the *exposure fusion* method for December 04, 2025 12:20 UTC

One issue with this method occurs for weather conditions with low clouds and higher wind speeds. Although most of those conditions were filtered before, there are still some critical conditioned images present in the dataset. The *exposure fusion* method expects every object to be at a constant position in the image, which is not the case for those specific timestamps. Figure 21 shows an example of such a problematic image.

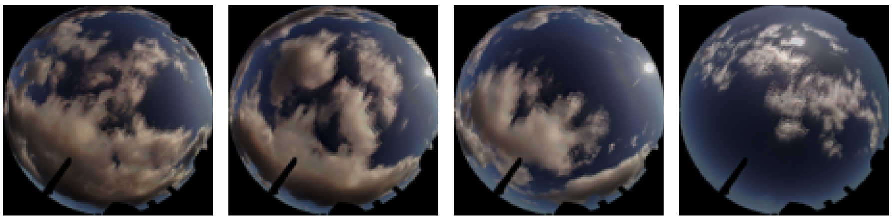

Figure 21: Example images where *exposure fusion* is flawed.

This bad behavior does not occur often and there was no directly related negative effect noticed when training the exposure fusion model. This might also be due to an already present offset between the visible light and the infrared image for those timestamps. Since both of them are vulnerable to higher wind speeds, the effects are expected to occur on similar timestamps.

This concludes all the preprocessing steps to obtain high quality input as well as the best possible output images. Now the dataset can be split into training, validation and test datasets.

3.5 Datasplit

Due to the nature of sky imagery timestamps close to each other can look similar if the clouds are not moving fast. The ASI system also captures a broad FOV such that pixels near the horizon do not change a lot even if the images are taken a few minutes apart. This makes it necessary to split the dataset by days to avoid the model seeing images during training that are very similar to the testing data.

However this datasplit by day is only done for the test set. This is due to the limited data availability so far. Often times cloud formations vary only in large scale time windows. Removing specific condition days from the training set could influence the models performance on those specific conditions negatively because those type of cloud formations could then be underrepresented or even never seen by the model. Hence for now the validation set is not sampled per day. This should not negatively impact the training, because the frequency is of at least 5 minutes, often times more due to the previously applied filtering process.

The test set consists of seven manually selected days with different conditions at PSA, to make an unbiased evaluation for all of them possible. Those include clear sky, high layer, low layer, and multi layer condition days.

Example images are shown in Figure 22.

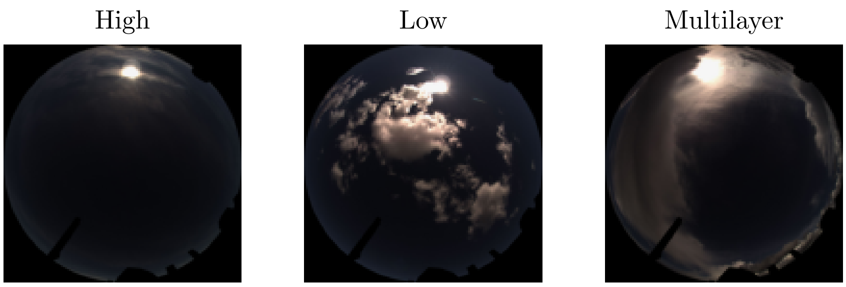

Figure 22: Example ASI images from the test set. High layer (left), low layer (middle), mixed conditions (right)

The remaining images are randomly partitioned into training and validation sets using an 80/20 split. This way the pixel-wise height distribution stays the same in the training and the validation split. Figure 23 shows the pixel-wise height distribution for the different splits for the combined PSA Trauen dataset, normalized with total pixel count of the dataset.

The manually selected test days do not follow the same distribution. This is because the split is created day-wise. Also there is a lack of low cloud images, because the test days are only selected from the PSA site.

This concludes all the data preparation steps for training a height regression model. The data can now be used to train a ML model which is thematized in the upcoming section after explaining the infrared segmentation preprocessing shortly.

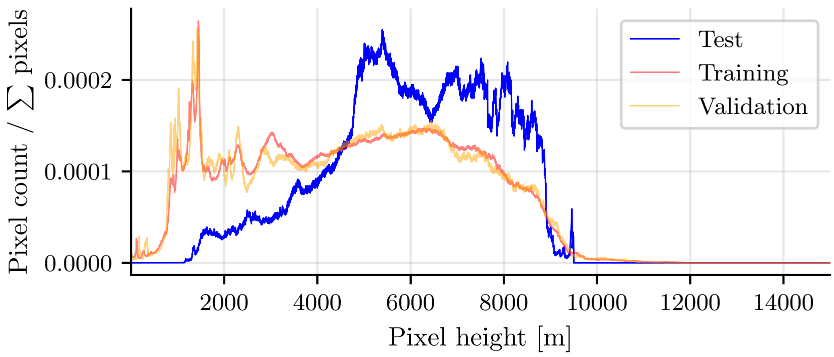

Figure 23: Pixel-wise height distribution normalized with total pixel count of the combined PSA and Trauen 160 exposure time data. Colors show the different splits.

3.6 Preprocessing IR segmentation

To prepare the infrared data to train a binary segmentation model, not as many steps are necessary. It is even preferable to not overload preprocessing of the input images for real time applicability. The necessary steps consist of masking, cropping, subtracting clear sky radiation, sun masking and, to create ground truth data, it is also unavoidable to transform the data to the visible light camera space. The data does not need to be transferred into height space.

All mentioned steps follow the same logic as described for the regression data preparation and are summarized here for completeness. First the image is masked and the zenith angle is cropped at $\theta = 80°$ Then the clear sky radiation is determined using the same criteria as previously and afterwards subtracted. The image is then transformed into the camera space of the mobotix camera and the images are resized to 512×512 pixels. To generate ground truth data, the current cloud segmentation model is applied to the respective visible light image. The infrared image is evaluated in the regions where the segmentation mask classified clear sky. Based on what percentage of the sky is classified as clear, a certain percentile is chosen and the infrared temperature is thresholded at that value. Everything below is set as clear sky and every pixel value above as cloud. This defines the ground truth data for the binary classification task.

The steps above are applied to all available dates between December 2025 and April 2026 for the data acquired at PSA. The test set consists of the same preselected days. Since the lighting conditions do not play a role in infrared imagery and the input and target images lie in the same camera space, additional data filtering is not needed.

After the successful preprocessing the data is now prepared to train a ML model.

4 Deep learning

In the previous section a lot of steps were taken to prepare a high quality dataset for deep learning methods. The ground truth is obtained from the temperature calibrated infrared images and the input images are prepared to obtain the best possible solution. In the following the chosen model architecture as well as the selected loss functions are described.

4.1 Model architecture

The previous best model trained for semantic cloud segmentation is a Convolutional Neural Network. CNNs are very wide-spread architectures and known for there excellent feature extraction capabilities. In general they consist of a combination of convolution and pooling layers chained after another. Each convolutional layer convolutes the image or the output from the previous layer to detect features. This happens by sliding one or usually multiple kernels over the matrix of numbers and returning the output of those predefined kernel operations. All of these operations are linear, however real world data is mostly non linear. To enable the model to learn complex patterns non-linearity is necessary. This usually happens by applying a rectified linear unit (ReLU) $ReLU(x) = \max(0, x)$. By mapping all negative values to zero, this operation introduces nonlinearity at $x = 0$. Afterwards the pooling layer reduces the contained information to the essential. This is used to reduce spatial size and make it translation invariant. The most common operation to do this is maxpooling, which simply slides a predefined sized window over the feature map and saves only the maximum of the window in a new matrix. The layer weights get adjusted during the training, such that the model learns, which *dimensions* and parts of the image are important and which structures are present.

To obtain a prediction of the same size as the input image the convoluted layers need to be upscaled again. This usually happens using a combination of interpolation upscaling and skip connections to keep structural information from early convolution layers. The exact architecture varies for specific use cases. This convolutional structure allows the model to learn spatial features and detect structures and edges. They are known to perform excellent for image classification and semantic segmentation tasks [75], which is no different for cloud segmentation. The current best model at DLR uses a DeepLabV3+ [76] architecture with a ResNet-50 [77] backbone. This is also the architecture used in this work. Because the approach in this work follows the concept of image regression, a slightly adapted DeeplabV3+ architecture is used. By deactivating the final sigmoid function in the classifier head, the model is fully functional for pixel-wise image regression tasks.

Nevertheless clear sky distinction can not solely be achieved using a single regression head. Therefore a second classifier head is used for binary cloud detection, and on all cloudy pixels the model is trained to learn the cloud height. This enables the model to distinguish between cloud and no cloud and additionally predict cloud height.

As a model output a mixed map is produced, giving a binary segmentation and predicting cloud height for every cloud pixel.

Figure 24 shows a general schematic of the used model. For the exact architectural design of the DeeplabV3+ model it is referred to the original publication [76].

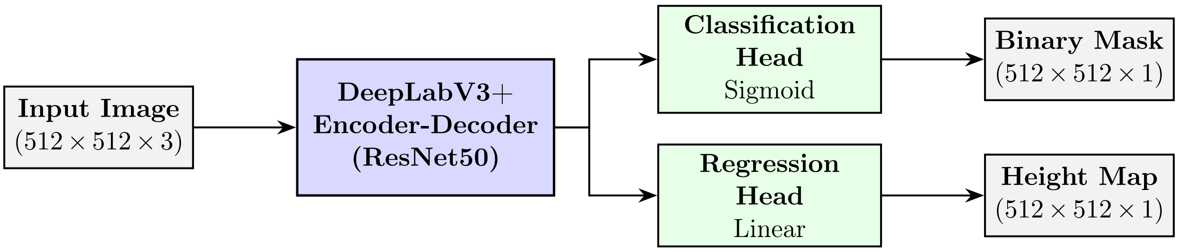

Figure 24: Schematic of the ML model in use.

For the binary cloud detection model from infrared images a similar architecture is used. The only difference is that the input images only have one channel and a regression head is not needed.

4.2 Loss functions

The learning process of the CNN is driven by the optimization of a loss function via an optimizer. By calculating the gradient of the loss with respect to each weight through backpropagation, the model iteratively updates its parameters in the direction that minimizes the cost function, thereby converging toward a more accurate representation of the target data. The goodness of the model therefore directly depends on the selection of a well-suited loss function. Those depend highly on the application and the goal of the model. In this case the model needs to perform two tasks, each needing its own loss. For classification or segmentation tasks like the binary detection of clouds a typical loss function is the binary cross entropy (BCE) loss. It encourages correct probability predictions pixel-wise and works really well when both classes are balanced. If there is a slight class imbalance it is often combined with the dice loss. It focuses on overlapping predicted and true masks. This means BCE ensures pixel-wise accuracy, while the dice loss gives good global mask shape results. They are defined as follows:

$$
\begin{aligned}
\mathcal{L}_{BCE} &= -\frac{1}{N}\sum_{i=1}^{N}\Big(\underbrace{y_i \log(p_i)}_{\text{positive class}} + \underbrace{(1-y_i)\log(1-p_i)}_{\text{negative class}}\Big), \\
\mathcal{L}_{Dice} &= 1 - \frac{2\sum_i p_i y_i}{\sum_i p_i + \sum_i y_i},
\end{aligned}
\tag{4.1}
$$

where p is the predicted probability, y the true label, and N the total number of pixels in the images in the batch.

For the regression task a different loss is needed. Since the clouds in input images and ground truth masks often do not match exactly a L1 or L2 loss penalizes inherent dataset problems strongly often leading to bad convergence. Hence an additional loss is chosen that tries to match structures, the so called structural similarity index (SSIM) loss [78]. It is combined with a gradient and L1 loss to improve cloud boundary regions and nonetheless provide good pixel-wise height information. The Regression loss definitions are as follows

$$
\begin{aligned}
\mathcal{L}_{L1} &= \frac{1}{N}\sum_{i}^{N} |p_i - g_i|, \\
\mathcal{L}_{grad} &= \frac{1}{N}\sum_{i}^{N} \Big| |\nabla P_i| - |\nabla G_i| \Big|, \\
\mathcal{L}_{SSIM} &= 1 - \frac{(2\mu_p\mu_g + C_1)(2\sigma_{pg} + C_2)}{(\mu_p^2 + \mu_g^2 + C_1)(\sigma_p^2 + \sigma_g^2 + C_2)},
\end{aligned}
\tag{4.2}
$$

where p is the predicted value for each pixel, g is the true value, N the number of pixels, μ_p, μ_g are local means, σ_p^2, σ_g^2 local variances and σ_{pg} the local covariance, C_1 and C_2 are numerical stability constants. $|\nabla P_i| = \sqrt{(G_x(p)_i)^2 + (G_y(p)_i)^2 + \epsilon}$, where G_x and G_y are the horizontal and vertical Sobel gradients and ϵ is added for numerical stability, analogous for $|\nabla G_i|$. The SSIM Loss is always calculated using Gaussian weighted sliding windows.

This combination with structural loss functions, based on gradients or structural similarity index prove to be effective especially for multilayer conditions where distinctions between lower and higher clouds are needed. Using only L1 Error tends to blur the clouds edges and the cloud borders are hard to distinguish.

The different loss functions are combined to a total loss

$$
\mathcal{L}_{total} = \alpha(\mathcal{L}_{BCE} + \mathcal{L}_{Dice}) + \beta\mathcal{L}_{L1} + \gamma\mathcal{L}_{SSIM} + \delta\mathcal{L}_{grad}, \tag{4.3}
$$

with coefficients $\alpha = 0.5$, $\beta = 1$, $\gamma = 1$ and $\delta = 0.2$.

Other hyperparameter combinations have not been fully tested, and there might be better choices available. However further investigation only makes sense, once the hardware synchronization is executed, such that cloud borders can be expected to be captured well by the model.

The binary cloud detection model is trained using a combination of BCE and dice loss $\mathcal{L}_{binary} = \mathcal{L}_{BCE} + \mathcal{L}_{Dice}$.

To accelerate convergence, a learning rate scheduler was employed in conjunction with the AdamW optimizer. During most training runs convergence is only reached after the learning rate decreases from its starting value of LR = 0.0001. The setup is hereby complete and in the next section the results will be discussed.

5 Experimental results

In this section the experimental results of the work are presented. The model is val To guarantee the robustness of the results, a multi-faceted validation approach was adopted. The direct model performance compared to the ground truth will be discussed. Furthermore the models performance will be evaluated in predicting the correct pixel-wise height value for the same timestamps used in 3 to validate the height conversion, based on ceilometer measurements. Afterwards the model output will be transformed to the WMO predefined height layers 1 and the performance will be evaluated on the original hand labeled dataset. Finally a short error analysis and discussion is performed to track down sources of error and get an estimation how many errors accumulated over the whole process. The infrared cloud detection model will be evaluated on the test set.

5.1 Model validation

To start of the model will be evaluated in how successful the training procedure is. To do this the following Error metrics are used. To determine if the clear sky regions are detected correctly the intersection over union is evaluated. In terms of binary classification components, this can be expressed as the ratio of True Positives (TP) to the sum of True Positives, False Positives (FP), and False Negatives (FN). For all cloud pixels the regression is then evaluated using the mean absolute error, the root mean square error and the bias. They are defined as follows

$$
\begin{aligned}
\mathrm{IoU} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}, \\
\mathrm{MAE} &= \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|, \\
\mathrm{RMSE} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \\
\mathrm{Bias} &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i),
\end{aligned}
\tag{5.1}
$$

where n is the number of samples, y_i represents the ground truth value, and $\hat{y}_i$ represents the predicted value for the i-th sample. The MAE measures the average magnitude of the errors. The RMSE is a quadratic scoring function that also measures the average magnitude of the error, but it penalizes outliers heavier. Bias indicates whether the model systematically over-predicts or under-predicts the target value. For pixel-wise regression tasks this is a common set of evaluation metrics that give good understanding about the models performance.

Five different models are evaluated in this section. They all share the same architecture and loss functions but differ in the datasets on which they are trained. The first one (PSA 160) is trained on all available images from PSA with exposure time 160 μs without applying timestamp filtering as described in section 3. The second one (PSA Trauen 160) additionally incorporated data from Trauen. The third (PSA Trauen 160 dt filtered) and fourth (PSA Trauen 160 wind filtered) one apply fixed time stamp filtering or wind speed filtering respectively. The last one uses all available images from both sites but using the exposure fusion input images (PSA Trauen HDR).

The MAE, RMSE and Bias for the different model runs on the predefined test set are shown in Table 5.

Model	IoU	MAE	RMSE	Bias
PSA 160	0.740	943.06	1171.63	-389.77
PSA Trauen 160	0.775	868.16	1070.94	-254.17
PSA Trauen 160 dt filtered	0.758	1109.2	1361.41	-734.16
PSA Trauen 160 wind filtered	0.773	949.03	1196.81	**-184.63**
PSA Trauen HDR	**0.777**	**834.90**	**1040.14**	-232.07

Table 5: Model performance of the different runs.

For some models the bias reaches high values, but a correction is not yet discussed or applied before evaluating the performance compared to ceilometer measurements. To analyze the sources of errors and get insight of potential problems multiple diagnostic plots are created. First of all the binary cloud detection is evaluated using a confusion matrix. It shows how well the model predicts which class, showing all combinations of false negatives, true negatives, false positives and true positives. The row-normalized (recall) version for the best model from Table 5 is shown in Figure 25. It shows, that the classes are detected with good accuracy.

To investigate the pixel-wise height prediction performance a hexbin plot, shown in Figure 26, is created, that shows predicted vs true heights for all pixels in the test set.

The general trend follows the ideal red line, which symbolizes the ideal ratio of one to one. Nevertheless there are a few problematic regions to be mentioned. First, the model seems to be incapable of predicting heights below 1000 meters. There are close to no bins in that region, while there are a lot close to 2000 meters predicted height. This suggests that all values below 1000 meters are predicted to be in the range of 1000 to 2500 meters. Interestingly, those bins do not appear in datasets without Trauen images, indicating that this might be due to a calibration step during preprocessing. This is likely related to the calibration curve 12 which uses only PSA data. Due to the unavailability of a ceilometer in Trauen, this is the best available option, but it might negatively influence the performance at other sites or using other hardware. This is to be investigated in the future. Second, there seems to be an accumulation of data-points at predicted height 8000 meters and above. This might be explained by wrongly misclassified non cloud pixels which are assigned high layer cloud heights. Furthermore, the ground truth as well as the model output never go above 10000 meters. This is likely due to the infrared data not following the ideal trend in 12. Even after the calibration, the pixels above 10000 meters are classified as clear sky and hence are not represented in the graph. Lastly, the spread from the red line can be quite big, especially in mid layer regions.

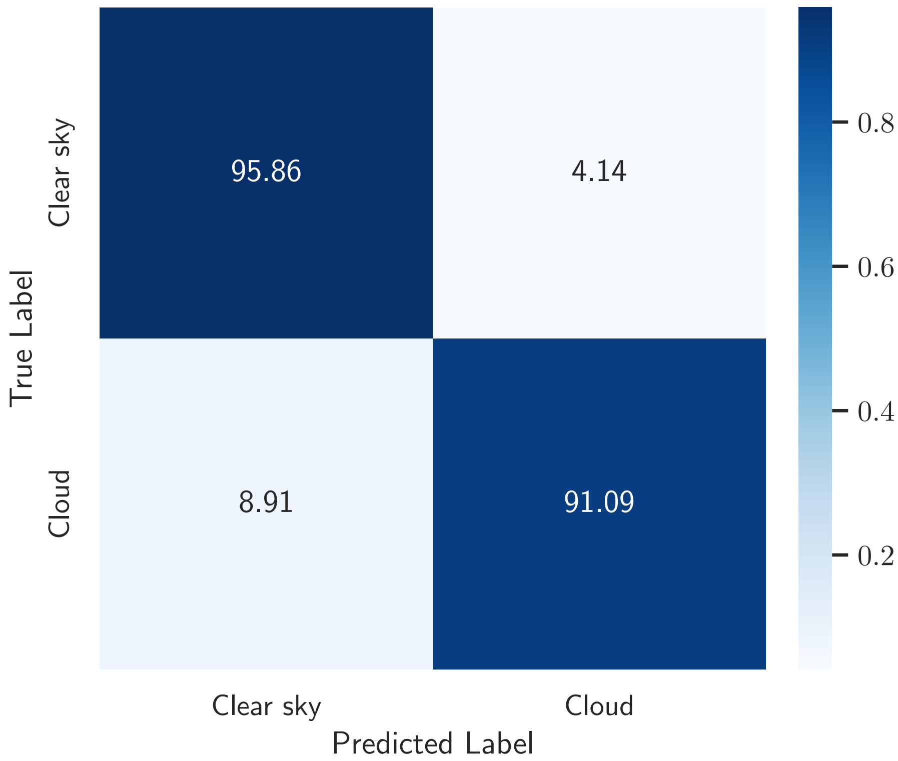

Figure 25: Row-normalized confusion matrix for the PSA Trauen HDR model

This might be explained with the gradient problem in the model output. Cloud border regions do not represent the cloud base height well. This is inherit to the data, because cloud borders thin out and hence are more transparent, yielding a smaller temperature response in the sensor and ultimately a greater height. But the bigger problem is the data offset from RGB to infrared image. This makes it impossible for the model to learn the borders correctly, and makes the prediction blurry near the cloud edges. The height prediction almost always has high gradients in border regions across all cloud heights. Since it learns that cloud border regions have high gradients, the model struggles to predict small clouds correctly. It detects them but does not always assign the correct heights. This is likely due to every cloud pixel being close to a border, hence over-predicting the whole cloud. To illustrate this a few examples from the test set are shown in Figure 27

It shows the input image, the target, the prediction as well as the error heatmap of target height minus predicted height. The height plots are normalized to enhance contrast.

As hypothesized the biggest errors occur near cloud borders or for very small clouds that get highly over-predicted. Especially isolated small low layer clouds yield high errors.

Generally speaking the model performs worse in less light conditions or image regions. This includes times with low sun elevation or fully covered skies with low layer clouds

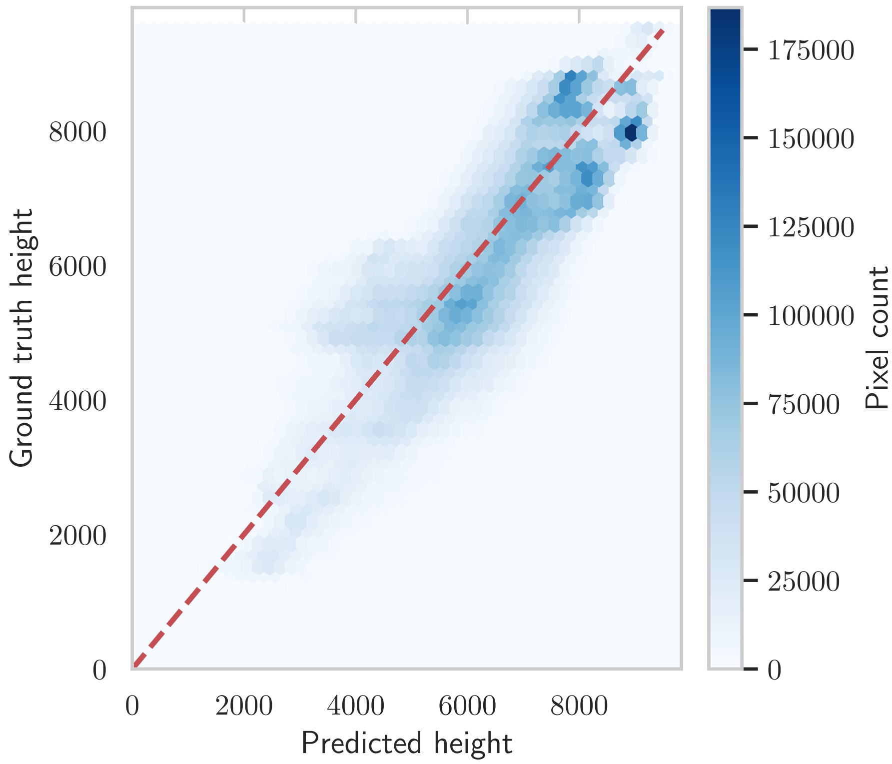

Figure 26: Hexbin plot for the PSA Trauen 160 model

as seen in the first row of Figure 27. The model can detect the low clouds fairly well, however since the whole image is quite dark it has problems distinguishing between the lower and mid-layer clouds. This is to be expected, since the model cannot extract as much detail and is one of the main reasons why the enhanced image model is trained in the first place. However for regions that remain without a lot of light even after enhancement, an accurate prediction is difficult.

To further analyze the bias a histogram of the error distribution is plotted for the test set. It is shown in Figure 28.

The peak of the histogram is slightly shifted to the negative side, yielding the bias of -232.07 meters. Interestingly the curve looks like a cutoff gaussian distribution. It is hypothesized that this is induced due to the cloud border gradients. They are higher in regions where a low layer cloud is adjacent to high layer clouds or clear sky, hence the model over-predicts low layer cloud borders. The predicted height is greater than the ground truth yielding negative bias. This effect is reduced if the adjacent clouds are high or mid layer, where the gradient does not have a big effect on the absolute height value. If the model under-predicts a cloud the cloud border gradients smoothen the output and the model artificially improves in border regions yielding a smaller overall error.

In the next step the model performance is evaluated in regards to predicting actual true heights as measured by the ceilometer.

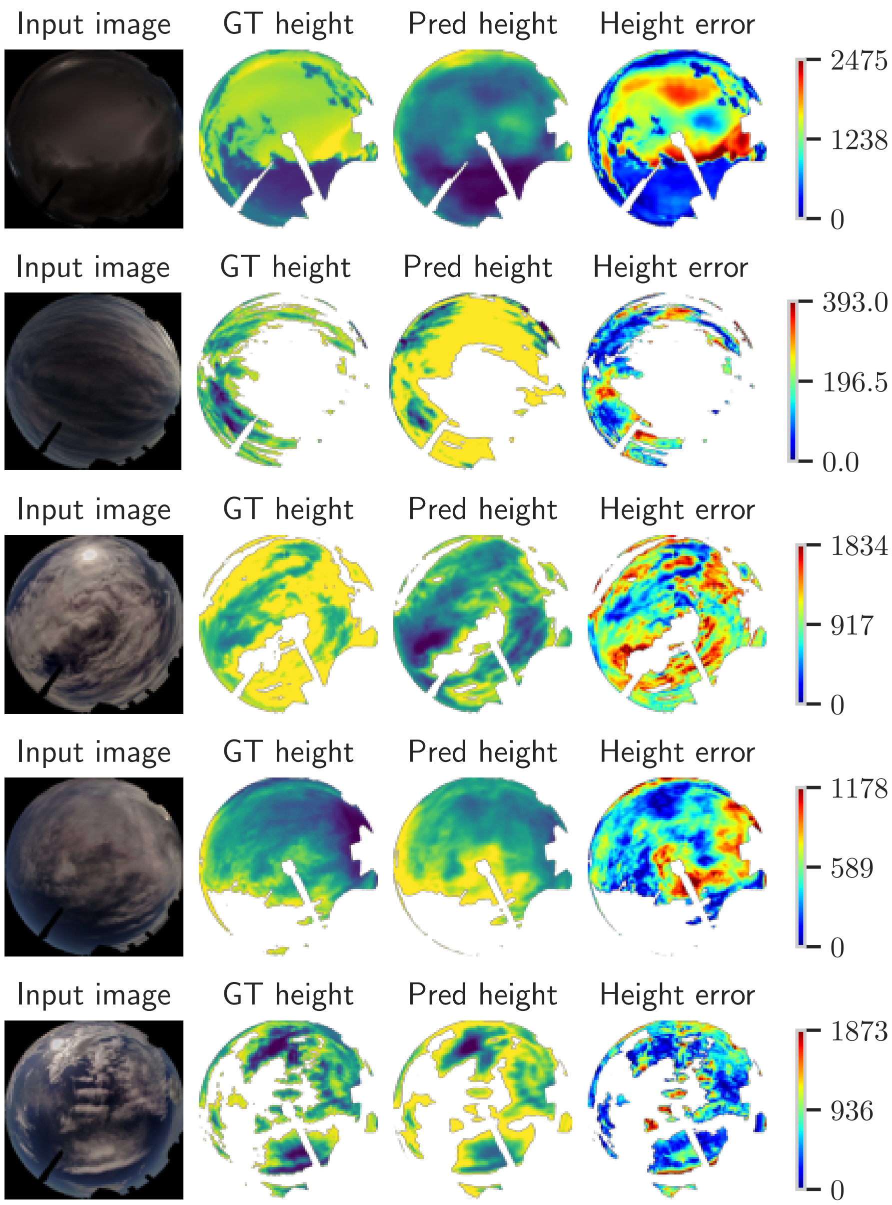

Figure 27: Error visualization, left to right: input image, normalized ground truth height, normalized predicted height and error heatmap, white pixels represent masked or clear sky pixels.

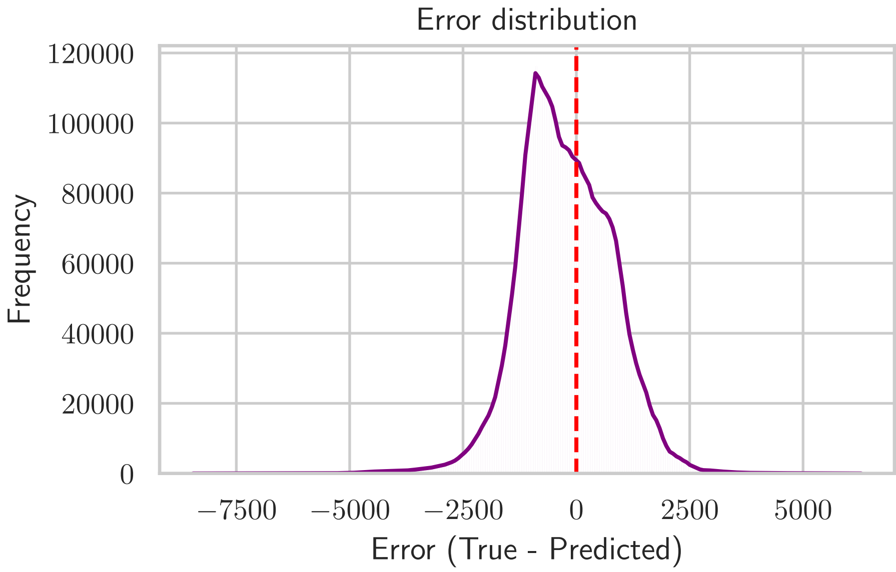

Figure 28: Histogram of errors for the PSA Trauen HDR model, red line indicates 0 error

5.2 Height validation

To get some information about the goodness of the predicted heights the different models are evaluated for the previously selected calibration timestamps in 3. Due to limited data availability this is the best possible validation procedure without risking big ceilometer errors. To begin it has to be mentioned that this comparison is far less than ideal. To compare a single pixel value from a single prediction is bound to have high errors, as there might be single pixels inducing great errors, as already seen in Figure 27. Nevertheless there is no better comparison available.

The same 5 models from the previous section are evaluated in MAE, RMSE and bias which are displayed in Table 6. The comparison follows the same procedure as during the calibration process. Timestamps at which the model predicts clear sky are omitted and removed from the comparison, reducing the available comparison windows even further. The table shows the number of evaluated timestamps after reduction for completeness.

Model	**# timestamps**	**MAE**	**RMSE**	**Bias**
PSA 160	276	1241.96	1569.47	**183.88**
PSA Trauen 160	265	1105.32	1402.04	184.76
PSA Trauen 160 dt filtered	284	1558.95	2025.26	-715.19
PSA Trauen 160 wind filtered	304	**1046.09**	**1375.08**	-399.76
PSA Trauen HDR	274	1691.4	2199.75	-880.91

Table 6: Number of evaluated timestamps, MAE, RMSE and bias compared to ceilometer measurements for the different models.

Interestingly the models trained using the 160 exposure time images outperform the model trained on the enhanced images if no additional bias correction is applied. To in-

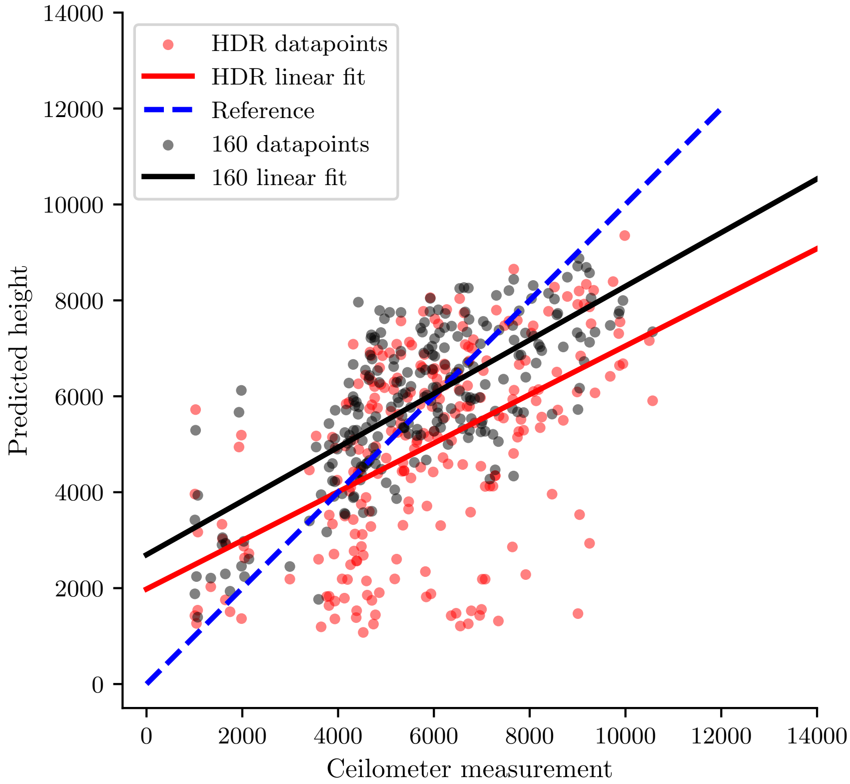

Figure 29: Ceilometer measurements vs predicted heights. Red: The data-points and linear regression line for the enhanced image model, Black: The data-points and regression line for the mixed 160 exposure time model, Blue: Reference

vestigate this the predicted height is plotted over the ceilometer measurements, displayed in Figure 29.

The pseudo HDR model often times underestimates the true height measured from the ceilometer. It is hypothesized that this is due to the cloud colors being more equal. In normal images the colors of clouds with different water content vary highly. This leads to lower layer clouds appearing darker while high layer clouds are often very bright. In the enhanced images this effect is softened. Since all the exposure times are combined, darker regions or shadows are overexposed compared to brighter regions where the underexposed setting is used. This shortens the overall contrast range. However since the previous validation on the test set suggest that the HDR model overall performs better, it might just be a difference in the pixel-wise ceilometer comparison. Unfortunately the true source can not yet be determined since not enough data has been gathered. A more relaxed time window selection and a repetition of the validation process with more data might bring more insight in the future.

The linear fitted polynomial shows a different slope than the ideal behavior. This is likely due to the data quality already observed in 3. Some bias was corrected before training, however since the spread is so large it could only be corrected to some degree.

An additional bias correction could be applied to the model results to minimize errors on the tested timestamps. However after closer investigation, the bias changes throughout the heights and also with time of the year. Therefore it is decided to not implement a postprocessed bias correction, to prevent overfitting on the few available comparison test timestamp data.

Again this might also be influenced by the limited data and pixel outliers that get weighted too strong. The model is most likely improvable by recalibrating with more available data.

5.3 Semantic cloud segmentation

Finally it is analyzed how well the height predictions can be used for semantic cloud segmentation in a similar manner as it has been done in the previous height layer approach.

To do this the model is applied to the original hand labeled test set. It consists of 153 images that represent all kinds of different cloud formations. One downside is, that at this point in time only a fixed exposure time of 160 μs was used to capture the all sky images meaning that no enhanced images can be created. As a consequence the pseudo HDR model can not be evaluated properly. The evaluation will therefore be limited to the non enhanced image models.

The models output height map is converted into different layers as defined by the WMO in Table 1. This is challenging since the layer definitions have big overlapping regions, however the height map does not give information about which of the layers is the one desired. This raises the challenge to which layer the pixel should be assigned. To minimize this problem, the height prediction is first converted into a soft classification map. This classification assigns the low, mid or high class if the pixel value does not lay inside an overlap region, like a normal classification. If the value is inside this overlap region, both classes are assigned and weighted based on the *distance* to the two classes. A few examples from the original test set are shown in 30. The model performs well in binary cloud detection, in some cases even better than the currently used model. The soft classification almost always assigns the mid-layer correctly. The general cloud detection also yields very good results. High layer clouds are often well presented too. However for a lot of low layer conditions and also some high layer conditions the high gradients in the prediction make a clear class assignment almost impossible. To clarify and discuss the issues Figure 31 displays some conditions in which a clear assignment fails.

As mentioned before, the conversion to layer classes often fails for the low layer class. This is mostly due to the border gradients. The model learned to predict high gradients at cloud borders because the clouds often do not align well due to the inaccurate synchronization of the timestamps. This behavior is worst for low layer clouds, because the movement induces a high change in the pixel position. Often times the Cloud is detected to be low layer in the center but then the gradients change that result to mid layer for outer pixels. This behavior rightly completely ruins class-wise evaluation metrics which can be seen in the confusion matrix in Figure 33. Some multilayer combinations of high

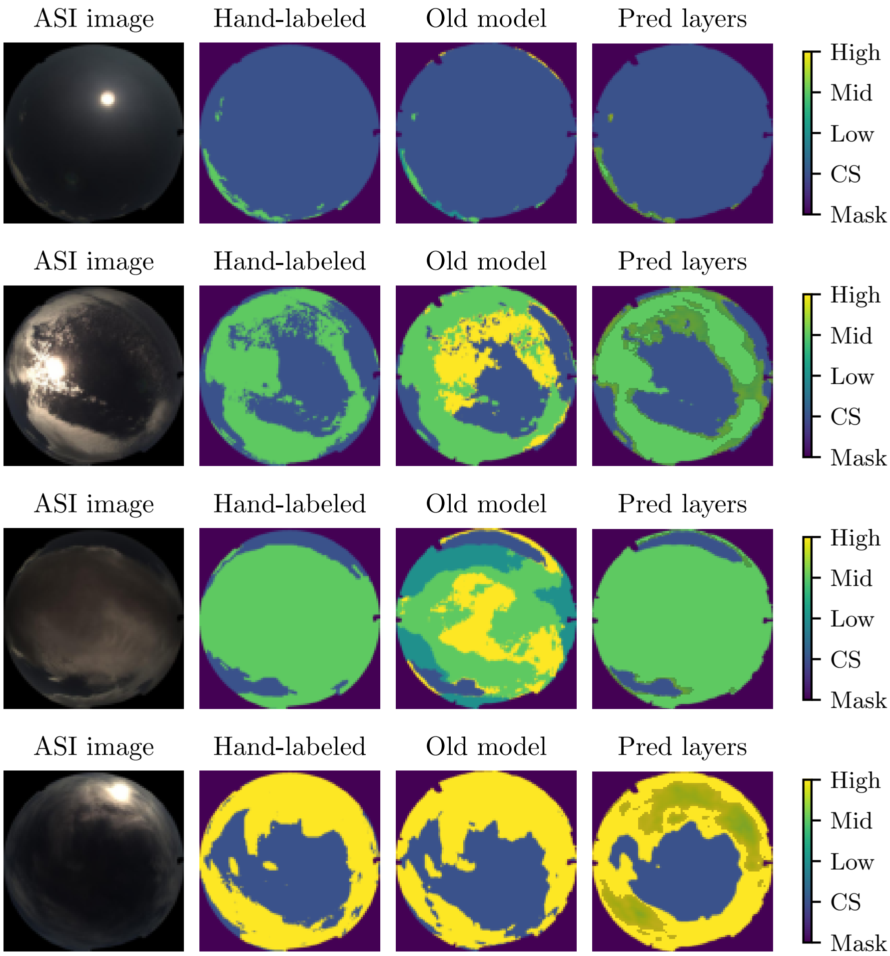

Figure 30: Good examples from the original manually labeled test set for semantic cloud segmentation using soft classification. From left to right: Input ASI image, hand labeled ground truth, old model prediction, height prediction from the PSA 160 model transformed using soft classification. Darker regions mark in between regions, to enhance contrast to them.

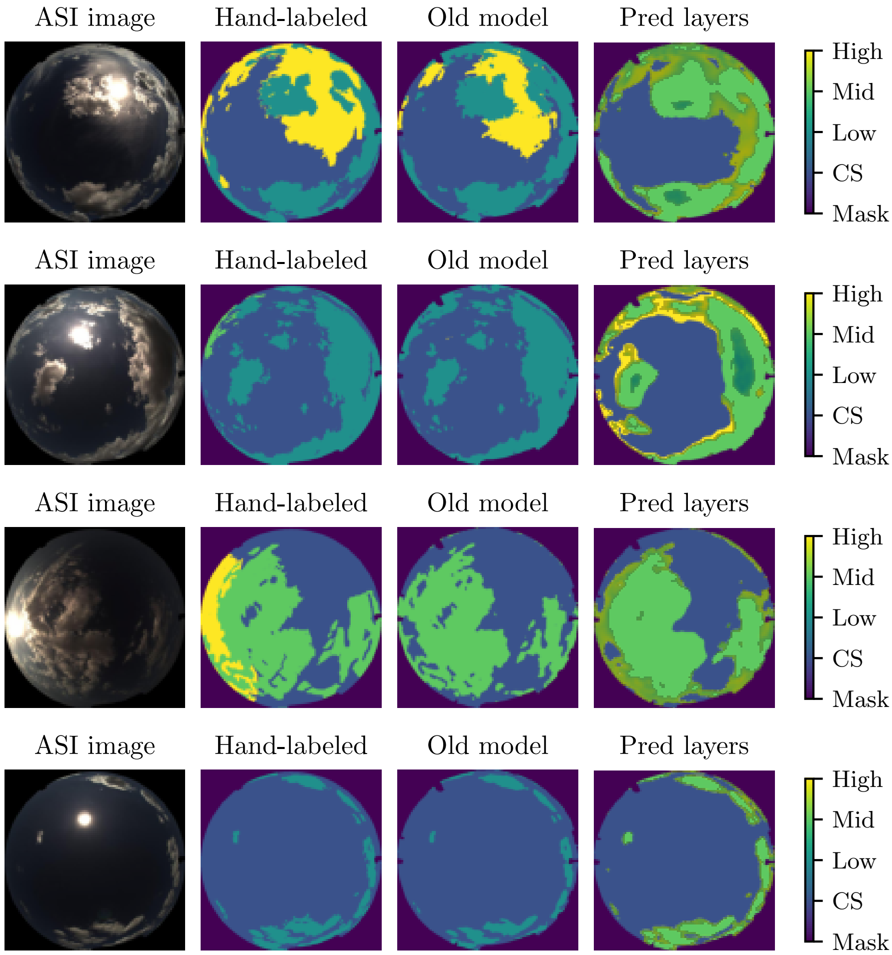

Figure 31: Bad examples from the original manually labeled test set for semantic cloud segmentation using soft classification. From left to right: Input ASI image, hand labeled ground truth, old model prediction, height prediction from the PSA 160 model transformed using soft classification. Darker regions mark in between regions, to enhance contrast to them.

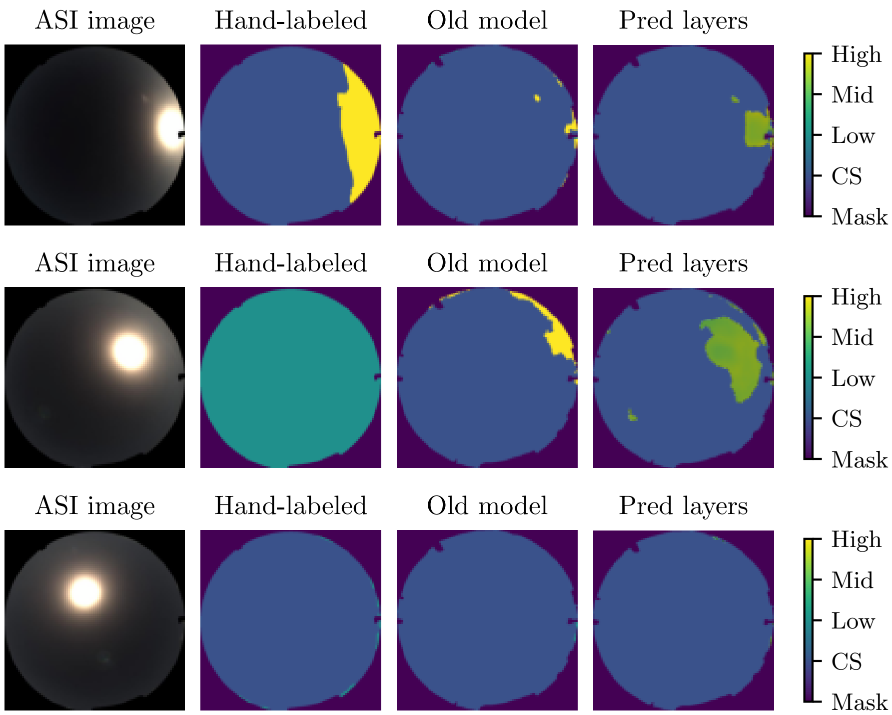

Figure 32: Bad ground truth examples from the manually labeled test set. From left to right: Input ASI image, hand labeled ground truth, old model prediction, height prediction from the PSA 160 model transformed using soft classification. Darker regions mark in between regions, to enhance contrast to them.

and low layer clouds are hance especially difficult, as can be seen in row one of Figure 31. Another issue to be discussed are the high elevation angles at which the infrared temperature obtained does not correspond to the cloud base temperature. The model adapted this knowledge and can therefore not predict low layer base heights for high elevation angles. As mentioned in previous evaluations small clouds are especially difficult as well, since they get dominated by the gradient problem.

Before talking about the quantitative evaluation metrics, it is mentioned that the hand-labeled ground truth data is also not perfect. Especially hazy conditions are hard to distinguish from clouds and especially hard to assign cloud layer labels. This however has an impact on the evaluation metrics and results. Figure 32 shows and example for the test set for similar conditions from the same day with differently assigned labels, where the models predict more consistent output. The distinction in when a model predicts hazy conditions as clouds also has strong influence on the evaluation. While the current model often interprets haze as high layer cloud, which corresponds to the hand labeled data, the new model often predicts clear sky.

The soft-classification is sufficient for visual comparison, however to get a quantitative value to compare to other models this soft classification map needs to be transformed further. TThe most straight forward approach is to assign the nearest class for all pixels inside the soft classification regions. This allows class-wise comparisons using metrics

Model	mean IoU	clear sky	low	mid	high
PSA 160	**0.372**	0.850	0.037	**0.356**	**0.245**
PSA Trauen 160	0.366	**0.851**	0.069	0.347	0.198
PSA 160 dt filtered	0.368	0.838	**0.155**	0.320	0.158
PSA Trauen 160 wind filtered	0.357	0.819	0.090	0.353	0.166

Table 7: Semantic segmentation mean IoU and class-wise IoUs for different models.

like IoU and confusion matrices. Table 7 shows the mean IoUs and class wise IoUs for the different model runs.

For the combined camera model the row-normalized confusion matrix is shown in Figure 33.

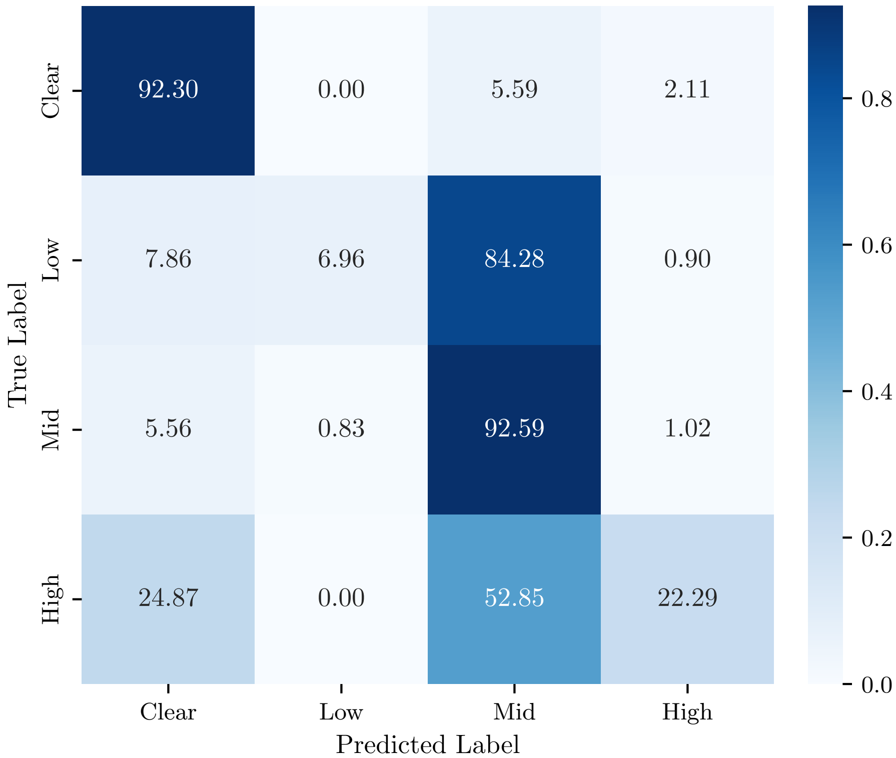

Figure 33: Row -normalized confusion matrix for the PSA Trauen 160 model for the semantic cloud segmentation

The results from the confusion matrix show good performance for clear sky and mid layer class. The weak performance for low layer is assumed to be mostly related to the cloud border gradients. Even though the middle of a cloud can be predicted well, the broad borders are false yielding bad IoU. Additionally the high layer clouds often get detected as mid layer too. This false prediction behavior aligns with the results from the height validation. The smaller slope of the regression line in 29 already suggested that most predictions will fall into the middle layer, over-predicting low and under-predicting high layer clouds.

A remark has to be made, that the height regression model might not preform as well on the original manually labeled test set because it has data spread around the year representing all possible sun elevation angles and maybe weather conditions that are only present in summer or autumn. Those angles or conditions are not present in the training data and have not been seen by the models.

5.4 Error analysis

Although the models yield good results for many different conditions, there are multiple errors present that accumulate throughout the data processing. Those include sensor noise, assumptions and estimations in the data preprocessing and model performance. Since there are many possible error sources it is important to try to narrow down the potential issues to there sources.

Previously it has been discussed that for large elevation angles the temperature measured by the infrared camera does not correspond to cloud base height because the cloud is observed from the side. Therefore the assigned temperature is lower than that of the cloud base. In this error analysis the large elevation angles will be excluded and only the error with respect to ceilometer measurements will be discussed. For large elevation angles it is open to discussion which behavior is desired. For shadow projection for example it would be more important to have the height of the obstacle in sight and not necessarily the cloud base height.

The sensor-induced temperature uncertainties range up to $+/-5°$C. Due to the non-linear relationship between temperature and geopotential height z(T), this thermal noise can translate into significant altitude errors; for the purposes of this analysis, a maximum sensor-induced error of $e_{sensor} \approx 1500$ m is assumed. It should be noted that potential calibration drifts at low temperatures may further increase this uncertainty for high-altitude cloud layers. The way the clear sky subtraction is handled it does not influence the temperature near zenith but might be a source of errors for higher elevation angles. If the clouds base height already has a high error or the clouds are extremely low, the transformation could theoretically move a pixel from higher elevation angles near the zenith. However this has not been observed for any of the visually inspected images, hence those cases are excluded from the error budget.

A further source of error stems from the temperature to height conversion. While ERA5 provides a highly consistent global dataset, its spatial resolution of ~31 km introduces representation errors for sub-grid scale phenomena such as the present localized cloud formations.

To evaluate the reliability of the atmospheric Temperature profiles used in this work, it is referred to the validation work by Hassan et al. [67], who compared ERA5 re-analysis against radiosonde observations across multiple isobaric levels. Their findings indicate that while ERA5 temperature remains highly accurate in the lower troposphere, uncertainty increases at higher geopotential heights due to a decrease in observational density. They analyzed this for different cites in eastern Mediterranean region and it is assumed that the accuracy is of the same level for the site in southern Spain. From the plots they provided a maximum RMSE for temperature of $e_{T,ERA5} \approx 1.35°C$ is provided, which corresponds to an altitude error of approximately$e_{T.ERA5_m} \approx 450$m which will be used as an upper bound for the absolute error. The maximum RMSE for geopotential height is $e_{Z,ERA5} = 30$m.

Finally the intrinsic validation error of the model training itself must be considered. Depending on the chosen configuration the mean absolute error ranges from 834.90 to 1109.2 meters before bias correction. Inside this analysis a representative error of $e_{model} \approx 1000$m is adopted.

To estimate the overall error, the independent sources are combined using the root sum square method:

$$\begin{aligned} e_{total} &= \sqrt{e^2_{sensor} + e^2_{T,ERA5\_m} + e^2_{Z,ERA5} + e^2_{model}} \\ &= \sqrt{(1500\text{m})^2 + (450\text{m})^2 + (30\text{m})^2 + (1000\text{m})^2} \\ &= 1858.33\text{m} \end{aligned} \tag{5.2}$$

This concludes the height error analysis and potential improvements are discussed in section 6.

5.5 IR detection model

Now the infrared binary cloud detection model is discussed. The validation is performed on the same predefined test set. As metrics the mean IoU, accuracy = (TP + TN)/(TP + TN + FP + FN), precision= TP/(TP + FP), recall= TP/(TP + FN) and F1-score as the harmonic mean of precision and recall are evaluated.

Table 8 summarizes the evaluation metrics for the test set.

mIoU	**Accuracy**	**Precision**	**Recall**	**F1**
0.835	0.910	0.915	0.901	0.908

Table 8: Evaluation metrics for the IR detection model

The model performs strongly in binary segmentation. The close alignment between precision and recall suggests a well-balanced classifier that effectively captures cloud boundaries without significantly overestimating their extent. The row-wise normalized confusion matrix is shown in Figure 34.

Figure 35 shows some example images for the binary infrared segmentation.

After analyzing a few images the highest errors occur for either high layer conditions, where infrared imagery comes to its limits or for complex multilayer conditions. In those multilayer scenarios where one cloud is very low and has a high temperature, the background is covered with a cloud of a lot lower temperature. An example is shown in the last row of Figure 35. For the model this appears quite similar as a low cloud in front of clear sky (compare to one row above), especially when we account for sensor uncertainties. To further improve the performance in those challenging conditions additional

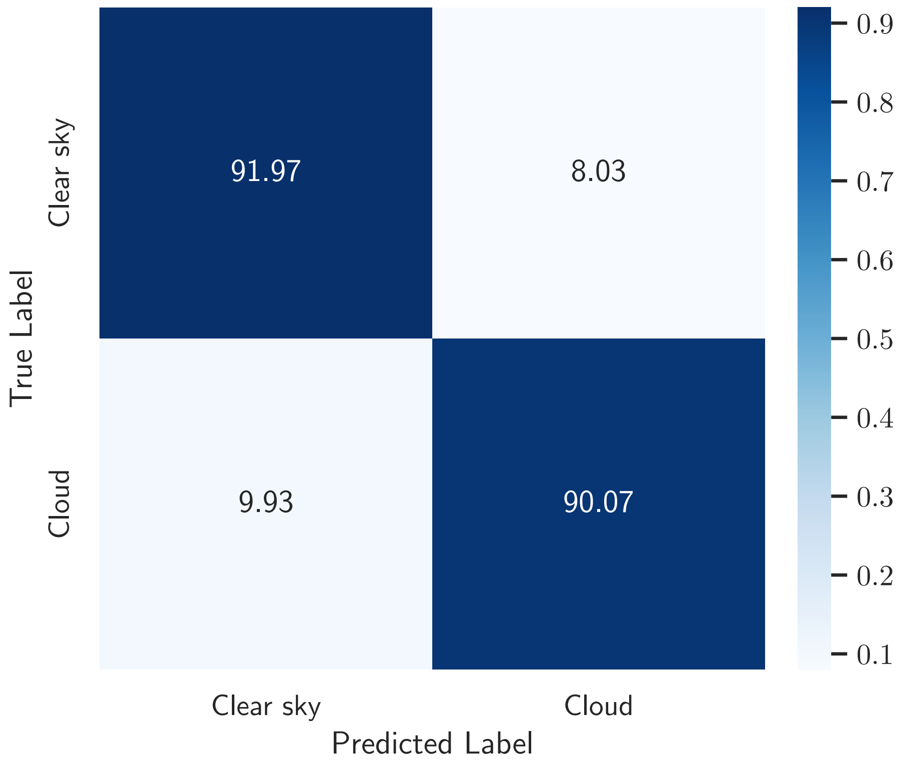

Figure 34: Row normalized confusion matrix for the IR segmentation model

postprocessing could be applied in the form of temperature thresholds. However this would most likely again worsen the results for high layer cloud conditions.

Unfortunately there is currently no validation data for nighttime images available. The model has been applied to randomly selected nighttime timestamps and visually observed with very good agreement. However to make an absolute and quantitative statement expert annotations for a few images are necessary as there are some atmospheric temperature changes at nighttime. It might come to temperature inversions when the surface air cools faster than higher layers. The temperature changes also differ depending on the height of a cloud. A combination of those effects lead to different temperature responses in the infrared sensors. Although visually the data looks similar, it likely differs in absolute values which are unknown to the model.

Nevertheless the results look very promising and suggest that cloud segmentation also works well for nighttime applications.

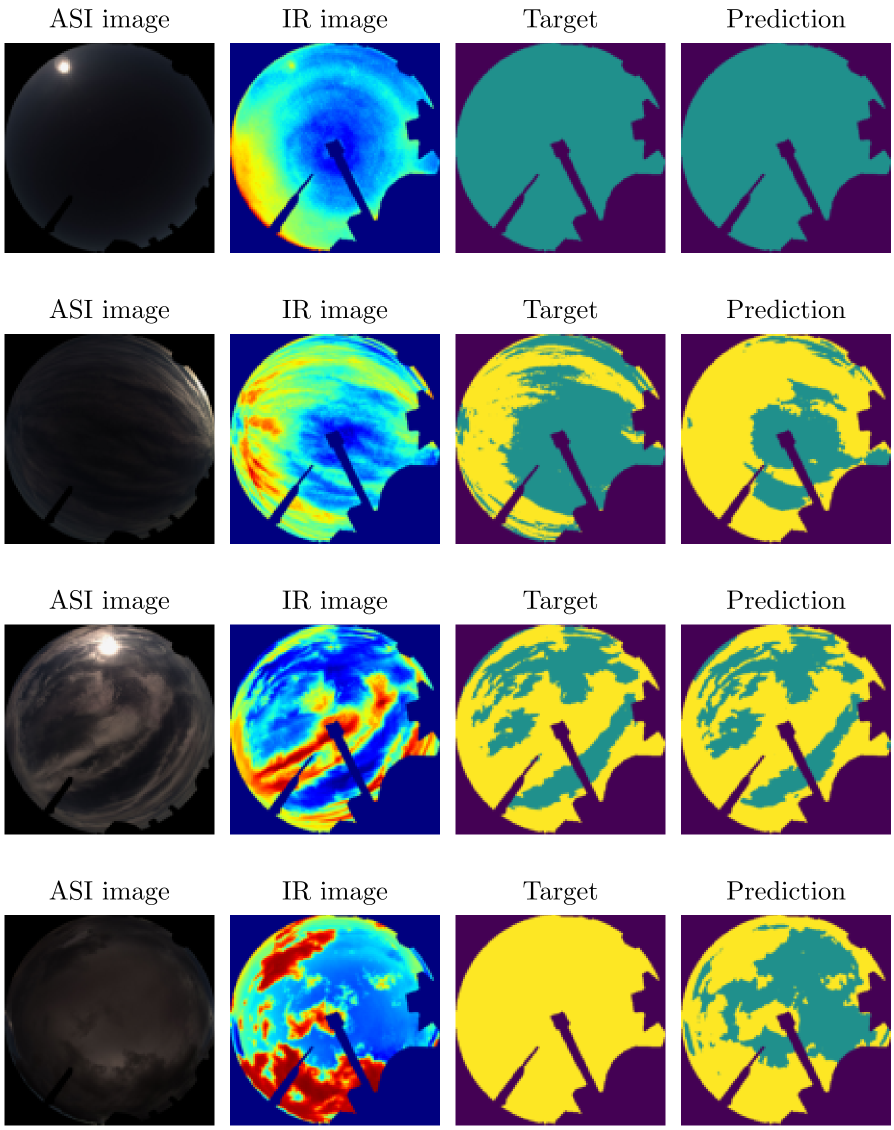

Figure 35: IR segmentation examples for different conditions. IR images normalized to enhance contrast, yellow: cloud, turquoise: clear sky

6 Discussion and conclusion

6.1 Discussion

In this work a lot of time was spent on preparing a high quality dataset, and it still contains data-inherent issues that require hardware recalibration. The low capture frequency and the bad synchronization in time and space of the infrared and visible light cameras, make training a good model very challenging. The work and its result can be improved upon by synchronizing the cameras so they take pictures at the same timestamps, increasing the amount of usable data. At the same time the capture frequency of the infrared camera should be increased to 30 seconds. After conversation with Miratlas the sensor itself and the internal image processing does not last over a second, so there is no hardware limit to the image capture frequency, theoretically allowing a higher sampling rate. This would further increase the dataset by factor ten. Additionally such a change in frequency enables the use of time information for layer detection and improved segmentation [79]. The clouds can then be distinguished in multilayer conditions by there movement vectors [14].

Furthermore it is not ideal, that the cameras at PSA are located at different weather stations about 270 meters apart from each other. This introduces errors in matching the images, especially when it is not known if the movement is due to perspective and the camera parallax or due to wind conditions and temporal offset. Object matching is already a challenge in cloud imagery, especially if the cameras capture different wavelengths.

All Error sources mentioned so far are theoretically easily preventable and the author did not know about all the complications before the start of this work.

Ideally a camera system is available to which full control is available. This would close loopholes that currently exist, when it comes to camera calibration procedures to which no insight is given. Miratlas states that their cameras are calibrated so they respond with accurate physical temperature values. They confirmed that the most known approach of using Planck's law is used to convert radiation to temperature however they did not provide information about how a clouds emissivity is obtained. Additionally cloud detection is easier achieved in infrared imagery if the camera is calibrated in multiple IR channels instead of the whole range $8 - 14\mu$m.

This so called split window technique is commonly used in satellite imagery to determine sea or land surface temperatures [80, 81]. By using multiple channels the effect of water vapor can be isolated accurately and hence could help distinguish thin ice clouds from clear sky in ground based infrared images. However this could only be achieved with a camera setup of which full calibration control is available.

Overall the height regression results presented in section 5 are satisfactory but far from perfect. Especially the cloud border gradients and the related detection problems for small clouds are expected to improve a lot with better matching data.

The infrared binary segmentation model performs very well and is expected to give similar results for nighttime data. Nevertheless some manual annotated dataset would be needed to put the night time performance into numbers.

6.2 Conclusion

This thesis was driven by the goal of using the newly available infrared all sky image data to improve the current cloud layer segmentation model. Additionally the CBH retrieval from LWIR data has been explored and a cloud detection system for infrared cameras has been developed.

The infrared image data was preprocessed to ground truth height maps using the temperature height relation obtained from ERA5 reanalysis data. Two different CNNs were trained, one to predict cloud heights from visible light images and one to detect clouds in infrared images.

The best height regression model used a data filter for high wind speeds and distinguishes between cloud and clear sky with IoU = 0.773. It predicts the pixel-wise cloud height with MAE = 1046.09m, RMSE = 1375.08m and bias = −399.76m for detected cloudy pixels compared to ceilometer measurements with no clear height specific error trend on the test set. In addition the possibility to combine multiple exposure times to create enhanced image contrasts and detect clouds more easily has been explored. Compared to a fixed exposure time, the enhanced image model performed better when it comes to pure model validation not against ceilometer measurements yielding IoU = 0.777, MAE = 834.90m, RMSE = 1040.14m and bias = −232.07m on the test set.

The predicted height maps were transformed to height layers using the WMO definitions from Table 1. The best performing model was the one trained only on PSA images and yielded mean IoU of 0.372 showing the limited layer distinction capabilities. Visualizing prediction height over ceilometer measurements showed, that most of the predictions fall into the mid to high layer range of 3000m − 8000m region. An error analysis using root sum square method gives an accumulated error of $e_{total} = 1858.33$m over all processing steps.

The binary infrared segmentation model performs strongly on the test set with mean IoU = 0.835 and F1 = 0.908. Visual observation over multiple nighttime images suggests applicability during night, providing a nighttime cloud detection tool.

6.3 Outlook

It should be noted that the author initially aimed to investigate most of the topics mentioned below. However, due to data acquisition challenges and time constraints, it was unfortunately not feasible to explore all these topics in depth.

As mentioned, there are hardware calibration steps, that need to be taken before further working on this topic. Those include time synchronization, spatial identity and higher image capture frequency. Once those problems are resolved and those errors are minimized the models should be retrained on the newly available data. During this work the focus was compulsorily set to obtain an automatic high quality data acquisition tool. This tool can be applied to newly generated data easily, so the dataset theoretically grows by day. Afterwards there are a lot of additional topics that can be tackled.

For once those changes increase the dataset size drastically enabling the possibility to use different model architectures. In recent years there have been many breakthroughs in image segmentation models one being the use of transformer based architectures [82, 83, 84]. A downside often times is the necessity of bigger datasets, but they generally yield more accurate results in different benchmarks. With more available data those architectures would be interesting to investigate.

Additionally the higher frequency capturing and use of cloud movement for layer distinction looks very promising [14] and should be investigated further.

Miratlas stated that they are working on a reliable LWIR clear sky detection method that should soon be available. Once this product is available it should be tested and compared to the current clear sky distinction method. Maybe this product could even replace the IR segmentation model.

Another interesting approach would be to incorporate cloud optical depth in the pipeline. This can be obtained from infrared imagery [85] as well and could be of use to improve cloud border detection.

Bibliography

[1] C. J. Hahn, W. B. Rossow, and S. G. Warren. "ISCCP cloud properties associated with standard cloud types identified in individual surface observations". In: *Journal of Climate* 14.1 (2001), pp. 11–28.

[2] J. T. Coyne, C. L. Baldwin, and K. A. Latorella. "Pilot weather assessment: Implications for visual flight rules flight into instrument meteorological conditions". In: *The International Journal of Aviation Psychology* 18.2 (2008), pp. 153–166.

[3] D. Giggenbach, M. T. Knopp, and C. Fuchs. "Link budget calculation in optical LEO satellite downlinks with on/off-keying and large signal divergence: A simplified methodology". In: *International Journal of Satellite Communications and Networking* 41.5 (2023), pp. 460–476.

[4] G. Schimmel et al. "Free space laser telecommunication through fog". In: *Optica* 5.10 (2018), pp. 1338–1341.

[5] N. B. Blum et al. "Analyzing spatial variations of cloud attenuation by a network of all-sky imagers". In: *Remote Sensing* 14.22 (2022), p. 5685.

[6] International Energy Agency. *Solar PV*. `https://www.iea.org/energy-system/renewables/solar-pv`. Accessed: 2026-06-16. 2026.

[7] J. A. Duffie and W. A. Beckman. *Solar engineering of thermal processes*. John Wiley & Sons, 2013.

[8] E. W. Law, M. Kay, and R. A. Taylor. "Evaluating the benefits of using short-term direct normal irradiance forecasts to operate a concentrated solar thermal plant". In: *Solar Energy* 140 (2016), pp. 93–108.

[9] N. Mahdavi et al. "Model predictive control of distributed air-conditioning loads to compensate fluctuations in solar power". In: *IEEE Transactions on Smart Grid* 8.6 (2017), pp. 3055–3065.

[10] S. Ghosh, S. Rahman, and M. Pipattanasomporn. "Distribution voltage regulation through active power curtailment with PV inverters and solar generation forecasts". In: *IEEE Transactions on Sustainable Energy* 8.1 (2016), pp. 13–22.

[11] B. Nouri and A. Hammer. "Advanced weather forecasting for Renewable Energy System applications: Multi-source observations to improve solar forecasting within the Smart4Res project". In: (2021).

[12] Z. Peng et al. "3D cloud detection and tracking system for solar forecast using multiple sky imagers". In: *Solar Energy* 118 (2015), pp. 496–519.

[13] G. Terrén-Serrano and M. Martınez-Ramón. "Comparative analysis of methods for cloud segmentation in infrared images". In: *arXiv preprint arXiv:2012.06930* (2020).

[14] G. Terrén-Serrano and M. Martínez-Ramón. "Detection of clouds in multiple wind velocity fields using ground-based infrared sky images". In: *Knowledge-Based Systems* 274 (Aug. 2023), p. 110628. ISSN: 0950-7051. DOI: 10.1016/j.knosys.2023.110628. URL: http://dx.doi.org/10.1016/j.knosys.2023.110628.

[15] K. Sommer, W. Kabalan, and R. Brunet. "Infrared radiometric image classification and segmentation of cloud structures using a deep-learning framework from ground-based infrared thermal camera observations". In: *Atmospheric Measurement Techniques* 18.9 (May 2025), pp. 2083–2101. ISSN: 1867-8548. DOI: 10.5194/amt-18-2083-2025. URL: http://dx.doi.org/10.5194/amt-18-2083-2025.

[16] J. A. Shaw and P. W. Nugent. "Physics principles in radiometric infrared imaging of clouds in the atmosphere". In: *European Journal of Physics* 34.6 (Oct. 2013), S111–S121. ISSN: 1361-6404. DOI: 10.1088/0143-0807/34/6/s111. URL: http://dx.doi.org/10.1088/0143-0807/34/6/S111.

[17] D. Klebe et al. "All-sky mid-infrared imagery to characterize sky conditions and improve astronomical observational performance". In: *Publications of the Astronomical Society of the Pacific* 124.922 (2012), pp. 1309–1317.

[18] Y. Fabel et al. "Applying self-supervised learning for semantic cloud segmentation of all-sky images". In: *Atmospheric Measurement Techniques* 15.3 (Feb. 2022), pp. 797–809. ISSN: 1867-8548. DOI: 10.5194/amt-15-797-2022. URL: http://dx.doi.org/10.5194/amt-15-797-2022.

[19] J. M. Wallace and P. V. Hobbs. *Atmospheric science: an introductory survey.* Vol. 92. Elsevier, 2006.

[20] World Meteorological Organization. *International Cloud Atlas.* Accessed: 2026-06-15. World Meteorological Organization. Geneva, Switzerland, 2017. URL: https://cloudatlas.wmo.int/.

[21] F. P. Incropera et al. *Fundamentals of Heat and Mass Transfer.* 7th. Hoboken, NJ: Wiley, 2011.

[22] A. Berk et al. "MODTRAN6: a major upgrade of the MODTRAN radiative transfer code". In: *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX.* Ed. by Miguel Velez-Reyes and Fred A. Kruse. Vol. 9088. SPIE, June 2014, 90880H. DOI: 10.1117/12.2050433. URL: http://dx.doi.org/10.1117/12.2050433.

[23] International Civil Aviation Organization. *International Standard Atmosphere.* Doc 7488/3. ICAO. Montreal, Canada, 1987.

[24] Geosciences LibreTexts. *5.2: Atmospheric Stability and Lapse Rates.* Accessed: 2026-06-14. 2023. URL: https://geo.libretexts.org/Bookshelves/Meteorology_and_Climate_Science/Atmospheric_Processes_and_Phenomena/05%3A_Atmospheric_Stability/5.02%3A_Atmospheric_Stability_and_Lapse_Rates.

[25] Copernicus Climate Change Service, Climate Data Store. *ERA5 hourly data on pressure levels from 1940 to present*. Accessed: 2026-05-01. 2023. DOI: 10.24381/cds.bd0915c6. URL: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels.

[26] A. Heinle, A. Macke, and A. Srivastav. "Automatic cloud classification of whole sky images". In: *Atmospheric Measurement Techniques* 3.3 (2010), pp. 557–567.

[27] Q. Li, W. Lu, and J. Yang. "A hybrid thresholding algorithm for cloud detection on ground-based color images". In: *Journal of atmospheric and oceanic technology* 28.10 (2011), pp. 1286–1296.

[28] S. Liu et al. "Ground-based cloud detection using automatic graph cut". In: *IEEE Geoscience and remote sensing letters* 12.6 (2015), pp. 1342–1346.

[29] C. N. Long et al. "Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images". In: *Journal of Atmospheric and Oceanic Technology* 23.5 (May 2006), pp. 633–652. ISSN: 0739-0572. DOI: 10.1175/jtech1875.1. URL: http://dx.doi.org/10.1175/JTECH1875.1.

[30] A. Kazantzidis et al. "Cloud detection and classification with the use of whole-sky ground-based images". In: *Atmospheric Research* 113 (2012), pp. 80–88. ISSN: 0169-8095. DOI: 10.1016/j.atmosres.2012.05.005. URL: http://dx.doi.org/10.1016/j.atmosres.2012.05.005.

[31] M. Hasenbalg et al. "Benchmarking of six cloud segmentation algorithms for ground-based all-sky imagers". In: *Solar Energy* 201 (2020), pp. 596–614.

[32] A. Taravat et al. "Neural Networks and Support Vector Machine Algorithms for Automatic Cloud Classification of Whole-Sky Ground-Based Images". In: *IEEE Geoscience and Remote Sensing Letters* 12.3 (Mar. 2015), pp. 666–670. ISSN: 1558-0571. DOI: 10.1109/lgrs.2014.2356616. URL: http://dx.doi.org/10.1109/LGRS.2014.2356616.

[33] L. Ye et al. "Supervised Fine-Grained Cloud Detection and Recognition in Whole-Sky Images". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.10 (Oct. 2019), pp. 7972–7985. ISSN: 1558-0644. DOI: 10.1109/tgrs.2019.2917612. URL: http://dx.doi.org/10.1109/TGRS.2019.2917612.

[34] W. Xie et al. "SegCloud: a novel cloud image segmentation model using a deep convolutional neural network for ground-based all-sky-view camera observation". In: *Atmospheric Measurement Techniques* 13.4 (Apr. 2020), pp. 1953–1961. ISSN: 1867-8548. DOI: 10.5194/amt-13-1953-2020. URL: http://dx.doi.org/10.5194/amt-13-1953-2020.

[35] D. Magiera et al. "Advancing semantic cloud segmentation in all-sky images: A semi-supervised learning approach with ceilometer-driven weak labels". In: *Solar Energy* 300 (2025), p. 113822.

[36] C Chow et al. "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed". In: *Solar Energy* 85.11 (2011), pp. 2881–2893.

[37] A. Taravat et al. "Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images". In: *IEEE Geoscience and remote sensing letters* 12.3 (2014), pp. 666–670.

[38] D. Magiera. "Semi-supervised Learning for Probabilistic Cloud Detection in Ground-based Imagery". PhD thesis. RWTH Aachen, 2024.

[39] R. Gut. "Development of advanced cloud classification and segmentation models for solar energy applications using deep learning and synthetic data augmentation". PhD thesis. Hochschule Kempten, 2025.

[40] R. J. Hogan, E. J. O'Connor, and A. J. Illingworth. "Verification of cloud-fraction forecasts". In: *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 135.643 (2009), pp. 1494–1511.

[41] D. Li et al. "Validation and revision of low latitudes cloud base height from ERA5". In: *Atmospheric Research* 309 (Oct. 2024), p. 107595. ISSN: 0169-8095. DOI: `10.1016/j.atmosres.2024.107595`. URL: `http://dx.doi.org/10.1016/j.atmosres.2024.107595`.

[42] M. Inoue et al. "An Assessment of Numerical Weather Prediction–Derived Low-Cloud-Base Height Forecasts". In: *Weather and Forecasting* 30.2 (Apr. 2015), pp. 486–497. ISSN: 1520-0434. DOI: `10.1175/waf-d-14-00052.1`. URL: `http://dx.doi.org/10.1175/WAF-D-14-00052.1`.

[43] D. Nguyen and J. Kleissl. "Stereographic methods for cloud base height determination using two sky imagers". In: *Solar Energy* 107 (2014), pp. 495–509. ISSN: 0038-092X. DOI: `10.1016/j.solener.2014.05.005`. URL: `http://dx.doi.org/10.1016/j.solener.2014.05.005`.

[44] N. B. Blum et al. "Cloud height measurement by a network of all-sky imagers". In: *Atmospheric Measurement Techniques* 14.7 (2021), pp. 5199–5224. ISSN: 1867-8548. DOI: `10.5194/amt-14-5199-2021`. URL: `http://dx.doi.org/10.5194/amt-14-5199-2021`.

[45] Z. Tan et al. "Improving Satellite-Retrieved Cloud Base Height with Ground-Based Cloud Radar Measurements". In: *Advances in Atmospheric Sciences* 41.11 (2024), pp. 2131–2140. ISSN: 1861-9533. DOI: `10.1007/s00376-024-4052-7`. URL: `http://dx.doi.org/10.1007/s00376-024-4052-7`.

[46] D. M. Romps and R. Öktem. "Observing Clouds in 4D with Multiview Stereophotogrammetry". In: *Bulletin of the American Meteorological Society* 99.12 (Dec. 2018), pp. 2575–2586. ISSN: 1520-0477. DOI: `10.1175/bams-d-18-0029.1`. URL: `http://dx.doi.org/10.1175/BAMS-D-18-0029.1`.

[47] Y. Dong, X. Sun, and Q. Li. "A Method for Retrieving Cloud-Top Height Based on a Machine Learning Model Using the Himawari-8 Combined with Near Infrared Data". In: *Remote Sensing* 14.24 (Dec. 2022), p. 6367. ISSN: 2072-4292. DOI: `10.3390/rs14246367`. URL: `http://dx.doi.org/10.3390/rs14246367`.

[48] E. Hirsch, E. Agassi, and I. Koren. "A novel technique for extracting clouds base height using ground based imaging". In: *Atmospheric Measurement Techniques* 4.1 (Jan. 2011), pp. 117–130. ISSN: 1867-8548. DOI: 10.5194/amt-4-117-2011. URL: http://dx.doi.org/10.5194/amt-4-117-2011.

[49] Z. Wang et al. "Cloud-Base Height Derived from a Ground-Based Infrared Sensor and a Comparison with a Collocated Cloud Radar". In: *Journal of Atmospheric and Oceanic Technology* 35.4 (Apr. 2018), pp. 689–704. ISSN: 1520-0426. DOI: 10.1175/jtech-d-17-0107.1. URL: http://dx.doi.org/10.1175/JTECH-D-17-0107.1.

[50] G. Terrén-Serrano and M. Martínez-Ramón. "Deep learning for intra-hour solar forecasting with fusion of features extracted from infrared sky images". In: *Information Fusion* 95 (2023), pp. 42–61.

[51] Y. Wang et al. "Day and Night Clouds Detection Using a Thermal-Infrared All-Sky-View Camera". In: *Remote Sensing* 13.9 (May 2021), p. 1852. ISSN: 2072-4292. DOI: 10.3390/rs13091852. URL: http://dx.doi.org/10.3390/rs13091852.

[52] B. Thurairajah and J. A. Shaw. "Cloud statistics measured with the infrared cloud imager (ICI)". In: *IEEE Transactions on Geoscience and Remote Sensing* 43.9 (2005), pp. 2000–2007.

[53] L. Liu et al. "Comparison of Cloud Properties from Ground-Based Infrared Cloud Measurement and Visual Observations". In: *Journal of Atmospheric and Oceanic Technology* 30.6 (2013), pp. 1171–1179. ISSN: 1520-0426. DOI: 10.1175/jtech-d-12-00157.1. URL: http://dx.doi.org/10.1175/JTECH-D-12-00157.1.

[54] Q. Luo et al. "Cloud classification of ground-based infrared images combining manifold and texture features". In: *Atmospheric Measurement Techniques* 11.9 (2018), pp. 5351–5361. ISSN: 1867-8548. DOI: 10.5194/amt-11-5351-2018. URL: http://dx.doi.org/10.5194/amt-11-5351-2018.

[55] Y. Fabel. "Cloud Segmentation and Classification from All-Sky Images Using Deep Learning". PhD thesis. TU München, 2020.

[56] K. Sassen and Z. Wang. "The Clouds of the Middle Troposphere: Composition, Radiative Impact, and Global Distribution". In: *Surveys in Geophysics* 33.3-4 (Dec. 2011), pp. 677–691. ISSN: 1573-0956. DOI: 10.1007/s10712-011-9163-x. URL: http://dx.doi.org/10.1007/s10712-011-9163-x.

[57] B. Nouri et al. "Cloud height and tracking accuracy of three all sky imager systems for individual clouds". In: *Solar Energy* 177 (Jan. 2019), pp. 213–228. ISSN: 0038-092X. DOI: 10.1016/j.solener.2018.10.079. URL: http://dx.doi.org/10.1016/j.solener.2018.10.079.

[58] F. Weiler. "Cloud Base Height Estimation Using Stereo All-Sky Imagery and Deep Learning". PhD thesis. TU München, 2025.

[59] Climate-Data.org. *Klima Trauen (Deutschland)*. https://de.climate-data.org/europa/deutschland/niedersachsen/trauen-184005/. Accessed: 2026-04-16.

[60] Climate-Data.org. *Klima Tabernas (Spanien).* `https://de.climate-data.org/europa/spanien/andalusien/tabernas-29965/`. Accessed: 2026-04-16.

[61] MiraAtlas. *Turbulence Monitoring.* `https://miratlas.com/turbulence-monitoring-seeing/`. Accessed: 2026-04-21.

[62] Lufft GmbH. *Ceilometer CHM 15k "NIMBUS".* `https://www.lufft.com/de-de/produkte/wolken-schneehoehensensoren-306/ceilometer-chm-15k-nimbus-2093/`. Accessed: 2026-04-16.

[63] S. Boyd et al. "Analysis algorithm for sky type and ice halo recognition in all-sky images". In: *Atmospheric Measurement Techniques* 12.8 (Aug. 2019), pp. 4241–4259. ISSN: 1867-8548. DOI: `10.5194/amt-12-4241-2019`. URL: `http://dx.doi.org/10.5194/amt-12-4241-2019`.

[64] F Kasten and A. T. Young. "Revised optical air mass tables and approximation formula". In: *Applied Optics* 28.22 (Nov. 1989), p. 4735. ISSN: 1539-4522. DOI: `10.1364/ao.28.004735`. URL: `http://dx.doi.org/10.1364/AO.28.004735`.

[65] D. Klebe, R. D. Blatherwick, and V. R. Morris. "Ground-based all-sky mid-infrared and visible imagery for purposes of characterizing cloud properties". In: *Atmospheric Measurement Techniques* 7.2 (2014), pp. 637–645.

[66] H. Hersbach et al. *ERA5 hourly data on pressure levels from 1940 to present.* Accessed: 2026-05-01. 2023. DOI: `10.24381/cds.bd0915c6`. URL: `https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels`.

[67] R. Hassan et al. "Assessment of reanalysis datasets against radiosonde observation over the Eastern Mediterranean region". In: *Acta Geophysica* 71.6 (2023), pp. 2989–3005. ISSN: 1895-7455. DOI: `10.1007/s11600-023-01121-5`. URL: `http://dx.doi.org/10.1007/s11600-023-01121-5`.

[68] T. Hastie. *The elements of statistical learning: data mining, inference, and prediction.* 2009.

[69] K. Hornik. "Approximation capabilities of multilayer feedforward networks". In: *Neural networks* 4.2 (1991), pp. 251–257.

[70] X. Sun, L. Liu, and S. Zhao. "Whole sky infrared remote sensing of cloud". In: *Procedia Earth and Planetary Science* 2 (2011), pp. 278–283.

[71] Z. Zou et al. "Object detection in 20 years: A survey". In: *Proceedings of the IEEE* 111.3 (2023), pp. 257–276.

[72] M. Kotthapalli, D. Ravipati, and R. Bhatia. "Yolov1 to yolov11: A comprehensive survey of real-time object detection innovations and challenges". In: *arXiv preprint arXiv:2508.02067* (2025).

[73] J. Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016, pp. 779–788.

[74] T. Mertens, J. Kautz, and F. Van Reeth. "Exposure fusion". In: *15th Pacific Conference on Computer Graphics and Applications (PG'07)*. IEEE. 2007, pp. 382–390.

[75] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[76] L. Chen et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 801–818.

[77] K. He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[78] Z. Wang et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE transactions on image processing* 13.4 (2004), pp. 600–612.

[79] D. Pathak et al. "Learning features by watching objects move". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2701–2710.

[80] E. P. McClain, W. G. Pichel, and C. C. Walton. "Comparative performance of AVHRR-based multichannel sea surface temperatures". In: *Journal of Geophysical Research: Oceans* 90.C6 (1985), pp. 11587–11601.

[81] J. A. Sobrino et al. "Improvements in the split-window technique for land surface temperature determination". In: *IEEE Transactions on Geoscience and Remote Sensing* 32.2 (1994), pp. 243–253.

[82] A. Kirillov et al. "Segment anything". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2023, pp. 4015–4026.

[83] E. Xie et al. *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*. 2021. DOI: `10.48550/ARXIV.2105.15203`. URL: `https://arxiv.org/abs/2105.15203`.

[84] Y. Fu, M. Lou, and Y. Yu. "SegMAN: Omni-scale context modeling with state space models and local attention for semantic segmentation". In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, pp. 19077–19087.

[85] J. A. Shaw et al. "Cloud optical depth measured with ground-based, uncooled infrared imagers". In: *Remote Sensing of the Atmosphere, Clouds, and Precipitation IV*. Ed. by Tadahiro Hayasaka, Kenji Nakamura, and Eastwood Im. Vol. 8523. SPIE, Nov. 2012, p. 85231D. DOI: `10.1117/12.981294`. URL: `http://dx.doi.org/10.1117/12.981294`.