

IRIS-CloudDeep: Infrared Radiometric Image classification and Segmentation of Cloud structure using Deep-learning framework for ground-based long-wave infrared thermal camera observations

Kélian Sommer  ^{1,*}, Wassim Kabalan^{2,*}, and Romain Brunet^{3,*}

¹Laboratoire Univers et Particules de Montpellier, Université de Montpellier, CNRS, Montpellier, France

²Université Paris Cité, CNRS, Astroparticule et Cosmologie, F-75013 Paris, France

³Laboratoire d’Astrophysique de Marseille, Marseille, France

Correspondence: Kélian Sommer (kelian.sommer@umontpellier.fr), Wassim Kabalan (wassim@apc.in2p3.fr) and Romain Brunet (romain.brunet@lam.fr)

Abstract. Infrared thermal cameras offer reliable means of assessing atmospheric conditions by measuring the downward radiance from the sky, facilitating their usage in cloud monitoring endeavors. Precise identification and detection of clouds in images poses great challenges stemming from the indistinct boundaries inherent to cloud formations. Various methodologies for segmentation have been previously suggested. Most of them rely on color as the distinguishing criterion for cloud identification in the visible spectral domain and thus lack the ability to detect cloud structure on gray-scaled images with satisfying accuracy. In this work, we propose a new complete deep-learning framework to perform image classification with Convolutional Neural Networks (CNNs) and image segmentation by exploring the use of U-Net model. We demonstrate the effectiveness of this technique by conducting a series of tests and validations on self-captured infrared sky images and transformed publicly available datasets. Our findings reveal that the models can effectively differentiate between image types and accurately capture detailed cloud structure information, even when trained with a single binary ground-truth mask per input sample. We also compare our framework with other state-of-the-art machine-learning methods and conclude that it outperforms them when applied to infrared data. The classifier model achieves an excellent accuracy of 93% in image types distinction, while the segmentation model attains a mean pixel accuracy of 94% on our own dataset. We emphasize that our framework exhibits strong viability and can be used for infrared thermal ground-based cloud monitoring operations over extended durations.

1 Introduction

Accurate and continuous monitoring of cloud properties contributes to a profound understanding of atmospheric processes and their subsequent impacts on various Earth systems (Liou, 1992). It provides essential insights for weather predictions and climate dynamics (Hu et al., 2004; Petzold et al., 2015). Many diverse instruments are dedicated towards cloud detection. Observation methods can be divided into two primary distinct categories: downward satellite-based observations (Roy et al., 2017; Martin, 2008) and upward ground-based observations with all-sky cameras, lidar, radar and other instruments (Wilczak et al., 1996). The principal aim of satellite-based observations is to investigate the upper regions of clouds, facilitating the examination and analysis of global atmospheric patterns and climate conditions over expansive geographical areas (Schiffer and Rossow, 1983; Boers et al., 2006; Geer et al., 2017; Várnai and Marshak, 2018). In contrast, ground-based cloud observation excels in the surveillance of localized regions, furnishing valuable data pertaining to the lower segments of clouds by giving information on cloud altitude, cloud extent, and cloud typology (Bower et al., 2000; Zhou et al., 2019). Combination of these two measurement techniques enhances our overall comprehension of cloud behavior (Mokhov and Schlesinger, 1994; Schreiner et al., 1993; Yamashita and Yoshimura, 2012; Yoshimura and Yamashita, 2013).

Ground-based observations have been extensively used in recent years and have become a viable means to detect, study and identify cloud formations (Paczyński, 2000; Skidmore et al., 2008; Tzoumanikas et al., 2016; Ugolnikov et al., 2017; Mommert, 2020; Tzoumanikas et al., 2016; Román et al., 2022). As technological evolution has ushered in a new era of monitoring methodologies (Mandat et al., 2014), the utilization of infrared thermal cameras has emerged as a promising avenue for atmospheric investigations through precise radiometric measurements (Szejwach, 1982; Shaw and Nugent, 2013; Liandrat et al., 2017b; Lopez et al., 2017; Klebe et al., 2014; Nikolenko and Maslov, 2021).

Because of their practical use, high sensitivity, low-cost, operating range and wide field-of-view (FOV) (Rogalski, 2011; Rogalski and Chrzanowski, 2014; Kimata, 2018), it makes them particularly useful for medicine (Ring and Ammer, 2012), agriculture (Ishimwe et al., 2014), aerial (Wilczak et al., 1996), defense (Gallo et al., 1993; Akula et al., 2011), surveillance (Wong et al., 2009), weather forecast (Sun et al., 2008; Liandrat et al., 2017a), or even astronomical related applications to determine the cloud cover fraction during operations and therefore assess the quality of scientific observations (Sebag et al., 2010; Lewis et al., 2010; Klebe et al., 2012, 2014; Reil et al., 2014). Indeed, uncooled 25 infrared microbolometers array sensors working in the 10–12 μm spectral band can directly detect the LWIR emission of both clouds and the atmospheric background, excluding the scattered light of the sun or starlight (Houghton and Lee, 1972). These LWIR sensors are able to provide high-contrast 30 images and allow fine radiometric measurements to detect low-emissivity cirrus clouds (Lewis et al., 2010; Shaw and Nugent, 2013).

Across recent years, multiple automatic ground-based observations systems have been developed. For example, the infrared cloud imager (ICI) (Thuraiyah and Shaw, 2005), can detect clouds and assess cloud coverage both in daylight and at nighttime with a dedicated infrared sensor. Sharma et al. (2015) designed an instrument to detect of the cloud infrared radiations to be used in search for a potential site 40 for India's National Large Optical Telescope project. The development of the Radiometric All-Sky Infrared Camera (RASICAM, referenced in Lewis et al. 2010 and Reil et al. 2014) was aimed at enabling automated, real-time quantitative evaluation of nighttime sky conditions for the Dark Energy Survey (DES see Collaboration 2005). This specialized camera is designed for the precise detection of the location, movement, and optical properties of thin, high-altitude cirrus clouds and contrails by measuring their brightness temperature against the sky background. The all-sky infrared visible analyzer (ASIVA, see Klebe et al. 2014) is a similar instrument whose primary goal is to provide radiometrically calibrated imagery in the LWIR band to estimate fractional sky cover and sky/cloud brightness temperature, emissivity and cloud height. The ASC-200 system (Wang et al., 2021b) 55 combines information from two all-sky cameras facing the

sky operating in both the visible spectrum (450–650 nm) and the LWIR band.

As next-generation cosmological surveys require more demanding precision on photometric observations (implying better characterization of the atmosphere), monitoring telescope instruments field-of-view (FOV) with LWIR thermal cameras may provide significant asset to ; (i) classify observations quality in real-time; (ii) evaluate potential cloud coverage (Smith and Toumi, 2008; Liandrat et al., 2017b; Aebi et al., 2018; Wang et al., 2021b); (iii) estimate precipitable water vapor content (PWV) (Kelsey et al., 2022; Hack et al., 2023; Salamatikis et al., 2023).

In this study, we plan to address the first objective. We use a LWIR thermal infrared camera with a specifically chosen narrower FOV that aims to image the surrounding area of the StarDICE telescope FOV. The StarDICE metrology experiment (Betoule et al., 2022) that aims at measuring CALSPEC (Bohlin, 2014) spectrophotometric standard stars absolute flux at the 0.1% relative uncertainty level. Enhanced characterization of atmospheric conditions are required to reach the target sensitivity. As a preliminary step, basic knowledge of the atmosphere conditions in the telescope FOV may provide valuable insights onto the quality of spectrophotometric measurements. However, these kind of infrared instruments operate at high framerate and produce considerable amounts of data which makes it extremely difficult to analyze by human observers. Therefore, to determine cloud presence in infrared images, deep convolutional neural networks appear to be a viable approach to process images in real-time. Multiple models relying on convolutional neural network (CNN) have been developed such as: CloudSegnet (Dev et al., 2019a), CloudU-Net (Shi et al., 2021b) CloudU-Netv2 (Shi et al., 2021a), SegCloud (Xie et al., 2020), TransCloudSeg (Liu et al., 2022), CloudDeepLabV3 (Li et al., 2023), ACLNet (Makwana et al., 2022), DeepCloud (Ye et al., 2017), CloudRaednet (Shi et al., 2022), DMNet (Zhao et al., 2022) and DPNet Zhang et al. (2022a). Nonetheless, these methodologies exclusively address RGB-colored images (Li et al., 2011; Dev et al., 2016). Colors or hue provides the essential of information for segmentation (especially red and blue channels). In the case of LWIR thermal images, we implement a model capable of achieving comparable accuracy for single-channel gray-scaled images. Inspired by their large successes in image classification and structure detection for various computer vision tasks, we propose a new dedicated deep-learning framework. Our approach is specifically designed towards gray-scaled infrared images and consists of: (i) classify images (e.g. detect if any cloud is present onto the image and discriminates between clear and cloudy images); (ii) identify cloud structure (e.g., generate a pixel-based probabilistic segmentation map and verify if the CCD camera FOV is impacted). The main contributions of this paper are threefold:

The remainder of the paper is structured as follows. Background about the scientific context and related works are

presented in Section 2. Section 3 details the experimental setup and dataset. Section 4 introduces the proposed framework, describing deep-learning architectures and training procedures. Experimental results and comparisons with other methods are provided in Section 5. Relevant matters and future perspectives are discussed in Section 6. Section 7 depicts a summary and finally concludes the paper.

2 Background

2.1 Motivation

¹⁰ KS : may need to shorten the motivation subsection to be more synthetic and straight to the point. Is it necessary to explain the StarDICE project principles ? Maybe not

StarDICE represents one of the initiatives focused on creating a measurement process that bridges the gap between laboratory flux standards (such as silicon photodiodes calibrated by NIST) and the stars found in the CALSPEC library of spectrophotometric references (Bohlin et al., 2020). Since type 1a supernovae and most astronomical surveys rely on the calibration of these standard stars for their measurements (Bohlin et al., 2011; Betoule et al., 2013; Rubin et al., 2015; Scolnic et al., 2015; Currie et al., 2020; Brout et al., 2022; Rubin et al., 2022), successfully establishing this connection with high precision effectively addresses the calibration challenge associated with the Hubble diagram. The StarDICE proposal encompasses a five-step sequence, as illustrated in Figure 1. This process hinges on the near-field calibration of a dim but stable light source, emitting less than $1 \mu\text{W}$ of optical power. It serves as a distant ($\sim 100 \text{ m}$) in-situ reference for a compact astronomical telescope. One of the largest remaining source of systematic uncertainty is Earth atmosphere transmission (Stubbs and Tonry, 2012; Stubbs and Brown, 2015; Li et al., 2016). It is dependent on many environmental conditions and processes, including: absorption and scattering by molecular constituents (O_2 , O_3 , and others), absorption by precipitable water vapor (PWV), scattering by aerosols, and shadowing by larger ice crystals and water droplets in clouds that is independent of wavelength and responsible for *gray extinction* (Burke et al., 2010, 2017). Current atmospheric transmission or extinction models do not integrate the possible impact of cloud. Indeed, formation of thin clouds through the condensation of water droplets and ice can result in clouds that are extremely faint and cannot be perceived in the visible spectrum with the naked eye. These clouds often exhibit complex spatial structures, as demonstrated in Burke et al. 2013. The condensation process takes place along well-defined boundaries in temperature and pressure, which are influenced by PWV. This phenomenon commonly categorizes observing conditions as either *photometric* or *non-photometric*. To address this issue, we propose to use an infrared thermal camera providing high-sensitivity radiometric measurements of the sky

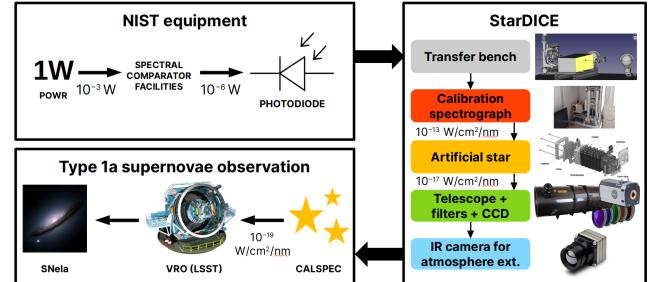


Figure 1. Schematics of the metrology chain of the StarDICE experiment. Each arrow represents a step in the chain and the label gives the order of magnitude of the beam intensity. The steps in upper left box are conducted at NIST (Larason and Houston, 2008) and result in a silicon photodiode calibrated against an electrical substitution cryogenic radiometer. The StarDICE collaboration builds the dedicated bench transfer system in the right box, designed to reach $10^{-19} \text{ W/cm}^2/\text{nm}$ flux sensitivity. Primary calibration stars catalog is then used by large cosmological surveys for absolute flux calibration.

radiance in the atmosphere transparency window ($10\text{-}12 \mu\text{m}$). Previous work demonstrated our ability to calibrate it with better than $0.1 \text{ W/m}^2/\text{sr}$ on each pixel, corresponding to the added radiance of a theoretical cirrus cloud at surface temperature of 205 K with a visible optical depth of $\tau = 0.02$. With the help of cloud spatial structure, this instrument may be the key to assess photometric observations quality and label science images with superior than state-of-the-art accuracy.

2.2 Related work

In recent years, numerous cloud sky/cloud segmentation algorithms have been introduced along with the increased development of all-sky ground-based cloud monitoring stations (Long et al., 2006; Yang et al., 2012; Krauz et al., 2020; Fa et al., 2019; Mommert, 2020; Li et al., 2022). Indeed, cloud segmentation is big challenge for remote sensing applications as clouds come in various shapes and forms. Therefore, the most modern common approach aims to use computer vision encoder-decoder architectures algorithms and train them onto very specific publicly available cloud image databases such as: SWIMSEG (Dev et al., 2016), SWIN-SEG (Dev et al., 2019b, 2017), SWINySEG (Dev et al., 2019a), WSISEG (Xie et al., 2020), HYTA (Li et al., 2011) and TLCDD. Most proposed solutions are focused on visible RGB images. CloudSegNet (Dev et al., 2019a) is a light-weight deep-learning encoder/decoder network that detects clouds onto daytime and nighttime visible color images. CloudU-Net (Shi et al., 2021b) modifies CloudSegNet architecture by adding dilated convolution, skip connection, and fully connected conditional random field (CRF, see McCallum 2012) layers to demonstrates better segmentation performance overall. It uses the powerful U-Net architecture (Ronneberger et al., 2015) originally applied to medi-

cal image segmentation. CloudU-Netv2 (Shi et al., 2021a) replaces the upsampling in CloudU-Net with bilinear up-sampling, improves discrimination ability of features representation and uses rectified Adam optimizer (rADAM is a variant of the Adam stochastic optimizer (Kingma and Ba, 2014) that introduces a term to rectify the variance of the adaptive learning rate, see Liu et al. 2019). SegCloud (Xie et al., 2020) has been trained onto 400 images and possesses a symmetric encoder-decoder structure and outputs low/high-level cloud feature maps to the same resolution of input images. TransCloudSeg (Liu et al., 2022) addresses the loss of global information due to limited receptive field size of the filters in CNN by proposing an hybrid model containing both the CNN and a transformer (Vaswani et al., 2023) as the encoders to obtain different features. Cloud-DeepLabV3+ (Li et al., 2023) designs a lightweight ground-based cloud image adaptive segmentation method that integrates multi-scale features aggregation and multi-level attention feature enhancement. ACLNet (Makwana et al., 2022) uses EfficientNet-B0 as the backbone, “à trous spatial pyramid pooling” (ASPP see Chen et al. 2017) to learn at multiple receptive fields, and global attention module (GAM see Liu et al. 2021) to extract fine-grained details from the image. It provides lower error rate, higher recall and higher F1-score than state-of-art cloud segmentation models. Deep-Cloud (Ye et al., 2017) uses the method of Fisher vector encoding which is applied to executing the spatial feature aggregation and high-dimensional feature mapping on the raw deep convolutional features. CloudRaednet (Shi et al., 2022) proposes a residual attention-based encoder-decoder network and train it over the SWINySEG dataset. Zhang et al. 2022b introduces a novel deep model named multiscale attention convolutional neural network (MACNN) to obtain different receptive fields by using different hole rates for the filters and propose the attention module to learn the attention coefficients in order to reflect different importance of pixels. DMNet (Zhao et al., 2022) proposes a novel cloud detection network that aims to achieve information complementarity by exploiting the different properties of the features at different levels of the encoder, so as to strengthen the detailed information of high-level features and make the low-level features have more semantics. DPNet (Zhang et al., 2022a) possesses an encoder-decoder structure with Dual Pyramid Pooling Module (DPPM). They process the feature maps of different scales in the encoder through a technique known as dual pyramid pooling. They also implement the Encoder-Decoder Constraint (EDC) to relieve information loss in the process of encoding and decoding.

Majority of these models are typically structured using an encoder-decoder architecture, which is the primary innovation brought forth by incorporating CNNs (O’Shea and Nash, 2015). The encoder is tailored to acquire representational features, facilitating the extraction of semantic information while the decoder reconstructs these representational features into the segmentation mask, allowing for pixel-level

classification (Badrinarayanan et al., 2017; Alzubaidi et al., 2021).

Others have proposed solutions for all-sky infrared image classification. Liu et al. (2011) applies pre-processing steps (smoothing noise reduction, enhancement through top-hat transformation and a high-pass filtering, edges detection) before extracting features that are useful for distinguishing cirriform, cumuliform, and waveform clouds. A simple rectangle method as supervised classifier is applied. They find a 90% agreement between a priori classification carried out manually by visual inspection and their algorithm on 277 images. Sun et al. (2011) suggested: (i) a method for determining clear sky radiance threshold; (ii) cloud identification combined threshold method with texture method; (iii) an algorithm to retrieve cloud base height from downwelling infrared radiance. They showed that structural features are better than texture features in classifying cloud. Luo et al. (2018) proposed a three-step process: (i) pre-processing; (ii) feature extraction; (iii) classification method to group images into five cloud categories (stratiform, cumuliform, waveform, cirriform and clear) based on manifold and texture features using support vector machine (SVM see Cortes and Vapnik 1995). Their experimental results demonstrate the higher recognition rate with an increase of 2%-10% on ground-based infrared images datasets. Nevertheless, these methods class clouds into separate categories based on their typology. Until now, all the previously examined approaches, while effective within their specific domains, proved to be unsuccessful when applied to our particular use case. Therefore, we propose a new deep-learning framework based on CNNs and U-Net architectures to identify cloud images and detect cloud structure in real-time.

3 Experimental setup and datasets

3.1 Description of the infrared thermal camera

Our instrument is an infrared thermal camera, specifically the FLIR Tau2, which operates in the long-wave infrared (LWIR) band, covering the $8\text{-}14 \mu\text{m}$ range. It features a focal plane array (FPA) consisting of 640×512 uncooled microbolometers, capturing images at a framerate of 9 Hz. To achieve a narrow field of view (FOV), we’ve paired the camera with a 60 mm F1.25 lens. The primary purpose of deploying this instrument on the equatorial mount adjacent to the StarDICE photometric telescope is to continuously assess the atmospheric conditions (specifically gray extinction) within the line of sight of the visible CCD camera during observations. Through meticulous calibration, radiative transfer calculations, and data analysis using simulations, we can extract valuable information about the sky to monitor real-time atmospheric conditions. In Figure 2, we show the instrument mounted on the equatorial mount inside the observatory dome, complete with the necessary command and

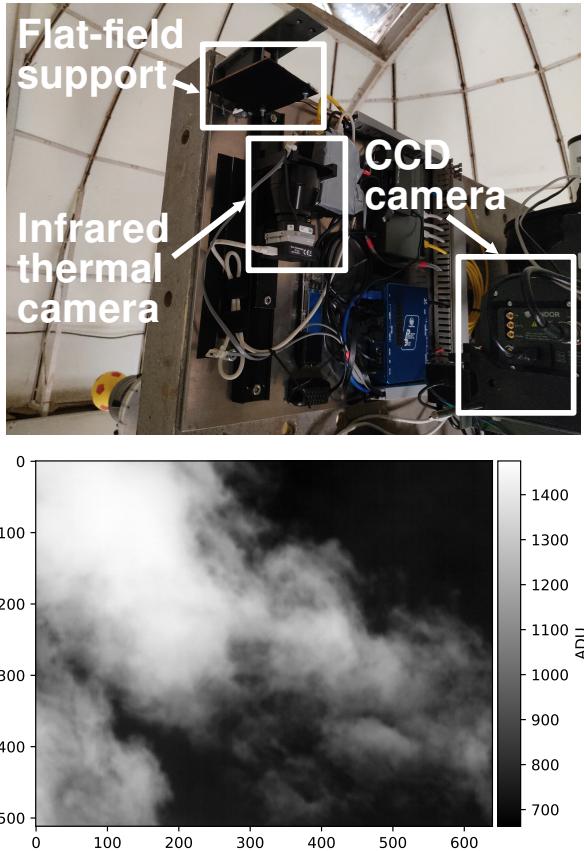


Figure 2. *Top:* infrared instrument installed onto the equatorial table of the StarDICE experiment at Observatoire de Haute-Provence. *Bottom:* sample of raw infrared thermal image in ADU.

control equipment. We also monitor the surrounding and internal temperatures of the camera in real-time to correct for temperature-related variations in sensor response. The device is controlled and commanded via the ThermalCapture ThermalGrabber USB 2.0 interface¹, which grants access to full 14-bit radiometric raw data. We have developed an open-access PYTHON program, available on GitHub², to control the camera's functions and capture images. These images are saved in FITS format. To correct for anisotropies caused by fixed pattern noise (FPN), we employ an enhanced flat-field calibration source at regular intervals of approximately 30 seconds. True scene radiances are computed from raw images with per-pixel coefficient matrices from pre-established calibration procedure.

3.2 Datasets and pre-processing

A substantial quantity of images is essential for the effective training and testing of both the classifier and segmentation algorithms. Our dataset comprises LWIR sky images that

we captured ourselves. It encompasses XXX sky images acquired by the infrared thermal camera. To speed up computations and minimize memory consumption, we downsampled the original-sized images (640×512) into a 160×128 resolution by binning them in a 4×4 format. Cloudy sky images were collected during a three-night period at Observatoire de Haute-Provence in January 2023 ($43^{\circ} 55' 51''$ N, $5^{\circ} 42' 48''$ E). Conversely, cloud-free images were obtained over a short time span during the same month.

To compensate for the lack of cloud-free images and prevent potential biases in training due to data imbalance, we generated synthetic cloud-free images to create a composite dataset containing as many images as the cloudy dataset. These synthetic images replicate realistic observations by simulating 2D horizontal gradients, mimicking the increase in sky downwelling radiance as the camera's field of view tilts toward high zenith angles (i.e., low elevation angles). Realistic sources of noises affecting uncooled infrared thermal cameras are introduced, including: read noise, fixed pattern noise, sky noise and narcissus effect. This addition ensures that the spatial noise in the synthetic images closely resembles that of actual cloud-free images. Figure 3 illustrates a typical cloud-free image alongside a synthetically generated one, with spatial noise indicated for each. It's worth noting that the absolute analog-to-digital unit (ADU) value has no impact, as the data is normalized prior to training

All images and masks are visually inspected. Samples presenting artifacts such as tree branches from surroundings or buildings in the FOV corners are discarded. As the camera acquisition framerate enables to record up to 9 images per second, the pre-processing algorithm included constraints on consecutive image selection based on their timeseries. Selected frames are taken from at least 2 seconds between each other to introduce a wider range of displayed clouds.

Ground-truth masks identifying cloud structure on cloud images were manually created through multiple steps of stretching procedures using ASTROPY (Astropy Collaboration, 2013, 2018) methods for each image in the dataset. They consist of boolean 2D array of the same image size, where *True* identified pixels represent cloud pixels and *False* identified pixels represent clear sky areas. This step could not be automated as cloud optical depths, brightness temperatures and structure differ significantly. Indeed, the segmentation model aims at automating this tedious time consuming operation in real-time during observations. Figure 4 depicts two raw images with their associated manually generated ground-truth cloud masks for training purposes.

Furthermore, we performed multiple random augmentations (e.g, flip, shear, rotate, shift and zoom) on each original image to artificially enlarge the size of each dataset and reduce overfitting (Perez and Wang, 2017; Mikolajczyk and Grochowski, 2018; Yang et al., 2022). All augmented images are produced through the random sequential applications of these five distinct operations to initial images. These operations are executed with a random varying degree of inten-

¹<https://thermalcapture.com/>

²https://github.com/Kelian98/tau2_thermalcapture

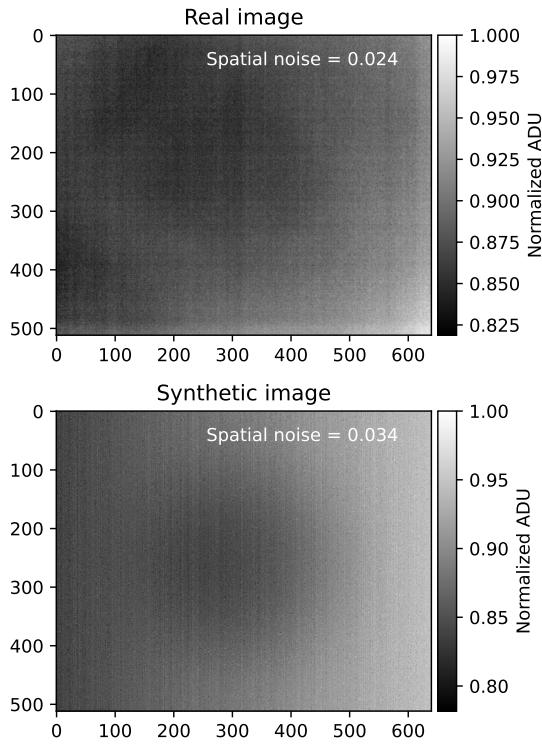


Figure 3. Comparison between real observed clear sky image (top) and synthetically generated realistic image (bottom). Spatial noise is marked on the top right corner of each image.

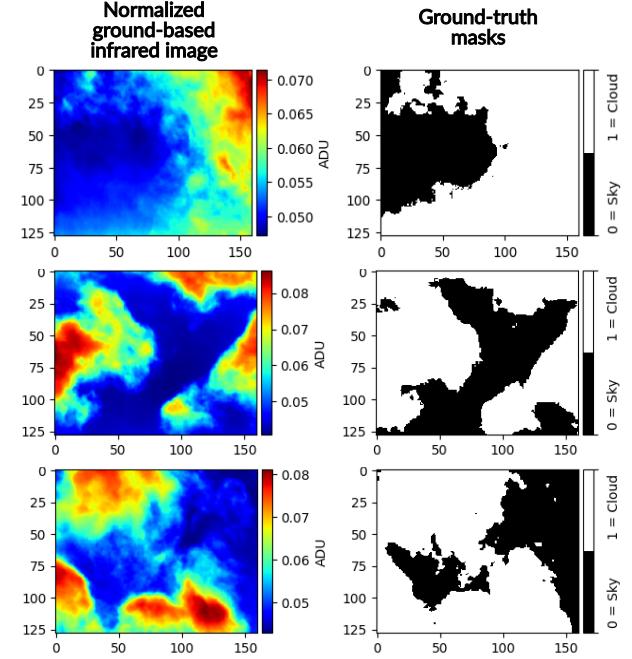


Figure 4. Representative infrared cloudy sky images and their corresponding manually created ground-truth masks. The masks are binary images, where zero represents clear sky and one represents cloud. **KS :** the last two rows are too similar images. Ideal would be to have the last one in the middle row and add a very cloudy image.

Table 1. Overview of the collected image datasets for classification and segmentation.

Image type	# training	# validation	# total
Clear	XXXX	XXXX	XXXX
Cloud	XXXX	XXXX	XXXX

sity contained in specific ranges. Random rotations are applied within an amplitude ranging from -45 to +45 degrees. Shear is introduced with a random magnitude ranging from -0.2 to +0.2. Shifting operations are carried out with a maximum ratio of 10% in both width and height directions to avoid generation of unrealistic symmetric structures. Zoom operation is applied within the range of 1 to 1.8. No other transformation such as histogram equalization or contrast enhancement is applied to prevent any bias or alteration in the segmentation performance. After the selection and augmentation procedures, we conducted a visual examination of all the created sky/cloud images to ensure that they appear realistic. Since all the parameters in the image augmentation process undergo controlled adjustments, our generated images closely mirror authentic sky/cloud scenes. Datasets are for both models are splitted into training, testing and validation subsets with ratio of in the ratio of 70%, 20% and 10% respectively. Table 1 depicts the numbers and properties of the datasets.

4 Methodology

4.1 Overall framework of IRIS-CloudDeep

In this section, we outline the architectural designs of two distinct deep-learning models tailored for cloud-related tasks. On the one hand, we implement a classifier for cloud classification using Convolutional Neural Networks (CNN) (LeCun et al., 1995; Krizhevsky et al., 2012), whose specific goal is to discriminate between cloud-free and cloudy images. On the other hand, the segmentation for cloud structure detection is performed via an optimized U-Net model (Ronneberger et al., 2015). The output probability map can later be thresholded according to the user needs in order to produce the desired predicted binary segmentation map. Fig-

ure 5 illustrates the proposed deep-learning framework compared to conventional segmentation algorithms.

4.2 Image classification

WK : review

For our image classification model, we employed a Convolutional Neural Network (CNN) architecture (Sharma et al., 2018) derived from the VGG-16 network (Simonyan and Zisserman, 2015), which has proven to be highly effective in image recognition tasks (Canziani et al., 2016) without introducing much complexity as ResNets (He et al., 2015). VGG-16 is a convolutional neural network that is 16 layers deep whereas ResNet generally have higher depths. This modified network was designed with the primary objective of distinguishing between images that contain clouds and those that do not and is similar to SegCloud (Xie et al., 2020).

The VGG-16 architecture serves as the backbone of our model. It consists of a series of convolutional layers, followed by rectified linear unit (ReLU) activation (Agarap, 2018) and max-pooling operations. This configuration facilitates the extraction of hierarchical features that are important for accurate classification. We retained the convolutional layers and their weight parameters from the original VGG-16 model to benefit from the network's ability to capture intricate visual patterns.

To adapt the network to our binary classification task, we made adjustments to the fully connected layers towards the end of the architecture. Specifically, we replaced the original fully connected layers with a custom set of fully connected layers. These modified layers were designed to map the learned features to the two classes of interest: images containing clouds and images without clouds. The final output layer consisted of two neurons, each representing one of the classes, and a softmax activation function was applied to obtain the class probabilities.

Furthermore, we incorporated dropout layers after the fully connected layers to mitigate overfitting and enhance the generalization capability of our model. This architectural modification helped us strike a balance between model complexity and performance, ensuring that the network could effectively differentiate between cloud and non-cloud images.

The model is trained on a comprehensive dataset encompassing both cloud and cloud-free infrared images, with corresponding ground truth labels. Figure 6 depicts the schematic diagram of the architecture.

4.2.1 Encoder block

Our modified VGG-16-derived encoder comprises a stack of 10 convolutional layers and 5 max-pooling layers, each meticulously designed to facilitate feature extraction. Within each convolutional layer, a sequence of operations unfolds, encompassing convolution, batch normalization (Ioffe and

Szegedy, 2015; Bjorck et al., 2018), and rectified linear unit (ReLU) activation (Agarap, 2018).

The initial step in this process is the convolution operation, where input feature maps are convolved with trainable filters featuring a 3×3 window size and a stride of 1. Subsequently, batch normalization is employed to normalize the obtained feature maps. This normalization step is pivotal in accelerating the convergence of our model during training and counteracting issues related to the vanishing gradient problem.

Following batch normalization, ReLU activation is applied, introducing non-linearity into the network and expanding its capacity for feature representation. Notably, the early convolutional layers focus on capturing fine-grained visual details, such as shapes and edges, while the deeper convolutional layers leverage these foundational features to compute higher-level and more complex semantic characteristics, aligning with the principles elucidated by Liang et al. in 2017.

Additionally, our encoder network integrates 5 max-pooling layers, thoughtfully positioned after the convolutional layers. These layers play a pivotal role in enhancing the network's translation invariance, a critical characteristic for robust image classification. Each max-pooling layer performs subsampling on the input feature maps using a 2×2 window size and a stride of 2, effectively reducing the size of output feature maps by half while capturing salient features. The incorporation of max-pooling layers further fortifies the network's ability to discern crucial information, ultimately contributing to the accuracy of our cloud image classification task.

4.2.2 Decoder block

The decoder network in our CNN classifier model, adapted from a modified VGG-16 architecture, plays a pivotal role in the restoration of high-level feature maps to the original image resolution, enabling precise cloud image classification.

Comprising 5 upsampling layers and 10 convolutional layers, the decoder network progressively increases the spatial resolution of feature maps while enhancing segmentation accuracy. Four of the upsampling layers utilize pooling indices from corresponding max-pooling layers in the encoder network, optimizing feature restoration with minimal computational overhead. However, this approach may slightly compromise cloud boundary details.

Recognizing the importance of preserving boundary information, the final upsampling layer employs a distinct strategy. It directly utilizes feature maps duplicated from the first max-pooling layers of the encoder network to improve cloud boundary recognition. This process involves bilinear interpolation, doubling the size of feature maps, and concatenating them with duplicated feature maps, ultimately achieving high-resolution cloud feature maps.

In summary, our decoder network systematically restores high-level cloud feature maps to the original image resolu-

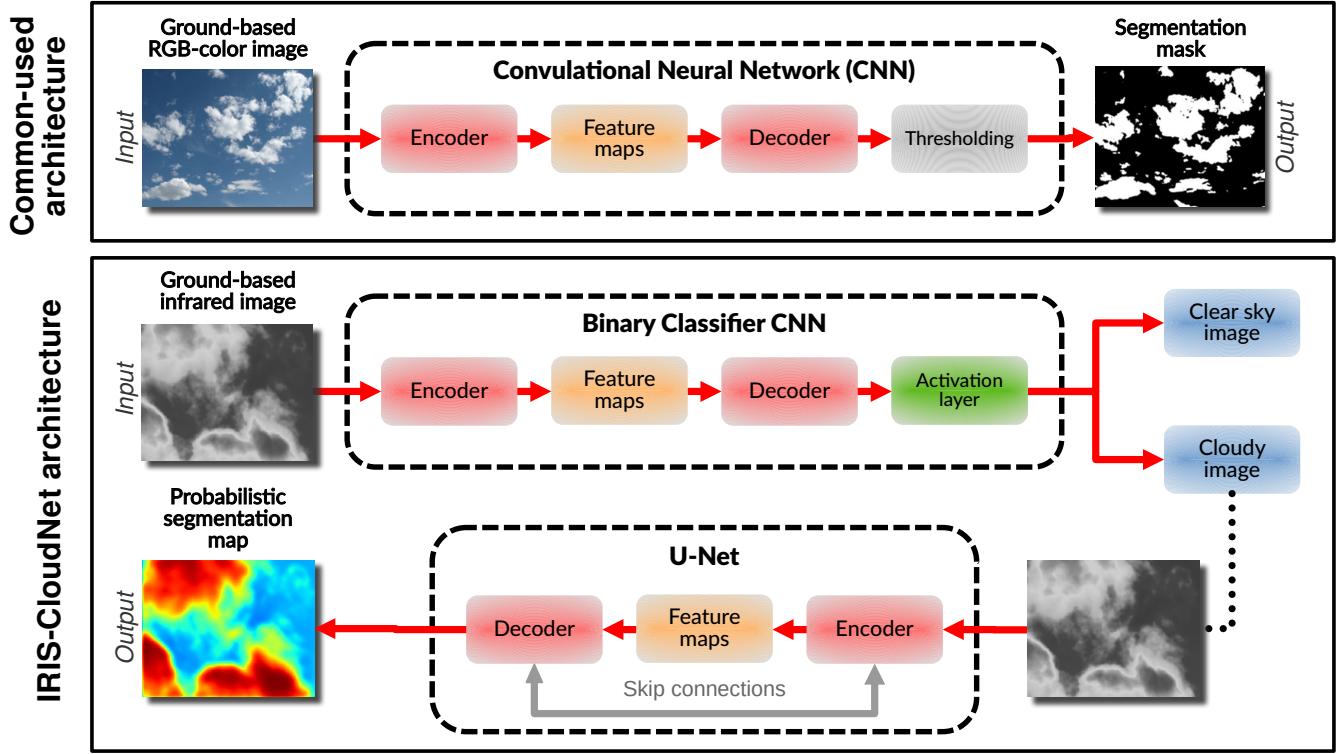


Figure 5. Schematic diagram of common-used deep-learning architectures (top) and this paper solution (bottom). The input image is a 160 × 128 radiometric grayscale image of the sky from the LWIR instrument. Classifier model output is a boolean.

tion through a series of upsampling layers. This process is vital for preserving details and enabling accurate cloud image classification within our modified VGG-16-based CNN classifier.

4.2.3 Binary classifier

The binary classifier is located after the decoder network to achieve final image classification. It is specifically designed for tasks where there are only two possible classes or categories, often labeled as 1 and 0, in this case "cloud" or "clear", contrary to the softmax classifier which applies to multiclass problems. The classification process is realized through a sigmoid activation function. The output is a single-channel probability image, where each pixel's value is interpreted as the probability of it belonging to the cloud class. In practical terms, the pixel-wise predictions are obtained by applying a threshold to the probabilities. Pixels with probabilities exceeding a predefined threshold are classified as belonging to the cloud class, while those below the threshold are classified as non-cloud.

4.3 Image segmentation

For cloud structure identification, the U-Net architecture is adopted owing to its efficiency in semantic segmentation tasks. The U-Net model comprises an encoder and a decoder,

facilitating the capturing of context-rich features and precise delineation of cloud structures. The encoder integrates convolutional and max-pooling layers to progressively downsample the input image, thereby capturing high-level features. These features are then decoded using up-convolutions and skip connections, enabling the reconstruction of the segmented cloud structures. Figure 7 illustrates some examples of infrared cloud images and their corresponding ground-truth masks and predictions. Figure 8 shows the architecture of the segmentation model.

4.3.1 Encoder block

The encoder block of IRIS-CloudDeep segmentation model consists of four convolution layers and four max-pooling layers. We feed a normalized and binned 4x4 radiometrically calibrated image of a fixed input size (160 × 128 pixels) into the model. The initial convolutional layer applies a set of learnable filters to the input image, extracting low-level features. Additional convolutional layers increase the complexity of learned features by applying convolutions to the feature map generated by the previous layer, creating a hierarchy of increasingly abstract features. After a set of convolutional layers, a max-pooling layer is applied to downsample the feature map. Max-pooling helps to reduce the spatial di-

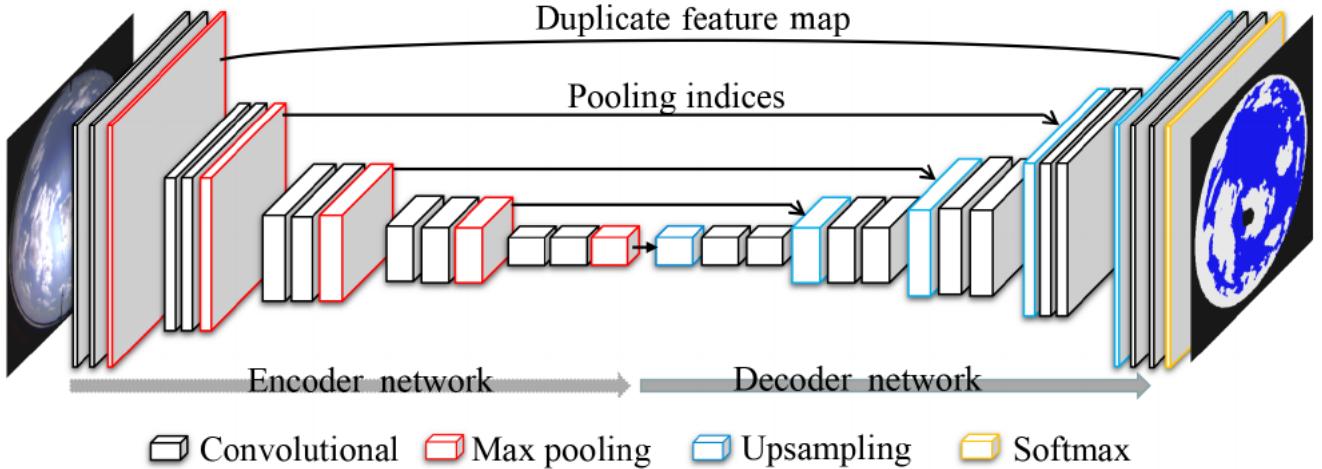


Figure 6. Update this plot for the actual classifier architecture. Keep one of the following caption written here and rearrange it so that it matches the diagram. Schematic diagram of the classifier model architecture. Boxes represent cross-sections of square feature maps. The networks contain an encoder network, a corresponding decoder network, and a final Softmax classifier. Each map's dimensions are indicated on its lower left, and its number of channels are indicated above it. Half-grey boxes represent maps for which half of their channels are copied. The input image is a 160×128 radiometrically calibrated grayscale image of the sky from the LWIR instrument. Classifier model output is a boolean. Arrows represent operations, specified by the legend—notably, blue arrows represent convolutions, while gray ones represent copying (skip connections). Tensor dimensions at the output of each block are specified. VGG-8 architecture for CIFAR-10/SVHN: We implement a non-standard VGG-8 architecture (without Batch Normalization), which is similar to the standard VGG-11 [40] network but with five convolution layers instead of eight. Each convolution layer has 3×3 filters with padding 1 and stride 1. The convolution layer is followed by ReLU non-linearity and a max pooling layer with 2×2 filter, no padding and stride 2 resulting in a downsampling factor of 2. The convolution layers are followed by 3 fully connected layers of dimensions 128, 128 and 10. We use ReLU non-linearity after each fc layer except the final one. CNN architecture for CIFAR-10/SVHN: The network consists of three convolution layers with 3×3 filters, 0 padding and stride 1. The convolution layers are followed by a ReLU non-linearity. We use max pooling in this work with a filter size of 2×2 , no padding and stride 2 which results in a downsampling of the features by a factor of 2. The three convolution layers have 6, 16 and 32 filters respectively. Finally, a Global Average Pooling (GAP) is applied and a fully connected (fc) outputs logits over the number of classes.

mensions of the feature map while retaining the most salient information.

4.3.2 Decoder block

The decoder takes the high-level features generated by the encoder and aims to restore the spatial resolution of the original input image. In the U-Net architecture, the decoder is connected to the encoder via skip connections, enabling the network to combine local and contextual information. The decoder starts with an upsampling operation to increase the spatial dimensions of the feature map. A skip connection connects the upsampled feature map from the decoder with the corresponding feature map from the encoder. This enables the network to leverage both local and global context information. Following the concatenation, a series of convolutional layers are applied. These layers refine the combined feature map, gradually transitioning from abstract features to more detailed information. The final convolutional layer produces the segmentation mask of the U-Net model.

4.3.3 Model output

The output the classifier is a boolean representing whether the image contains cloud or not. On the contrary, the image segmentation model output is a probability mask that assigns a probability value to individual pixels, indicating their potential belonging in the cloud category. Subsequently, a straightforward thresholding technique is applied to transform the probability mask into a binary map. The threshold for labeling is obtained from the Receiver Operating Curve (ROC) analysis specific to our experimental conditions.

4.4 Training procedure and implementation details

The CNN classifier is trained on the dataset containing both cloudy and synthetic clear sky images. The U-Net model is trained on cloudy images only which is the labeled dataset containing infrared images with pixel-wise cloud structure annotations (ground-truth masks). The loss function employed for training is the binary cross-entropy, which quantifies the difference between predicted probabilities and actual binary class labels for each instance in the dataset. Mathematically, given an instance's true binary label y (0 or 1) and

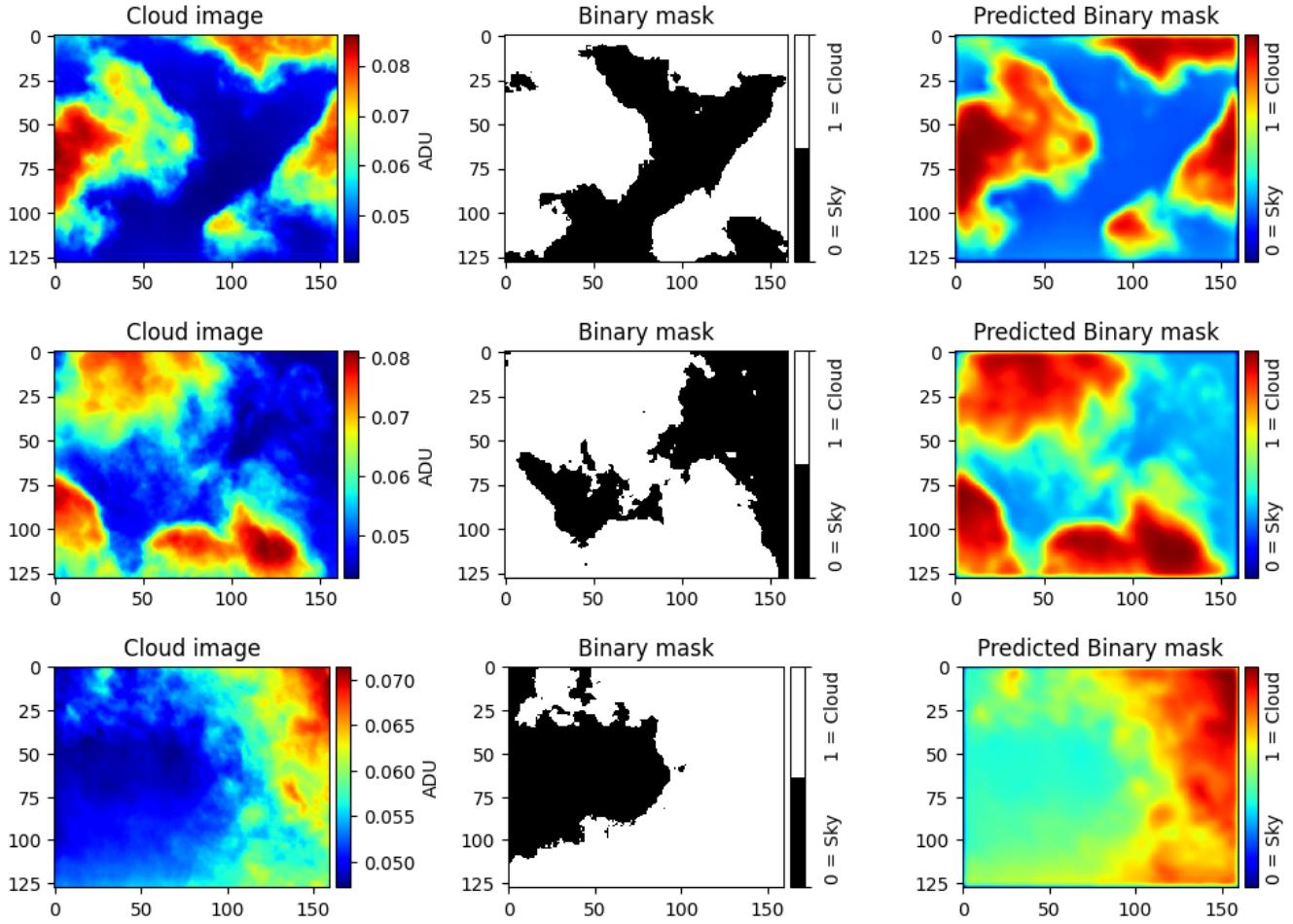


Figure 7. Sample of input images (left), associated ground-truth binary masks (center) and probabilistic map predictions (right) of IRIS-CloudDeep segmentation model. Use other images to not be redundant with precent figures + only titles for the first row : cloud image & ground-truth binary mask & model prediction. Change the colormaps so that the left and right columns are not made from the same colors.

the predicted probability p of it belonging to class 1, the binary cross-entropy loss \mathcal{L} is calculated as:

$$\mathcal{L} = -\frac{1}{N} \sum_i y_i \cdot \log(f_w(x_i)) + (1-y_i) \cdot \log(1-f_w(x_i)) \quad (1)$$

where \mathcal{L} is the binary cross-entropy loss. N is the total number of instances in the dataset, i index represents an individual instance, y_i is the i -th true binary label (0 or 1) and $f_w(x_i)$ is the predicted probability that belongs to class 1, based on the model with parameters w . The goal of training is to minimize this loss function by adjusting the model parameters weights w to better align the predicted probabilities $f_w(x_i)$ with the true labels y_i .

Both architectures are implemented in PYTHON with the KERAS (Chollet et al., 2015) subpackage of TENSORFLOW framework (Abadi et al., 2015) and FLAX (Heek et al., 2023) neural network library of JAX (Bradbury et al., 2018). IRIS-CloudDeep is trained on the GPU cluster infrastructure of

the MESO@LR³ high-performance computing center with an NVIDIA Quadro RTX 6000. Images are normalized and downsampled into a fixed 160×128 resolution format to speed up computations and to capture the global trend. Models are trained over 50 and 100 epochs respectively using the ADAM optimizer (Kingma and Ba, 2014). We set the batch size to 16 images with a dynamic learning rate initial value of $\lambda = 0.001$ that decreases through the training epochs following the relation : $\lambda(\text{epoch}) = 0.001 \cdot 0.5^{\lfloor \frac{\text{epoch}}{10} \rfloor}$. We also set an earlystopping watchdogs that interrupts the training if the loss value doesn't decrease below a certain threshold after 15 epochs. Hyperparameters are fine tuned with the OPTUNA framework (Akiba et al., 2019) using the sampling strategy algorithm.

20

25

30

³<https://meso-lr.umontpellier.fr/>

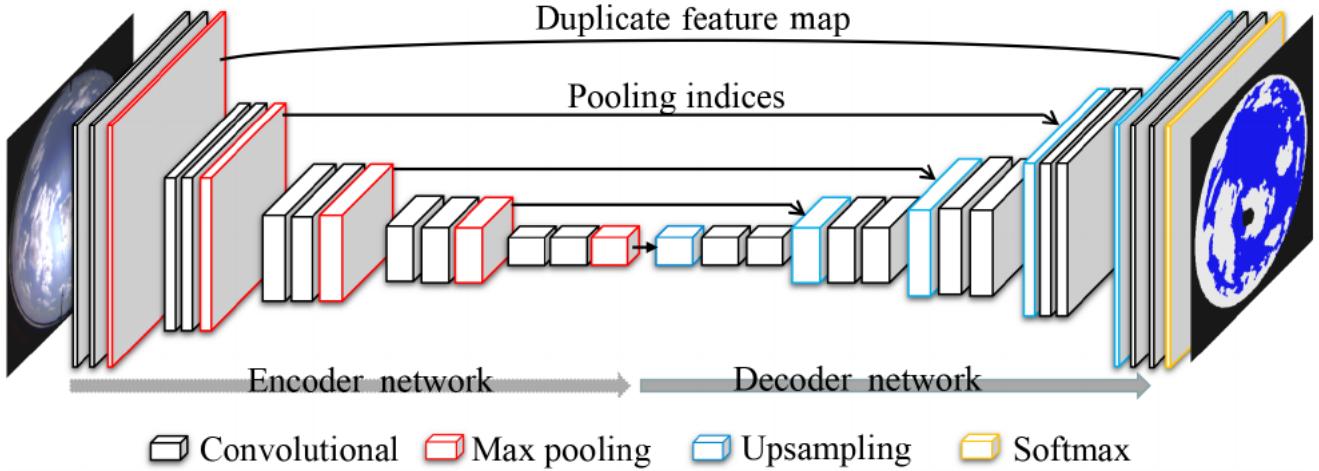


Figure 8. Update this plot for the actual segmentation architecture. Keep one of the following caption written here and rearrange it so that it matches the diagram. Schematic diagram of U-Net based segmentation model architecture. Boxes represent cross-sections of square feature maps. The U-net model contains two parts: down-sampling (left half) and up-sampling (right half). After each convolutional layer, ReLU activation function (improves the computational speed of the training stage)and BN function were applied to effectively capture non-linearities in data and speedup the training Each map's dimensions are indicated on its lower left, and its number of channels are indicated above it. Half-grey boxes represent maps for which half of their channels are copied. The input image is a 160×128 radiometrically calibrated grayscale image of the sky from the LWIR instrument. The output image is a probabilistic mask prediction of pixels being cloudy or clear. Arrows represent operations, specified by the legend-notably, blue arrows represent convolutions, while gray ones represent copying (skip connections). Tensor dimensions at the output of each block are specified. Classifier model output is a boolean.

5 Experiments

5.1 Performance metrics

KS : question to WK : do you see any other relevant performance metrics to add ?

- 5 In order to evaluate the performance of the proposed models, we adopt six metrics: precision (P), accuracy (A), recall (R), F1-score (F1), intersection over union (IoU) and error rate (ER) between the ground-truths and the predictions. For the segmentation model, accuracy and intersection over 10 union are averaged over the pixels to get the mean pixel accuracy (mA) and the mean intersection over union (mIoU). All of these parameters are defined in the following equations,

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (5)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (6)$$

$$ER = \frac{FP + FN}{TP + TN + FP + FN} \quad (7)$$

with True Positives (TP) the number of correctly classified positive instances; False Positives (FP) the number of negative instances that were incorrectly classified as positive; False Negatives (FN) the number of positive instances that were incorrectly classified as negative; True Negatives (TN) the number of correctly classified negative instances.

5.2 Results

KS: the text in red is copy-pasted from relevant articles to give some ideas for the writing of the associated sections. I took what I found to be the best written of all. We should present the results in a similar manner.

Table 1 reports that SegCloud achieves a high average accuracy of 96.24%, which further objectively proves its effectiveness. Moreover, SegCloud performs well on whole-sky images with different cloud cover conditions and achieves 96.98% accuracy on clear-sky images, 95.2% accuracy on partial-cloud images, and 99.44% near-perfect accuracy on overcast-sky images. These experimental results show that

the SegCloud is effective and accurate for cloud segmentation and can provide a reference for future cloud segmentation research.

5.2.1 Classifier

The performance of the ResNet model is somewhat sensitive to the learning rate and momentum, but outcomes are very similar for learning rates of the order of (1-3)% and momentum values 0.7-0.9. However, we do find significant variations between independent training runs despite the use of manual random seeding, which we attribute to random scheduling during the GPU acceleration and the relatively small training sample size for this type of model. In the following, we report on the results of the best of five independent training runs. Figure 2 shows that we find validation sample accuracies of the order of $\sim 80\%$, peaking around 87% for individual training epochs. After ~ 20 epochs, the training sample loss becomes mostly stationary, meaning that the model does not improve. The test sample loss, however, is subject to significant variations, which we attribute to the relatively small sample size. Training of the ResNet model leads to rather high validation sample accuracies of the order of 85% after only ~ 10 training epochs. We adopt this accuracy and number of epochs in our further analysis. We find f1-scores of the order of 0.88. The f1 score is defined as the harmonic mean of precision and recall and serves as a measure for the overall performance of a binary classifier, where 1 denotes a flawless classification and lower values denote flawed classification results. Training our ResNet adaptation for 100 epochs takes 6.9 hr, 10 epochs of training takes accordingly 41 minutes.

5.2.2 Segmentation

We adopt the following set of hyperparameters for our light-GBM model: a maximum depth of each tree of 5, 500 estimators, a learning rate of 0.25, 30 leaves per tree, 100 examples required to form a leaf, $a = 10$, and $b = 100$. This configuration leads to a training sample accuracy of 96% and a test sample accuracy of 95%. The accuracy on the validation sample, which was neither used in the training of the model nor in the tuning of the hyperparameters, is 95% too. The f1 score on the validation sample is 0.94, underlining the good performance of the trained model. The training of the entire training sample using the selected hyperparameters and a five-fold cross-validation takes 12 s on a standard desktop computer. Figure 3 shows the feature importances extracted from the final trained model. The feature importance used here is defined as the number of times a feature is used in this model throughout all individual decision trees. The comparison shows that environmental parameters that affect sky brightness are extremely important, followed by the subregion location. Actual subregion properties and their time dif-

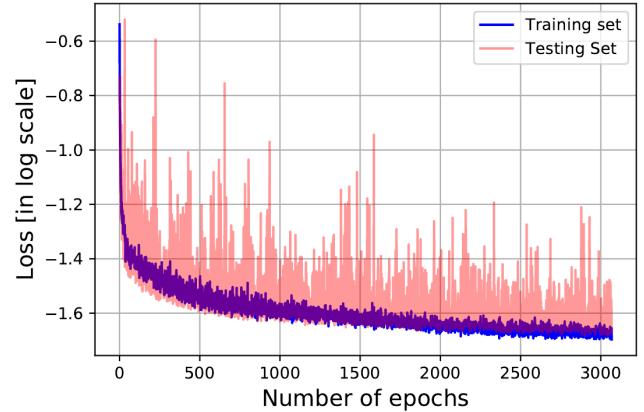


Figure 9. Keep the caption but transform the plot following it : Training and testing losses of IRIS-CloudDeep models over epochs. The dashed curve is for training whereas the continuous curve is for testing. Blue and red curves denote the classifier and segmentation models respectively.

ferentials follow, the latter of which only have a small-but not negligible impact on the model results.

5.3 Framework effectiveness/performance

5.3.1 Losses trend

The trend of binary cross-entropy loss for training and testing sets are shown in Fig. 2. We observe that the loss saturates after a few thousand iterations, and the model exhibits comparable loss performance for both training and testing sets. We choose the CloudSegNet model with the lowest validation loss for our subsequent experiments.

5.3.2 Receiver Operating Curve (ROC)

As discussed above, the output of CloudSegNet is a probability mask, wherein each pixel indicates the degree of belongingness to the cloud category. Since the ground-truth maps are binary in nature, it is necessary to convert the probabilistic output into binary maps as well. We employ a Receiver Operating Curve (ROC) technique to understand the impact of the threshold on the performance. We vary the threshold from 0 to 1 in steps of 0.01, and record the False Positive Rate (FPR) and True Positive Rate (TPR) of cloud detection. Figure 3 shows the resulting ROC curve. The area under the ROC curve (AUC) is 0.97, indicating the competitive performance of CloudSegNet. The ROC curve provides an opportunity to choose a threshold, based on the trade-off between FPR and TPR. In our experiments, we choose the threshold of 0.5 to convert the probabilistic map into a binary sky/cloud image, which is very close to equal true and false positive rates. Of course, this threshold can be further adjusted by the user, based on the specific requirements for TPR and/or FPR. Figure 4 shows some sample outputs of our pro-

55

60

65

70

75

80

Table 2. Evaluation metrics for the proposed IRIS-CloudDeep framework. (A = Accuracy, P = precision, R = Recall, F1 = F1-score, IoU= intersection over union, ER = error-rate, BC Loss = BinaryCrossEntropy Loss, AUC = Area under the curve)

Model	A [%]	P [%]	R [%]	F1 [%]	ER [%]	BC Loss	IoU	AUC
Segmentation	0.9434	0.9474	0.9921	0.9692	0.0566	0.13	0.9402	0.7547
Classifier	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	N/A	XXXX

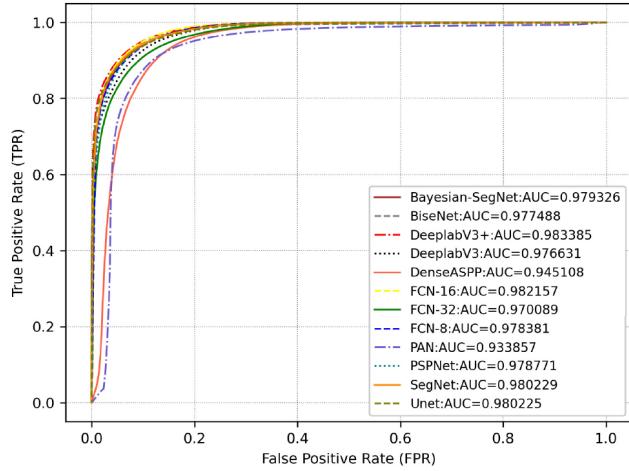


Figure 10. ROC curve comparison of different segmentation models. The AUC value represents the area under the curve. Update this plot for the classifier and segmentation algorithms. Use one color of choice for each. Be consistent with colors from the loss plot above.

posed approach. Visual inspection of additional images from our SWINySeg dataset confirms that CloudSegNet can successfully identify cloud pixels from nychthemeron images.

5.3.3 Confusion matrices

The cloud detection probability for a single subregion is $\sim 85\%$ using ResNet and $\sim 95\%$ using lightGBM. Since clouds typically cover more than one subregion, the probability that any subregion in a set of N subregions that actually include clouds increases exponentially with N. In the same way, the probability to miss clouds decreases. For example, the probability to miss the detection of clouds in three different subregions with the lightGBM classifier is $\sim 0.053 = 10^{-4}$. Hence, the confidence in detecting the presence of clouds anywhere on the sky is much higher than the probability to detect them in a single subregion, supporting the usefulness of this machine-learning approach. We further investigate the performance of our models using a confusion matrix, which not only provides information on the overall classification accuracy, but also additional information on the rate of false-positive and false-negative classifications. Here, a false-positive classification means a subregion that

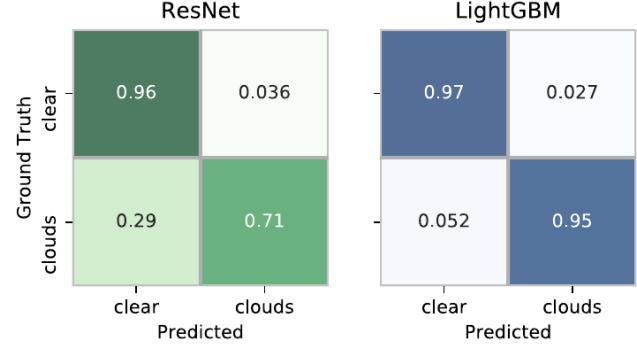


Figure 11. Confusion matrices for the methods considered in the framework. The normalized confusion matrix demonstrates the performance of the proposed model in classifying tiles containing obscurant clouds. The model tends to miss-classify tiles with tiles near the edge of the obscurant clouds since the human labels considered the edges clear although containing very light occlusion. Again update this plot. Also keep the same color fashion as the 2 previous plot for the classifier and segmentation models to not be confusing. Maybe red and blue will do the trick.

has been predicted to contain clouds, although this is not the case. A false-negative classification refers to a subregion that contains clouds that are not identified by the classifier. The confusion matrices for both methods used here are shown in Figure 4. As Figure 4 shows, the false-negative and false-positive rates using the lightGBM classifier are rather small at 5.2% and 2.7%, respectively. We point out that the false-negative rate is roughly a factor 2 higher than the false-positive rate, which might be slightly affected by the class imbalance inherent to the training data sample (see Section 2.2), but is mostly likely due to the classifier's inability to identify non-opaque clouds that were labeled in the training data set. This effect, as well as additional shortcomings potentially related to the insufficient size of the training sample (see Section 5.2.1), is much more obvious in the results of the ResNet classifier, which achieves a false-negative rate of 29% and a false-positive rate of 3.6%, underlining the insufficient performance of this classifier. The comparison of these numbers support the suitability of the lightGBM approach for this task, which is able to identify clouds with high confidence.

We use this augmented composite datasets of 10,000 images respectively to train the classifier and identifier. Random sampling is applied on each composite dataset, and standard train-test-validation split is performed. Figure ?? depicts trends for the binary cross-entropy losses for training and testing subsets with various hyperparameters. OPTUNA framework allows ease of hyperparameters optimization. We observe that the loss saturates after a few hundred iterations for all attempts, and the model exhibits comparable loss performance for both training and testing sets. We select the model and layer parameters with the lowest validation loss for our subsequent experiments.

6 Discussion

6.1 Limited benchmarking comparison

Comparing machine-learning models optimized for different types of input data can be meaningful in certain contexts, but it requires careful considerations. The most critical factor is the nature of the data. As state-of-the art models presented in Section 2 are optimized for different type of data (3-channels RGB images) and it may seem uncoherent to compare the performance with single-channel grayscale infrared images. Indeed, these data types have distinct characteristics, and models may perform differently based on these differences. Some models Sun et al. (2011); Liu et al. (2011); Luo et al. (2018) have been proposed to target infrared images with categorization tasks. Consider whether the models can be adapted or fine-tuned to work with both RGB and infrared data is challenging. This might involve multi-modal learning approaches (Liu et al., 2018; Li et al., 2020; Wei et al., 2023) or transfer learning techniques (Manzo and Pellino, 2021; Wang et al., 2021a; Zhou et al., 2021) which are not the intended purpose of this work.

Nevertheless, we attempt to evaluate the robustness of our segmentation model by testing its ability to generalize to other datasets including: SWIMSEG (Dev et al., 2016), SWINSEG (Dev et al., 2019b, 2017), SWINySEG (Dev et al., 2019a), WSISEG (Xie et al., 2020), HYTA (Li et al., 2011) and TLCDD. Images containing elements other than clouds or sky are cropped to prevent the model to learn unwanted features and avoid subsequent misinterpretations. We transform RGB images into gray-scaled images with OPENCV (Bradski, 2000) color conversion method COLOR_RGB2GRAY defined by the following equation,

$$\text{RGB} \rightarrow \text{Gray} = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \quad (8)$$

where R, G and B are respectively red, green and blue channels of the input color image. Metrics for each dataset are summarized in Table 3.

Although results show that most cloud structures in the images can be accurately recognized by models, we find this method of transforming color images to grayscale not

optimal. Due to Rayleigh scattering producing strong blue color channels (Bates, 1984), our framework is less efficient on the transformed publicly available datasets. Additional work may be needed to improve conversion of RGB color to grayscale images with similar contrast to infrared thermal images. Nevertheless, we confirm its ability to generalize on unseen data.

6.2 Impact of training data quality and quantity

We investigate the impact of the training subset quality and quantity. It is often challenging to produce non-ambiguous ground-truth masks as it is subject to biases with manual work. As clouds can take on various forms in images, establishing a precise definition of cloud presence and differentiation from sky background is difficult. Machine learning model's susceptibility to cirrus clouds and those that are not completely opaque is greatly influenced by the quality of the training data. Contrary to Mommert (2020), we do not expect clouds to be entirely opaque. Indeed, as we image in LWIR band, the sensor is not susceptible to traditional visible spectrum phenomena such as air glow. This consideration is apparent in the false-positive (FP) rates of classifiers and segmentation models which sometimes indicates cloud presence where the expected airmass gradient is located.

To evaluate the effect of the training set size, we train the exact same classifier and segmentation architectures while selecting different number of images. Results for different subset sizes are listed in Table ??.

Furthermore, small number of nights leads to less variety in cloud shapes and structures.

In the case of our ResNet approach, a steady rise of both accuracy and F1 score can be observed through all training sample sizes used in this analysis. The fact that neither metric plateaus indicates that the training sample size required to max out the performance of the ResNet model has not been reached and that this model will benefit from additional training data. We fit a power-law function of the form $f(x; a, b)$ to the ResNet accuracy values in Table 1 and find a saturation accuracy of only 92%. Furthermore, we find through extrapolation that our ResNet approach requires of the order of 20,000 training samples to achieve an accuracy of 90%. We acknowledge that this extrapolation may not be highly accurate, but it certainly provides reasonable estimates of the orders of magnitudes for both the maximum accuracy that can be expected and the training sample size. Based on these estimates, we conclude that our ResNet approach in this form is extremely expensive compared to our lightGBM model. We find lightGBM performances that are comparable to those reported in Section 4.2 for training sample sizes of the order of 1000 image examples. Even in the case of only 100 image examples, an accuracy above 90% can be reached. This result implies that the lightGBM approach is useful even if only a small training sample is available. We furthermore conclude

Table 3. Evaluation metrics for the proposed segmentation model on publicly available state-of-the-art datasets. Note that RGB color images are transformed into gray-scale images as IRIS-CloudDeep segmentation model is optimized for this type of data. Best values are denoted in bold font. (A = Accuracy, P = Precision, R = Recall, F1 = F1-score, ER = Error Rate, BC Loss = BinaryCrossEntropy Loss, IoU = Intersection Over Union, AUC = Area Under the Curve, FL = Used LWIRISEG model FT = Finetuned from LWIRISEG mode, FS = Trained from Scratch).

Model	A [%]	P [%]	R [%]	F1 [%]	ER [%]	BC Loss	IoU	AUC
HYTA FL	0.4585	0.4317	0.9607	0.5957	0.5415	8.1540	0.4242	0.5313
HYTA FT	0.7753	0.8066	0.6444	0.7164	0.2247	0.5010	0.5582	0.7614
HYTA FS	0.5599	1.0000	0.0012	0.0025	0.4401	0.6376	0.0012	0.5006
SWIMSEG FL	0.5751	0.5633	0.9883	0.7176	0.4249	6.4259	0.5595	0.5332
SWIMSEG FT	0.8456	0.8500	0.8709	0.8603	0.1544	0.3425	0.7549	0.8430
SWIMSEG FS	0.8456	0.8500	0.8709	0.8603	0.1544	0.3425	0.7549	0.8430
SWINSEG FL	0.6706	0.5827	0.9639	0.7263	0.3294	4.7272	0.5703	0.6955
SWINSEG FT	0.9164	0.9363	0.8753	0.9047	0.0836	0.1999	0.8261	0.9129
SWINSEG FS	0.9325	0.8843	0.9793	0.9293	0.0675	0.1670	0.8680	0.9364
LWIRISEG	0.9434	0.9474	0.9921	0.9692	0.0566	0.1300	0.9402	0.7547

that more than 1000 training examples will not significantly improve the performance of this model.

6.3 Comparison of libraries: FLAX vs KERAS

When comparing training deep learning models with KERAS and FLAX, we observed notable distinctions in terms of speed and data loading capacity. Keras, being a high-level framework, offers a straightforward and user-friendly experience, but its training speed may be limited, particularly with larger datasets and batch sizes. In contrast, FLAX, with its low-level nature and support for hardware acceleration, demonstrated impressive speed improvements, especially when dealing with substantial datasets or larger batch sizes. The ability of FLAX to efficiently leverage hardware resources, coupled with its flexibility in managing data loading, enables smoother training on extensive datasets and better overall training. These factors make FLAX an appealing choice for tasks demanding scalability and computational efficiency. Table ?? depicts speed comparison and model size for the same segmentation model architecture. Results show a speed increase of XXX% and smaller model disk size being XXX% smaller than the KERAS trained model.

6.4 Future perspectives

The framework established in this paper is planned to be used in real-time with dedicated computing module. Functions of this module will include the following aspects: (i) classifying infrared sky images obtained by the imaging system in real-time, (ii) analyzing cloud-labeled sky images and deriving the corresponding cloud structure and cover using the

Table 4. Comparison of model size and computation speed for the same segmentation model implemented with KERAS and FLAX.

Library	Run-time [s]	Size [MB]
Keras	XXXX	24
Flax	XXXX	14

segmentation algorithm; (iii) generating alerts/flag in accordance with the results; (iv) fully-autonomous processing.

Training on a larger dataset may be an additional step towards improving the accuracy and robustness of our framework. As next-year will bring large amount of data thanks to the telescope system remote operations capabilities, we have the potential to capture a broader range of sky atmospheric conditions, making our models more adept at handling real-world variations. However, it's crucial to maintain a balance between the model's complexity and the available computational resources, as larger datasets can lead to increased training times and resource demands. Additionally, staying vigilant for shifts in data distribution over time is essential to ensure that our models remain adaptive and responsive to evolving experimental conditions.

7 Conclusion

In this paper, we proposed IRIS-CloudDeep, a deep-learning framework for classification and segmentation of LWIR

ground-based thermal images. As far as we know, it is the framework that attempts to apply two sequential models for complementary tasks on single-channel gray-scaled infrared images. Specifically, we presented the CNN-based classifier and the U-Net based segmentation model tailored to extract cloud structures on pre-identified cloud images. Extensive experimental results on a combination of self-acquired data and transformed publicly available datasets have demonstrated the effectiveness and performances of the proposed framework.

KS : Remind some key metrics of the classifier and segmentation algorithms WK : Classifier => A P R F1 Loss WK : Seg => A P R F1 IoU Loss

We successfully increased the size of training, testing and validation subsets with random application of augmentation methods. We developed an accurate simulation tool to produce realistic clear sky images. In the future, additional data will be collected by the infrared instrument, capturing various weather conditions. The framework may be re-trained on heavier datasets that will probably increase its accuracy. Furthermore, if enough data is collected with many different cloud categories and proven-to-be accurate radiometric calibration, we will be able to expand the segmentation model to perform cloud typology through multi-label segmentation. Finally, the current trained model is expected to process data in real-time for the StarDICE experiment in order to: (i) give live alerts to remote observers in the case of cloud detection; (ii) set a flag in CCD images for quality sorting in post-processing.

Code and data availability. The source code is freely available online at <https://github.com/kelian98/IRIS-CloudDeep>. The datasets and other supporting materials will be made available from corresponding author Kélian Sommer (kelian.sommer@umontpellier.fr) upon request.

Author contributions. KS conceived the instrument, collected data, pre-processed the dataset, created ground truths masks for segmentation and realistic synthetic data for classification. WK, RB and KS designed the framework. WK and RB performed the experiments. KS wrote the paper and collected relevant literature. WK and RB revised the manuscript. AB, JCT and BP proposed constructive suggestions on the revision of the article.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work has been realized with the support of MESO@LR-Platform at the University of Montpellier. Some of the results in this paper have been derived using ASTROPY (Astropy Collaboration, 2013, 2018), FLAX (Heek et al., 2023), JAX (Bradbury et al., 2018), KERAS (Chollet et al., 2015), MAT-

PLOTLIB (Hunter, 2007), NUMPY (Harris et al., 2020), OPENCV-PYTHON (Olli-Pekka and OpenCV, 2016), PANDAS (Pandas Development Team, 2020), SCIPY (Virtanen et al., 2020) and TENSORFLOW (Abadi et al., 2015). Add thanks for the people that revised the manuscript : JCT and AB; on the same model as Betoule et al. (2022) acknowledgements

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, <https://www.tensorflow.org/>, software available from tensorflow.org, 2015.
- Aebi, C., Gröbner, J., and Kämpfer, N.: Cloud fraction determined by thermal infrared and visible all-sky cameras, Atmospheric Measurement Techniques, 11, 5549–5563, <https://doi.org/10.5194/amt-11-5549-2018>, 2018.
- Agarap, A. F.: Deep learning using rectified linear units (relu), arXiv preprint arXiv:1803.08375, 2018.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-generation Hyperparameter Optimization Framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.
- Akula, A., Ghosh, R., and Sardana, H.: Thermal imaging and its application in defence systems, in: AIP conference proceedings, vol. 1391, pp. 333–335, American Institute of Physics, 2011.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A. Q., Duan, Y., Al-Shamma, O., Santamaría, J. I., Fadhel, M. A., Al-Amidie, M., and Farhan, L.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, Journal of Big Data, 8, <https://api.semanticscholar.org/CorpusID:232434552>, 2021.
- Astropy Collaboration, T.: Astropy: A community Python package for astronomy, Astronomy and Astrophysics, 558, A33, <https://doi.org/10.1051/0004-6361/201322068>, 2013.
- Astropy Collaboration, T.: Astropy: A community Python package for astronomy, AJ, 156, 123, <https://doi.org/10.3847/1538-3881/aac387>, 2018.
- Badrinarayanan, V., Kendall, A., and Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE transactions on pattern analysis and machine intelligence, 39, 2481–2495, 2017.
- Bates, D.: Rayleigh scattering by air, Planetary and Space Science, 32, 785–790, 1984.
- Betoule, M., Marriner, J., Regnault, N., Cuillandre, J.-C., Astier, P., Guy, J., Balland, C., Hage, P. E., Hardin, D., Kessler, R., Guillou, L. L., Mosher, J., Pain, R., Roccia, P.-F., Sako, M., and Schahmaneche, K.: Improved photometric calibration of the SNLS and the SDSS supernova surveys, Astronomy and Astrophysics, 552, A124, <https://doi.org/10.1051/0004-6361/201220610>, 2013.
- Betoule, M., Antier, S., Bertin, E., Éric Blanc, P., Bongard, S., Cohen Tanugi, J., Dagoret-Campagne, S., Feinstein, F., Hardin, 105

- D., Juramy, C., Le Guillou, L., Le Van Suu, A., Moniez, M., Neveu, J., Nuss, É., Plez, B., Regnault, N., Sepulveda, E., Sommer, K., Souverin, T., and Wang, X. F.: StarDICE I: sensor calibration bench and absolute photometric calibration of a Sony IMX411 sensor, arXiv e-prints, arXiv:2211.04913, <https://doi.org/10.48550/arXiv.2211.04913>, 2022.
- Bjorck, N., Gomes, C. P., Selman, B., and Weinberger, K. Q.: Understanding batch normalization, Advances in neural information processing systems, 31, 2018.
- Boers, R., Acarreta, J. R., and Gras, J. L.: Satellite monitoring of the first indirect aerosol effect: Retrieval of the droplet concentration of water clouds, Journal of Geophysical Research: Atmospheres, 111, 2006.
- Bohlin, R. C.: HUBBLE SPACE TELESCOPE CALSPEC FLUX STANDARDS: SIRIUS (AND VEGA), The Astronomical Journal, 147, 127, <https://doi.org/10.1088/0004-6256/147/6/127>, 2014.
- Bohlin, R. C., Gordon, K. D., Rieke, G. H., Ardila, D., Carey, S., Deustua, S., Engelbracht, C., Ferguson, H. C., Flanagan, K., Kalirai, J., Meixner, M., Noriega-Crespo, A., Su, K. Y. L., and Tremblay, P.-E.: ABSOLUTE FLUX CALIBRATION OF THE IRAC INSTRUMENT ON THE SPITZER SPACE TELESCOPE USING HUBBLE SPACE TELESCOPE FLUX STANDARDS, The Astronomical Journal, 141, 173, <https://doi.org/10.1088/0004-6256/141/5/173>, 2011.
- Bohlin, R. C., Hubeny, I., and Rauch, T.: New Grids of Pure-hydrogen White Dwarf NLTE Model Atmospheres and the HST/STIS Flux Calibration, The Astronomical Journal, 160, 21, <https://doi.org/10.3847/1538-3881/ab94b4>, 2020.
- Bower, K., Choularton, T., Gallagher, M., Beswick, K., Flynn, M., Allen, A., Davison, B., James, J., Robertson, L., Harrison, R., et al.: ACE-2 HILLCLOUD. An overview of the ACE-2 ground-based cloud experiment, Tellus B, 52, 750–778, 2000.
- Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q.: JAX: composable transformations of Python+NumPy programs, <http://github.com/google/jax>, 2018.
- Bradski, G.: The OpenCV Library, Dr. Dobb's Journal of Software Tools, 2000.
- Brout, D., Taylor, G., Scolnic, D., Wood, C. M., Rose, B. M., Vincenzi, M., Dwomoh, A., Lidman, C., Riess, A., Ali, N., Qu, H., and Dai, M.: The Pantheon+ Analysis: SuperCal-fragilistic Cross Calibration, Retrained SALT2 Light-curve Model, and Calibration Systematic Uncertainty, The Astrophysical Journal, 938, 111, <https://doi.org/10.3847/1538-4357/ac8bcc>, 2022.
- Burke, D. L., Axelrod, T., Blondin, S., Claver, C., Željko Ivezić, Jones, L., Saha, A., Smith, A., Smith, R. C., and Stubbs, C. W.: PRECISION DETERMINATION OF ATMOSPHERIC EXTINCTION AT OPTICAL AND NEAR-INFRARED WAVELENGTHS, The Astrophysical Journal, 720, 811, <https://doi.org/10.1088/0004-637X/720/1/811>, 2010.
- Burke, D. L., Saha, A., Claver, J., Axelrod, T., Claver, C., De-Poy, D., Ivezić, Ž., Jones, L., Smith, R. C., and Stubbs, C. W.: ALL-WEATHER CALIBRATION OF WIDE-FIELD OPTICAL AND NIR SURVEYS, The Astronomical Journal, 147, 19, <https://doi.org/10.1088/0004-6256/147/1/19>, 2013.
- Burke, D. L., Rykoff, E. S., Allam, S., Annis, J., Bechtol, K., Bernstein, G. M., Drlica-Wagner, A., Finley, D. A., Gruendl, R. A., James, D. J., Kent, S., Kessler, R., Kuhlmann, S., Lasker, J., Li, T. S., Scolnic, D., Smith, J., Tucker, D. L., Wester, W., Yanny, B., Abbott, T. M. C., Abdalla, F. B., Benoit-Lévy, A., Bertin, E., Rosell, A. C., Kind, M. C., Carretero, J., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., Desai, S., Diehl, H. T., Doel, P., Estrada, J., García-Bellido, J., Gruen, D., Gutierrez, G., Honscheid, K., Kuehn, K., Kuropatkin, N., Maia, M. A. G., March, M., Marshall, J. L., Melchior, P., Menanteau, F., Miquel, R., Plazas, A. A., Sako, M., Sanchez, E., Scarpine, V., Schindler, R., Sevilla-Noarbe, I., Smith, M., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Tarle, G., and and, A. R. W.: Forward Global Photometric Calibration of the Dark Energy Survey, The Astronomical Journal, 155, 41, <https://doi.org/10.3847/1538-3881/aa9f22>, 2017.
- Canziani, A., Paszke, A., and Culurciello, E.: An analysis of deep neural network models for practical applications, arXiv preprint arXiv:1605.07678, 2016.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L.: DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, 2017.
- Chollet, F. et al.: Keras, <https://keras.io>, 2015.
- Collaboration, T. D. E. S.: The Dark Energy Survey, 2005.
- Cortes, C. and Vapnik, V.: Support-vector networks, Machine learning, 20, 273–297, 1995.
- Currie, M., Rubin, D., Aldering, G., Deustua, S., Fruchter, A., and Perlmutter, S.: Evaluating the Calibration of SN Ia Anchor Datasets with a Bayesian Hierarchical Model, 2020.
- Dev, S., Lee, Y. H., and Winkler, S.: Color-based Segmentation of Sky/Cloud Images From Ground-based Cameras, 2016.
- Dev, S., Savoy, F. M., Lee, Y. H., and Winkler, S.: Nighttime sky-cloud image segmentation, 2017.
- Dev, S., Nautiyal, A., Lee, Y. H., and Winkler, S.: CloudSegNet: A deep network for nychthemeron cloud image segmentation, IEEE Geoscience and Remote Sensing Letters, 16, 1814–1818, 2019a.
- Dev, S., Savoy, F., Lee, Y. H., and Winkler, S.: Singapore Whole sky Nighttime Image SEGmentation Database, <https://doi.org/10.21227/jsf0-ga67>, 2019b.
- Fa, T., Xie, W., Yiren, W., and Xia, Y.: Development of an all-sky imaging system for cloud cover assessment, Applied Optics, 58, 5516, <https://doi.org/10.1364/AO.58.005516>, 2019.
- Gallo, M. A., Willits, D. S., Lubke, R. A., and Thiede, E. C.: Low-cost uncooled IR sensor for battlefield surveillance, in: Infrared Technology XIX, vol. 2020, pp. 351–362, SPIE, 1993.
- Geer, A., Baordo, F., Bormann, N., Chambon, P., English, S., Kazumori, M., Lawrence, H., Lean, P., Lonitz, K., and Lupu, C.: The growing impact of satellite observations sensitive to humidity, cloud and precipitation, Quarterly Journal of the Royal Meteorological Society, 143, 3189–3206, 2017.
- Hack, E. D., Pauliquevis, T., Barbosa, H. M. J., Yamasoe, M. A., Klebe, D., and Correia, A. L.: Precipitable water vapor retrievals using a ground-based infrared sky camera in subtropical South America, Atmospheric Measurement Techniques, 16, 1263–1278, <https://doi.org/10.5194/amt-16-1263-2023>, 2023.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Rio, J. F., Wiebe, M., Peterson, P., Gerard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi,

- H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, *Nature*, 585, 357–362, <https://doi.org/10.1038/s41586-020-2649-2>, 2020.
- He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, 2015.
- Heek, J., Levskaya, A., Oliver, A., Ritter, M., Rondepierre, B., Steiner, A., and van Zee, M.: Flax: A neural network library and ecosystem for JAX, <http://github.com/google/flax>, 2023.
- Houghton, J. T. and Lee, A. C. L.: Atmospheric Transmission in the 10–12 μm Window, *Nature Physical Science*, 238, 117–118, <https://doi.org/10.1038/physci238117a0>, 1972.
- Hu, Y., Wielicki, B. A., Yang, P., Stackhouse, P. W., Lin, B., and Young, D. F.: Application of deep convective cloud albedo observation to satellite-based study of the terrestrial atmosphere: Monitoring the stability of spaceborne measurements and assessing absorption anomaly, *IEEE Transactions on Geoscience and Remote Sensing*, 42, 2594–2599, 2004.
- Hunter, J. D.: CSE, 9, 90, <https://doi.org/10.1109/MCSE.2007.55>, 2007.
- Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015.
- Ishimwe, R., Abutaleb, K., Ahmed, F., et al.: Applications of thermal imaging in agriculture—A review, *Advances in remote Sensing*, 3, 128, 2014.
- Kelsey, V., Riley, S., and Minschwaner, K.: Atmospheric precipitable water vapor and its correlation with clear-sky infrared temperature observations, *Atmospheric Measurement Techniques*, 15, 1563–1576, <https://doi.org/10.5194/amt-15-1563-2022>, 2022.
- Kimata, M.: Uncooled infrared focal plane arrays, *IEEJ Transactions on Electrical and Electronic Engineering*, 13, 4–12, <https://doi.org/https://doi.org/10.1002/tee.22563>, 2018.
- Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*, 2014.
- Klebe, D., Sebag, J., Blatherwick, R. D., and Zimmer, P. C.: All-Sky Mid-Infrared Imagery to Characterize Sky Conditions and Improve Astronomical Observational Performance, *Publications of the Astronomical Society of the Pacific*, 124, 1309, <https://doi.org/10.1086/668866>, 2012.
- Klebe, D. I., Blatherwick, R. D., and Morris, V. R.: Ground-based all-sky mid-infrared and visible imagery for purposes of characterizing cloud properties, *Atmospheric Measurement Techniques*, 7, 637–645, <https://doi.org/10.5194/amt-7-637-2014>, 2014.
- Krauz, L., Janout, P., Blažek, M., and Páta, P.: Assessing Cloud Segmentation in the Chromacity Diagram of All-Sky Images, *Remote Sensing*, 12, <https://doi.org/10.3390/rs12111902>, 2020.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, in: *Advances in Neural Information Processing Systems*, edited by Pereira, F., Burges, C., Bottou, L., and Weinberger, K., vol. 25, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper_2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf, 2012.
- Larason, T. and Houston, J.: Spectroradiometric Detector Measurements: Ultraviolet, Visible, and Near Infrared Detectors for Spectral Power, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=841061, 2008.
- LeCun, Y., Bengio, Y., et al.: Convolutional networks for images, speech, and time series, *The handbook of brain theory and neural networks*, 3361, 1995, 1995.
- Lewis, P. M., Rogers, H., and Schindler, R. H.: A radiometric all-sky infrared camera (RASICAM) for DES/CTIO, in: *Ground-based and Airborne Instrumentation for Astronomy III*, vol. 7735, pp. 1307–1318, SPIE, 2010.
- Li, M., Liu, S., and Zhang, Z.: Deep tensor fusion network for multimodal ground-based cloud classification in weather station networks, *Ad hoc networks*, 96, 101 991, 2020.
- Li, Q., Lu, W., and Yang, J.: A Hybrid Thresholding Algorithm for Cloud Detection on Ground-Based Color Images, *Journal of Atmospheric and Oceanic Technology*, 28, 1286 – 1296, <https://doi.org/https://doi.org/10.1175/JTECH-D-11-00009.1>, 2011.
- Li, S., Wang, M., Wu, J., Sun, S., and Zhuang, Z.: Cloud-DeepLabV3+: a lightweight ground-based cloud segmentation method based on multi-scale feature aggregation and multi-level attention feature enhancement, *International Journal of Remote Sensing*, 44, 4836–4856, <https://doi.org/10.1080/01431161.2023.2240034>, 2023.
- Li, T. S., DePoy, D. L., Marshall, J. L., Tucker, D., Kessler, R., Annis, J., Bernstein, G. M., Boada, S., Burke, D. L., Finley, D. A., James, D. J., Kent, S., Lin, H., Marriner, J., Mondrik, N., Nagasawa, D., Rykoff, E. S., Scolnic, D., Walker, A. R., Wester, W., Abbott, T. M. C., Allam, S., Benoit-Lévy, A., Bertin, E., Brooks, D., Capozzi, D., Rosell, A. C., Kind, M. C., Carretero, J., Crocce, M., Cunha, C. E., D’Andrea, C. B., da Costa, L. N., Desai, S., Diehl, H. T., Doel, P., Flaugher, B., Fosalba, P., Friedman, J., Gaztanaga, E., Goldstein, D. A., Gruen, D., Gruendl, R. A., Gutierrez, G., Honscheid, K., Kuehn, K., Kuropatkin, N., Maia, M. A. G., Melchior, P., Miller, C. J., Miquel, R., Mohr, J. J., Nielsen, E., Nichol, R. C., Nord, B., Ogando, R., Plazas, A. A., Romer, A. K., Roodman, A., Sako, M., Sanchez, E., Scarpine, V., Schubnell, M., Sevilla-Noarbe, I., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Tarle, G., Thomas, D., and V. V.: ASSESSMENT OF SYSTEMATIC CHROMATIC ERRORS THAT IMPACT SUB-1% PHOTOMETRIC PRECISION IN LARGE-AREA SKY SURVEYS, *The Astronomical Journal*, 151, 157, <https://doi.org/10.3847/0004-6256/151/6/157>, 2016.
- Li, X., Wang, B., Qiu, B., and Wu, C.: An all-sky camera image classification method using cloud cover features, *Atmospheric Measurement Techniques*, 15, 3629–3639, <https://doi.org/10.5194/amt-15-3629-2022>, 2022.
- Liandrat, O., Cros, S., Braun, A., Saint-Antonin, L., Decroix, J., and Schmutz, N.: Cloud cover forecast from a ground-based all sky infrared thermal camera, p. 10, <https://doi.org/10.1117/12.2278636>, 2017a.
- Liandrat, O., Cros, S., Braun, A., Saint-Antonin, L., Decroix, J., and Schmutz, N.: Cloud cover forecast from a ground-based all sky infrared thermal camera, in: *Remote Sensing of Clouds and the Atmosphere XXII*, vol. 10424, pp. 19–31, SPIE, 2017b.
- Liou, K.-N.: Radiation and cloud processes in the atmosphere. Theory, observation, and modeling, New York, NY (United States); Oxford University Press, 1992.
- Liu, L., Sun, X., Chen, F., Zhao, S., and Gao, T.: Cloud Classification Based on Structure Features of Infrared Images, *Journal of Atmospheric and Oceanic Technology*, 28, 410 – 417,

- <https://doi.org/https://doi.org/10.1175/2010JTECHA1385.1>, 2011.
- Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., and Han, J.: On the variance of the adaptive learning rate and beyond, arXiv preprint arXiv:1908.03265, 2019.
- Liu, S., Li, M., Zhang, Z., Xiao, B., and Cao, X.: Multimodal ground-based cloud classification using joint fusion convolutional neural network, *Remote Sensing*, 10, 822, 2018.
- Liu, S., Zhang, J., Zhang, Z., Cao, X., and Durrani, T. S.: Tran-
10 sCloudSeg: Ground-Based Cloud Image Segmentation With Transformer, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 6121–6132, <https://doi.org/10.1109/JSTARS.2022.3194316>, 2022.
- Liu, Y., Shao, Z., and Hoffmann, N.: Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions, 2021.
- Long, C. N., Sabburg, J. M., Calbó, J., and Pagès, D.: Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images, *Journal of Atmospheric and Oceanic Technology*, 23, 633 – 652, <https://doi.org/https://doi.org/10.1175/JTECH1875.1>, 2006.
- Lopez, T., Antoine, R., Baratoux, D., and Rabinowicz, M.: Contribution of thermal infrared images on the understanding of the subsurface/atmosphere exchanges on Earth., in: EGU General Assembly Conference Abstracts, p. 11811, 2017.
- Luo, Q., Meng, Y., Liu, L., Zhao, X., and Zhou, Z.: Cloud classification of ground-based infrared images combining manifold and texture features, *Atmospheric Measurement Techniques*, 11, 5351–5361, <https://doi.org/10.5194/amt-11-5351-2018>, 2018.
- Makwana, D., Nag, S., Susladkar, O., Deshmukh, G., Teja R, S. C., Mittal, S., and Mohan, C. K.: ACLNet: an attention and clustering-based cloud segmentation network, *Remote Sensing Letters*, 13, 865–875, 2022.
- Mandat, D., Pech, M., Hrabovsky, M., Schovanek, P., Palatka, M.,
35 Travnicek, P., Prouza, M., and Ebr, J.: All Sky Camera instrument for night sky monitoring, arXiv preprint arXiv:1402.4762, 2014.
- Manzo, M. and Pellino, S.: Voting in transfer learning system for ground-based cloud classification, *Machine Learning and Knowledge Extraction*, 3, 542–553, 2021.
- Martin, R. V.: Satellite remote sensing of surface air quality, *Atmospheric environment*, 42, 7823–7843, 2008.
- McCallum, A.: Efficiently inducing features of conditional random fields, arXiv preprint arXiv:1212.2504, 2012.
- Mikołajczyk, A. and Grochowski, M.: Data augmentation for im-
45 proving deep learning in image classification problem, in: 2018 International Interdisciplinary PhD Workshop (IIPhDW), pp. 117–122, <https://doi.org/10.1109/IIPHDW.2018.8388338>, 2018.
- Mokhov, I. L. and Schlesinger, M. E.: Analysis of global cloudiness: 2. Comparison of ground-based and satellite-based cloud climatologies, *Journal of Geophysical Research: Atmospheres*, 99, 17 045–17 065, 1994.
- Mommert, M.: Cloud Identification from All-sky Camera Data with Machine Learning, *The Astrophysical Journal*, 159, 178, <https://doi.org/10.3847/1538-3881/ab744f>, 2020.
- Nikolenko, I. and Maslov, I.: Infrared (thermal) camera for monitoring the state of the atmosphere above the sea horizon of the Simeiz Observatory INASAN, *INASAN Science Reports*, 6, 85–87, 2021.
- Olli-Pekka, H. and OpenCV, t.: OpenCV Python packages, <https://github.com/opencv/opencv-python>, 2016.
- O’Shea, K. and Nash, R.: An Introduction to Convolutional Neural Networks, 2015.
- Paczyński, B.: Monitoring All Sky for Variability1, *Publications of the Astronomical Society of the Pacific*, 112, 1281, 2000.
- Pandas Development Team, T.: pandas-dev/pandas: Pandas, <https://doi.org/10.5281/zenodo.3509134>, 2020.
- Perez, L. and Wang, J.: The Effectiveness of Data Augmentation in Image Classification using Deep Learning, 2017.
- Petzold, A., Thouret, V., Gerbig, C., Zahn, A., Brenninkmeijer, C. A., Gallagher, M., Hermann, M., Pontaud, M., Ziereis, H., Boulanger, D., et al.: Global-scale atmosphere monitoring by in-service aircraft—current achievements and future prospects of the European Research Infrastructure IAGOS, *Tellus B: Chemical and Physical Meteorology*, 67, 28 452, 2015.
- Reil, K., Lewis, P., Schindler, R., and Zhang, Z.: An update on the status and performance of the Radiometric All-Sky Infrared Camera (RASICAM), in: *Observatory Operations: Strategies, Processes, and Systems V*, vol. 9149, pp. 321–331, SPIE, 2014.
- Ring, E. and Ammer, K.: Infrared thermal imaging in medicine, *Physiological measurement*, 33, R33, 2012.
- Rogalski, A.: Recent progress in infrared detector technologies, *Infrared Physics & Technology*, 54, 136–154, <https://doi.org/10.1016/j.infrared.2010.12.003>, 2011.
- Rogalski, A. and Chrzanowski, K.: Infrared Devices And Techniques (Revision), *Metrology and Measurement Systems*, 21, 565–618, <https://doi.org/10.2478/mms-2014-0057>, 2014.
- Román, R., Antuña-Sánchez, J. C., Cachorro, V. E., Toledano, C., Torres, B., Mateos, D., Fuertes, D., López, C., González, R., Lapionok, T., et al.: Retrieval of aerosol properties using relative radiance measurements from an all-sky camera, *Atmospheric Measurement Techniques*, 15, 407–433, 2022.
- Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv preprint arXiv:1505.04597, 2015.
- Roy, P., Behera, M., and Srivastav, S.: Satellite remote sensing: sensors, applications and techniques, *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, 87, 465–472, 2017.
- Rubin, D., Aldering, G., Amanullah, R., Barbary, K., Dawson, K. S., Deustua, S., Faccioli, L., Fadeyev, V., Fakhouri, H. K., Fruchter, A. S., Gladders, M. D., de Jong, R. S., Koekemoer, A., Krechmer, E., Lidman, C., Meyers, J., Nordin, J., Perlmutter, S., Ripoche, P., Schlegel, D. J., Spadafora, A., and Suzuki, N.: A CALIBRATION OF NICMOS CAMERA 2 FOR LOW COUNT RATES*, *The Astronomical Journal*, 149, 159, <https://doi.org/10.1088/0004-6256/149/5/159>, 2015.
- Rubin, D., Aldering, G., Antilogus, P., Aragon, C., Bailey, S., Baltay, C., Bongard, S., Boone, K., Buton, C., Copin, Y., Dixon, S., Fouchez, D., Gangler, E., Gupta, R., Hayden, B., Hillebrandt, W., Kim, A. G., Kowalski, M., Kuesters, D., Leget, P. F., Mondon, F., Nordin, J., Pain, R., Pecontal, E., Pereira, R., Perlmutter, S., Ponder, K. A., Rabinowitz, D., Rigault, M., Runge, K., Saunders, C., Smadja, G., Suzuki, N., Tao, C., Taubenberger, S., Thomas, R. C., Vincenzi, M., and Factory, T. N. S.: Uniform Recalibration of Common Spectrophotometry Standard Stars onto the CALSPEC System using the SuperNova Integral Field Spectrograph, 2022.

- Salamalikis, V., Tzoumanikas, P., Argiriou, A. A., and Kazantzidis, A.: Estimation of Precipitable Water Using Thermal Infrared Images, *Environmental Sciences Proceedings*, 26, <https://doi.org/10.3390/environsciproc2023026033>, 2023.
- Schiffer, R. A. and Rossow, W. B.: The International Satellite Cloud Climatology Project (ISCCP): The first project of the world climate research programme, *Bulletin of the American Meteorological Society*, 64, 779–784, 1983.
- Schreiner, A. J., Unger, D. A., Menzel, W. P., Ellrod, G. P., Strabala, K. I., and Pellet, J. L.: A comparison of ground and satellite observations of cloud cover, *Bulletin of the American Meteorological Society*, 74, 1851–1862, 1993.
- Scolnic, D., Casertano, S., Riess, A., Rest, A., Schlafly, E., Foley, R. J., Finkbeiner, D., Tang, C., Burgett, W. S., Chambers, K. C., Draper, P. W., Flewelling, H., Hodapp, K. W., Huber, M. E., Kaiser, N., Kudritzki, R. P., Magnier, E. A., Metcalfe, N., and Stubbs, C. W.: SUPERCAL: CROSS-CALIBRATION OF MULTIPLE PHOTOMETRIC SYSTEMS TO IMPROVE COSMOLOGICAL MEASUREMENTS WITH TYPE Ia SUPERNOVAE, *The Astrophysical Journal*, 815, 117, <https://doi.org/10.1088/0004-637X/815/2/117>, 2015.
- Sebag, J., Andrew, J., Klebe, D., and Blatherwick, R.: LSST all-sky IR camera cloud monitoring test results, *Proceedings of SPIE - The International Society for Optical Engineering*, 7733, <https://doi.org/10.1117/12.856337>, 2010.
- Sharma, N., Jain, V., and Mishra, A.: An Analysis Of Convolutional Neural Networks For Image Classification, *Procedia Computer Science*, 132, 377–384, <https://doi.org/https://doi.org/10.1016/j.procs.2018.05.198>, international Conference on Computational Intelligence and Data Science, 2018.
- Sharma, T., Parihar, P., and Kemkar, M.: All sky scanning cloud monitor for NLOT site survey, *Journal of Physics: Conference Series*, 595, 012032, <https://doi.org/10.1088/1742-6596/595/1/012032>, 2015.
- Shaw, J. A. and Nugent, P. W.: Physics principles in radiometric infrared imaging of clouds in the atmosphere, *European Journal of Physics*, 34, S111–S121, <https://doi.org/10.1088/0143-0807/34/6/s111>, 2013.
- Shi, C., Zhou, Y., and Qiu, B.: CloudU-Netv2: A cloud segmentation method for ground-based cloud images based on deep learning, *Neural Processing Letters*, 53, 2715–2728, 2021a.
- Shi, C., Zhou, Y., Qiu, B., Guo, D., and Li, M.: CloudU-Net: A Deep Convolutional Neural Network Architecture for Daytime and Nighttime Cloud Images' Segmentation, *IEEE Geoscience and Remote Sensing Letters*, 18, 1688–1692, <https://doi.org/10.1109/LGRS.2020.3009227>, 2021b.
- Shi, C., Zhou, Y., and Qiu, B.: CloudRaednet: residual attention-based encoder-decoder network for ground-based cloud images segmentation in nycthemeron, *International Journal of Remote Sensing*, 43, 2059–2075, 2022.
- Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, 2015.
- Skidmore, W., Schöck, M., Magnier, E., Walker, D., Feldman, D., Riddle, R., Els, S., Travouillon, T., Bustos, E., Seguel, J., et al.: Using All Sky Cameras to determine cloud statistics for the Thirty Meter Telescope candidate sites, in: *Ground-based and Airborne Telescopes II*, vol. 7012, pp. 862–870, SPIE, 2008.
- Smith, S. and Toumi, R.: Measuring Cloud Cover and Brightness Temperature with a Ground-Based Thermal Infrared Camera, *Journal of Applied Meteorology and Climatology*, 47, 683 – 693, <https://doi.org/https://doi.org/10.1175/2007JAMC1615.1>, 2008.
- Stubbs, C. W. and Brown, Y. J.: Precise astronomical flux calibration and its impact on studying the nature of the dark energy, *Modern Physics Letters A*, 30, 1530030, <https://doi.org/10.1142/S021773231530030X>, 2015.
- Stubbs, C. W. and Tonry, J. L.: Addressing the Photometric Calibration Challenge: Explicit Determination of the Instrumental Response and Atmospheric Response Functions, and Tying it All Together, 2012.
- Sun, X., Liu, L., and Zhao, S.: Whole Sky Infrared Remote Sensing of Cloud, *Procedia Earth and Planetary Science*, 2, 278–283, <https://doi.org/https://doi.org/10.1016/j.proeps.2011.09.044>, the Second International Conference on Mining Engineering and Metallurgical Technology (MEMT 2011), 2011.
- Sun, X.-J., Gao, T.-C., Zhai, D.-L., Zhao, S.-J., and Lian, J.-G.: Whole sky infrared cloud measuring system based on the uncooled infrared focal plane array, 2008.
- Szejwach, G.: Determination of Semi-Transparent Cirrus Cloud Temperature from Infrared Radiances: Application to METEOSAT, *Journal of Applied Meteorology and Climatology*, 21, 384 – 393, [https://doi.org/https://doi.org/10.1175/1520-0450\(1982\)021<0384:DOSTCC>2.0.CO;2](https://doi.org/https://doi.org/10.1175/1520-0450(1982)021<0384:DOSTCC>2.0.CO;2), 1982.
- Thurairajah, B. and Shaw, J.: Cloud statistics measured with the infrared cloud imager (ICI), *IEEE Transactions on Geoscience and Remote Sensing*, 43, 2000–2007, <https://doi.org/10.1109/TGRS.2005.853716>, 2005.
- Tzoumanikas, P., Nikitidou, E., Bais, A., and Kazantzidis, A.: The effect of clouds on surface solar irradiance, based on data from an all-sky imaging system, *Renewable energy*, 95, 314–322, 2016.
- Ugolnikov, O. S., Galkin, A. A., Pilgaev, S. V., and Roldugin, A. V.: Noctilucent cloud particle size determination based on multi-wavelength all-sky analysis, *Planetary and Space Science*, 146, 10–19, 2017.
- Várnai, T. and Marshak, A.: Satellite observations of cloud-related variations in aerosol properties, *Atmosphere*, 9, 430, 2018.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I.: Attention Is All You Need, 2023.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, 17, 261–272, <https://doi.org/10.1038/s41592-019-0686-2>, 2020.
- Wang, M., Zhuang, Z., Wang, K., Zhou, S., and Liu, Z.: Intelligent classification of ground-based visible cloud images using a transfer convolutional neural network and fine-tuning, *Opt. Express*, 29, 41 176–41 190, <https://doi.org/10.1364/OE.442455>, 2021a.
- Wang, Y., Liu, D., Xie, W., Yang, M., Gao, Z., Ling, X., Huang, Y., Li, C., Liu, Y., and Xia, Y.: Day and Night Clouds Detection Us-

- ing a Thermal-Infrared All-Sky-View Camera, *Remote Sensing*, 13, <https://doi.org/10.3390/rs13091852>, 2021b.
- Wei, L., Zhu, T., Guo, Y., and Ni, C.: MMST: A Multi-Modal Ground-Based Cloud Image Classification Method, *Sensors*, 23, 5, <https://doi.org/10.3390/s23094222>, 2023.
- Wilczak, J., Gossard, E., Neff, W., and Eberhard, W.: Ground-based remote sensing of the atmospheric boundary layer: 25 years of progress, *Boundary-Layer Meteorology*, 78, 321–349, 1996.
- Wong, W. K., Tan, P. N., Loo, C. K., and Lim, W. S.: An effective surveillance system using thermal camera, in: 2009 international conference on signal acquisition and processing, pp. 13–17, IEEE, 2009.
- Xie, W., Liu, D., Yang, M., Chen, S., Wang, B., Wang, Z., Xia, Y., Liu, Y., Wang, Y., and Zhang, C.: SegCloud: a novel 15 cloud image segmentation model using a deep convolutional neural network for ground-based all-sky-view camera observation, *Atmospheric Measurement Techniques*, 13, 1953–1961, <https://doi.org/10.5194/amt-13-1953-2020>, 2020.
- Yamashita, M. and Yoshimura, M.: Ground-based cloud observation 20 for satellite-based cloud discrimination and its validation, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39, 137–140, 2012.
- Yang, J., Lu, W., Ma, Y., and Yao, W.: An Automated Cirrus Cloud Detection Method for a Ground-Based Cloud Image, *Journal of Atmospheric and Oceanic Technology*, 29, 25 527 – 537, <https://doi.org/10.1175/JTECH-D-11-0002.1>, 2012.
- Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., and Shen, F.: Image Data Augmentation for Deep Learning: A Survey, 2022.
- Ye, L., Cao, Z., and Xiao, Y.: DeepCloud: Ground-Based Cloud Image Categorization Using Deep Convolutional Features, *IEEE Transactions on Geoscience and Remote Sensing*, 55, 5729–5740, <https://doi.org/10.1109/TGRS.2017.2712809>, 2017.
- Yoshimura, M. and Yamashita, M.: Contribution of ground-based 35 cloud observation to satellite-based cloud discrimination, *Journal of Environmental Science and Engineering. A*, 2, 487, 2013.
- Zhang, Z., Yang, S., Liu, S., Cao, X., and Durrani, T. S.: Ground-Based Remote Sensing Cloud Detection Using Dual Pyramid Network and Encoder-Decoder Constraint, *IEEE 40 Transactions on Geoscience and Remote Sensing*, 60, 1–10, <https://doi.org/10.1109/TGRS.2022.3163917>, 2022a.
- Zhang, Z., Yang, S., Liu, S., Xiao, B., and Cao, X.: Ground-Based 45 Cloud Detection Using Multiscale Attention Convolutional Neural Network, *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5, <https://doi.org/10.1109/LGRS.2021.3106337>, 2022b.
- Zhao, C., Zhang, X., Luo, H., Zhong, S., Tang, L., Peng, J., and Fan, J.: Detail-Aware Multiscale Context Fusion Network for Cloud Detection, *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5, <https://doi.org/10.1109/LGRS.2022.3207426>, 2022.
- Zhou, Q., Zhang, Y., Li, B., Li, L., Feng, J., Jia, S., Lv, S., Tao, F., and Guo, J.: Cloud-base and cloud-top heights determined from a ground-based cloud radar in Beijing, China, *Atmospheric environment*, 201, 381–390, 2019.
- Zhou, Z., Zhang, F., Xiao, H., Wang, F., Hong, X., Wu, K., and 55 Zhang, J.: A novel ground-based cloud image segmentation method by using deep transfer learning, *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5, 2021.