Review

An Introduction to Machine and Deep Learning Methods for Cloud Masking Applications

Anna Anzalone [1,2,3,*], Antonio Pagliaro [1,2,3] and Antonio Tutone [1]

1   Istituto Nazionale di Astrofisica INAF IASF Palermo, Via Ugo La Malfa 153, 90146 Palermo, Italy;
    antonio.pagliaro@inaf.it (A.P.); antonio.tutone@inaf.it (A.T.)
2   Istituto Nazionale di Fisica Nucleare Sezione di Catania, Via Santa Sofia 64, 95123 Catania, Italy
3   ICSC—Centro Nazionale di Ricerca in HPC, Big Data e Quantum Computing, 40121 Bologna, Italy
*   Correspondence: anna.anzalone@inaf.it

Abstract: Cloud cover assessment is crucial for meteorology, Earth observation, and environmental monitoring, providing valuable data for weather forecasting, climate modeling, and remote sensing activities. Depending on the specific purpose, identifying and accounting for pixels affected by clouds is essential in spectral remote sensing imagery. In applications such as land monitoring and various remote sensing activities, detecting/removing cloud-contaminated pixels is crucial to ensuring the accuracy of advanced processing of satellite imagery. Typically, the objective of cloud masking is to produce an image where every pixel in a satellite spectral image is categorized as either clear or cloudy. Nevertheless, there is also a prevalent approach in the literature that yields a multi-class output. With the progress in Machine and Deep Learning, coupled with the accelerated capabilities of GPUs, and the abundance of available remote sensing data, novel opportunities and methods for cloud detection have emerged, improving the accuracy and the efficiency of the algorithms. This paper provides a review of these last methods for cloud masking in multispectral satellite imagery, with emphasis on the Deep Learning approach, highlighting their benefits and challenges.

Keywords: cloud masking; deep learning; machine learning; image analysis

1. Introduction

Cloud cover evaluation is a crucial task in meteorology and Earth observation, as it provides valuable information for weather prediction, climate modeling, and environmental monitoring [1–3]. Depending on the application, cloud contaminated pixels must be recognized and valued to contribute in the scientific understanding of the atmospheric processes, or they must be removed to not affect advanced processing of the satellite imagery e.g., for land monitoring objectives and most remote sensing activities.

The importance of cloud screening is not limited to the most common contexts; in fact, considering that clouds are present on almost two-thirds of the Earth [4], cloud recognition is also an essential task in the detection of ultra-high-energy cosmic rays (UHECRs) from ground observatories and space missions [5–9]. One of the main concerns is the presence of clouds in the ultraviolet telescope field of view that can affect night-time indirect measurements of the UHECRs and Cherenkov radiation, and UHECR energy/direction reconstruction phases. Meaningful parameters, like cloud cover and cloud top/bottom height, can be retrieved from the data of the purpose-built spectral sensors [10,11].

There has also been a growing interest in characterizing cloud distribution in order to optimize the management of photovoltaic (PV) systems. The presence of clouds causes fluctuations in the efficiency of PV by attenuating the solar irradiance reaching the panels. Cloud cover and solar irradiance nowcasting allows the estimation of PV power generation, optimal grid and battery management, and the optimal scheduling and use of the power generated by the different units [12–14].

In general, the cloud masking process aims to generate an image in which each pixel in a spectral satellite image is classified as either clear or cloudy. However, a multi-class output is also common in the literature, where image pixels are classified according to the type of clouds or the type of objects: land, cloud, clear sky, cloud shadow, haze, ice/snow.

In the past years, when the public access to the remote sensing imagery was limited to a very reduced number of images, cloud cover retrieval was carried out with the intervention of a human operator. Approaches that rely on manually classified images are not only time consuming, depend on the expertise and discretion of the operator, on the criteria used to distinguish ambiguous cases such as clouds over ice/snow, thin clouds versus haze/aerosol, and on how to deal with intrinsic cloud problems such as the blurred nature of cloud edges.

Subsequently, as a consequence of the large amount of remote sensing data available, such as Terra-Aqua, Landsat, Sentinel open access archives (e.g., Landsat since 2008 [15]), more accurate automatic cloud masking codes have been developed, proposing algorithms for single or multi-temporal scenes. The most widespread traditional methods are known as rule based or thresholding techniques. These approaches apply thresholds to a selected set of image spectral bands according to fixed rules linked to physical properties of clouds.

Advancements in artificial intelligence methods and the acceleration of GPU, have provided new opportunities and methods with improved accuracy and efficiency. In this paper, we introduce the application of Machine and in particular Deep Learning methods for cloud masking, highlighting their advantages, challenges, and future prospects. However, some traditional methods will be briefly presented below, as they represent the reference methods for comparing the results of the latest approaches.

This paper is structured as follows. Section 2 introduces the approach utilized by classic thresholding methods that rely on spectral information. Section 3 presents the most common Machine Learning algorithms employed in cloud detection. Section 4 focuses on Deep Learning techniques, providing an examination of their role in cloud detection for the generation of dense masks. Section 5 critically addresses challenges associated with data quality and labelled data, emphasizing their pivotal role in enhancing the accuracy and reliability of Machine and Deep Learning models. Finally, Section 6 summarizes the challenges and future prospects in cloud masking applications.

2. Thresholding Methods

Traditional cloud detection approaches generally use combinations of multi-spectral information derived from solar reflectance and thermal emissions. These methods are based on the general spectral properties of clouds compared to those of underneath elements. Cloud reflectances are generally higher in visible spectral bands than most cloud-free Earth's surfaces, while cloud-top thermal radiances (Brightness Temperatures, BT), when available, are lower than the cloud-free surface temperatures [16,17].

Fixed or adaptive thresholds on BTs, BT-Differences and BT ratio, are applied to perform multi tests, which when combined with the results of those on solar reflectance bands, according to fixed rules, contribute to determine the final output mask. The band tests take into account different cloud and scene types [18–20]; however, there are conditions where cloud detection is not a straightforward process, such as clouds over ice/snow due to their similar radiative appearance, or very transparent cirrus (Figure 1), whose temperature is affected by the underlying atmosphere. In general, the radiance BT received by a sensor in the thermal infrared (TIR) band would not be exactly that emitted by the cloud top, especially in the case of cirrus, very high thin clouds, where the measured radiance represents the temperature of the underlying objects (lower thick clouds or surfaces). Therefore, in this case, some correction algorithms based on the radiative transfer process are needed to account for the atmospheric effects and retrieve the true cloud top temperature [10,21,22]. Combining information from multiple bands is then useful for resolving uncertain cases. In [23], it is reported how the TIR band can help to improve the performance of a combined radiative transfer and Machine Learning method in doubtful cases for cloud and haze

discrimination. In [24], the water vapour absorption band, used also for the Moderate Resolution Imaging Spectroradiometer (MODIS) masks [19], is a way to identify cloud cirrus in Landsat 8 images. The solar radiation, in fact, is partially absorbed in that band so that the typical bright appearance of the cloud is maintained.
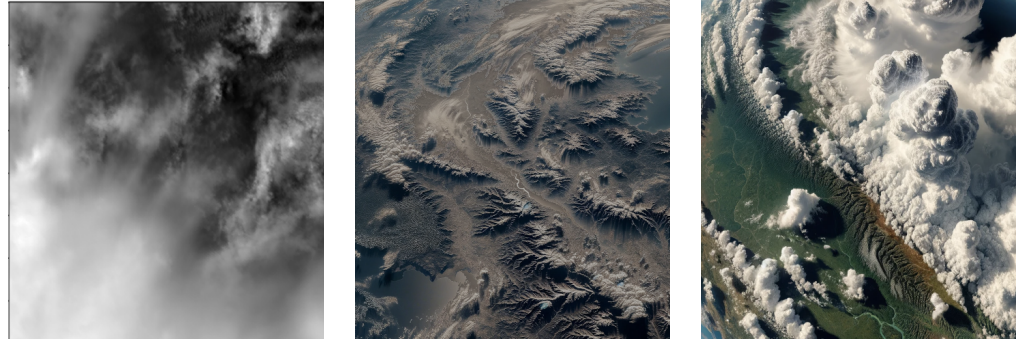
Figure 1. (**Left**) Thick cloud and thin cirrus real image from the Sentinel 3-SLSTR data archive (S1 band). (**center** cirrus) and (**right** cumulus) generated through advanced artificial intelligence models (developed by OpenAI, specifically DALL·E), illustrate idealized representations of these cloud types to enhance specific features of cumulus and cirrus clouds, providing clearer insights into their properties.

Hence, the availability of instruments with a large range of spectral bands is helpful in dealing with some extensively studied issues of cloud identification e.g., clouds over sand, snow or ice, whose spectra, in general, are similar in reflectance, cloud–haze discrimination, semi-transparent thin cloud detection [25–27]. Nevertheless, traditional cloud detection methods show difficulty to perform with high accuracy. They are limited to instruments that are equipped with the required bands, are more effective at detecting high, thick clouds than low, thin clouds. Thin cloud can be omitted above all cloud edges or when surfaces have smooth texture or high reflectance [28,29].

Some relevant operational algorithms that provide advanced products for Landsat 8 and Sentinel-2 satellite imagery as well as NASA and ESA missions dedicated to Earth-monitoring for environment surveillance and security [30,31], respectively, are listed below. These two satellites and the methods developed are mentioned here because the results of the latest Machine Learning algorithms are mainly compared with them. Furthermore, the annotated cloud masks for their image archives are publicly available and used to train and validate the more recent Deep Learning algorithms.

FMask [26,32–34] is an operational algorithm initially developed to discriminate between cloud and cloud shadows, and successively improved to include snow and water detection. It is based on a single scene analysis where tests and thresholds are applied to the image spectral bands according to the brightness and temperature characteristics of clouds and cloud-free surfaces, and the darkness characteristics of cloud shadows. The method was initially developed for Landsat 4 to 7 data, but was subsequently improved and developed for Landsat 8 and Sentinel-2 data.

Another single-scene approach is Sen2Cor. This is an ESA processor consisting of five modules [35] designed to correct Sentinel-2 Top-Of-Atmosphere products for the effects of the atmosphere, and to output various second-level products, including scene classification maps with cloud and snow probabilities. Labels are assigned to each pixel as a combination of the confidence level of a set of threshold tests on spectral reflectance, ratios, and indices.

MAJA (Maccs-Atcor Joint Method) [36,37] was proposed by the National Centre for Space Studies (CNES) in collaboration with Center d'Etudes Spatiales de la BIOsphère (CESBIO). This is an example of a multi-temporal method that uses image time series to detect cloud-covered pixels. The frequent satellite revisit of an area allows a local correlation of pixel features. The algorithm is mainly based on a threshold of the temporal variation of the reflectance in the selected bands, as the position and characteristics of clouds vary

considerably over time in images over the same areas compared to surfaces with cloud-free pixels. Some updates of the method are described in [38] where the major changes regard cloud shadow detection. Cloud shadows are usually more evident in the IR band, the variation of the vegetation cause also greater variations over time of the surface reflectances in the NIR and SWIR bands than in the shorter wavelengths. Consequently, the authors decided to use the red band that provided better results.

To exploit temporal information, addressing the inherent complexity and the need for clear-sky pixel reference, long-term missions with open access to the satellite image archive and adequate computing power are required.

Although the physical properties on which the rules are based allow the methods to be easily applied to instruments with similar characteristics [27], all these methods generally suffer from limitations as they rely on a priori knowledge to determine the test sets and thresholds, and on the spectral difference between cloudy and clear sky.

3. Machine Learning Methods

3.1. Machine Learning Algorithms

Machine Learning (ML) has significantly advanced cloud masking, offering efficient and accurate identification of complex cloud patterns in vast datasets like satellite imagery and climate data [3,39]. These ML models, unlike traditional programming, autonomously learn from data inputs and corresponding outputs, uncovering hidden correlations. This learning process enhances cloud masking precision and efficiency, reducing the need for extensive programming and deep expert knowledge.

Key to the success of ML in cloud masking is the use of labelled data, vital for training models to discern between different cloud types and clear-sky regions. Labelled data allows for data training and validation, conferring the output model knowledge inference capabilities onto the application domain. On top of that, it aims to provide accurate predictions for new, unseen data, increasing cloud mask reliability across various settings. However, challenges persist in their application, as summarized in Section 6.

ML methods usually apply to cloud detection as a single-pixel supervised classification problem, identifying classes like cloud, cloud-free, cloud shadow, land, or specific cloud types. Common algorithms include artificial neural networks (ANN) [40–42], Support Vector Machines (SVM) [43–46], and Random Forest (RF) methods [47,48], each capturing intricate relationships between input features and output labels.

In the traditional ML algorithms, prior to analysis, satellite data can undergo preprocessing to extract pertinent spectral bands and indices, which are then used as input for these models. The final masks consist of pixel probabilities of belonging to predefined classes or labels corresponding to the categories.

The features used are crucial for a successful classification, as well as the choice of a representative training data-set. Observed radiative variables inherited by thresholding approaches, spectral signatures (BTs, temperatures, derived indices as Normalized Difference Vegetation/Snow/Water indices), and cloud texture are some of the main characteristics to identify differences between the radiative properties of clear and cloudy skies [44]. As an example the 10 Sentinel-2 spectral bands are used to train the single-scene S2cloudless method [49] with a pixel-based strategy, using gradient boosting decision tree. The final cloud probability map can be converted to a cloud mask by fixing a threshold.

In [3], Li et al., deal extensively with the use of spectral, spatial, temporal, and multi-source features in cloud detection algorithms of various types, with a focus on cloud shadow detection. In [50], the authors use an SVM classifier to distinguish between cloud types and land surfaces proposing a method to detect and remove thick clouds, which constitute a significant portion of remote sensing data. Brightness intensity combined with gradient of texture features are the main characteristics used to detect samples to train the classifier. In [45], the SVM model was applied to high-resolution image data and this allowed the use of more defined texture features in addition to spectral characteristics. Ishida et al. [46] developed an SVM method to treat various cloudy combinations, and

investigate on the effectiveness of the use of reduced amount of typical data to train the model. The adaptability of the method to multiple conditions was demonstrated with data from MODIS. Gomez-Chova et al. [51] introduced a semi-supervised SVM method that showed good performance when the reference data did not adequately describe the target classes. Random Forest algorithms have been applied to cloud identification by using different splitting criteria, such as entropy, Gini index, and classification error [52,53]. Hollstein et al. [54] applied decision trees and Bayesian models for a ready-to-use cloud detection algorithm using Sentinel-2 images. As an example of neural network applied to clouds, in [55] the authors presented the Spatial Procedures for Automated Removal of Cloud and Shadow (SPARCS) algorithm that uses neural networks and rule-based post-processing for a multi-class cloud detection on Landsat images.

3.2. Advantages and Challenges of Traditional Machine Learning for Cloud Masking

One notable advantage of using traditional ML methods over thresholding-based ones is their ability to manage complex and nonlinear patterns and provide good generalization performance. This proves particularly valuable when dealing with intricate and variable cloud patterns that defy simple rule-based models [56]. Using labelled data for training enables these algorithms to understand cloud and clear-sky characteristics and generalize this knowledge to new data. ML models, can handle noisy and large data-sets and high-dimensional feature spaces. The examination of high-resolution satellite images and the incorporation of multiple input features, permit the detection of small-scale cloud features that can be overlooked by traditional rule-based methods.

However, there are challenges in their application. In very-high-resolution imagery, the finer resolution, while providing more detail and thus greater spectral heterogeneity in neighbouring features, can also increase the problematic nature of cloud edges and thin clouds, as well as highlighting small bright objects on the ground that have the same brightness as the clouds. In this case, for example, RF algorithms for cloud segmentation, which are trained from datasets of individual pixels with unrelated features in their neighbourhood, perform worse than a convolutional neural network (CNN) that inherently utilizes spatial context in addition to the spectral features of pixels.

Preparing extensive labelled data for algorithm training can be resource-intensive and time-consuming, limiting its applicability in certain scenarios [57,58]. In addition, the creation of adequate training data-sets is a significant challenge. It requires knowledge of spectral properties of clouds as well as of image analysis methods. Furthermore, it is difficult to consider all possible cloudy combinations in advance of constructing a classifier, particularly for rare or localized conditions. Satellite images present various spectral information, different spatial resolutions, and a great variety of complex land types, which should be taken into account for the adaptability of the model. Additionally, the performance of the sensor can limit the effectiveness of the model. For instance, the lack of key wavelengths can make it difficult to find features that can reduce the frequency of incorrect cloud discrimination. Finally, the lack of spatial information, in terms of spatial correlation between neighbouring pixels, that is typical of single-pixel methods, can significantly reduce accuracy.

These critical aspects of the traditional ML models are improved or overcome by the application of Deep Learning methods. Deep Learning algorithms have excelled in image classification due to their superior feature representation, enhancing the final accuracy. Instead of manually choosing the appropriate characteristics of reference data, these algorithms autonomously learn spatial and semantic features directly from training data, reducing the subjectivity. The multi-layered architecture significantly amplifies the richness and variety of extracted features, thereby elevating classification accuracy, as discussed in the next sections.

4. Deep Learning Methods

4.1. Deep Learning for Semantic Segmentation

Deep Learning enables the development of computational models structured with several layers of processing. These models are capable of learning and identifying data representations through various levels of abstraction [59]. Deep Learning architectures such as CNNs (Convolutional Neural Networks) and Fully Convolutional Neural Network (FCN) [60,61] have significantly improved the results in cloud detection and masking [39,62]. Specifically, their ability to effectively capture spatial patterns and structures from images, in addition to spectral features, has led to more accurate image classification and segmentation algorithms [63–65]. In this subsection it is a brief overview of state-of-the-art Deep Learning methods for semantic segmentation.

The hierarchical structure of CNNs, including convolutional, pooling, and fully connected layers, enables them to learn increasingly complex features from the input data by increasing the number of convolutional layers, automatically capturing in the cloud case, spatial, and spectral characteristics, with the notable advantage of not depending on domain experts for manual feature selection. The loss of resolution, i.e., of local and global information due to pooling and subsampling operations of a basic CNN, make it less suitable for semantic segmentation, which requires knowledge of more details to achieve accurate pixel-level classification.

However, the research in this area has produced CNN-based architectures in the last decade [65] that minimize the loss of global, local, and spatial information. Considerable progress has been made in Fully Supervised methods whose main limitation is the need for large volumes of annotated data for training purposes. Weakly supervised learning approaches are a growing area of research that tends to overcome this limitation and speed up the process, although their performance is currently inferior to that of supervised methods, partly due to the coarser granularity of data annotations [66–68]. Generative Adversarial Networks or GANs are generative/discriminative models capable of generating input-consistent images and acting as [69,70] classifiers, used as a weakly supervised method to tackle the laborious task of generating pixel labels. Unsupervised Domain Adaptation method, UDA, can be considered a special case of transfer learning, and the strategy consists of adapting existing models from a source domain to a target domain for which no labelled data exists. UDA approaches differ according to the adaptation techniques used, e.g., self-training or Generative Adversarial Networks-based [65,71–73].

In supervised learning, fully convolutional neural networks (FCNs) represent a milestone for pixel-wise end-to-end prediction masks. In FCNs, the substitution of fully connected layers into convolutional layers enabled to obtain segmentation maps with the same size as the input image but in general, spatially detailed information is lost.

The encoder–decoder two-stage structure is extensively used where the spatial information is recovered after having downsampled the feature map size, by gradually resizing the spatial features in the upsample stage. Classic examples are the well-known architectures U-Net [74] (Figure 2) and SegNet [75], designed for segmenting biomedical images and natural images, respectively, and the more recent SFANet [76] and CANet [77]. The U-Net model has given rise to several variations in different fields such as UNet++, a nested U-Net architecture proposed in [78] for medical images. High-Resolution Network, HRNet, and its evolution Hrnetv2, which is more specific for semantic segmentation applications [79], are designed with two multiresolution connected sub-nets to maintain the high-resolution image information through the parallel high-to-low convolution streams. DeepLab and its later changes such as DeepLabv3+ are representative of a class of models where dilated/atrous convolutions are introduced to enlarge the receptive field of convolutions while maintaining the computational cost, enhancing the ability to capture multi-scale contextual information critical for detailed segmentation tasks [80]. Differently from the previous examples, there are models that include the generation of multi-resolution pyramid-based feature representations to better learn the global context and detect objects at different scale: FPN, Feature Pyramid Network [81], and PSPNet,

Pyramid Scene Parsing Network [82], are examples. They are based on multi-resolution image analysis approach that is well suited to the multi-scale hierarchical architecture of a deep CNN. In attention-based models, attention mechanisms are used to selectively focus on relevant input information, even at different scales and positions as in [83]. Transformer architectures [84], which have been successful in natural language processing and are based on self-attention mechanisms, have also been introduced for semantic segmentation with the goal of capturing more contextual information than the receptive field of an FCN allows. They can capture global interactions between elements in a scene, but at a high computational cost when applied to the raw image. Vision Transformer (ViT) like in [85] and Segmenter [86] are applied directly to sequences of image patches, with efficient performance on reference image segmentation datasets and without the use of convolutions. In contrast, other proposed architectures are combined with FCN, such as Swin Transformer [87] and SETR [88].
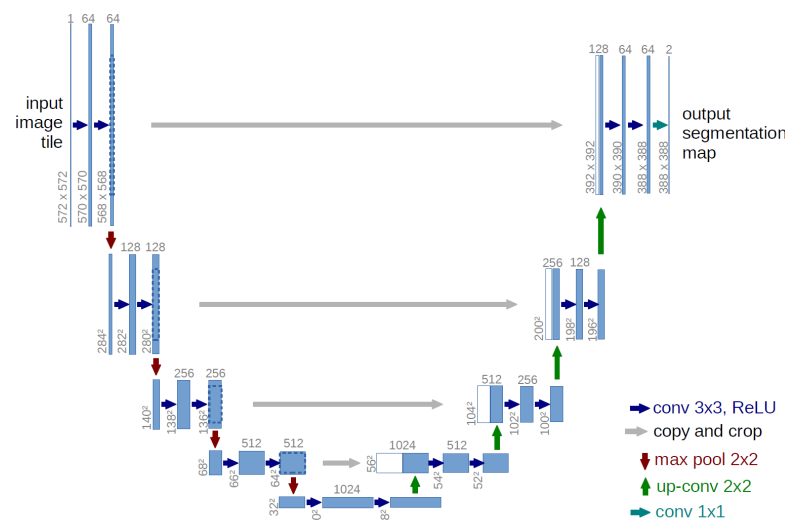
Figure 2. U-Net architecture overview. The U-Net design, characterized by a contracting path on the left and an expansive path on the right, forms a U-shaped structure. The contracting path involves convolutional and pooling layers for feature extraction, while the expansive path employs transposed convolutions to enable precise localization and reconstruction. Skip connections bridge corresponding layers, aiding in the retention of fine-grained details. Courtesy of O. Ronneberger [74].

A large collection of datasets is available to train and evaluate the performance of methods and models for semantic segmentation, that are used also for other computer vision tasks such as classification and object recognition. The annotated data include 2D, RGB-D and 3D images, indoor/outdoor scenes, general and street scenes with a large number of different object classes. The review [65] shows that the Intersection of Union, IoU, that quantifies the overlap between the segmented region and the ground truth, averaged (mIoU) over all methods included in the articles exceeded 70% on the PASCAL VOC datasets [89] and Cityscapes [90] out of 14 data collections. Both datasets are considered the gold standard for image segmentation. Furthermore, from both studies [64,65] came the result that the state-of-the-art models led to a relative improvement of approximately 25% of the mIoU accuracy on the two datasets compared to the first FCN model.

This is a non-exhaustive overview of approaches developed in the domain of semantic segmentation; deep learning cloud masking methods originate from many of them.

4.2. Deep Learning for Cloud Masking

The problem of generating pixel-wise cloud masks can be regarded as a semantic segmentation task, whose algorithms are highly concerned with the semantic content of an image and aim to classify each pixel into meaningful object classes (specifically cloud, clear

sky, land, water, snow, haze and so on). Deep Learning architectures developed to segment natural images form the basis for addressing this topic.

Some early works proved the potentiality of CNN for this task. For instance, the article in [91] is an example of a study for the application of a simple CNN architecture to the task of cloud masking; performed on multispectral Proba-V images, it was shown that the approach was promising. Generally, CNNs are used for classification aims, but also for segmentation tasks approached as a classification task when a pixel level is not required. As an example, in [92] an improved version of the Simple Linear Iterative Clustering (SLIC [93]) method is used to generate superpixels, on the basis of value similarity and proximity. Patches extracted from the superpixels are fed into a specifically designed deep CNN, dual-branch PCANet in [94] combined with an SVM classifier, responsible for predicting their categorization a cloud/non-cloud. This approach strongly relies on the accuracy of the image segmentation in superpixels.

Deep Learning algorithms for cloud detection at pixel-level provide labelled masks typically employing techniques that learn and infer class labels based on high-level semantic features directly from the raw input image. See Figure 3 as an example of masks.
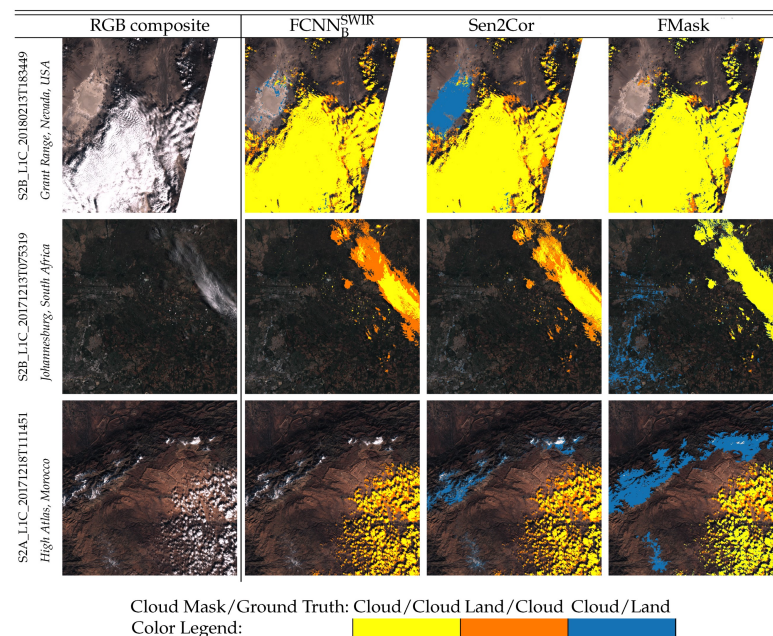
Figure 3. Cloud masks for different scenes (rows) with varying conditions, obtained by three different methods (columns). First column shows the original RGB images. Predicted masks and ground truth are overlapped. Color legend for predicted cloud masks vs ground truth: yellow indicates areas where cloud mask agrees with the ground truth, designating cloudy pixels (True Positives). Orange denotes regions where cloud mask erroneously identifies cloud-covered pixels as clear land (False Negatives). Blue highlights zones where cloud mask inaccurately classifies clear land as cloud-covered areas (False Positives). Reflective pixels are left white in masks and ground truth when they are not misclassified. Courtesy of López-Puigdollers [95].

U-Net-like architectures are well-suited for segmenting large images such as satellite data and they are extensively used in cloud masking. U-Net architecture involves feature detection and up-sampling, with the feature extraction phase growing in scale with each max-pooling operation. In recent years, a considerable amount of work has been presented in the literature where several methods for cloud masking are based on improved U-Net and different semantic segmentation models. CS-CNN, a Cloud Segmentation model for cloud masking was trained on MSG-SEVIRI images, in [96], and compared with RF models; U-Net variations were proposed for cloud/snow and cloud/cloud-shadow segmentation, respectively, in [97,98] and in [99] where the RS-Net model performance, for Landsat 8 images, was comparable to the reference FMask algorithm. In [100], the authors proposed

the specialized architecture Cloud-Net, trained on Landsat 8 38Cloud dataset (labelled by the authors), to capture global and local cloud features using specifically designed convolutional blocks, leading to a better Overall Accuracy (proportion of correct predictions out of all the predictions made by a model) than FMask. Francis et al. in [101] introduced the CloudFCN, a U-Net model that was trained and evaluated on a high-resolution Carbonite-2 dataset (80 cm), manually annotated by the authors, and on the L8 Biome dataset [27,102] from Landsat 8 level-1 images with annotations for different terrain types. CloudFCN predicts cloudy/clear pixel masks, but can also provide mask of pixel cloudiness for cloud amount evaluation, according to the user need. Relying on the texture characteristics of the surface, its performance decreases when clouds are over ice/snow if high-resolution bands are not available or only low-resolution RGB bands are included. Domnich et al. [103] present the KappaMask processor based on a 5-level U-Net, designed to provide 10 m resolution masks for Sentinel-2 images, at Northern European latitudes. It is an example of multi-class segmentation (clear, cloud, cloud shadow, thin cloud, and invalid). KappaMask was trained on KappaZeta Sentinel-2 annotated masks created by the authors and is freely available, and on Sentinel-2 CloudCatalogue [104]. The authors provide a comprehensive analysis of KappaMask's performance compared to state-of-the-art rule-based and Machine Learning methods. Key findings include KappaMask's superiority in cloud and cloud shadow detection with finer resolution (10 m) and a consistent reduction in false negatives for the clear class. CDUNet (Cloud Detection UNet) is proposed in [105] to cope with a better definition of cloud boundaries and fragmented clouds in cloud/cloud-shadow masking. The model improves detail feature extraction and feature fusion in upsampling steps, achieving the best results in various standard performance indices among a significant set of methods on SPARCS [55] datasets and selected Google Earth images. There are also methods utilizing, as post-processing algorithms, the conditional random field (CRF)-based method to improve boundaries via local features and distance information. The authors in [106] applied a slightly modified version of the U-Net architecture to obtain cloud detection masks for very-high-resolution images. Manually annotated images over three regions were used to train and validate the predicted cloud/non-cloud masks. Particular attention was paid to the case of terrain characteristics that can lead to misleading detection at this resolution; seasonality by choosing images from sites with different climate characteristics. SegNet, is adopted in [107] to effectively process Landsat imagery, including cloud shadow detection. Various architectures, such as MSCFF (Multi-Scale Convolutional Feature Fusion) [108], MF-CNN (Multiscale Features-CNN) [109], CD-FM3SF [110], were designed to work with multiscale feature extraction that, combined with FCN, performed more accurately than single-scale feature-based methods. Moreover, the incorporation of augmented atrous spatial pyramid pooling and fully connected fusion paths in FCN architectures was presented for high-resolution remote sensing imagery in [111]. Li et al. [39] provide a review of the most popular models and their modified versions for a more general cloud detection objective and focusing on the architectural aspects. In order to capture a wider information about global context, the authors in [112] present the U-shape Attention-based Transformer Net, (UATNet). It includes two transformer structures that integrate spatial and multi-spectral features. The model was trained and tested on the CRMSCD dataset, China Region Meteorological Satellite Cloud images, which was created by the authors taking images covering China, from the L1 data product of FY-4A labelled according to the synchronized Himawari-8 image classification. The overall performance was comparable to state-of-the-art FCNs-based and transformer-based models, on the same dataset, but it was not tested on other datasets.

Some authors have focused their attention on the use of transfer learning technique. When employing Deep Learning-based methods, it is essential to use an extensive, independent set of labeled samples (ground truths) for training and testing purposes. This training dataset must be sufficiently diverse, encompassing various characteristics relating to geographical locations, different seasonality, terrain types, balanced and class-representative data, and, in addition, must have a considerable volume to achieve the full potential of

Deep Learning. Consequently, the process of creating a verified dataset and developing a cloud detection algorithm via Deep Learning is extensive. Transfer learning techniques can capitalize pre-trained models on other datasets or tasks to initialize or fine-tune the Deep Learning model, which can reduce the amount of labelled data and accelerate the training process. Exploiting the existing labelled datasets of one satellite to train Deep Learning models suitable for other satellites is the subject of the articles in [95,113,114]. The transfer learning approach in [113] is applied between Landsat 8 and Proba-V sensors, that have different but similar spatial and spectral features, taking advantage of the existing labeled datasets. Different transfer learning strategies were experimented based on the distinction between transductive/inductive [115] transfer learning, after having applied a domain adaptation transformation from Landsat 8 to Proba V data. The results were compared with FMask, MSCFF, and RS-Net reporting remarkable detection accuracy despite the sensor differences. In [95,116], the authors trained U-Net based architectures, on Landsat 8 dataset to transfer the model to Sentinel-2 images. In [95] the authors present a modified U-Net architecture with a moderate number of parameters, that gave rise to CD-FCNN (CD-FCNN: U-Net with two different Sentinel-2 band combinations: VNIR, RGBISWIR trained on the L8 Biome), which was part of the 10 cloud masking methods evaluated in the first Mask Intercomparison eXercise (CMIX) [117]. The aim of the work in [95] was twofold: to provide a model obtained from training on Landsat 8 data, to be applied to Sentinel 2 images, and to highlight the dependence of the performance evaluation on the datasets used for validation. The proposed model performance was competitive versus the reference methods for Landsat 8 and Sentinel-2. In [118], an extension of GANs was used to adapt the real Proba-V images to match the distribution of the transformed Landsat 8 data. These data and their ground-truths are used to predict clouds in Proba-V.

The different specifications of the sensors on board the various satellites generate data that, even when the sensors are similar, are characterized differently. The previous works have shown the need to adapt the data in terms of spatial and spectral characteristics, e.g., by resampling operations or by discarding some spectral information, which leads to a lack of immediacy in the transfer of models from one sensor to another, as well as a lack of bands that may be meaningful for cloud detection. The sensor dependence of a model is addressed by Francis et al. in [119]. The proposed Spectral Encoder for Sensor Independence (SEnSeI) new architecture, represents a notable advancement in remote sensing applications and in cloud masking in particular, functioning as a preprocessing module that translates data from various sensors into a unified format. In this way, a model can operate on the data regardless of the type of sensor that produced it. Unlike methods like transfer learning, SEnSeI does not adapt pre-trained features but rather integrates data from multiple sources, creating a broadly applicable model across different sensors. This approach is particularly useful in standardizing multi-spectral data, thus overcoming the challenge of sensor-specific data variability. This approach can have an immediate spin-off in the design of supervised cloud detection models prior to the launch of new satellites. The pre-existence of data and their ground truths, is a necessary requirement to develop Deep Learning cloud segmentation models for new satellites. This limits the application of such methods to the time of the data availability and implies the use of more traditional methods that have a higher portability, although they are less accurate, until the satellite becomes operational. Alternatively, one could consider models trained on sensor data with similar characteristics to those to be launched, using a transfer learning approach, or exploit the advantages offered by SEnSeI by taking the data regardless of the type of source sensors.

4.3. Datasets and Performance Metrics

It is common to assess the result of a masking method by measuring its agreement with reference ground truths, which can be manually or tool-supported annotated masks, the most popular of which are summarized in Table 1, or validated cloud masks released as official satellite products.

Table 1. Summary of the most used datasets.

Dataset	Geo-Distribution	Number of Scenes
Landsat 8 (30 m)		
L8_SPARCS [55]	Worldwide	80
L8_Biome [27]	Worldwide	96
L8-38Cloud [100]	USA	38
L8-95Cloud [120]	USA	95
Sentinel-2 (10 m)		
S2-Hollstein [54]	Europe	59
S2-BaetensHagolle [121]	Worldwide	35
CESBIO [38]	Europe	30
S2-CloudCatalogue (20 m) [104]	Worldwide	513
KappaZeta [103]	Northern Europe	150
CloudSEN12 [122]	Worlwide	49,400
WHUS2-CD [123]	China	32
Gaofen1 (16 m)		
GF1_WHU [124]	Worldwide	108
Levir_CS [125]	Worldwide	4168

Training Deep Learning models requires sets of annotated data, not only for training purposes but also for evaluating and comparing their performance. However, the datasets currently available have inconsistent characteristics. They are often created without standard protocols, and experts annotate images either by visual inspection or with the aid of tools (e.g., eCognition, Active Learning for Cloud Detection (ALCD) [38], Intelligently Reinforced Image Segmentation (IRIS) graphical user interface [126]). Moreover, there is no universal definition of a cloud, leading to variations in handling ambiguous scenarios like cloud boundaries, thin clouds, and cloud shadows across different teams. Factors like the unit of labeled sample, geographic coverage, number of annotated classes, and spatial resolution vary, impacting the evaluation and comparison of models or masks with different datasets.

CloudSEN12 [122] represents a significant advancement in addressing these issues. It is a recent, global, multitemporal dataset designed for Sentinel data. Its strengths lie in its extensive data volume, variety, inclusion of different cloud types including thin clouds and shadows, and a meticulous annotation process with quality control. Additionally, it provides temporal images of the same locations under varying weather conditions, enhancing its utility and overcoming many limitations of existing datasets.

An extensive overview of current annotated data collections, in addition to those mentioned in Table 1, can be found in [3,39]. A different approach to the creation of training datasets is presented in [127]. The authors introduce an open-source tool for the generation of various synthetic cloud types that is useful for different tasks, such as masking and cloud removal, without annotation costs. It has been shown that, in terms of performance, a good strategy is to perform training and evaluation using both simulated and real data.

To evaluate, validate and compare the performance of predicted masks and segmentation models, a framework of metrics is needed in addition to ground truth masks. The most popular metrics to quantify the accuracy of the segmentation models are summarized in Table 2. They are derived from the Confusion Matrix, determined on all pixels of the images in the dataset.

Table 2. Summary of evaluation metrics.

Metric	Formula	Description
Overall accuracy (OA)	$\frac{TP+TN}{TP+TN+FP+FN}$	Proportion of TP and TN detected out of all the predictions made by the model.
Balanced OA	$0.5 * \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right)$	Index of the model's ability to predict both classes.
Precision (UA)	$\frac{TP}{TP+FP}$	Proportion of TP detected out of all positive predictions made by the model.
Recall (PA)	$\frac{TP}{TP+FN}$	Proportion of TP that have been detected out of all actual positives.
F1-score	$\frac{2 \times Precision \times Recall}{Precision+Recall}$	Harmonic mean of recall and precision.
Intersection of Union (IoU) or Jaccard Index	$\frac{TP}{TP+FP+FN}$	Widely used for semantic segmentation, measures the overlap between predicted and actual positive.
Omission Error	$\frac{FN}{TP+FN}$	Proportion of actual positives missed by the model out of all actual positives.
Commission Error	$\frac{FP}{FP+TN}$	Proportion of negative instances incorrectly identified as positive by the model (overdetection).

TP (True Positive) = # of pixels for which the model produced a correct prediction of a positive class. TN (True Negative) = # of pixels for which the model correctly predicted a negative class. FP (False Positive) = # of pixels for which the model incorrectly predicted a positive class, when it was negative. FN (False Negative) = # of pixels for which the model incorrectly predicted a negative class when it was positive.

OA might be biased when the dataset is not balanced; in this case BOA or mIoU are preferred. BOA is the average of true positive rate and true negative rate. mIoU is computed by the average of the IoU (also Jaccard index) on the dataset images. This metric takes into account false positives and false negatives contrary to the OA. Recall PA (also Producer's Accuracy) is a complement to the Omission Error, while precision UA (also User's Accuracy) is a complement to the Commission Error. High PA values are obtained by methods that are negative conservative that for cloud/cloud-free classification, it means that the masks tend to be cloud-free conservative, while for high UA values they tend to be positive conservative (cloud conservative). F1 takes into account both precision and recall, its a good index when both metrics have the same weight.

It is worth noting that the quantitative assessment of the accuracy of methods are highly dependent on the reference datasets, as was highlighted in studies carried out in [95] and in CMIX [117]. In CMIX is also shown that the accuracy achieved by reference algorithms for cloud masking proves to be quite satisfactory although variable across all datasets. In CMIX, 10 algorithms, including traditional thresholding, Machine Learning and FCN-based methods for Sentinel-2 and Landsat 8 images (including Fmask, Sen2core, MAJA, S2cloudless, CD-FCNN), were compared on five cloud reference datasets. The BOA of the methods across the datasets ranged between $80.0 \pm 5.3\%$ and $89.0 \pm 2.4\%$ for Sentinel-2 and between $79.8 \pm 7.1\%$ and $97.6 \pm 0.8\%$ for Landsat 8, with an improvement

in the algorithm OA from +1.5% to 7.4% when more challenging cases were omitted from the datasets.

5. Most Significant Challenges Related to Data Quality and Labelled Data

The efficacy of Machine Learning and Deep Learning in cloud detection and masking is intrinsically linked to the use of labelled data. These data are crucial for training and testing models to differentiate cloud types and cloudy/clear-sky regions. Enhanced generalization abilities of models trained on labelled data ensure accurate predictions on unseen data, bolstering the reliability of cloud masks in various scenarios.

However, obtaining and labeling extensive datasets is resource-intensive. Techniques like transfer learning [115], data augmentation [128], and active learning [129] have emerged as solutions to mitigate this challenge. They reduce the need for large labelled datasets, enhance the diversity of training data, and optimize the labeling process.

Despite these advancements, several significant challenges persist:

- Quality of satellite imagery: Factors such as atmospheric interference and sensor noise affect the accuracy and reliability of satellite imagery. Pre-processing techniques improve input data quality but can introduce uncertainties [130,131].
- Variability of the publicly available annotated datasets: Publicly available annotated datasets are used for the training and testing of models. The development of large, high-quality datasets requires expertise and time-consuming processes; inconsistent labeling, especially in ambiguous situations, complicates model training and evaluation [95]. In addition, the resulting annotated masks are generally specific to a particular sensor. Different formats, resolution, variability of class types, different data volume, non exhaustive geographic scene distributions, are some factors that characterize most of the currently available annotated masks, besides the lack of thin clouds or cloud shadows labels. This leads to an uneasy utilization of the available amount of data. Finally, not all the reference data used by some developers are freely available.
- Diversity of labelled data: The diversity of the labelled data is another important factor that affects the accuracy and generalization performance of the model. The labelled data should be diverse enough to capture the variability and complexity of cloud patterns and atmospheric conditions. However, it can be challenging to obtain diverse labelled data, especially for rare or extreme cloud events.
- Data imbalance: Another challenge is the imbalance between cloud and non-cloud regions in the input data. Cloud regions are often less frequent than non-cloud regions, which can lead to bias and poor model performance. Techniques such as data augmentation and sampling can help to address this issue, but they can also introduce new sources of error and uncertainty.
- Sensor-dependent data: each dataset is linked to the spatial and spectral characteristics of the source sensor.

Future research should focus on improving and developing new Deep Learning techniques that address these challenges.

6. Challenges and Future Prospects in Cloud Masking Applications

Cloud detection and masking play pivotal roles in the process of cloud and atmospheric correction for remote sensing data. Employing techniques based on spectral, spatial, temporal, or contextual criteria (or their combinations) is commonplace for identifying and eliminating pixels affected by clouds, cloud shadows, or haze. However, this intricate process is susceptible to errors and uncertainties, presenting challenges such as cloud misjudgement with other elements. These challenges can significantly impact subsequent analyses and applications. Despite these difficulties, there exists a promising avenue for enhancing cloud masking techniques. Future prospects involve the development of advanced and robust cloud detection and masking methods capable of handling complex cloud scenarios by leveraging multiple sensors and platforms.

The accuracy of cloud segmentation has been improved with the introduction of Machine Learning methods, but in particular with the exploitation of the advantages offered by the inherent properties of Deep Learning. The dependence of these methods on the availability of large reference datasets for training and testing curbs the potential of the approach to date. Furthermore, the development of cloud segmentation methods for multi-sensor images is currently slowed down by the close relationship between the images and the characteristics of each sensor. It requires extensive reference data to train the models. Improvements in transfer learning techniques and data adaptability can benefit to address the problem. An important step forward in this direction is represented by the work in [119], where the sensor independence of a model is highlighted to make better use of existing data, increase the efficiency and applicability of models, and ultimately contribute to more effective and versatile remote sensing analysis.

Finally, the implementation of rigorous and reliable quality assessment and validation techniques emerges as a crucial aspect to measure the quality of corrected data and support informed decision-making processes.

To conclude, some considerations that recent studies have highlighted in assessing environmental and economic costs. The energy consumption of Deep Learning methods, which are crucial for processing large datasets, is often assumed to increase exponentially with the growth in the number of parameters [132]. However, this perspective shifts when focusing on inference costs rather than training costs. Inference, due to its multiplicative factors, accounts for a significant portion of computing effort in AI applications, including cloud masking. Contrary to expectations, for a sustained increase in AI performance, particularly in areas such as computer vision and natural language processing relevant to remote sensing, there is a much softer growth in energy consumption [133]. This can be attributed to both algorithmic innovations and advancements in hardware, which not only offer higher FLOPS but also bring significant energy efficiency optimizations. These developments suggest that while the initial environmental impact and costs of advanced techniques might be high, their long-term efficiency and cost-effectiveness are enhanced by these technological improvements. Despite this, the increasing pervasiveness of AI in various applications, including remote sensing, could potentially multiply the overall energy consumption, a factor that must be carefully considered in the broader environmental impact assessment [133]. Future work in cloud masking should increase performance evaluations from this point of view as well, and not only from the point of view of quantitative accuracy estimation, which is the most popular way so far.

7. Conclusions

The integration of Machine Learning and Deep Learning methods for cloud masking holds significant promise, offering considerable advantages over traditional approaches such as the thresholding-based method. Ongoing advancements in Deep Learning, including unsupervised and semi-supervised learning techniques, along with the creation of specialized architectures designed for cloud detection tasks, suggest a trajectory toward even more precise and reliable results. The utilization of Deep Learning has already demonstrated substantial enhancements in accuracy, with further refinements expected as novel techniques and architectures emerge. However, a critical consideration lies in the usability, amount, and quality of labelled data. The process of acquiring and labeling substantial datasets can be resource-intensive, potentially limiting the method's applicability. Additionally, the efficacy of the model can be influenced by their quality, as inaccurate or inconsistent labels may introduce noise and bias during training. Consequently, ongoing research is imperative to devise strategies that address challenges related to the data. Investigations should explore the potential of transfer learning techniques, exploiting pre-existing labelled datasets or models. This approach holds promise in diminishing the labelled data requirements for model training while enhancing accuracy and efficiency. Weakly supervised and unsupervised learning is a promising area of research for general image segmentation, which may also have positive implications in the field of remote sensing, overcoming the

problematic need for ground truth. Another avenue for research involves the development of ways to make the existing and new models, nonspecific to a given sensor. In summary, as Deep Learning techniques advance and specialized architectures for cloud detection evolve, these methods are expected to become more accurate, efficient, and widely applicable in the future.

References

1. Lefebvre, A.; Sannier, C.; Corpetti, T. Monitoring Urban Areas with Sentinel-2A Data: Application to the Update of the Copernicus High Resolution Layer Imperviousness Degree. *Remote Sens.* **2016**, *8*, 606. [CrossRef]
2. Zhao, S.; Liu, M.; Tao, M.; Zhou, W.; Lu, X.; Xiong, Y.; Li, F.; Wang, Q. The role of satellite remote sensing in mitigating and adapting to global climate change. *Sci. Total. Environ.* **2023**, *904*, 166820. [CrossRef] [PubMed]
3. Li, Z.; Shen, H.; Weng, Q.; Zhang, Y.; Dou, P.; Zhang, L. Cloud and cloud shadow detection for optical satellite imagery: Features, algorithms, validation, and prospects. *ISPRS J. Photogramm. Remote Sens.* **2022**, *188*, 89–108. [CrossRef]
4. King, M.D.; Platnick, S.; Menzel, W.P.; Ackerman, S.A.; Hubanks, P.A. Spatial and Temporal Distribution of Clouds Observed by MODIS Onboard the Terra and Aqua Satellites. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3826–3852. [CrossRef]
5. Pierre Auger Observatory. Available online: https://www.auger.org/ (accessed on 1 March 2024).
6. JEM-EUSO Joint Experiment Missions for Extreme Universe Space Observatory. Available online: https://www.jemeuso.org/ (accessed on 1 March 2024).
7. Mini-EUSO. Available online: http://jem-euso.roma2.infn.it (accessed on 1 March 2024).
8. Aielli, G.; Bacci, C.; Bartoli, B.; Bernardini, P.; Bi, X.J.; Bleve, C.; Branchini, P.; Budano, A.; Bussino, S.; Melcarne, A.C.; et al. Highlights from the ARGO-YBJ experiment. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip.* **2012**, *661* (Suppl. S1), S50–S55. [CrossRef]
9. Vercellone, S.; Bigongiari, C.; Burtovoi, A.; Cardillo, M.; Catalano, O.; Franceschini, A.; Lombardi, S.; Nava, L.; Pintore, F.; Stamerra, A.; et al. ASTRI Mini-Array core science at the Observatorio del Teide. *J. High Energy Astrophys.* **2022**, *35*, 1–42. [CrossRef]
10. Anzalone, A.; Bertaina, M.E.; Briz, S.; Cassardo, C.; Cremonini, R.; de Castro, A.J.; Ferrarese, S.; Isgrò, F.; López, F.; Tabone, I. Methods to Retrieve the Cloud-Top Height in the Frame of the JEM-EUSO Mission. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 304–318. [CrossRef]
11. Anzalone, A.; Bruno, A.; Isgrò, F. Measurements of High Energy Cosmic Rays and Cloud presence: A method to estimate Cloud Coverage in Space and Ground-Based Infrared Images. *Nucl. Part. Phys. Proc.* **2019**, *306–308*, 116–123. [CrossRef]
12. Fu, C.L.; Cheng, H.Y. Predicting solar irradiance with all-sky image features via regression. *Sol. Energy* **2013**, *97*, 537–550. [CrossRef]
13. Park, S.; Kim, Y.; Ferrier, N.J.; Collis, S.M.; Sankaran, R.; Beckman, P.H. Prediction of Solar Irradiance and Photovoltaic Solar Energy Product Based on Cloud Coverage Estimation Using Machine Learning Methods. *Atmosphere* **2021**, *12*, 395. [CrossRef]
14. Mustaza, M.S.; Latip, M.F.A.; Zaini, N.; Asmat, A.; Norhazman, H. Cloud Cover Profile using Cloud Detection Algorithms towards Energy Forecasting in Photovoltaic (PV) Systems. In Proceedings of the 2019 IEEE 7th Conference on Systems, Process and Control (ICSPC), Melaka, Malaysia, 13–14 December 2019; pp. 102–107. [CrossRef]
15. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2008**, *122*, 2–10. [CrossRef]
16. Rossow, W.B.; Garder, L.C. Cloud detection using satellite measurements of infrared and visible radiances for ISCCP. *J. Clim.* **1993**, *6*, 2341–2369. [CrossRef]

17. Saunders, R.W.; Kriebel, K.T. An improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. J. Remote Sens.* **1988**, *9*, 123–150. [CrossRef]

18. Ackerman, S.A.; Strabala, K.I.; Menzel, W.P.; Frey, R.A.; Moeller, C.C.; Gumley, L.E. Discriminating clear sky from clouds with MODIS. *J. Geophys. Res.* **1998**, *103*, 32141–32157. [CrossRef]

19. Ackerman, S.A.; Frey, R.A.; Strabala, K.; Liu, Y.; Gumley, L.E.; Baum, B.; Menzel, P.. Discriminating Clear-Sky from Cloud with MODIS Algorithm Theoretical Basis Document (MOD35 v.6.1). Available online: https://atmosphere-imager.gsfc.nasa.gov/sites/default/files/ModAtmo/MOD35_ATBD_Collection6_1.pdf (accessed on 1 March 2024).

20. Sakaida, F.; Hosoda, K.; Moriyama, M.; Murakami, H.; Mukaida, A.; Kawamura, H. Sea surface temperature observation by Global Imager (GLI)/ADEOS-II: Algorithm and accuracy of the product. *J. Oceanogr.* **2006**, *62*, 311–319. [CrossRef]

21. Inoue, T. On the Temperature and Effective Emissivity Determination of Semi-Transparent Cirrus Clouds by Bi-Spectral Measurements in the 10 μm Window Region. *J. Meteorol. Soc. Jpn.* **1985**, *63*, 88–99. [CrossRef]

22. Heidinger, A.K.; Pavolonis, M.J. Gazing at Cirrus Clouds for 25 Years through a Split Window. Part I: Methodology. *J. Appl. Meteorol. Climatol.* **2009**, *48*, 1100. [CrossRef]

23. Jiao, Y.; Zhang, M.; Wang, L.; Qin, W. A New Cloud and Haze Mask Algorithm from Radiative Transfer Simulations Coupled with Machine Learning. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [CrossRef]

24. Qiu, S.; Zhu, Z.; Woodcock, C.E. Cirrus clouds that adversely affect Landsat 8 images: What are they and how to detect them? *Remote Sens. Environ.* **2020**, *246*, 111884. [CrossRef]

25. Istomina, L.; Marks, H.; Huntemann, M.; Heygster, G.; Spreen, G. Improved cloud detection over sea ice and snow during Arctic summer using MERIS data. *Atmos. Meas. Tech.* **2020**, *13*, 6459–6472. [CrossRef]

26. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7. 8, and Sentinel 2 images. *Remote Sens. Environ.* **2015**, *159*, 269–277. [CrossRef]

27. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Hughes, M.J.; Laue, B. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [CrossRef]

28. Stillinger, T.; Roberts, D.A.; Collar, N.M.; Dozier, J. Cloud Masking for Landsat 8 and MODIS Terra Over Snow-Covered Terrain: Error Analysis and Spectral Similarity between Snow and Cloud. *Water Resour. Res.* **2019**, *55*, 6169–6184. [CrossRef] [PubMed]

29. Melchiorre, A.; Boschetti, L.; Roy, D.P. Global evaluation of the suitability of MODIS-Terra detected cloud cover as a proxy for Landsat 7 cloud conditions. *Remote Sens.* **2020**, *12*, 202. [CrossRef]

30. Landsat Satellites. Available online: https://landsat.gsfc.nasa.gov/satellites/ (accessed on 1 March 2024).

31. Sentinel Satellites. Available online: https://sentinels.copernicus.eu/web/sentinel/ (accessed on 1 March 2024).

32. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [CrossRef]

33. Frantz, D.; Haß, E.; Uhl, A.; Stoffels, J.; Hill, J. Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. *Remote Sens. Environ.* **2018**, *215*, 471–481. [CrossRef]

34. Qiu, S.; Zhu, Z.; He, B. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sens. Environ.* **2019**, *231*, 111205. [CrossRef]

35. Main-Knorn, M.; Pflug, B.; Louis, J.M.B.; Debaecker, V.; Müller-Wilm, U.; Gascon, F. Sen2Cor for Sentinel-2. *Remote Sens.* **2017**, *10*, 3. Available online: https://api.semanticscholar.org/CorpusID:64430252 (accessed on 1 March 2024).

36. Hagolle, O.; Huc, M.; Pascual, D.V.; Dedieu, G. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* **2010**, *114*, 1747–1755. [CrossRef]

37. MAJA-Github. 2023. Available online: https://github.com/CNES/MAJA (accessed on 1 March 2024).

38. Baetens, L.; Desjardins, C.; Hagolle, O. Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure. *Remote Sens.* **2019**, *11*, 433. [CrossRef]

39. Li, L.; Li, X.; Jiang, L.; Su, X.; Chen, F. A review on Deep Learning techniques for cloud detection methodologies and challenges. *SIViP* **2021**, *15*, 1527–1535. [CrossRef]

40. Kristollari, V.; Karathanassi, V. Artificial neural networks for cloud masking of Sentinel-2 ocean images with noise and sunglint. *Int. J. Remote Sens.* **2020**, *41*, 4102–4135. [CrossRef]

41. Poulsen, C.; Egede, U.; Robbins, D.; Sandeford, B.; Tazi, K.; Zhu, T. Evaluation and comparison of a Machine Learning cloud identification algorithm for the SLSTR in polar regions. *Remote Sens. Environ.* **2020**, *248*, 111999. [CrossRef]

42. Liu, C.; Yang, S.; Di, D.; Yang, Y.; Zhou, C.; Hu, X.; Sohn, B.J. A Machine Learning-based Cloud Detection Algorithm for the Himawari-8 Spectral Image. *Adv. Atmos. Sci.* **2022**, *39*, 1994–2007. [CrossRef]

43. Joshi, P.P.; Wynne, R.H.; Thomas, V.A. Cloud detection algorithm using SVM with SWIR2 and tasseled cap applied to Landsat 8. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101898. [CrossRef]

44. Lee, K.-Y.; Lin, C.-H. Cloud Detection of Optical Satellite Images using Support Vector Machine. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B7*, 289–293. [CrossRef]

45. Bai, T.; Li, D.; Sun, K.; Chen, Y.; Li, W. Cloud Detection for High-Resolution Satellite Imagery Using Machine Learning and Multi-Feature Fusion. *Remote. Sens.* **2016**, *8*, 715. [CrossRef]

46. Ishida, H.; Oishi, Y.; Morita, K.; Moriwaki, K.; Nakajima, T.Y. Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions. *Remote Sens. Environ.* **2018**, *205*, 390–407. [CrossRef]

47. Ghasemian, N.; Akhoondzadeh, M. Introducing two Random Forest based methods for cloud detection in remote sensing images. *Adv. Space Res.* **2018**, *62*, 288–303. [CrossRef]

48. Thampi, B.V.; Wong, T.; Lukashin, C.; Loeb, N.G. Determination of CERES TOA fluxes using Machine Learning algorithms. Part I: Classification and retrieval of CERES cloudy and clear scenes. *J. Atmos. Oceanic Technol.* **2017**, *34*, 2329–2345. [CrossRef]

49. Sentinel Hub Cloud Detector for Sentinel-2 Images in Python. Available online: https://github.com/sentinel-hub/sentinel2-cloud-detector (accessed on 1 June 2023).

50. Li, P.; Dong, L.; Xiao, H.; Xu, M. A cloud image detection method based on SVM vector machine. *Neurocomputing* **2015**, *169*, 34–42. [CrossRef]

51. Gomez-Chova, L.; Camps-Valls, G.; Bruzzone, L.; Calpe-Maravilla, J. Mean Map Kernel Methods for Semisupervised Cloud Classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 207–220. [CrossRef]

52. Shiffman, S.; Nemani, R. Evaluation of decision trees for cloud detection from AVHRR data. In Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium, IGARSS '05, Seoul, Korea, 29 July 2005; Volume 8, pp. 5610–5613. [CrossRef]

53. De Colstoun, E.C.B.; Story, M.H.; Thompson, C.; Commisso, K.; Smith, T.G.; Irons, J.R. National Park vegetation mapping using multi-temporal Landsat 7 data and a decision tree classifier. *Remote Sens. Environ.* **2003**, *85*, 316–327. [CrossRef]

54. Hollstein, A.; Segl, K.; Guanter, L.; Brell, M.; Enesco, M. Ready-to-Use Methods for the Detection of Clouds, Cirrus, Snow, Shadow, Water and Clear Sky Pixels in Sentinel-2 MSI Images. *Remote Sens.* **2016**, *8*, 666. [CrossRef]

55. Hughes, M.J.; Hayes, D.J. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. *Remote Sens.* **2014**, *6*, 4907–4926. [CrossRef]

56. Cilli, R.; Monaco, A.; Amoroso, N.; Tateo, A.; Tangaro, S.; Bellotti, R. Machine Learning for Cloud Detection of Globally Distributed Sentinel-2 Images. *Remote Sens.* **2020**, *12*, 2355. [CrossRef]

57. Singh, R.; Biswas, M.; Pal, M. Cloud detection using sentinel 2 imageries: A comparison of XGBoost, RF, SVM, and CNN algorithms. *Geocarto Int.* **2022**, *38*, 1–32. [CrossRef]

58. Gómez-Chova, L.; Mateo-García, G.; Muñoz-Marí, J.; Camps-Valls, G. Cloud detection Machine Learning algorithms for PROBA-V. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 2251–2254. [CrossRef]

59. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

60. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [CrossRef]

61. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]

62. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]

63. Le Goff, M.; Tourneret, J.; Wendt, H.; Ortner, M.; Spigai, M. Deep learning for cloud detection. In Proceedings of the ICPRS 8th International Conference of Pattern Recognition Systems, Madrid, Spain, 11–13 July 2017; pp. 1–6. [CrossRef]

64. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 3523–3542. [CrossRef]

65. Soylu, B.E.; Guzel, M.S.; Bostanci, G.E.; Ekinci, F.; Asuroglu, T.; Acici, K. Deep-Learning-Based Approaches for Semantic Segmentation of Natural Scene Images: A Review. *Electronics* **2023**, *12*, 2730. [CrossRef]

66. Song, C.; Huang, Y.; Ouyang, W.; Wang, L. Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3136–3145.

67. Sun, K.; Shi, H.; Zhang, Z.; Huang, Y. Ecs-net: Improving weakly supervised semantic segmentation by using connections between class activation maps. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 20–25 June 2021; pp. 7283–7292.

68. Ma, T.; Wang, Q.; Zhang, H.; Zuo, W. Delving deeper into pixel prior for box-supervised semantic segmentation. *IEEE Trans. Image Process.* **2022**, *31*, 1406–1417. [CrossRef] [PubMed]

69. Hung, W.-C.; Tsai, Y.-H.; Liou, Y.-T.; Lin, Y.-Y.; Yang, M.-H. Adversarial learning for semi-supervised semantic segmentation. *arXiv* **2018**, arXiv:1802.07934.

70. Xue, Y.; Xu, T.; Zhang, H.; Long, L.R.; Huang, X. Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation. *Neuroinformatics* **2018**, *16*, 383–392. [CrossRef] [PubMed]

71. Toldo, M.; Maracani, A.; Michieli, U.; Zanuttigh, P. Unsupervised Domain Adaptation in Semantic Segmentation: A Review. *arXiv* **2020**, arXiv:2005.10876. [CrossRef]

72. Huo, X.; Xie, L.; Zhou, W.; Li, H.; Tian, Q. Focus on Your Target: A Dual Teacher-Student Framework for Domain-adaptive Semantic Segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 18981–18992.

73. Hoyer, L.; Dai, D.; Gool, L.V. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9924–9935.

74. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351. [CrossRef]

75. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]

76. Weng, X.; Yan, Y.; Chen, S.; Xue, J.-H.; Wang, H. Stage-aware feature alignment network for real-time semantic segmentation of street scenes. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 4444–4459. [CrossRef]

77. Tang, Q.; Liu, F.; Zhang, T.; Jiang, J.; Zhang, Y.; Zhu, B.; Tang, X. Compensating for Local Ambiguity With Encoder-Decoder in Urban Scene Segmentation. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 19224–19235.

78. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2018; p. 11045.

79. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Xiao, B. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3349–3364. [CrossRef]

80. Chen, L.C.; Papreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]

81. Lin, T.-Y.; Dollàr, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

82. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.

83. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.

84. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, 5998–6008. [CrossRef]

85. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.

86. Strudel, R.; Garcia, R.; Laptev, I.; Schmid, C. Segmenter: Transformer for Semantic Segmentation, In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 7242–7252. [CrossRef]

87. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv* **2021**, arXiv:2103.14030.

88. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 21–24 June 2021; pp. 6881–6890.

89. Everingham, M.; Winn, J. The PASCAL visual object classes challenge 2012 (VOC2012) development kit. *Pattern Anal. Stat. Model. Comput. Learn. Tech. Rep.* **2012**, *2007*, 1–45.

90. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3213–3223.

91. Mateo-García, G.; Gómez-Chova, L.; Camps-Valls, G. Convolutional neural networks for multispectral image cloud masking. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 2255–2258. [CrossRef]

92. Xie, F.; Shi, M.; Shi, Z.; Yin, J.; Zhao, D. Multi-level Cloud Detection in Remote Sensing Images Based on Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3631–3640. [CrossRef]

93. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]

94. Zi, Y.; Xie, F.; Jiang, Z. A Cloud Detection Method for Landsat 8 Images Based on PCANet. *Remote Sens.* **2018**, *10*, 877. [CrossRef]

95. López-Puigdollers, D.; Mateo-García, G.; Gómez-Chova, L. Benchmarking Deep Learning Models for Cloud Detection in Landsat 8 and Sentinel-2 Images. *Remote Sens.* **2021**, *13*, 992. [CrossRef]

96. Drönner, J.; Korfhage, N.; Egli, S.; Mühling, M.; Thies, B.; Bendix, J.; Freisleben, B.; Seeger, B. Fast Cloud Segmentation Using Convolutional Neural Networks. *Remote Sens.* **2018**, *10*, 1782. [CrossRef]

97. Zhan, Y.; Wang, J.; Shi, J.; Cheng, G.; Yao, L.; Sun, W. Distinguishing Cloud and Snow in Satellite Images via Deep Convolutional Network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1785–1789. [CrossRef]

98. Wieland, M.; Li, Y.; Martinis, S. Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. *Remote Sens. Environ.* **2019**, *230*, 111203. [CrossRef]

99. Jeppesen, J.H.; Jacobsen, R.H.; Inceoglu, F.; Toftegaard, T.S. A cloud detection algorithm for satellite imagery based on Deep Learning. *Remote Sens. Environ.* **2019**, *229*, 247–259. [CrossRef]

100. Mohajerani, S.; Saeedi, P. Cloud-Net: An end-to-end Cloud Detection Algorithm for Landsat 8 Imagery. *arXiv* **2019**, arXiv:1901.10077.
101. Francis, A.; Sidiropoulos, P.; Muller, J.-P. CloudFCN: Accurate and Robust Cloud Detection for Satellite Imagery with Deep Learning. *Remote Sens.* **2019**, *11*, 2312. [CrossRef]
102. L8 Biome Cloud Validation Masks. U.S. Geological Survey. 2016. Available online: https://landsat.usgs.gov/landsat-8-cloud-cover-assessment-validation-data (accessed on 1 March 2024).
103. Domnich, M.; Sünter, I.; Trofimov, H.; Wold, O.; Harun, F.; Kostiukhin, A.; Järveoja, M.; Veske, M.; Tamm, T.; Voormansik, K.; et al. KappaMask: AI-Based Cloudmask Processor for Sentinel-2. *Remote Sens.* **2021**, *13*, 4100. [CrossRef]
104. Francis, A.; Mrziglod, J.; Sidiropoulos, P.; Muller, J.-P. Sentinel-2 Cloud Mask Catalogue. 2020. Available online: https://zenodo.org/records/4172871 (accessed on 1 March 2024).
105. Hu, K.; Zhang, D.; Xia, M. CDUNet: Cloud Detection UNet for Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 4533. [CrossRef]
106. Caraballo-Vega, J.A.; Carroll, M.L.; Neigh, C. Optimizing WorldView-2,-3 cloud masking using Machine Learning approaches. *Remote Sens. Environ.* **2023**, *284*, 113332. [CrossRef]
107. Chai, D.; Newsam, S.; Zhang, H.K.; Qiu, Y.; Huang, J. Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks. *Remote Sens. Environ.* **2019**, *225*, 307–316. [CrossRef]
108. Li, Z.; Shen, H.; Cheng, Q.; Liu, Y.; You, S.; He, Z. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 197–212. [CrossRef]
109. Shao, Z.; Pan, Y.; Diao, C.; Cai, J. Cloud Detection in Remote Sensing Images Based on Multiscale Features-Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4062–4076. [CrossRef]
110. Li, J.; Wu, Z.; Hu, Z.; Jian, C.; Luo, S.; Mou, L.; Zhu, X.X.; Molinier, M. A lightweight deep learning-based cloud detection method for Sentinel-2A imagery fusing multiscale spectral and spatial features. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5401219. [CrossRef]
111. Chen, G.; Li, C.; Wei, W.; Jing, W.; Woźniak, M.; Blažauskas, T.; Damaševičius, R. Fully Convolutional Neural Network with Augmented Atrous Spatial Pyramid Pool and Fully Connected Fusion Path for High Resolution Remote Sensing Image Segmentation. *Appl. Sci.* **2019**, *9*, 1816. [CrossRef]
112. Wang, Z.; Zhao, J.; Zhang, R.; Li, Z.; Lin, Q.; Wang, X. UATNet: U-Shape Attention-Based Transformer Net for Meteorological Satellite Cloud Recognition. *Remote Sens.* **2022**, *14*, 104. [CrossRef]
113. Mateo-García, G.; Laparra, V.; López-Puigdollers, D.; Gómez-Chova, L. Transferring Deep Learning models for cloud detection between Landsat 8 and Proba-V. *ISPRS J. Photogramm. Remote Sens.* **2020**, *150*, 1–17. [CrossRef]
114. Mateo-García, G.; Laparra, V.; López-Puigdollers, D.; Gómez-Chova, L. Cross-Sensor Adversarial Domain Adaptation of Landsat-8 and Proba-V Images for Cloud Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 747–761. [CrossRef]
115. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
116. Pang, S.; Sun, L.; Tian, Y.; Ma, Y.; Wei, J. Convolutional Neural Network-Driven Improvements in Global Cloud Detection for Landsat 8 and Transfer Learning on Sentinel-2 Imagery. *Remote Sens.* **2023**, *15*, 1706. [CrossRef]
117. Skakun, S.; Wevers, J.; Brockmann, C.; Doxani, G.; Aleksandrov, M.; Batič, M.; Žust, L. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sens. Environ.* **2022**, *274*, 112990. [CrossRef]
118. Mateo-García, G.; Laparra, V.; Gómez-Chova, L. Domain Adaptation of Landsat-8 and Proba-V Data Using Generative Adversarial Networks for Cloud Detection. In Proceedings of the IGARSS 2019—IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 712–715. [CrossRef]
119. Francis, A.; Mrziglod, J.; Sidiropoulos, P.; Muller, J.P. SEnSeI: A Deep Learning Module for Creating Sensor Independent Cloud Masks. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5406121. [CrossRef]
120. Mohajerani, S.; Saeedi, P. Cloud-Net+: A cloud segmentation CNN for landsat 8 remote sensing imagery optimized with 393 filtered jaccard loss function. *arXiv* **2020**, arXiv:2001.08768.
121. Baetens, L.; Hagolle, O. Sentinel-2 Reference Cloud Masks Generated by an Active Learning Method. 2018. Available online: https://zenodo.org/records/1460961 (accessed on 1 March 2024).
122. Aybar, C.; Ysuhuaylas, L.; Loja, J.; Gonzales, K.; Herrera, F.; Bautista, L.; Gómez-Chova, L. CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2. *Sci. Data* **2022**, *9*, 782. [CrossRef] [PubMed]
123. Wu, Z.; Li, J.; Wang, Y.; Hu, Z.; Molinier, M. Self-attentive generative adversarial network for cloud detection in high resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1792–1796. [CrossRef]
124. Li, Z.; Shen, H.; Li, H.; Xia, G.; Gamba, P.; Zhang, L. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. *Remote Sens. Environ.* **2017**, *191*, 342–358. [CrossRef]
125. Wu, X.; Shi, Z.; Zou, Z. A geographic information-driven method and a new large scale dataset for remote sensing cloud/snow detection. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 87–104. [CrossRef]
126. Mrziglod, J.; Francis, A. Intelligently Reinforced Image Segmentation graphical user interface (IRIS) 2019.
127. Czerkawski, M.; Atkinson, R.; Michie, C.; Tachtatzis, C. SatelliteCloudGenerator: Controllable Cloud and Shadow Synthesis for Multi-Spectral Optical Satellite Images. *Remote Sens.* **2023**, *15*, 4138. [CrossRef]
128. Alhassan M.; Fuseini M. Data augmentation: A comprehensive survey of modern approaches. *Array* **2022**, *16*, 100258. [CrossRef]

129. Ghasemi, A.; Rabiee, H.R.; Fadaee, M.; Manzuri, M.T.; Rohban, M.H. Active Learning from Positive and Unlabelled Data. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, Canada, 11 December 2011. [CrossRef]

130. Zeng, Y.; Hao, D.; Park, T.; Zhu, P.; Huete, A.; Myneni, R.; Knyazikhin, Y.; Qi, J.; Nemani, R.; Fa, L.; et al. Structural complexity biases vegetation greenness measures. *Nat. Ecol. Evol.* **2023** *7*, 1790–1798. [CrossRef]

131. Dozier, J.; Bair, E.; Baskaran, L.; Brodrick, P.; Carmon, N.; Kokaly, R.; Miller, C.; Miner, K.R.; Painter, T.; Thompson, D. Error and Uncertainty Degrade Topographic Corrections of Remotely Sensed Data. *J. Geophys. Res. Biogeosci.* **2022**, *127*, e2022JG007147. [CrossRef]

132. Lazzaro, D.; Cinà, A.E.; Pintor, M.; Demontis, A.; Biggio, B.; Roli, F.; Pelillo, M. Minimizing Energy Consumption of Deep Learning Models by Energy-Aware Training. *arXiv* **2023**, arXiv:2307.00368.

133. Desislavov, R.; Martínez-Plumed, F.; Hernández-Orallo, J. Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. *Sustain. Comput. Inform. Syst.* **2023**, *38*, 100857. [CrossRef]