remote sensing

MDPI

Systematic Review

Deep Learning-Based Cloud Detection for Optical Remote Sensing Images: A Survey

Zhengxin Wang [1,2], Longlong Zhao [1,3], Jintao Meng [1], Yu Han [1,3], Xiaoli Li [1,3], Ruixia Jiang [1,4], Jinsong Chen [1,3] and Hongzhong Li [1,3,*]

[1] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; zx.wang5@siat.ac.cn (Z.W.); ll.zhao@siat.ac.cn (L.Z.); jt.meng@siat.ac.cn (J.M.); yu.han@siat.ac.cn (Y.H.); xl.li2@siat.ac.cn (X.L.); rx.jiang1@siat.ac.cn (R.J.); js.chen@siat.ac.cn (J.C.)
[2] University of Chinese Academy of Sciences, Beijing 100094, China
[3] Shenzhen Engineering Laboratory of Ocean Environmental Big Data Analysis and Application, Shenzhen 518055, China
[4] School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China
* Correspondence: hz.li@siat.ac.cn

Abstract: In optical remote sensing images, the presence of clouds affects the completeness of the ground observation and further affects the accuracy and efficiency of remote sensing applications. Especially in quantitative analysis, the impact of cloud cover on the reliability of analysis results cannot be ignored. Therefore, high-precision cloud detection is an important step in the preprocessing of optical remote sensing images. In the past decade, with the continuous progress of artificial intelligence, algorithms based on deep learning have become one of the main methods for cloud detection. The rapid development of deep learning technology, especially the introduction of self-attention Transformer models, has greatly improved the accuracy of cloud detection tasks while achieving efficient processing of large-scale remote sensing images. This review provides a comprehensive overview of cloud detection algorithms based on deep learning from the perspective of semantic segmentation, and elaborates on the research progress, advantages, and limitations of different categories in this field. In addition, this paper introduces the publicly available datasets and accuracy evaluation indicators for cloud detection, compares the accuracy of mainstream deep learning models in cloud detection, and briefly summarizes the subsequent processing steps of cloud shadow detection and removal. Finally, this paper analyzes the current challenges faced by existing deep learning-based cloud detection algorithms and the future development direction of the field.

Keywords: cloud detection; deep learning; semantic segmentation; optical satellite imagery; remote sensing

Citation: Wang, Z.; Zhao, L.; Meng, J.; Han, Y.; Li, X.; Jiang, R.; Chen, J.; Li, H. Deep Learning-Based Cloud Detection for Optical Remote Sensing Images: A Survey. *Remote Sens.* **2024**, *16*, 4583. https://doi.org/10.3390/rs16234583

Academic Editors: Seongwook Lee and Byung-Kwan Kim

Received: 28 October 2024
Revised: 4 December 2024
Accepted: 4 December 2024
Published: 6 December 2024

1. Introduction

With the continuous advancement of remote sensing technology, an increasing number of countries are processing in Earth observation projects. According to the American Satellite Industry Association, the global satellite count in orbit has witnessed a 43% increase since 2016, and as of January 2023, there exist 1193 remote sensing satellites worldwide [1,2], thereby providing substantial data for Earth surface observation. The timeliness and accuracy of remote sensing applications such as land resource exploration, natural disaster warning, environmental monitoring, and military reconnaissance have also been enhanced within the framework of advanced big data technology in remote sensing. However, based on findings from the International Satellite Cloud Climatology Project (ISCCP), average Earth cloud cover surpasses 66% [3–7], impeding optical satellites from acquiring valuable information about the Earth's surface due to numerous clouds present in optical remote sensing images. Consequently, this affects the image utilization

rate. Therefore, cloud detection and removal are important steps in optical remote sensing image pre-processing.

Cloud detection of optical remote sensing images is one of the research hot spots in the field of remote sensing image recognition [8]. Since its introduction in 1982, cloud detection has been considered as one of the fundamental components in ISCCP [9]. In recent decades, with the continuous development of satellite remote sensing image processing technology, many cloud detection methods have been proposed to reduce the burden of labor costs. Traditional cloud detection methods are pixel-level recognition tasks, aiming to label each pixel in the scene as cloud or non-cloud. In general, the most commonly used traditional methods can be mainly divided into three categories: threshold-based algorithms [10–13], multi-temporal algorithms [14–16], and machine learning-based algorithms [17,18].

Threshold-based algorithms utilize threshold functions of one or more spectral bands to determine the classification of each pixel. These functions primarily utilize physical attributes such as brightness, color, and elevation to determine the optimal thresholds. These algorithms have relatively strong performance and low computational cost, so they have been widely used in image quality assessment and cloud detection of various satellite missions, including ACCA [11] and Fmask [19] for the Landsat series satellites, Sen2Cor [20] for Sentinel-2, CloudMask [21] for MODIS, and MFC [12] for GaoFen-1. Although these algorithms are among the earliest and most commonly used methods, they rely on empirical judgment to set thresholds, hindering adaptive adjustment for dynamic or complex scenarios and affecting the accuracy of cloud detection. In addition, for high-resolution optical remote sensing images with limited spectral information or physical attributes, the validity of the threshold functions is reduced. Although several studies [22,23] have proposed cloud detection methods with limited physical features, these algorithms lack generality and still require a significant amount of prior knowledge when applied to different sensors [24,25]. The multi-temporal algorithms use images acquired at different times to construct time series cloud-free reference data as an approximation of true surface reflectance. Cloud detection is achieved by analyzing the spatial features within the time series of reflectance, including differences, trends, or abrupt changes [14,26]. The advantage of these algorithms is their sensitivity to changes in surface reflectance, making them particularly effective for detecting thick clouds. However, the robustness of these algorithms depends on a large number of high-quality images. In the case of insufficient or low-quality data, the models may fail to achieve precise cloud detection results. In addition, generating cloud-free reference data requires manual labeling of pixel validity, thereby increasing workload [15]. Machine learning-based algorithms treat cloud detection as an image classification problem, where the spectral features of individual pixels or the textural features of locally segmented regions are used as inputs for classical machine learning algorithms, such as Support Vector Machines (SVM) [27], Random Forest (RF) [28,29], Bayes classifier [30], and fuzzy clustering [31]. Incorporating additional information such as elevation and land cover type to these algorithms enhances the diversity of features and thus improves the accuracy of cloud detection [32,33].

The three traditional algorithms can obtain high-precision cloud detection results with sufficient training data and have been widely studied and applied over the years, but there are still some limitations. First, the existing cloud detection methods rely on pixel-level predictions, which tend to produce poor detection results in high-resolution imagery due to the presence of "salt-and-pepper" noise [34]. Secondly, due to factors such as land cover type and climate conditions [7], the effectiveness of traditional methods in distinguishing cloud boundaries, especially for thin clouds, is limited. Third, the ability of these algorithms to extract spatial features is limited. This limitation hinders their effectiveness in dealing with complex scenarios [35]. In addition, the traditional methods can not meet the strict demands of large-scale data production [36].

In recent years, deep learning has achieved rapid development. Thanks to deep learning's powerful feature extraction capabilities, the ability to predict masks using the most representative features in the data allows deep learning to achieve state-of-the-art

performance in many benchmarks. Researchers are turning to deep learning techniques for cloud detection. Broadly speaking, deep learning-based cloud detection algorithms belong to the category of semantic segmentation tasks [37]. The finely annotated cloud detection datasets are input into the deep feature extractors. The model parameters are fine-tuned during the training process to optimize the parameter combination, and finally achieve high-precision cloud detection results. In 2017, Xie et al. [38] introduced Convolutional Neural Network (CNN) into cloud detection for the first time. The algorithm used simple linear iterative clustering to segment images into superpixels, then used two-branched CNN to extract multi-scale features, and finally predicted all superpixels to obtain cloud segmentation results. With the continuous progress of deep learning technology, more and more deep learning models are used for cloud detection. Depending on the different network structures, these models can be broadly divided into categories. (1) Fully Convolutional Neural Network (FCN): FCN extracts image features through a series of convolutional and pooling layers, and then restores the feature maps to the same resolution as the input image through deconvolution layers (or upsampling layers). This allows FCN to classify each pixel for precise cloud detection [39]. As a milestone in deep learning-based semantic segmentation, FCN and its variants have been widely used for remote sensing image cloud detection [40–42]. (2) Encoder–decoder network: The encoder–decoder network extracts multi-level features of the image through the encoder and gradually reconstructs the image through the decoder, and finally generates an output image with the same size as the input image to obtain pixel-wise classification. This network structure has become an important tool for remote sensing cloud detection due to its superior performance in image segmentation tasks [43]. (3) Attention-based network: The attention mechanism assigns different weights to different parts of the input data, enabling the model to focus on more important regions or features, thereby improving overall performance. In cloud detection, the attention mechanism helps the model better identify and distinguish clouds from the background [44–46]. (4) Generative adversarial network (GAN): GAN consists of two parts—the generator and the discriminator. Through adversarial training, the generator learns to produce more realistic data. By generating different types and shapes of cloud images, GAN can enrich the training samples of the model and improve the accuracy of cloud detection [47–49]. In general, the deep learning-based cloud detection algorithms are entirely data-driven and do not require manually setting thresholds or handcrafted features, reducing the need for prior knowledge. Additionally, these algorithms do not rely on multi-temporal data; combining multi-temporal data does not significantly improve the accuracy of the model, and too many convergent data will lead to overfitting of the model. Compared with traditional algorithms, deep learning-based methods offer superior detection efficiency. These methods can quickly process data in batches without processing each scene image separately, which can meet the requirements of large-scale data production.

With the continuous development of deep learning, a large number of excellent network structures and modules have been proposed to improve the accuracy of cloud detection across various scenarios. Cloud detection methods have become diverse and complex, especially with the incorporation of Transformer; remote sensing image cloud detection methods have stepped into a new era, and it is very necessary to classify and summarize these deep learning-based methods. Several reviews have been conducted on cloud detection algorithms. For instance, Li et al. [50] provide a comprehensive overview of cloud and cloud shadow detection, covering features, algorithms, and validation. They also compile open-source tools and datasets and offer insights into the future of cloud and cloud shadow detection. Similarly, Sawant et al. [51] focus on the types of clouds and imaging systems, reviewing cloud detection algorithms from the perspectives of satellite and ground imagery. However, these reviews do not provide a systematic overview of deep learning-based cloud detection algorithms, nor do they explore the future directions and potential developments of deep learning in this field. Therefore, a thorough classification and analysis of deep learning-based cloud detection methods are essential to guide future research and applications in remote sensing.

The structure of this survey is as follows: Section 2 reviews the development of deep learning-based cloud detection using the Web Of Science (WOS) literature analysis tool. Section 3 introduces the public datasets and evaluation metrics for remote sensing image cloud detection, which facilitates a quantitative comparison of the accuracy differences of deep learning methods on different datasets. Section 4 details the research progress and current status of different categories of deep learning cloud detection algorithms, and comparatively analyzes the advantages and shortcomings of deep learning algorithms compared with classical algorithms and machine learning-based algorithms. Section 5 describes the subsequent processing of remote sensing images after cloud detection, such as cloud shadow detection and cloud removal. Section 6 examines the current challenges faced by deep learning-based cloud detection algorithms and anticipates future development directions. Finally, Section 7 provides concluding remarks for the survey.

2. Literature Analysis

Based on the WOS database, this paper investigates the papers on the topic of deep learning-based cloud detection. The keywords used to refine the survey were "Deep Learning", "Cloud Detection", "Cloud Mask", "Cloud and Cloud Shadow Detection", and "Remote Sensing". In order to improve the accuracy of search results, we excluded "point clouds" from the database because point cloud detection involves Lidar data and can interfere with search results. As of 1 December 2024, a total of 931 relevant papers were collected. After manual screening to exclude irrelevant papers and conference papers, 169 officially published journal articles were included in the follow-up analysis. Figure 1 shows the number of published articles on deep learning-based cloud detection from 2016 to 2024, with the 2024 data counted up to December 1st. The data reveal a gradual increase in the number of published articles, with an increase of about 10 papers per year from 2017 to 2021 and a notable increase of 23 papers in 2022 over 2021. This trend highlights the growing interest of researchers in deep learning cloud detection. In 2023, the growth trend slowed, with the number of related papers falling from 52 in 2022 to 34 in 2023. All 200 published papers were statistically analyzed, and the results were visualized according to journals, institutions, and data sources. Figure 2 lists the major publishing journals by number of published articles (>7). Eight of them have published more than five articles each, and all of them are international journals in the field of remote sensing and geosciences. These eight journals published 152 articles, accounting for 76% of the total number of articles. *Remote Sensing*, *IEEE Transactions on Geoscience and Remote Sensing*, and *IEEE Geoscience and Remote Sensing Letters* are the three journals with the highest number of articles published, with *Remote Sensing* publishing more than 50 articles in the field.

Figure 3 shows the major types of optical remote sensing data utilized in deep learning-based cloud detection studies. The top eight types of remote sensing data, each referenced in more than six papers, are represented by bar charts, while the data cited in fewer than six papers but more than one paper are represented by pie charts. The most commonly used data types are Landsat-8, Gaofen-1, and Sentinel-2, from the United States, China, and Europe, respectively. Due to the open access capabilities of the three datasets, they are used significantly more frequently than other data sources.

Compared to the previous review on cloud detection [50], in which the literature survey mainly focused on traditional cloud detection methods, there are two obvious differences. First, the proportion of papers using low-resolution data has significantly decreased. In the previous review [50], MODIS ranked first in usage; however, in the field of deep learning-based cloud detection, it has only been cited in eight papers, ranking seventh in the current analysis. In addition, AVHRR, which ranked third in the previous review [50], no longer exists in the new literature statistics, as shown in Figure 3. This indicates that as the focus of research shifts, the attention to low-resolution data in the field of cloud detection has gradually declined. In contrast, the application of high-resolution and multi-source data has increased, reflecting the stronger adaptability and higher research value of deep learning methods for high-resolution data.

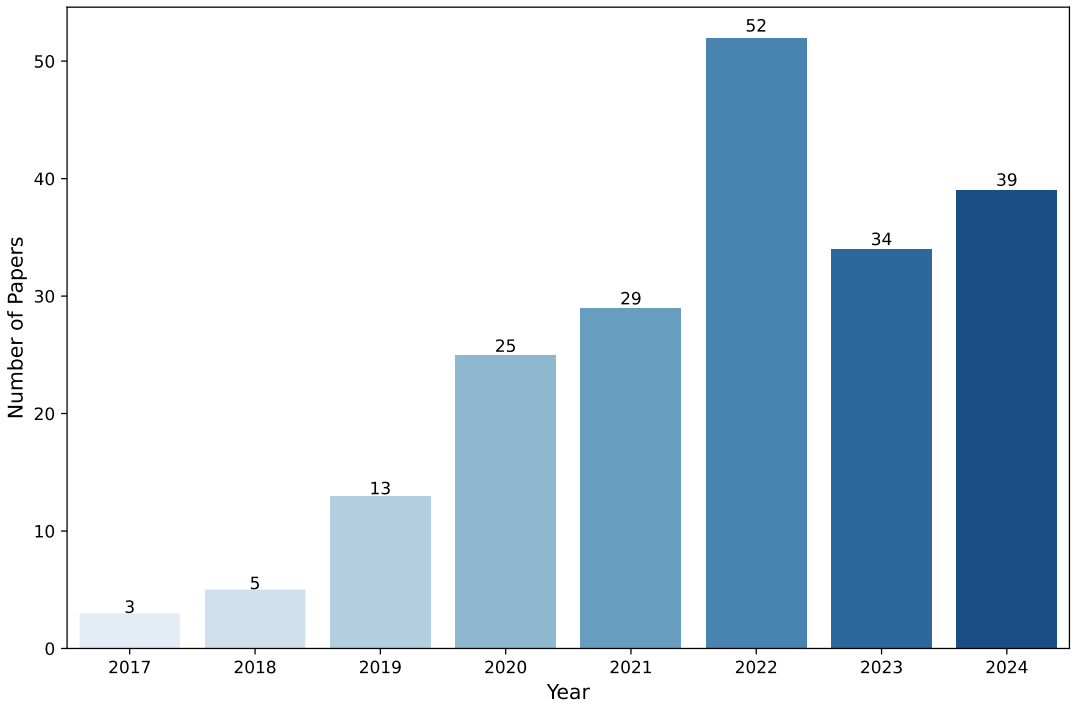


Figure 1. Number of published papers on deep learning-based cloud detection during 2016~2024.

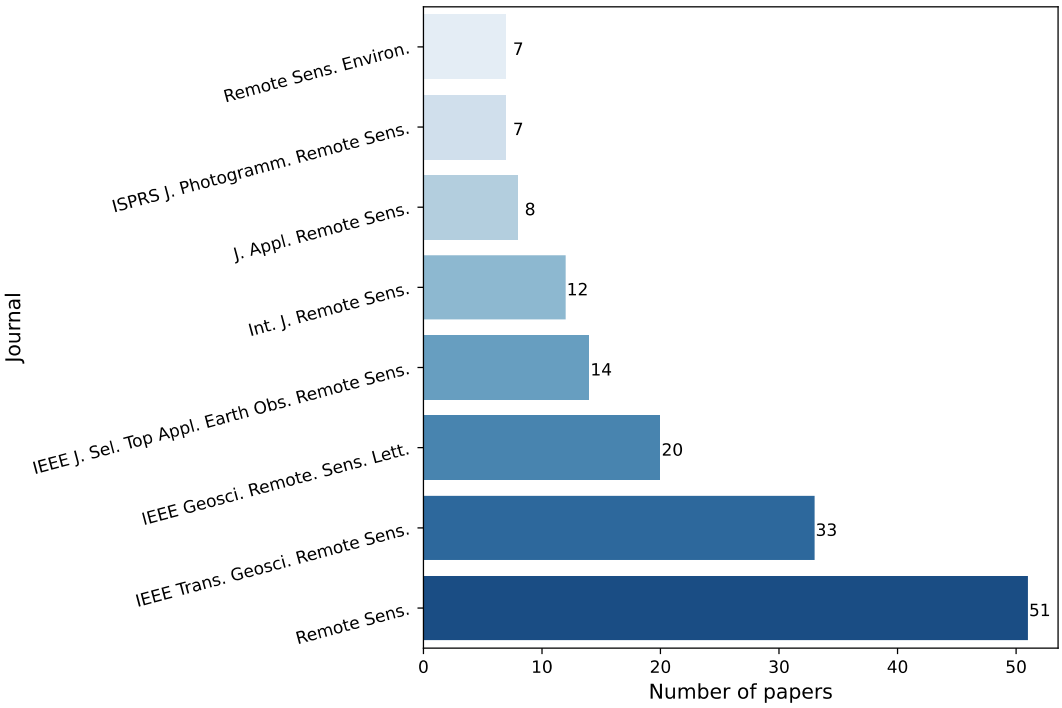


Figure 2. Major journals (published > 7 papers) in the field.

Second, in Figure 3, the types of high-resolution meter-level data have increased significantly. Among them, Google earth, Gaofen-2, and Ziyuan-3 have been utilized in 22, 10, and 10 papers respectively. In addition, newly launched remote sensing satellite data, such as SPOT-6 [52–54], Fengyun-4A/B [55,56], Gaofen-5 [36,45], and Perusat-1 [57], are also being used to verify the feasibility of deep learning in cloud detection. This difference is mainly due to the increasing availability of high-resolution remote sensing data and the

great potential of deep learning to achieve high accuracy and efficiency in cloud detection for high-resolution images.

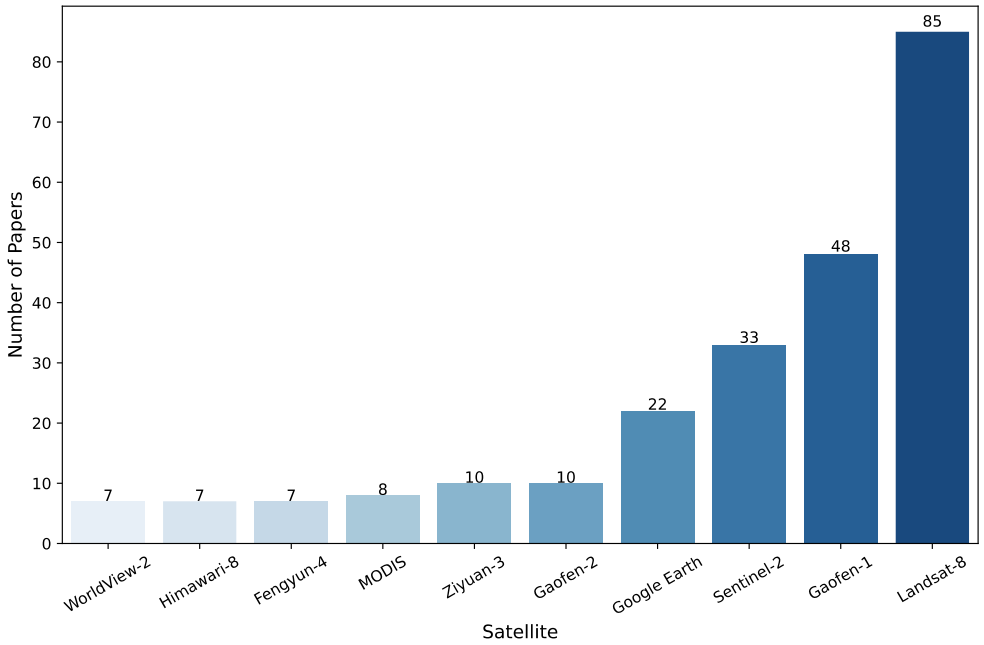


Figure 3. Major types of satellite images used.

The major institutions in the field are listed in Figure 4, highlighting those that published more than four papers. The top eight institutions are all from China. Among them, Nanjing University of Information Science and Technology, Wuhan University, and the Chinese Academy of Sciences lead the field, with 22, 19, and 15 papers, respectively. Following these are Beihang University, Harbin Institute Technology, Xidian University, China University of Petroleum (East China), and Anhui University, each contributing more than five papers to this research field. National Aeronautics and Space Administration (NASA) from USA, University of Valencia from Spain, and Simon Fraser University from Canada, are the top three institutions outside China, each contributing two papers.

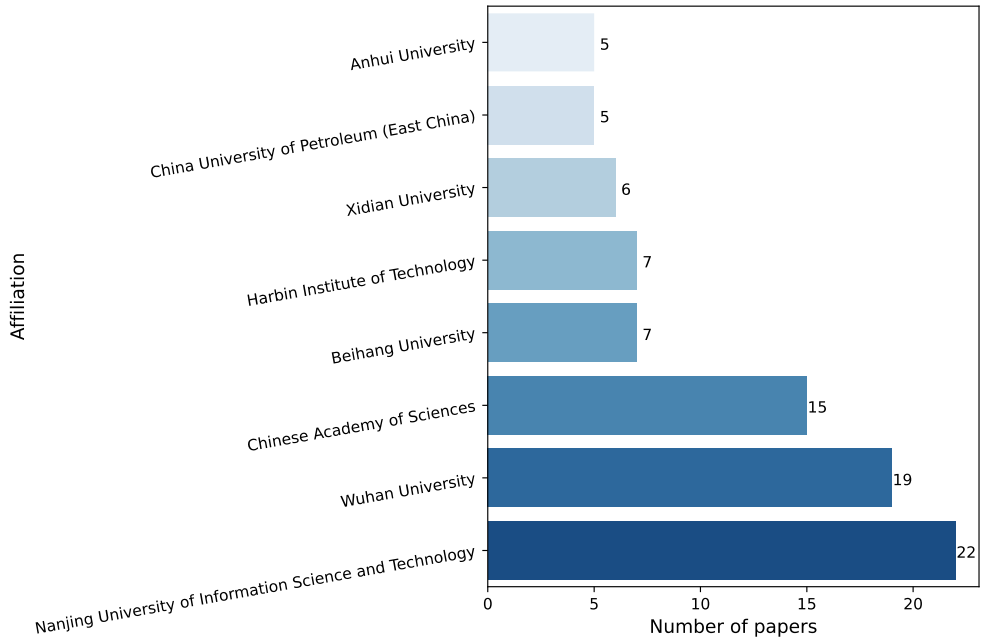


Figure 4. Major institutions (published > 5 papers) in the field.

3. Cloud Detection Datasets and Evaluation Metrics

3.1. Open-Source Datasets for Cloud Detection

The accuracy of deep learning models depends heavily on the quality of the training data. To ensure its robustness and ability to generalize, the training dataset should contain a wide range of scenarios so that different patterns and variations can be captured. In addition, the size of the dataset should be proportional to the number of model parameters. This relationship is crucial to prevent problems such as schema collapse (where models fail to learn representative distributions) and overfitting (where models become too specialized and perform poorly on new data). Proper cloud detection datasets are key to achieving optimal deep learning model performance.

With the growing interest and demand for cloud detection technologies, more and more open-source datasets have become available, greatly advancing research in this field. These datasets are very diverse and include data from different meteorological conditions and geographical regions. This not only improves the effectiveness of researchers' algorithm training, but also facilitates the validation and application of different models in different scenarios. A summary of the current open-source datasets for cloud detection is provided in Table 1. Table 1 is a continuation and supplement of the datasets listed in [50]. Compared with previous statistics [50], the open-source datasets collected in this article are more comprehensive, with a total of 26 datasets collected. Satellite data sources include Landsat-7/8, Sentinel-2, GF-1/2, Ziyuan-3, PlanetScope, SPOT-6, and Google Earth. It should be noted that in CloudSEN12 [58], the Synthetic Aperture Radar (SAR) Sentinel-1 Interferometric Wide Swath (IW) data with two polarization channels (VV and VH) were also included as special "spectral" band. Table 1 lists the attributes of various open-source datasets, including data sources, resolution, data bands, etc. Some datasets retain all the bands of the original data, such as L7 Irish [10], L8 Biome [59], S2-Hollstein [60], and CloudSEN12 [58], while others only preserve the visible and near-infrared bands, such as 38-Cloud [61], 95-Cloud [61], ADTCS [52], and T-S2 [62].

These datasets were released in two formats: scene and patch. Among them, the datasets that provide whole scene labels have more flexibility and can dynamically adjust the image size of the input model according to the number of model parameters and the performance of the processor. However, datasets in the form of patches offer greater advantages in terms of efficiency, flexibility, and fine-grained analysis, especially for large-scale image processing and deep learning. The column of "Number of image scenes" shows the number of image scenes used for dataset creation. If the dataset is released in the format of patch by cropping the whole image scene, Table 1 shows the size and number of patches; otherwise, a hyphen is used. The column of data scope indicates the coverage area of the selected image scenes used in the dataset creation. The last three columns of the table list the dataset providers, countries, and release dates. Furthermore, some datasets not only include cloud labels but also provide annotations for other types of ground cover, offering a more comprehensive set of labeled data for various land surface features. For example, S2-Hollstein [60] supports the detection of clouds, cloud shadows, snow, and water, and the labels supported by the dataset are listed in the table "Other Support".

Table 1. Summary of the current open-source datasets for cloud detection.

Dataset Name	Data Source	Resolution	Number of Image Scenes	Patch Size	Number of Patches	Bands	Data Scope	Institution	Country	Year
L7 Irish [10]	Landsat-7	30	207	-	-	Landsat-7 bands	Global	U.S. Geological Survey	USA	2012
L8 SPARCS [63]	Landsat-8	30	24	1000 × 1000	80	Landsat-8 bands	Global	Oregon State University	USA	2014
L8 Biome [59]	Landsat-8	30	96	-	-	Landsat-8 bands	Global	U.S. Geological Survey	USA	2016
38-Cloud [61]	Landsat-8	30	38	384 × 384	17,601	R,G,B,NIR	USA	Simon Fraser University	Canada	2017
95-Cloud [61]	Landsat-8	30	95	384 × 384	43,902	R,G,B,NIR	-	Simon Fraser University	Canada	2018
Snow-Cloud Validation Masks [64]	Landsat-8	30	13	-	-	Landsat-8 bands	Global	University of California	USA	2019
RICE1 [65]	Google Earth	-	-	512 × 512	500	R,G,B	-	Chinese Academy of Sciences	China	2019
RICE2 [65]	Landsat-8 Look Images	-	-	512 × 512	450	SWIR1,NIR,Red	-	Chinese Academy of Sciences	China	2019
WHU Cloud Dataset [66]	Landsat-8	30	7	512 × 512	859	R,G,B	China and North Korea	Wuhan University	China	2021
S2-Hollstein [60]	Sentinel-2	20	5647725 pixels.in a HDF5file	-	-	Sentinel-2 bands	Global	Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences	Germany	2016
S2-BaetensHagoll [67]	Sentinel-2	10	38	-	-	Sentinel-2 bands	Global	CESBIO, Université de Toulouse	France	2019
CloudSEN12 [58]	Sentinel-2	10	-	512 × 512	49,400	Sentinel-2 bands	Global	University of Valencia	Spain	2022
	Sentinel 1					S1 VV VH				
T-S2 [62]	Sentinel-2	3.125	-	128 × 128	4993	R,G,B,NIR	Wet Tropics	CSIRO, Agriculture and Food	Australia	2019
T-PS [62]	PlanetScope	3.125	-	128 × 128	4943	R,G,B,NIR	Wet Tropics	CSIRO, Agriculture and Food	Australia	2019
Sentinel-2 Cloud Mask Catalogue [68]	Sentinel-2	20	-	1022 × 1022	513	Sentinel-2 bands	Global	University College London	UK	2020
Sentinel-2 KappaZeta [69]	Sentinel-2	10	155	512 × 512	4403	Sentinel-2 bands	Northern Europe	University of Tartu	Estonia	2021
WHUS2-CD [70]	Sentinel-2	10	32	-	-	R,G,B,NIR	China	Wuhan University	China	2021
GF1_WHU [12]	GF1 WFV	16	108	-	-	R,G,B,NIR	Global	Wuhan University	China	2017
Levir_CS [71]	GF1 WFV	160	4168	1320 × 1200	-	R,G,B,NIR	Global	Beihang University	China	2021
GF1MS-WHU [72]	GF1 PMS	8	33	-	-	R,G,B,NIR,Pan	China	Wuhan University	China	2024
GF2MS-WHU [72]	GF2 PMS	4	29	-	-	R,G,B,NIR,Pan	China	Wuhan University	China	2024
WDCD-GF1 [73]	GF1 PMS	8	622	250 × 250	206,384	R,G,B,NIR	Global	Wuhan University	China	2020
WDCD-ZY3 [73]	ZiYuan-3	5.8	6	-	-	R,G,B,NIR	Global	Wuhan University	China	2020
AIR-CD [4]	GF2	4	34	-	-	R,G,B,NIR	China	Chinese Academy of Sciences	China	2021
HRC_WHU [74]	Google Earth	0.5–15	-	1280 × 720	150	R,G,B	Global	Wuhan University	China	2019
ADTCS [52]	SPOT-6 SPOT-7	1.5m	9	512 × 512	32,065	R,G,B	Global	National University of Defense Technology	China	2024

3.2. Cloud Detection Evaluation Metrics

Effective evaluation metrics are essential to accurately evaluate the performance of cloud detection models. At present, mainstream cloud detection evaluation usually regard cloud detection as a pixel-by-pixel classification task, relying on standard classification metrics such as True Positive (*TP*), True Negative (*TN*), False Positive (*FP*), and False Negative (*FN*), as well as derived metrics such as Precision (*P*) and Recall (*R*). Three widely used metrics to evaluate the accuracy of cloud detection are Overall Accuracy (*OA*) [4,70], Intersection-over-Union (*IoU*) [66,70], and $F1-score$ [75,76]. These metrics are calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$F1 - score = 2\frac{P \times R}{p + R} \tag{3}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{4}$$

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

Additionally, some studies have used the mean Intersection-over-Union (*mIoU*) [40,77] and mean Accuracy (*mA*) [78] to provide a more comprehensive assessment of a model's classification capabilities. By combining these metrics, a clearer understanding of both individual pixel classification and overall detection performance is achieved, ensuring a more robust evaluation of cloud detection algorithms across different datasets. The metrics *mIoU* and *mA* are calculated as follows:

$$mA = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{FP + TN}\right) \tag{6}$$

$$mIoU = \frac{1}{2}\left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FP + FN}\right) \tag{7}$$

4. Deep Learning-Based Models

Deep learning-based cloud detection in optical remote sensing falls within the domain of semantic segmentation in computer vision. Therefore, this paper provides an overview and analysis of cloud detection algorithms based on deep learning from the perspective of semantic segmentation.

Semantic segmentation is a computer vision task that classifies each pixel of an image into a specific category, thus segmenting the image into distinct regions. It is widely used in applications such as object detection, medical imaging, and remote sensing, enabling more detailed analysis by associating each pixel with a label. Before deep learning was used, semantic segmentation mainly relied on classical algorithms [79] and machine learning techniques [28,80–84]. These methods were based on features such as thresholds, colors, and textures, or handcrafted features based on expert knowledge. In 2012, Krizhevsky et al. [85] made a groundbreaking contribution to the introduction of the deep learning model AlexNet in the ImageNet challenge, marking the beginning of a new era in semantic segmentation and computer vision. Although their work focused primarily on image classification, their contributions laid the groundwork for semantic segmentation. In 2015, Long et al. [39] introduced a Fully Convolutional Network (FCN) for semantic segmentation, achieving end-to-end pixel-level classification, which is widely regarded as a milestone in the field. Minaee et al. [86] provided a comprehensive review of deep learning-based semantic and instance segmentation, categorizing deep learning models into 10 categories and discussing the pioneering work in each category. Garcia-Garcia et al. [87] also reviewed

semantic segmentation methods in different fields, and presented key datasets, common loss functions, and standard evaluation metrics. Semantic segmentation has been widely used in various fields of remote sensing imagery. Yuan et al. [88] comprehensively reviewed the latest progress of deep learning methods for semantic segmentation of remote sensing images. Cloud detection uses semantic segmentation technology to distinguish and label cloud and non-cloud pixels in images, which is a typical application of semantic segmentation in the remote sensing field. In this Section, we summarize the advancements in deep learning-based cloud detection according to the types of semantic segmentation networks, including FCN networks, encoder–decoder networks, neural networks based on attention mechanisms, and generative adversarial networks.

4.1. FCN-Based Models

The advancements in deep convolutional neutral networks have significantly changed the landscape of semantic segmentation and pushed their performance to unprecedented levels. In 2015, Long et al. [39] introduced FCN specifically for semantic segmentation tasks. The model adopted an end-to-end training framework, and the results showed that FCN was superior to traditional methods relying on handcrafted features. On the basis of FCN, Liu et al. proposed ParseNet [89], which enhances the model's receptive field by integrating global contextual information. ParseNet effectively utilizes the average features of each layer of the neural network to enhance the representation of individual pixels, aggregating feature maps of different scales into context vectors, and then upsampling the context vectors to their original dimensions. Chen et al. [90] further extended this work and improved FCN's global perception capabilities by introducing Conditional Random Fields (CRF) [91]. They developed a series of DeepLab models based on dilated convolutions, including DeepLabV1 [92], DeepLabV2 [90], DeepLabV3 [93], and DeepLabV3+ [94]. The use of dilated convolutions allows these models to maintain spatial resolution in feature maps while eliminating the need for pooling layers. The empirical results show that the proposed method significantly improves the performance of FCN. Zhao et al. [95] developed PSPNet and further improved the context extraction capability of FCN by introducing pyramid pool module. The model performs well in both object and context analysis in complex scenarios, achieving superior accuracy in semantic segmentation tasks in a variety of applications.

The iterations of FCNs have significantly enhanced cloud detection capability in remote sensing by automatically extracting spatial features of clouds and accurately distinguishing them. Li et al. [78] proposed a remote sensing cloud detection model AM-CDN, which integrated attention mechanism and residual connection to prevent degradation, and replaced standard convolution with expanded convolution to realize adaptive sensitivity field. This model combined the lightness of FCN with the high precision of PSPNet, and solved the difficulty of depth feature extraction of FCN and the computational complexity of PSPNet. On the Landsat-8 infrared RSI dataset, its accuracy was 10.17% higher than that of FCN, while the computational complexity was only 7.63% that of PSPNet. Shao et al. [96] proposed the MF-CNN model, which used the fully connected characteristics of FCN to combine high-level semantics with low-level spatial features. This model simultaneously realized the detection of thin-cloud, thick-cloud, and non-cloud pixels in remote sensing images.

Inspired by the pyramid pool architecture of PSPNet, Li et al. [97] proposed Pyramid Context Network (PCNet) for multi-scale feature extraction. PCNet used Dilated Residual Blocks with dilated convolution and residual connections to expand the receptive field, thus overcoming the limitations of traditional convolution cores. In addition, pyramid context blocks were introduced to enhance the model's ability to perceive features at different scales and improve the detection capability of thin clouds. Experimental results on Fengyun-3D MERSI-II images showed that the overall accuracy is 97.1% and the recall rate is 93.2%. Zhang et al. [42] proposed a cloud detection model that integrated FCN with a voting ensemble strategy, consisting of an encoder, feature fusion module, and voting

ensemble strategy. The encoder used an FCN backbone to generate 8 × 8 × 8 feature maps, while the feature fusion module preserved spatial and contextual semantic information. The voting ensemble module enhanced the model's prediction accuracy and robustness. Experiments on the FY-2 cloud detection dataset showed that the accuracy is improved by 7.15% and 8.46%, respectively, compared with the local FCN model and DeepLabV3 model. Zhan et al. [34] proposed an FCN model specifically for cloud and snow detection. They added a multi-scale learning module into the network, enabling it to utilize both low-level spatial information and high-level semantic information simultaneously. Compared to the original FCN model, the model with the multi-scale module achieved a 6.1% improvement in *mIoU* for cloud and snow detection tasks and demonstrated higher inference efficiency than CRF-based methods.

Advanced FCN-based models further extend cloud detection capabilities through innovative designs. Wu et al. [98] designed a Scene Information Aggregation module (SIA) to integrate different scene information from remote sensing imagery into the Scene Aggregation Network (SAN) for learning. This model generated cloud mask results and scene classification results at the same time. By effective use of scene information, the *mIoU* of the SAN model is 10.29% higher than that of CloudFCN [41]. Chen et al. [99] proposed an Automatic Cloud Detection Neural Network (ACDnet) that integrated remote sensing images with geospatial data. The feature extraction network consisted of a spectral spatial information extraction module and a geographic information extraction module, aiming to extract cloud spectral spatial and geographic semantic information from remote sensing images and large geospatial datasets, so as to improve the cloud detection accuracy under the coexistence of cloud and snow. Wu and Shi [100] designed a feature cascade cloud detection model based on FCN networks and specific filters, and achieved 85.38% *mIoU* on the GF-1 test set. The cloud detection models based on FCN networks are shown in Table 2, which also lists the datasets used, model evaluation metrics, and accuracy for each model. The comparison results of cloud and cloud shadow detection based on FCN structure are shown in Figure 5, which shows the visualization results of different FCN-based cloud detection networks (CDN) and cloud removal networks (CRN) in Landsat-8 datasets [66].

Table 2. Quantitative comparison of FCN-based cloud detection models.

Methods	Datasets	Evaluation Metrics	Values
Origin FCN [39]	L8 Biome [59]	*mIoU*	69.19%
DeepLabV3+ [94]			75.87%
PSPNet [95]			79.94%
AM-CDN [78]			79.43%
CloudFCN [41]			77.89%
SAN [34]			88.18%
MF-CNN [96]	Landsat-8	*F1-score*	89.20%
SVM [27]			83.03%
RF [28]			84.06%
Fused FCN [42]	FY-2G	*OA*	60.2%
Origin FCN [39]			43.1%
SegNet [96]			36.5%
DeepLabV3 [93]			47.1%
Origin FCN [39]	GF1_WHU [12]	*mIoU*	84.50%
Zhan's FCN [34]			90.60%
ACDnet [99]			91.04%

The progress of FCN improves the efficiency and reliability of remote sensing cloud detection. Through its end-to-end structure, FCN expands the model's receptive field and enhances its ability to express multi-scale features, allowing it to better capture the details of complex clouds. However, FCN and its extended models still have some limitations. First, a large number of fully connected layers greatly increases the model parameters

and complexity, which reduces the inference speed. This computational burden makes FCN less efficient in processing high-resolution remote sensing images, and this limitation is even more pronounced in real-time applications. Second, although FCN expands the receptive field and enhances its ability to capture large-scale features, it has also weakened FCN's ability to detect thin and fragmented cloud structures, making it less effective at distinguishing these complex cloud types. Third, FCN uses deconvolution to restore image size, which can retain global information, but at the same time, this process will cause cloud boundary blurring and affect the accuracy of cloud detection results. Due to the above limitations, research on remote sensing cloud detection based on FCN models has decreased since 2022. However, despite these challenges, some researchers are actively improving FCN models to meet specific needs and applications. Ni et al. [101] utilized deep learning to introduce cloud detection in infrared hyperspectral remote sensing based on three end-to-end models: Deep Neural Network (DNN), Convolutional Neural Network (CNN), and CNN integrated with Long Short-Term Memory (LSTM). These models showed good generalization ability and high detection accuracy in various cloud types and scenarios. Therefore, although the overall trend indicates that the number of studies is decreasing, the FCN model is still an important basis for remote sensing cloud detection research and deserves further exploration.

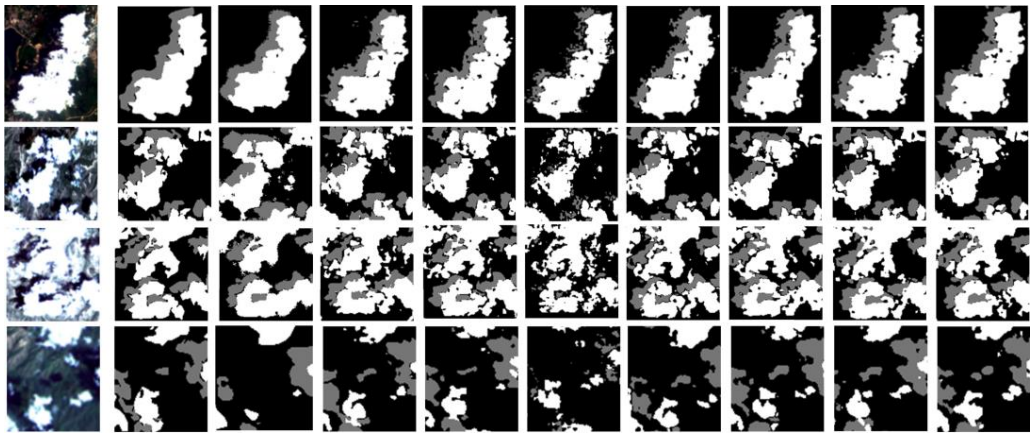


Figure 5. Comparison of different FCN structures in cloud and shadow detection [66]. (**a**) Cloudy image. (**b**) Label. (**c**) U-Net. (**d**) CRN(VGG). (**e**) CRN(ResNet). (**f**) CRN(ResNet-BN). (**g**) CRN(DenseNet). (**h**) CRN(VGG) with CAM. (**i**) CRN(ResNet) with CAM. (**j**) CRN(DenseNet) with CAM.

4.2. Encoder–Decoder-Model-Based Models

The encoder–decoder model consists of two parts: encoder and decoder. The encoder compresses the input data into a latent representation to extract data features, while the decoder uses that representation to generate the output. This architecture was initially applied to time series data processing, such as natural language processing and machine translation, and later extended to the field of computer vision, especially semantic segmentation tasks.

Noh et al. [102] were the first to introduce the encoder–decoder architecture in the domain of semantic segmentation, leveraging the powerful feature extraction capabilities of the VGG-16 model. The model achieved 72.5% accuracy on the PASCAL VOC 2012 dataset, the best performance at the time without relying on external data. Subsequently, Badrinarayanan et al. [103] proposed SegNet, another segmentation model based on encoder–decoder architecture, whose innovation lied in its upsampling method within the decoder. Nonlinear upsampling was performed using the pooled index generated at the encoder stage, thus avoiding the noise or loss of detail associated with transposed convolution. This approach reduced the need for explicit upsampling during inference and greatly reduced the number of trainable parameters, making SegNet suitable for resource-constrained applications. In addition, they introduced a Bayesian version of SegNet to

model uncertainty within the architecture, further improving its performance in scene segmentation. Ronneberger et al. [43] introduced U-Net, which directly transmitted the high-resolution feature map extracted by the encoder to the corresponding layer of the decoder by skipping the connection to prevent information loss. The model also used data enhancement technology to improve the learning efficiency in the case of limited labeled data. Although skip connections in U-Net facilitate the rapid transfer of high-resolution feature maps, they can also lead to misalignment, where feature maps with significant semantic differences may become confused during connection. To solve this problem, Zhou et al. [104] proposed the U-Net++ architecture, which used nested residual connections to reduce the semantic gap between encoder and decoder feature maps, employing dense convolution blocks to improve feature consistency, thereby capturing fine-grained details in two-dimensional images. Based on the advantage of U-Net architecture, other researchers have proposed extended versions [105–107].

Chai et al. [37] proposed a SegNet model with a 13-layer encoder and a 13-layer decoder structure for cloud and cloud shadow segmentation. This model introduced residual connections to accelerate model inference and added feature connections between encoder and decoder at the same level, preserving feature details without reducing resolution. The experimental results for cloud and cloud shadow segmentation on Landsat-7 and Landsat-8 datasets achieved accuracy rates of 95.26% and 95.47%, respectively. Hu et al. [108] drew inspiration from U-Net and introduced CDUNet for cloud detection. CDUNet incorporated a High-Frequency Feature Extractor and Multi-Scale Convolution to refine cloud boundaries and predict fragmented clouds. Additionally, to enhance thin-cloud detection accuracy, it introduced a Spatial Prior Self-Attention mechanism for spatial position information. CDUNet incorporated a Parallel-Channel Spatial Attention Block in the feature fusion stage to capture effective information rapidly, improving prediction accuracy. CDUNet achieved an average IoU of 93.52% on the L8 SPARCS dataset. Francis et al. [41] combined Inception modules and the U-Net model to propose CloudFCN. In the encoder part, CloudFCN used a Fully Convolutional Neural Network and employed a class-weighted scheme to encourage local information sharing between adjacent pixels, resulting in smooth output edges. CloudFCN showed robustness in multi-class segmentation on Landsat-8 images. Wieland et al. [109] introduced a cloud and cloud shadow segmentation model suitable for multi-target domains. This model improved upon U-Net to adapt to multi-source data. Experimental results demonstrated that this model outperformed Fmask and Random Forest-based algorithms in terms of inference speed and accuracy. Liu et al. [110] proposed DCNet to extract both contextual information and cloud internal geometric features simultaneously. The model incorporated deformable convolution blocks on top of an encoder–decoder structure, enabling adaptive spatial context capture based on cloud morphology. In the implementation, it borrowed strengths from other models like ResNet's residual connections and U-Net's guidance of low-level features for high-level feature recovery. Experiments on GF-1 images showed an accuracy of 96.94% and an average IoU of 89.19%, surpassing advanced methods like SegNet [103] and DeepLabV3 [93]. Pang et al. [111] and Pešek et al. [112] compared the accuracy of U-Net, FCN, SegNet, and DeepLab models across different satellite datasets. Across various surface scenarios, U-Net demonstrated superior accuracy in capturing cloud shapes and edges, especially in detecting thin and scattered cloud scenes, outperforming other deep learning models. Zhang et al. [113] integrated U-Net with an attention mechanism, using an asymmetric encoder–decoder structure. Specifically, the encoder incorporated the ActiveOrNot activation function and residual units, while the decoder mirrored U-Net, with skip connections incorporating attention gates to reduce interference from different land cover types during cloud detection. The model achieved 87.11% overall accuracy on the HRC WHU dataset, a 3.52% improvement over the original U-Net. Luo et al. [114] designed a lightweight encoder–decoder network for cloud detection, employing a dual-path architecture in the encoder to extract both spatial and semantic information. The detail branch used minimal parameters to capture low-level spatial features, while the

semantic branch introduced a Dense Pyramid Module for multi-scale context extraction, significantly reducing parameters and computations through feature reuse. Experiments on Landsat-8 and MODIS datasets demonstrated that ECDNet achieved comparable accuracy to state-of-the-art cloud detection methods with reduced computational costs. Jeppesen et al. [75] proposed RS-Net, which combines spatial and spectral features. In challenging scenarios like snow and ice, RS-Net outperformed the Fmask algorithm on the Landsat-8 dataset with higher accuracy. Li et al. [70] introduced the lightweight CDFM model, based on an encoder–decoder structure, and created the WHUS2-CD cloud detection dataset, covering 32 land cover types and all seasons. Shi et al. [115] integrated dilated convolutions and fully connected Conditional Random Field layers into U-Net, presenting CloudU-Net. Results on the SWINySEG dataset showed that CloudU-Net effectively identified clouds during both day and night. Li et al. [116] embedded Global Context Dense Blocks into U-Net to enhance thin-cloud detection. These blocks improve the importance of thin-cloud features and consist of non-local self-attention units and squeeze-excitation units. Li et al. also created a cloud detection dataset based on MODIS, where GCDB-UNet achieved 89.5% *mIoU*. Table 3 presents cloud detection models based on encoder–decoder architectures, along with the datasets used and their accuracy. The comparison results of cloud and cloud shadow detection based on U-shape structure are shown in Figure 6, which shows the results of U-Net, PSPNet, SegNet, HRNet, and AFMUNet models in the GF1_WHU dataset [117].

Table 3. Quantitative comparison of encoder–decoder-structure-based cloud detection models.

Methods	Datasets	Evaluation Metrics	Values
Chai's SegNet [37]	L7 Irish [118]	*OA*	94.33%
Wieland's U-Net [109]	L8 Biome [119]		94.00%
CAA-UNet [113]	L8 SPARCS [120]		89.00%
	HRC WHU [74]		87.11%
CDUNet [108]	GF1_WHU [12]	*mIoU*	93.33%
AFMUNet [117]			93.28%
DCNet [110]			89.19%
CloudU-Net [115]			89.09%
CDNet [40]			88.02%
SegNet [103]			88.28%
DABNet [4]			86.53%
CloudSegNet [121]			85.29%
Origin U-Net [43]			81.38%
ECDnet [114]	95-Cloud [61]	*mIoU*	91.6%
CDFM [70]			91.0%
RS-Net [75]			89.5%
GCDB-UNet [116]			91.9%
Origin U-Net [43]			88.9%
DCNet [110]			89.7%
GCDB-UNet [116]	L8 SPARCS [120]	*mIoU*	89.1%
Origin U-Net [43]			81.0%
RS-Net [75]			86.8%
DCNet [110]			82.6%
CDUNet [108]			93.52%
GCDB-UNet [116]	MODIS	*mIoU*	89.5%
Origin U-Net [43]			86.8%
RS-Net [75]			86.6%
DCNet [110]			87.0%

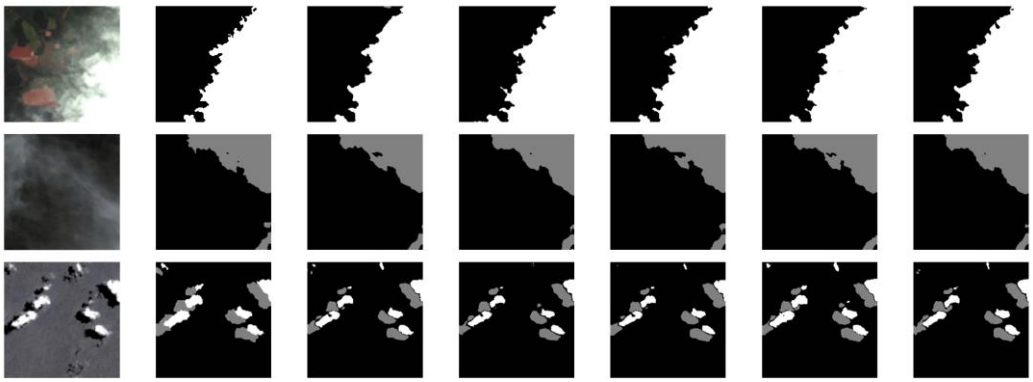


Figure 6. Comparison of visualization results for different models [117]: (**a**) the original image; (**b**) the corresponding label; (**c**) the prediction of U-Net; (**d**) the prediction of PSPNet; (**e**) the prediction of SegNet; (**f**) the prediction of HRNet; (**g**) the prediction of the proposed AFMUNet.

The encoder–decoder structure makes comprehensive use of residual connections, transposed convolutions, and dilated convolutions, and has powerful feature extraction and multi-scale information fusion capabilities, greatly improving feature resolution, and can retain cloud details of different scales, showing significant advantages in cloud detection. Compared with FCN, the encoder–decoder model performs better in cloud edge detection. On the one hand, the encoder gradually extracts high-level features from the input image, while the decoder restores the feature map resolution layer by layer, achieving finer segmentation while retaining details. On the other hand, the residual connection prevents information loss by fusing shallow and deep features, which is crucial for accurately delineating cloud boundaries. In terms of performance, the combination of transposed convolutions and dilated convolutions expands the receptive field of the model, ensuring accurate cloud detection while improving inference speed. Moreover, the modular design of the encoder–decoder structure also provides low coupling, allowing different parts to flexibly combine the latest advances in deep learning, such as more complex convolution operations or optimization algorithms, to further improve performance. Therefore, in the cloud detection task, the structure can not only effectively identify and segment clouds under complex weather conditions, but also significantly improve the accuracy of cloud edge detection, which is very suitable for real-time meteorological monitoring applications.

4.3. Attention-Mechanism-Based Models

The attention mechanism [122,123] is a technique inspired by cognitive attention in biology. It involves adaptively weighting certain parts of input data to extract key information. From an operational perspective, attention mechanisms can be categorized into spatial attention and channel attention.

Chen et al. [123] proposed an attention mechanism that adaptively assigns weights to multi-scale features of each pixel, enhancing the model's ability to capture details in semantic segmentation. By integrating this attention mechanism into a Deeplab-based model, they trained it with multi-scale images. The results showed that the attention mechanism enabled the model to effectively evaluate the correlation of features at different locations and scales. Fu et al. [124] introduced the Dual Attention Network (DANet) for scene segmentation. DANet utilized self-attention to capture rich contextual information and extended this by incorporating spatial and channel attention modules. These modules enhanced the model's ability to simulate the semantic interdependence between different image regions. On the Cityscapes dataset [125], DANet achieved an IoU score of 81.5%, demonstrating the critical role of attention mechanisms in improving scene segmentation accuracy.

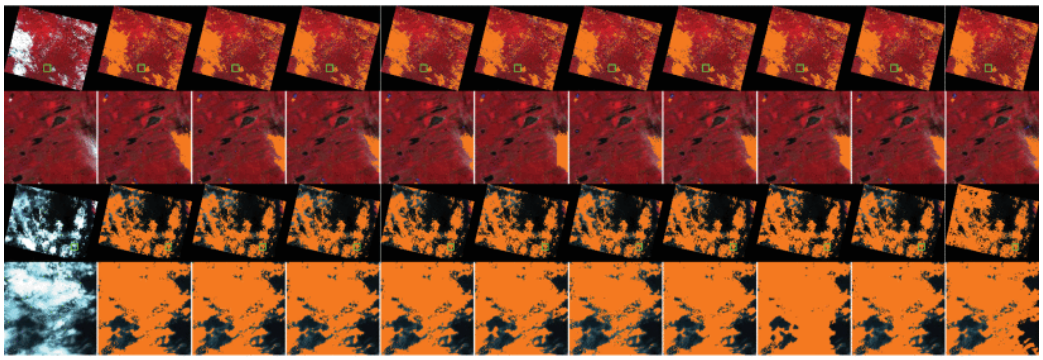
Recently, the self-attention-mechanism-based Transformer was introduced by Vaswani et al. [126] for sequence-to-sequence tasks in natural language processing. By utilizing self-

attention, Transformer can effectively capture long-range dependencies and relationships within data. In 2021, Dosovitskiy et al. [127] adapted the Transformer architecture to computer vision, where it has demonstrated outstanding performance. This success has generated significant research interest, leading to the application of Transformers across various computer vision tasks, where their ability to model complex patterns and contextual information has proven highly effective [128–131].

Zhang et al. [113] proposed the CAA-UNet, a new network structure based on the classical encoder–decoder U-Net, which integrates residual connections with an attention mechanism. This network maximizes cloud feature preservation during feature fusion and introduces adaptive activation, reducing parameters and accelerating inference. The Cloud-AttU model proposed by Guo et al. [132] is based on the U-Net architecture and incorporates an attention mechanism. Experimental results on the Landsat-8 dataset demonstrated significant accuracy improvements, allowing the model to effectively distinguish between cloud and non-cloud regions, even in noisy backgrounds. Liu et al. proposed TransCloudSeg [133], introducing Transformer to ground-based cloud remote sensing for the first time, and the model uses CNN and Transformer to form a dual pathway, each of which contains an encoder and decoder, finally applying a heterogeneous fusion module to fuse the results in the pathway to generate the final segmentation results. Zhang et al. [134] proposed Cloudformer, a cloud detection framework that combines CNN and Transformer, applied in the CNN module with a pyramidal architecture encoder to extract multi-scale features. Cloudformer uses asynchronous location coding to enhance the sensitivity of Transformer to contextual location information. The *mIoU* of Cloudformer reached 96.56%, a 7.6% improvement compared to DeepLabV3+ [94], but the model inference speed was reduced by 34% compared to DABNet [4]. Zhang et al. [135] proposed a lightweight Vision Transformer model (CloudViT) for satellite image cloud detection. CloudViT utilizes a multi-scale dark channel extractor to predict dark channels, which are then combined with image features and fed into an attention-based model. Results on the L7 Irish dataset showed that the dark channel-guided context aggregation module enhanced image features, achieving a cloud detection *mIoU* accuracy of 86.12%. CloudViT also incorporates a channel-adaptive module, with validation on the L8 Biome, WHUS2-CD, and AIR-CD datasets demonstrating significant improvements in cross-sensor cloud detection accuracy. Ge et al. [136] introduced a lightweight CNN–Transformer cloud detection model, referred to as CD-CTFM, which employs a lightweight CNN–Transformer backbone and a feature pyramid module to fuse multi-scale and global context information. Experiments on the 38-Cloud and MODIS datasets revealed that CD-CTFM effectively reduces computational complexity while maintaining accuracy. Yao et al. [137] proposed CD-AttDLV3+ based on DeepLabV3+ and a channel attention module, achieving efficient thin-cloud detection using only the common RGB and near-infrared bands. Gong et al. [138] combined Swin Transformer and CNN structures in their model STCCD, incorporating an attention-based feature fusion module and an aggregated multi-scale feature module to leverage interactions between different feature levels. STCCD achieved high accuracy across datasets, including GF-1 WHU, SPARCS, AIR-CCD, and L8-Biome. Singh et al. [139] designed the SSATR-CD model with a spatial–spectral attention module, improving average accuracy by 6.7% over ViT-based models on the WHUS2-CD dataset. Tan et al. [140] explored the cloud and cloud shadow detection capabilities of Swin-UNet. Validation on the GF1 WHU public dataset showed that Swin-UNet outperformed CNN-based models in both accuracy and robustness. Table 4 provides an overview of cloud detection models utilizing attention mechanisms, including the datasets employed and their respective accuracy results. The comparison results of cloud detection based on Transformer structure are shown in Figure 7, which shows the results of CloudViT, Segformer, LPGT, CloudNet, and other models in GF1 WFV data [141].

Table 4. Quantitative comparison of attention-mechanism-based cloud detection models.

Methods	Datasets	Evaluation Metrics	Values
CloudViT [135]	L7 Irish [118]	*mIoU*	86.12%
CDNet [40]			85.67%
DANet [124]			84.08%
CD-CTFM [136]	38-Cloud [61]	*mIoU*	84.77%
Cloud-AttU [132]			84.73%
CD-AttDLV3+ [137]			81.24%
Cloudformerv1 [134]			90.71%
STCCD [138]	GF1_WHU [12]	*mIoU*	91.96%
Swin-UNet [140]			90.34%
BuildFormer			90.15%
Origin PSPNet [95]			86.54%
LPGT [141]			86.06%
SegFormer [142]			85.72%
CloudViT [135]			85.24%
Origin DeepLabV3+ [94]			83.82%
Origin U-Net [43]			78.38%
STCCD [138]	L8 SPARCS [120]	*mIoU*	91.48%
Swin-UNet [140]			86.21%
BuildFormer			89.44%
STCCD [138]	AIR-CD [4]	*mIoU*	90.47%
Swin-UNet [140]			87.43%
BuildFormer			86.72%
SSATR-CD [139]	WHUS2-CD [70]	*mIoU*	76.17%
ViT [127]			69.47%
CloudViT [135]			58.48%



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**) (**g**) (**h**) (**i**) (**j**) (**k**)

Figure 7. Comparison with different models on GF-1 WFV validation set [141]. (**a**) False-color image. (**b**) FCNet. (**c**) U-Net. (**d**) PSPNet. (**e**) SegNet. (**f**) DeepLabV3+. (**g**) CloudNet. (**h**) SegFormer. (**i**) CloudViT. (**j**) LPGT. (**k**) GT mask.

The above research shows that the introduction of the attention mechanism and Transformer architecture significantly improves the performance of the cloud detection model. Similar to human cognitive attention, the core idea of the attention mechanism is to enable the model to adaptively focus on important features within the input data. In cloud detection, the attention mechanism strengthens the learning of cloud-related features by dynamically weighting different pixels or feature maps. This dynamic weighting not only helps capture the subtle structures of clouds, but also effectively suppresses background noise, thereby improving the model's ability to identify thin and fragmented clouds. Building on this, the Transformer model further leverages their advantages. The self-attention module in Transformer allows the model to consider all other inputs while

processing each individual input, facilitating the capture of long-range dependencies. This feature is especially important for cloud detection because cloud shapes, textures, and distributions can exhibit complex spatial relationships within images. By learning these complex relationships, Transformer can identify various types of clouds. Moreover, the Transformer architecture offers substantial parallel computing capabilities, making models more efficient in both training and inference phases. Compared to traditional CNNs, Transformers are better able to handle high-resolution image data. In cloud detection tasks, this high-resolution processing capability enables models to detect cloud details at fine scales, further improving accuracy and reliability. Literature analysis shows a great increase in Transformer-based cloud inspection papers since 2023, with studies demonstrating Transformer's performance across multiple cloud inspection datasets.

However, there are some limitations to the Transformer models. Transformer models typically contain millions of parameters, which makes it extremely difficult to train effectively on small datasets. For small datasets, the models are prone to overfitting, making the trained model poorly generalizable in real applications. This problem occurs because the model cannot capture the general characteristics of the data, but instead memorizes the specific details of the training set, which is definitely undesirable for actual cloud detection tasks. Moreover, the computational complexity of Transformer is also a concern. The self-attention mechanism needs to calculate the relationship between each pair of elements in the input data, resulting in a computational cost that grows quadratically with the input size. This high computational overhead not only increases training time, but may also limit the deployment of the model on edge devices or real-time applications.

4.4. Generative Adversarial Network-Based Models

Generative Adversarial Network (GAN) is a generative model introduced by Goodfellow et al. [143] in 2014. GANs consist of two primary components: a generator and a discriminator. The generator learns the latent distribution of real data samples to produce new data instances, while the discriminator functions as a binary classifier, determining whether the input data are real or generated. Both the generator and discriminator can be effectively implemented using deep neural networks, allowing for sophisticated modeling of complex data distributions.

Luc et al. [144] first introduced GANs into semantic segmentation by optimizing an objective function that merges traditional multi-class cross-entropy loss with an adversarial term. This innovative approach resulted in an adversarial segmentation network capable of distinguishing between authentic segmentation maps and those generated by the segmentation model, thereby improving the quality of the generated outputs. Souly et al. [145] proposed a semi-supervised semantic segmentation technique that leverages GANs. Their method incorporates a generator network that generates additional training examples for the discriminator within the GAN framework. In this setup, the discriminator functions as a multi-class classifier assigning labels 'y' from K possible classes to samples or marking them as fake samples. Hung et al. [146] developed a semi-supervised semantic segmentation framework using adversarial networks, focusing on improving the model's sensitivity to spatial resolution. They designed an FCN discriminator that differentiates between predicted probability maps and the actual label distribution. Their model's loss function incorporates three terms: cross-entropy loss for segmentation's true labels, adversarial loss for the discriminator network, and semi-supervised loss based on confidence maps. This multi-faceted loss function allows for a more nuanced learning process, contributing to improved segmentation performance.

Li et al. [73] proposed two GAN-based models for cloud detection, CR-GAN-PM and GAN-CDM, with the latter being an enhanced version. Unlike end-to-end methods, GAN-CDM eliminates the need for manually labeled cloud masks, using image blocks and block-level labels instead. Labeling at the block level requires only 0.7% of the time compared to pixel-level labeling. Experiments on Landsat-8 show that GAN-CDM surpasses baseline deep learning models when using the same input bands, and it demonstrates

superior transferability on sensors like Sentinel-2. Zou et al. [147] introduced a weakly supervised cloud detection model using GAN, comprising a cloud generator, discriminator, and dissipation network. The generator and discriminator are adversarially trained to produce realistic cloud images, while the dissipation network predicts cloud reflectance and attenuation. Experiments on the GF-1 dataset demonstrate that the model achieves accuracy comparable to or exceeding fully supervised methods, without requiring manual labels during training. Joachim and Ira [148] introduced FCD, a weakly supervised cloud detection method based on Fixed-Point GAN (FPGAN), which requires only block-level labels for training. Utilizing the strong transformation capabilities of FPGAN, the generated segmentation map primarily influences cloud regions. Experimental results show that FCD surpasses existing CAM-based methods in weakly supervised cloud detection on the Landsat-8 Biome dataset [119]. Yang et al. [149] proposed a weakly supervised adversarial training model for cloud and snow detection, referred to as WCSD. WCSD includes cloud GAN and snow GAN, which can generate remote sensing images of clouds and snow along with their corresponding pixel-level labels. The input to cloud GAN consists of randomly selected cloud and clear remote sensing images, producing results for cloud reflectance and opacity. By overlaying the cloud reflectance with clear remote sensing images according to cloud opacity, the model can generate cloud images and corresponding pixel-level pseudo-labels for cloud detection. Snow GAN performs similar operations, and the combination of the two GANs can produce mixed cloud–snow images and pixel-level labels. Validation results on the Levir_CS dataset demonstrate that WCSD outperforms other weakly supervised methods in cloud–snow detection. Wu et al. [150] combined attention mechanisms with GANs to propose the semi-supervised SAGAN model. The training strategy of SAGAN encompasses three fundamental principles: 1) expanding the attention map of cloud regions and replacing it with translated clear images; 2) reducing the attention map to align with cloud boundaries; and 3) optimizing the self-attention network to handle extreme situations. Experiments on the Sentinel-2 dataset indicate that SAGAN exhibits high label efficiency. Table 5 offers a summary of cloud detection models that utilize GAN, detailing the datasets used and their corresponding accuracy outcomes. The comparison results of cloud and snow detection based on GAN structure are shown in Figure 8, which shows the results of GradCAM, GAN-CDM, CloudMatting, and WCSD models in the Levir_CS dataset [149].

Table 5. Quantitative comparison of GAN-based cloud detection models.

Methods	Datasets	Evaluation Metrics	Values
WCSD [149]	Levir_CS [71]	*OA*	91.38%
CloudMatting [147]			90.32%
GAN-CDM [73]			71.61%
SAGAN [150]	Sentinel-2A	*OA*	99.38%
Origin DeepLabV3+ [92]			99.37%
Origin U-Net [43]			99.12%
CLDiff [151]	GF1_WHU [12]	*OA*	96.68%
DDP [152]			94.66%
GAN-CDM [73]			90.20%
FCD [148]			87.60%
CAM			82.90%
GradCAM			78.60%
WDCD [149]			72.09%
CloudMatting [147]	Gaofen-1	*OA*	95.38%
Origin U-Net [43]			96.13%
Origin DeepLabV3+ [94]			88.04%

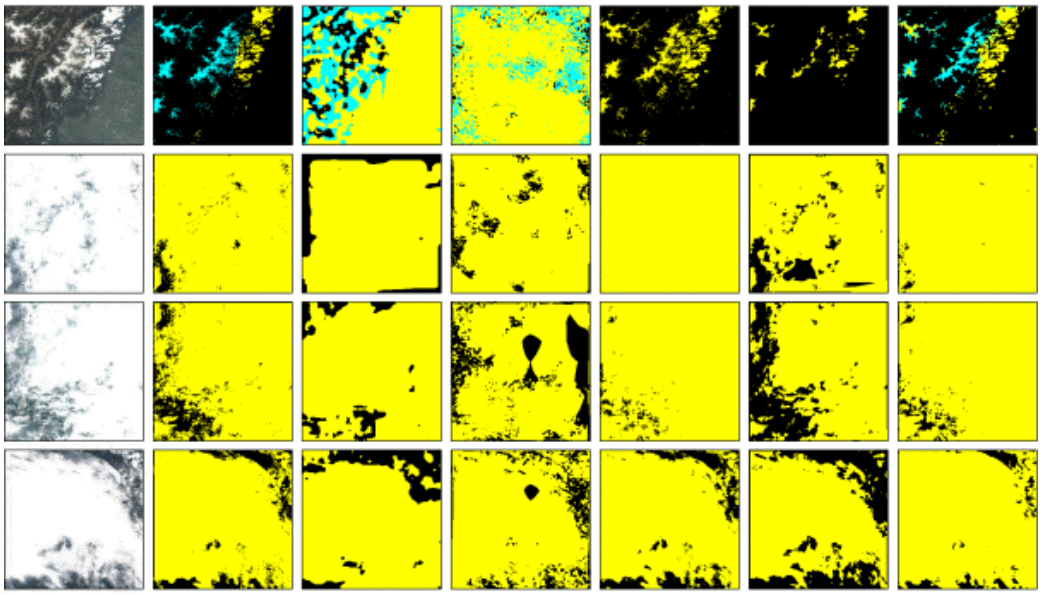


Figure 8. Detection results of different weakly supervised learning methods on cloud and snow detection task [149]. (**a**) Image. (**b**) Label. (**c**) GradCAM. (**d**) WDCD. (**e**) GAN-CDM. (**f**) CloudMatting. (**g**) WCSD.

The above research fully utilizes the unique advantages of GAN as a generative model and opens up new ideas for cloud detection. A key advantage of GAN is its ability to operate effectively with limited labeled data. Based on their adversarial training mechanism, GANs are able to generate synthetic training samples, which enhances the model's ability to learn meaningful features. This ability is particularly important for cloud detection, where obtaining labeled datasets is often time-consuming and laborious. In cloud detection, GANs are able to generate labeled samples that closely resemble real cloud patterns, where the generator creates cloud-like samples and the discriminator evaluates the reliability of these generated samples against real cloud samples. Another significant advantage of GAN in cloud detection is its understanding of cloud boundaries and structure. The combination of adversarial loss and standard segmentation loss enables the model to gradually enhance its understanding of cloud shape, boundaries, and subtle structures, thereby promoting a more robust representation of cloud features. This enables the model to more effectively distinguish between clouds of different types and conditions, resulting in more accurate cloud detection results.

However, in cloud detection, GANs also face issues such as non-convergence and mode collapse. In the case of non-convergence, the adversarial game between the generator and the discriminator may lead to the inability to generate high-quality images. Mode collapse means that the generator may only learn certain patterns within the data distribution, thereby failing to generate diverse cloud samples, which negatively impacts the accuracy and robustness of cloud detection. Research has shown that while GAN has higher detection accuracy than FCN when dealing with known cloud categories, it still suffers from missed detections and poor segmentation when faced with ambiguous or unclear cloud categories. Although GANs can generate images similar to real cloud patterns through adversarial training mechanisms, the lack of diversity in the generated samples may hinder the recognition of complex cloud structures in practical applications. In addition, the training process of GANs is relatively complex and requires careful hyperparameter adjustment to ensure the stability of the model. Without proper tuning and training strategies, GANs may fall into an unstable state, resulting in performance degradation. Compared with models based on attention mechanisms and Transformers, GANs have certain deficiencies in model stability and convergence. This is also an important reason why the number of

research papers based on Transformers in this field is significantly higher than that based on GANs.

5. Deep Learning in Cloud Detection Follow-Up Workflow

Cloud shadow detection and cloud/shadow removal are follow-up workflows to cloud detection and are crucial components of the complete cloud contamination elimination process. Deep learning algorithms not only play an important role in cloud detection but also demonstrate high precision and transferability in cloud shadow detection and cloud/shadow removal, showing great potential.

5.1. Cloud Shadow Detection

Compared to cloud detection, accurately identifying cloud shadows presents greater challenges. The complexity of land cover, along with the unpredictable spectral, radiometric, and thermal characteristics of cloud shadows, adds to the difficulty of cloud shadow detection [153]. Overall, cloud shadow detection primarily faces three key challenges [154]: Firstly, dark surfaces such as water bodies and wetlands have similar reflective properties to cloud shadows, making them hard to distinguish. Secondly, areas covered by cloud shadows often contain various types of land cover, leading to a broader spectral range within the shadows. Lastly, thin clouds have high radiation transmittance, and their shadow characteristics can easily mix with those of clear pixels, complicating the detection process.

Cloud shadow detection is closely related to cloud detection, and many algorithms have been proposed for this purpose. Traditional cloud shadow detection can mainly be categorized into the following types: projection matching methods [155,156], a combination of spectral analysis and geometric methods [157], and machine learning-based algorithms [158]. Projection matching methods utilize prior knowledge, such as solar azimuth angle, solar elevation angle, and sensor altitude, to calculate the cloud projection direction and employ matching filters to identify cloud shadows. Spectral analysis, combined with geometric methods, uses spectral tests to measure radiance and determine the locations of cloud shadows, while geometric methods fill in the gaps in spectral testing. Machine learning-based algorithms mainly rely on handcrafted features for pixel clustering. However, these methods do not take into account the significant spectral variations within cloud shadow regions, and the choice of spectral values and thresholds is heavily reliant on prior knowledge, making them susceptible to various interferences, leading to false detections and omissions.

In recent years, deep learning algorithms have gradually emerged in the field of cloud shadow detection, significantly improving detection accuracy and efficiency. Compared to traditional methods, deep learning algorithms do not require manual selection of spectral and threshold values, greatly enhancing their flexibility and adaptability in practical applications. Cloud shadow detection is a complex task influenced by various factors, such as the complexity of land cover and the similarity between cloud shadows and their backgrounds. The powerful features of deep learning enable researchers to more accurately identify and segment clouds and their shadows. Currently, several datasets have been annotated for clouds and cloud shadows, including L7 Irish, L8 SPARCS, WHU Cloud Dataset, S2-Hollstein, and GF1 WHU. These datasets provide rich samples for training deep learning models, facilitating improved model performance. For example, Lu et al. [77] proposed an innovative dual-branch network that combines the advantages of Transformers and Convolutional Neural Networks (CNNs). This network features a Mutual Guidance Module (MGM) that effectively enhances the model's ability to extract multi-scale semantic information, resulting in a more precise segmentation of clouds and cloud shadows. Wieland et al. [109] introduced an improved U-Net structure specifically designed for segmenting clouds, cloud shadows, ice, snow, water bodies, and land in multi-sensor images. Experiments conducted on Landsat and Sentinel data demonstrate that this model exhibits excellent transferability and inference speed, achieving an overall accuracy of 89%. This efficient model design not only improves the accuracy of cloud shadow detection but also

provides strong support for other remote sensing applications. Yan et al. [159] addressed the challenges of extracting multi-scale clouds and cloud shadows during rapid model inference by proposing a multi-level feature fusion segmentation network. This network maintains good detection performance even when spectral bands available are limited. Inspired by the feature pyramid structure in PSPNet [95], they designed a new pyramid aggregation module that effectively extracts contextual information and significantly improves cloud shadow detection capabilities under dark land cover conditions.

The introduction of deep learning has significantly reduced the reliance on prior knowledge in cloud shadow detection. Traditional methods typically depend on extensive domain expertise and physical models to accurately detect cloud shadows. In contrast, deep learning automatically learns patterns from data, reducing the need for such prior knowledge. Currently, there are many well-established semantic segmentation models capable of accurately segmenting both clouds and cloud shadows simultaneously. These models can not only recognize complex cloud structures but also handle cloud shadows generated under various lighting conditions. Particularly for satellite data with limited spectral bands, such as GaoFen-1 and Proba-V, which have only three visible bands and one near-infrared band, deep learning-based segmentation methods still achieve high accuracy despite the limited spectral information. This capability surpasses traditional algorithms and machine learning-based models, which typically struggle to achieve comparable results with restricted spectral data. Thus, the application of deep learning not only enhances segmentation accuracy but also significantly broadens the range of usable data.

5.2. Cloud Removal

Cloud removal aims to eliminate the loss of information caused by cloud cover in images and restore surface details to enhance the usability of remote sensing images. Thin clouds and thick clouds differ significantly in terms of spatial morphology and optical thickness, and their impact on surface information varies as well. Thin clouds are semi-transparent, allowing some ground information to pass through. As a result, when removing thin clouds, a large amount of valuable surface detail, such as image color and structure, can still be preserved, making the removal of thin clouds crucial for restoring image details. In contrast, thick clouds, due to their higher opacity, completely block the acquisition of ground information, leading to irreversible data loss in the covered areas.

Before deep learning became widely used, cloud removal algorithms were typically classified into four main types [160]. Spatial-based algorithms rely on image spatial autocorrelation to fill in cloud-covered areas but perform poorly without a reference image, especially when large cloud patches are present. Spectral-based algorithms utilize the complementary spectral information of different wavelengths to penetrate clouds, but are only effective for thin-cloud removal. Multi-temporal-based algorithms use cloud-free images from different time periods to replace cloud-contaminated areas, working best when there are minimal ground changes between the images. Finally, hybrid methods combine the strengths of these approaches to achieve more precise cloud removal. These four types of algorithms primarily address missing pixels from the perspectives of spatial, spectral, and temporal information. Among them, algorithms based on temporal information are widely used because they can effectively remove both thin and thick clouds. However, obtaining cloud-free images is quite challenging and computationally expensive.

Remote sensing cloud removal falls under the category of image restoration. Currently, models based on GANs and Transformers are widely applied in this field. For instance, Zhang et al. [161] proposed an end-to-end generative single-image cloud removal network called SCR-Net. This network employs a generator network G to learn the conditional distribution of cloud-free data and utilizes a multi-scale discriminator D to distinguish between real and fake images. Through supervised training, the generator G can generate cloud-free images based on input noise and cloudy images. Experimental results show that SCR-Net effectively removes both uniform and non-uniform thin clouds from remote sensing images while maintaining good color consistency. In another study, Xu et al. [162] introduced

an attention-based GAN called AMGAN-CR, which is trained using pairs of cloudy and cloud-free images. This model learns the feature correlations between the two types of images, allowing it to recover lost information during the image reconstruction process.

The incorporation of Transformers further enhances image restoration performance. For example, Huang and Wu [163] proposed an end-to-end training model called CTGAN, which combines Transformers and GANs. The use of attention mechanisms enables the model to identify the most crucial features before inpainting. Han et al. [164] introduced a cloud removal model named Former-CR based on Transformers, which reconstructs cloud-free images from Synthetic Aperture Radar (SAR) and multi-cloud optical images, ensuring global consistency in color, structure, and texture information in the reconstructed areas.

Deep learning has demonstrated outstanding performance in cloud removal tasks, particularly with the integration of Transformers and GANs, accelerating the automation of the cloud removal process. Compared to traditional algorithms, deep learning-based approaches reduce the need for manual intervention and possess a natural advantage in multi-source data fusion. Currently, many mature datasets [164,165] are available for research on cloud removal tasks. However, deep learning-based models are susceptible to challenges such as mode collapse and convergence issues. Addressing these challenges is crucial for improving the reliability and robustness of cloud removal algorithms. A major difficulty lies in effectively utilizing information from thin-cloud regions and multi-source data to inpaint thick-cloud areas. This remains an active research direction in remote sensing image processing, with researchers continuously exploring new methods and techniques to overcome these challenges and further enhance the capabilities of deep learning-based cloud removal models.

6. Existing Problems and Future Directions

By organizing and summarizing the tasks related to cloud detection and follow-up processing based on deep learning, it is evident that the continuous development of semantic segmentation in deep learning has significantly improved the accuracy and automation capabilities of cloud detection. Mainstream cloud detection algorithms are gradually shifting from traditional methods to deep learning-based models [50,166,167]. Researchers have conducted extensive work in the field of cloud detection, proposing many excellent models and processing workflows. However, there are still several issues within current cloud detection methods that warrant further investigation.

Efficient cloud detection models: The improvement in accuracy in deep learning often comes with an explosive increase in model parameters. High-accuracy models require significant computational power and large amounts of data for training [168]. The expansion of model parameters leads to longer training and inference times. Researchers have been exploring methods to reduce model parameters while maintaining accuracy or designing lightweight networks [70,137,169] to address this issue. However, these efforts may not yet meet the requirements for real-time processing in satellite deployments. Therefore, a promising future research direction is how to enhance model inference speed without sacrificing accuracy, to meet the demands of real-time processing in satellite-based applications [170–172].

Cross-sensor cloud detection: Currently, most cloud detection algorithms are trained for specific sensors or datasets, lacking a general cloud detection algorithm that can be applied across different sensors [173]. There have been efforts in the field of cross-sensor or multi-satellite cloud detection, but challenges related to model accuracy and inference speed still persist [111,174–177]. Addressing the issue of cross-sensor cloud detection could benefit from exploring how deep learning can be leveraged to enhance model generalization and transferability across different sensors. This could potentially lead to the development of more versatile and efficient cloud detection solutions.

Large-size finely labeled datasets: With the integration of Transformers, semantic segmentation models have experienced a significant increase in the number of parameters. Small datasets may not provide enough training data to effectively train all the parameters

of these models [168]. Meanwhile, powerful deep learning models such as ViT [127] and Swin Transformer [128] can process large image sizes, even up to 1024 × 1024, and achieve high-precision predictions. However, the lack of large, accurately annotated datasets remains a bottleneck in the development of cross-sensor cloud detection [52]. Therefore, expanding large-scale finely annotated datasets is a crucial area of work to address this limitation and advance cross-sensor cloud detection capabilities.

Reduce manual intervention: The mainstream deep learning-based cloud detection algorithms currently use supervised learning training strategies, although these methods reduce the need for a priori knowledge, but a large amount of repetitive work such as parameter adjustment and data labeling still exists; there are already some automated labeling tools such as [178,179], but this does not solve the fundamental problem of supervised learning [149,151]. How to reduce the reliance of the model on manual labeling data from the perspective of semi-supervised learning, weakly supervised learning, and self-supervised learning is the key to reducing repetitive work.

7. Conclusions

The primary goal of this survey is to provide a comprehensive overview of the current landscape of cloud detection algorithms used in optical remote sensing, with a particular focus on those utilizing deep learning techniques. We highlight the limitations of traditional cloud detection methods and explain how deep learning models offer promising solutions to address these challenges. A systematic analysis is conducted on various categories of deep learning-based cloud detection algorithms, including models based on FCN, encoder–decoder architectures, attention mechanisms, and GAN. By comparing their performance with classical and machine learning-based approaches, we identify the strengths and weaknesses of each methodology. Additionally, we explore the availability of publicly accessible datasets for cloud detection, as well as the necessary post-detection processing steps, such as cloud shadow detection and cloud removal. These steps are critical for achieving accurate and reliable results in remote sensing applications. In conclusion, this survey serves as a valuable resource for researchers and practitioners, offering insights into the advancements, limitations, and potential areas for improvement in cloud detection using deep learning models. Furthermore, we identify key challenges and suggest future directions for the development of deep learning-based cloud detection algorithms. By addressing these challenges and exploring new avenues, the accuracy and efficiency of cloud detection in optical remote sensing can be significantly enhanced.

Author Contributions: Conceptualization, Z.W. and H.L.; methodology, Z.W., H.L. and L.Z.; investigation, J.M. and Y.H.; writing—original draft preparation, Z.W., H.L., X.L. and J.C.; writing—review and editing, Z.W., R.J. and H.L.; supervision, J.C.; funding acquisition, H.L., L.Z. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Shenzhen Science and Technology Program (No. JCYJ20220818101617038), the Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515011858), the National Natural Science Foundation of China (No. 42271353), and Scientific research project of Ecology Environment Bureau of Shenzhen Municipality (No. SZDL2023001387).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, Z.; Lu, L.; Zhao, Y.; Wang, Y.; Wei, D.; Wu, X.; Ma, X. Recent Advances in Using Chinese Earth Observation Satellites for Remote Sensing of Vegetation. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 393–407. [CrossRef]
2. Zhao, Q.; Yu, L.; Du, Z.; Peng, D.; Hao, P.; Zhang, Y.; Gong, P. An Overview of the Applications of Earth Observation Satellite Data: Impacts and Future Trends. *Remote Sens.* **2022**, *14*, 1863. [CrossRef]
3. Li, Y.; Yu, R.; Xu, Y.; Zhang, X. Spatial Distribution and Seasonal Variation of Cloud over China Based on ISCCP Data and Surface Observations. *J. Meteorol. Soc. Jpn.* **2004**, *82*, 761–773. [CrossRef]

4. He, Q.; Sun, X.; Yan, Z.; Fu, K. DABNet: Deformable Contextual and Boundary-Weighted Network for Cloud Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
5. Li, J.; Wu, Z.; Sheng, Q.; Wang, B.; Hu, Z.; Zheng, S.; Camps-Valls, G.; Molinier, M. A Hybrid Generative Adversarial Network for Weakly-Supervised Cloud Detection in Multispectral Images. *Remote Sens. Environ.* **2022**, *280*, 113197. [CrossRef]
6. Zhang, J.; Wu, J.; Wang, H.; Wang, Y.; Li, Y. Cloud Detection Method Using CNN Based on Cascaded Feature Attention and Channel Attention. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [CrossRef]
7. Wu, K.; Xu, Z.; Lyu, X.; Ren, P. Cloud Detection with Boundary Nets. *ISPRS J. Photogramm. Remote Sens.* **2022**, *186*, 218–231. [CrossRef]
8. Liu, Z.; Yang, J.; Wang, W.; Shi, Z. Cloud Detection Methods for Remote Sensing Images: A Survey. *Chin. Space Sci. Technol.* **2023**, *43*, 1–17. [CrossRef]
9. Schiffer, R.A.; Rossow, W.B. The International Satellite Cloud Climatology Project (ISCCP): The First Project of the World Climate Research Programme. *Bull. Am. Meteorol. Soc.* **1983**, *64*, 779–784. [CrossRef]
10. Scaramuzza, P.L.; Bouchard, M.A.; Dwyer, J.L. Development of the Landsat Data Continuity Mission Cloud-Cover Assessment Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1140–1154. [CrossRef]
11. Irish, R.R.; Barker, J.L.; Goward, S.N.; Arvidson, T. Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 1179–1188. [CrossRef]
12. 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]
13. Kazakov, E.E.; Borisova, Y.I. Open-Source Software Implementation and Validation of the Split-Window Method for Automated Land Surface Temperature Retrieval from Landsat 8 Data. *Izv. Atmos. Ocean. Phys.* **2021**, *57*, 1171–1178. [CrossRef]
14. Zhang, H.; Huang, Q.; Zhai, H.; Zhang, L. Multi-Temporal Cloud Detection Based on Robust PCA for Optical Remote Sensing Imagery. *Comput. Electron. Agric.* **2021**, *188*, 106342. [CrossRef]
15. 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]
16. Sun, L.; Mi, X.; Wei, J.; Wang, J.; Tian, X.; Yu, H.; Gan, P. A Cloud Detection Algorithm-Generating Method for Remote Sensing Data at Visible to Short-Wave Infrared Wavelengths. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 70–88. [CrossRef]
17. Chen, X.; Liu, L.; Gao, Y.; Zhang, X.; Xie, S. A Novel Classification Extension-Based Cloud Detection Method for Medium-Resolution Optical Images. *Remote Sens.* **2020**, *12*, 2365. [CrossRef]
18. Yuan, Y.; Hu, X. Bag-of-Words and Object-Based Classification for Cloud Extraction From Satellite Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4197–4205. [CrossRef]
19. Zhu, Z.; Woodcock, C.E. Object-Based Cloud and Cloud Shadow Detection in Landsat Imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [CrossRef]
20. Main-Knorn, M.; Pflug, B.; Louis, J.; Debaecker, V.; Müller-Wilm, U.; Gascon, F. Sen2Cor for Sentinel-2. In Proceedings of the Image and Signal Processing for Remote Sensing XXIII, Warsaw, Poland, 11–13 September 2017; SPIE: Bellingham, WA, USA, 2017; Volume 10427, pp. 37–48.
21. Luo, Y.; Trishchenko, A.P.; Khlopenkov, K.V. Developing Clear-Sky, Cloud and Cloud Shadow Mask for Producing Clear-Sky Composites at 250-m Spatial Resolution for the Seven MODIS Land Bands over Canada and North America. *Remote Sens. Environ.* **2008**, *112*, 4167–4185. [CrossRef]
22. Zhu, Z.; Woodcock, C.E. Automated Cloud, Cloud Shadow, and Snow Detection in Multitemporal Landsat Data: An Algorithm Designed Specifically for Monitoring Land Cover Change. *Remote Sens. Environ.* **2014**, *152*, 217–234. [CrossRef]
23. 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]
24. 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]
25. Zhang, J.; Zhou, Q.; Shen, X.; Li, Y. Cloud Detection in High-Resolution Remote Sensing Images Using Multi-Features of Ground Objects. *J. Geovis. Spat. Anal.* **2019**, *3*, 14. [CrossRef]
26. Zhu, X.; Helmer, E.H. An Automatic Method for Screening Clouds and Cloud Shadows in Optical Satellite Image Time Series in Cloudy Regions. *Remote Sens. Environ.* **2018**, *214*, 135–153. [CrossRef]
27. Wang, X.-Y.; Wang, T.; Bu, J. Color Image Segmentation Using Pixel Wise Support Vector Machine Classification. *Pattern Recognit.* **2011**, *44*, 777–787. [CrossRef]
28. Shotton, J.; Johnson, M.; Cipolla, R. Semantic Texton Forests for Image Categorization and Segmentation. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
29. Wei, J.; Huang, W.; Li, Z.; Sun, L.; Zhu, X.; Yuan, Q.; Liu, L.; Cribb, M. Cloud Detection for Landsat Imagery by Combining the Random Forest and Superpixels Extracted via Energy-Driven Sampling Segmentation Approaches. *Remote Sens. Environ.* **2020**, *248*, 112005. [CrossRef]
30. Heidinger, A.K.; Evan, A.T.; Foster, M.J.; Walther, A. A Naive Bayesian Cloud-Detection Scheme Derived from CALIPSO and Applied within PATMOS-x. *J. Appl. Meteorol. Climatol.* **2012**, *51*, 1129–1144. [CrossRef]
31. Ping, B.; Su, F.; Meng, Y. A Cloud and Cloud Shadow Detection Method Based on Fuzzy C-Means Algorithm. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1714–1727. [CrossRef]

32. 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]
33. Zhang, Q.; Yu, Y.; Zhang, W.; Luo, T.; Wang, X. Cloud Detection from FY-4A's Geostationary Interferometric Infrared Sounder Using Machine Learning Approaches. *Remote Sens.* **2019**, *11*, 3035. [CrossRef]
34. 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]
35. Buttar, P.K.; Sachan, M.K. Semantic Segmentation of Clouds in Satellite Images Based on U-Net plus plus Architecture and Attention Mechanism. *Expert Syst. Appl.* **2022**, *209*, 118380. [CrossRef]
36. Yu, J.; Li, Y.; Zheng, X.; Zhong, Y.; He, P. An Effective Cloud Detection Method for Gaofen-5 Images via Deep Learning. *Remote Sens.* **2020**, *12*, 2106. [CrossRef]
37. 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]
38. Xie, F.; Shi, M.; Shi, Z.; Yin, J.; Zhao, D. Multilevel Cloud Detection in Remote Sensing Images Based on Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3631–3640. [CrossRef]
39. 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.
40. Yang, J.; Guo, J.; Yue, H.; Liu, Z.; Hu, H.; Li, K. CDnet: CNN-Based Cloud Detection for Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6195–6211. [CrossRef]
41. 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]
42. Zhang, H.; Wang, Y.; Yang, X. Cloud Detection for Satellite Cloud Images Based on Fused FCN Features. *Remote Sens. Lett.* **2022**, *13*, 683–694. [CrossRef]
43. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Springer International Publishing: Berlin/Heidelberg, Germany, 2015.
44. Pu, W.; Wang, Z.; Liu, D.; Zhang, Q. Optical Remote Sensing Image Cloud Detection with Self-Attention and Spatial Pyramid Pooling Fusion. *Remote Sens.* **2022**, *14*, 4312. [CrossRef]
45. Yan, Q.; Liu, H.; Zhang, J.; Sun, X.; Xiong, W.; Zou, M.; Xia, Y.; Xun, L. Cloud Detection of Remote Sensing Image Based on Multi-Scale Data and Dual-Channel Attention Mechanism. *Remote Sens.* **2022**, *14*, 3710. [CrossRef]
46. Zhang, J.; Zhou, Q.; Wu, J.; Wang, Y.; Wang, H.; Li, Y.; Chai, Y.; Liu, Y. A Cloud Detection Method Using Convolutional Neural Network Based on Gabor Transform and Attention Mechanism with Dark Channel Subnet for Remote Sensing Image. *Remote Sens.* **2020**, *12*, 3261. [CrossRef]
47. Xie, W.; Yang, J.; Li, Y.; Lei, J.; Zhong, J.; Li, J. Discriminative Feature Learning Constrained Unsupervised Network for Cloud Detection in Remote Sensing Imagery. *Remote Sens.* **2020**, *12*, 456. [CrossRef]
48. Wen, X.; Pan, Z.; Hu, Y.; Liu, J. Generative Adversarial Learning in YUV Color Space for Thin Cloud Removal on Satellite Imagery. *Remote Sens.* **2021**, *13*, 1079. [CrossRef]
49. Guo, J.; Xu, Q.; Zeng, Y.; Liu, Z.; Zhu, X. Semi-Supervised Cloud Detection in Satellite Images by Considering the Domain Shift Problem. *Remote Sens.* **2022**, *14*, 2641. [CrossRef]
50. 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]
51. Sawant, M.; Shende, M.K.; Feijóo-Lorenzo, A.E.; Bokde, N.D. The State-of-the-Art Progress in Cloud Detection, Identification, and Tracking Approaches: A Systematic Review. *Energies* **2021**, *14*, 8119. [CrossRef]
52. He, M.; Zhang, J.; He, Y.; Zuo, X.; Gao, Z. Annotated Dataset for Training Cloud Segmentation Neural Networks Using High-Resolution Satellite Remote Sensing Imagery. *Remote Sens.* **2024**, *16*, 3682. [CrossRef]
53. Le Goff, M.; Tourneret, J.-Y.; Wendt, H.; Ortner, M.; Spigai, M. Deep Learning for Cloud Detection. In Proceedings of the 8th International Conference of Pattern Recognition Systems (ICPRS 2017), Madrid, Spain, 11–13 July 2017; pp. 1–6.
54. He, M.; Zhang, J. Radiation Feature Fusion Dual-Attention Cloud Segmentation Network. *Remote Sens.* **2024**, *16*, 2025. [CrossRef]
55. Zhao, Z.; Zhang, F.; Wu, Q.; Li, Z.; Tong, X.; Li, J.; Han, W. Cloud Identification and Properties Retrieval of the Fengyun-4A Satellite Using a ResUnet Model. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–18. [CrossRef]
56. Zhang, J.; He, M. Methodology for Severe Convective Cloud Identification Using Lightweight Neural Network Model Ensembling. *Remote Sens.* **2024**, *16*, 2070. [CrossRef]
57. 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*, 1–21. [CrossRef]
58. Aybar, C.; Ysuhuaylas, L.; Loja, J.; Gonzales, K.; Herrera, F.; Bautista, L.; Yali, R.; Flores, A.; Diaz, L.; Cuenca, N.; et al. CloudSEN12, a Global Dataset for Semantic Understanding of Cloud and Cloud Shadow in Sentinel-2. *Sci. Data* **2022**, *9*, 782. [CrossRef] [PubMed]
59. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Joseph Hughes, M.; Laue, B. Cloud Detection Algorithm Comparison and Validation for Operational Landsat Data Products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [CrossRef]

60. 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]
61. Mohajerani, S.; Saeedi, P. Cloud-Net: An End-To-End Cloud Detection Algorithm for Landsat 8 Imagery. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1029–1032.
62. Shendryk, Y.; Rist, Y.; Ticehurst, C.; Thorburn, P. Deep Learning for Multi-Modal Classification of Cloud, Shadow and Land Cover Scenes in PlanetScope and Sentinel-2 Imagery. *ISPRS J. Photogramm. Remote Sens.* **2019**, *157*, 124–136. [CrossRef]
63. 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]
64. Stillinger, T.; Collar, N. Snow-Cloud Validation Masks for Multispectral Satellite Data. Available online: https://zenodo.org/records/3240937 (accessed on 1 June 2024).
65. Lin, D.; Xu, G.; Wang, X.; Wang, Y.; Sun, X.; Fu, K. A Remote Sensing Image Dataset for Cloud Removal. *arXiv* **2019**, arXiv:1901.00600.
66. Ji, S.; Dai, P.; Lu, M.; Zhang, Y. Simultaneous Cloud Detection and Removal From Bitemporal Remote Sensing Images Using Cascade Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 732–748. [CrossRef]
67. 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]
68. Francis, A.; Mrziglod, J.; Sidiropoulos, P.; Muller, J.-P. Sentinel-2 Cloud Mask Catalogue. Available online: https://zenodo.org/records/4172871 (accessed on 1 June 2024).
69. Domnich, M.; Voormansik, K.; Wold, O.; Harun, F.; Sünter, I.; Trofimov, H.; Kostiukhin, A.; Järveoja, M. Sentinel-2 KappaZeta Cloud and Cloud Shadow Masks. Available online: https://zenodo.org/records/5095024 (accessed on 1 June 2024).
70. 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.* **2022**, *60*, 1–19. [CrossRef]
71. 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]
72. Zhu, S.; Li, Z.; Shen, H. GF1MS-WHU and GF2MS-WHU. Available online: https://ieee-dataport.org/documents/gf1ms-whu-and-gf2ms-whu (accessed on 1 June 2024).
73. Li, Y.; Chen, W.; Zhang, Y.; Tao, C.; Xiao, R.; Tan, Y. Accurate Cloud Detection in High-Resolution Remote Sensing Imagery by Weakly Supervised Deep Learning. *Remote Sens. Environ.* **2020**, *250*, 112045. [CrossRef]
74. 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]
75. 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]
76. Jiao, L.; Huo, L.; Hu, C.; Tang, P. Refined UNet v3: Efficient End-to-End Patch-Wise Network for Cloud and Shadow Segmentation with Multi-Channel Spectral Features. *Neural Netw.* **2021**, *143*, 767–782. [CrossRef]
77. Lu, C.; Xia, M.; Qian, M.; Chen, B. Dual-Branch Network for Cloud and Cloud Shadow Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
78. Li, L.; Li, X.; Liu, X.; Huang, W.; Hu, Z.; Chen, F. Attention Mechanism Cloud Detection With Modified FCN for Infrared Remote Sensing Images. *IEEE Access* **2021**, *9*, 150975–150983. [CrossRef]
79. Jiang, B.; An, X.; Xu, S.; Chen, Z. Intelligent Image Semantic Segmentation: A Review Through Deep Learning Techniques for Remote Sensing Image Analysis. *J. Indian Soc. Remote Sens.* **2023**, *51*, 1865–1878. [CrossRef]
80. Guo, Y.; Nie, G.; Gao, W.; Liao, M. 2D Semantic Segmentation: Recent Developments and Future Directions. *Future Internet* **2023**, *15*, 205. [CrossRef]
81. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
82. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [CrossRef]
83. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. BRIEF: Binary Robust Independent Elementary Features. In Proceedings of the Computer Vision—ECCV 2010; Daniilidis, K., Maragos, P., Paragios, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 778–792.
84. Dhanachandra, N.; Manglem, K.; Chanu, Y.J. Image Segmentation Using K-Means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Comput. Sci.* **2015**, *54*, 764–771. [CrossRef]
85. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
86. 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] [PubMed]
87. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation. *Appl. Soft Comput.* **2018**, *70*, 41–65. [CrossRef]

88. Yuan, X.; Shi, J.; Gu, L. A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery. *Expert Syst. Appl.* **2021**, *169*, 114417. [CrossRef]
89. Liu, W.; Rabinovich, A.; Berg, A.C. ParseNet: Looking Wider to See Better. *arXiv* **2015**, arXiv:1506.04579.
90. Chen, L.-C.; Papandreou, 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]
91. Lafferty, J.; McCallum, A.; Pereira, F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *Dep. Pap. (CIS)* **2001**, *1*, 3.
92. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *arXiv* **2014**, arXiv:1412.7062. [CrossRef]
93. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
94. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv* **2018**, arXiv:1802.02611.
95. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
96. 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]
97. Li, W.; Sun, K.; Du, Z.; Hu, X.; Li, W.; Wei, J.; Gao, S. PCNet: Cloud Detection in FY-3D True-Color Imagery Using Multi-Scale Pyramid Contextual Information. *Remote Sens.* **2021**, *13*, 3670. [CrossRef]
98. Wu, X.; Shi, Z. Scene Aggregation Network for Cloud Detection on Remote Sensing Imagery. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
99. Chen, Y.; Weng, Q.; Tang, L.; Liu, Q.; Fan, R. An Automatic Cloud Detection Neural Network for High-Resolution Remote Sensing Imagery With Cloud-Snow Coexistence. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
100. Wu, X.; Shi, Z. Utilizing Multilevel Features for Cloud Detection on Satellite Imagery. *Remote Sens.* **2018**, *10*, 1853. [CrossRef]
101. Ni, Z.; Wu, M.; Lu, Q.; Huo, H.; Wang, F. Research on Infrared Hyperspectral Remote Sensing Cloud Detection Method Based on Deep Learning. *Int. J. Remote Sens.* **2024**, *45*, 7497–7517. [CrossRef]
102. Noh, H.; Hong, S.; Han, B. Learning Deconvolution Network for Semantic Segmentation. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015; pp. 1520–1528.
103. 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]
104. Zhou, Z.; Rahman Siddiquee, M.M.; 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*; Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J.M.R.S., Bradley, A., Papa, J.P., Belagiannis, V., et al., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–11.
105. Abraham, N.; Khan, N.M. A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 683–687.
106. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [CrossRef]
107. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
108. Hu, K.; Zhang, D.; Xia, M. CDUNet: Cloud Detection UNet for Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 4533. [CrossRef]
109. 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]
110. Liu, Y.; Wang, W.; Li, Q.; Min, M.; Yao, Z. DCNet: A Deformable Convolutional Cloud Detection Network for Remote Sensing Imagery. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
111. 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]
112. Pesek, O.; Segal-Rozenhaimer, M.; Karnieli, A. Using Convolutional Neural Networks for Cloud Detection on VEN Mu S Images over Multiple Land-Cover Types. *Remote Sens.* **2022**, *14*, 5210. [CrossRef]
113. Zhang, L.; Sun, J.; Yang, X.; Jiang, R.; Ye, Q. Improving Deep Learning-Based Cloud Detection for Satellite Images With Attention Mechanism. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
114. Luo, C.; Feng, S.; Li, X.; Ye, Y.; Zhang, B.; Chen, Z.; Quan, Y. ECDNet: A Bilateral Lightweight Cloud Detection Network for Remote Sensing Images. *Pattern Recognit.* **2022**, *129*, 108713. [CrossRef]
115. Shi, C.; Zhou, Y.; Qiu, B.; Guo, D.; Li, M. CloudU-Net: A Deep Convolutional Neural Network Architecture for Daytime and Nighttime Cloud Images' Segmentation. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 1688–1692. [CrossRef]

116. Li, X.; Yang, X.; Li, X.; Lu, S.; Ye, Y.; Ban, Y. GCDB-UNet: A Novel Robust Cloud Detection Approach for Remote Sensing Images. *Knowl.-Based Syst.* **2022**, *238*, 107890. [CrossRef]
117. Du, W.; Fan, Z.; Yan, Y.; Yu, R.; Liu, J. AFMUNet: Attention Feature Fusion Network Based on a U-Shaped Structure for Cloud and Cloud Shadow Detection. *Remote Sens.* **2024**, *16*, 1574. [CrossRef]
118. Scaramuzza, P.; Dwyer, J. *L7 Irish Cloud Validation Masks*; USGS: Sioux Falls, SD, USA, 2016. [CrossRef]
119. Scaramuzza, P.; Foga, S.; Dwyer, J. *L8 Biome Cloud Validation Masks*; USGS: Sioux Falls, SD, USA, 2016. [CrossRef]
120. Joseph Hughes, M. *L8 SPARCS Cloud Validation Masks*; USGS: Sioux Falls, SD, USA, 2016.
121. Dev, S.; Nautiyal, A.; Lee, Y.H.; Winkler, S. CloudSegNet: A Deep Network for Nychthemeron Cloud Image Segmentation. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1814–1818. [CrossRef]
122. Zhong, Z.; Lin, Z.Q.; Bidart, R.; Hu, X.; Daya, I.B.; Li, Z.; Zheng, W.-S.; Li, J.; Wong, A. Squeeze-and-Attention Networks for Semantic Segmentation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
123. Chen, L.-C.; Yang, Y.; Wang, J.; Xu, W.; Yuille, A.L. Attention to Scale: Scale-Aware Semantic Image Segmentation. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3640–3649.
124. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual Attention Network for Scene Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3141–3149.
125. 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 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3213–3223.
126. Vaswani, A.; Shazeer, N.M.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
127. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
128. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
129. Xu, X.; Feng, Z.; Cao, C.; Li, M.; Wu, J.; Wu, Z.; Shang, Y.; Ye, S. An Improved Swin Transformer-Based Model for Remote Sensing Object Detection and Instance Segmentation. *Remote Sens.* **2021**, *13*, 4779. [CrossRef]
130. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 1–41. [CrossRef]
131. Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H.R.; Xu, D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In *Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*; Crimi, A., Bakas, S., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 272–284.
132. Guo, Y.; Cao, X.; Liu, B.; Gao, M. Cloud Detection for Satellite Imagery Using Attention-Based U-Net Convolutional Neural Network. *Symmetry* **2020**, *12*, 1056. [CrossRef]
133. Liu, S.; Zhang, J.; Zhang, Z.; Cao, X.; Durrani, T.S. TransCloudSeg: Ground-Based Cloud Image Segmentation With Transformer. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 6121–6132. [CrossRef]
134. Zhang, Z.; Xu, Z.; Liu, C.; Tian, Q.; Wang, Y. Cloudformer: Supplementary Aggregation Feature and Mask-Classification Network for Cloud Detection. *Appl. Sci.* **2022**, *12*, 3221. [CrossRef]
135. Zhang, B.; Zhang, Y.; Li, Y.; Wan, Y.; Yao, Y. CloudViT: A Lightweight Vision Transformer Network for Remote Sensing Cloud Detection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
136. Ge, W.; Yang, X.; Jiang, R.; Shao, W.; Zhang, L. CD-CTFM: A Lightweight CNN-Transformer Network for Remote Sensing Cloud Detection Fusing Multiscale Features. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 4538–4551. [CrossRef]
137. Yao, X.; Guo, Q.; Li, A. Light-Weight Cloud Detection Network for Optical Remote Sensing Images with Attention-Based DeeplabV3+Architecture. *Remote Sens.* **2021**, *13*, 3617. [CrossRef]
138. Gong, C.; Long, T.; Yin, R.; Jiao, W.; Wang, G. A Hybrid Algorithm with Swin Transformer and Convolution for Cloud Detection. *Remote Sens.* **2023**, *15*, 5264. [CrossRef]
139. Singh, R.; Biswas, M.; Pal, M. A Transformer-Based Cloud Detection Approach Using Sentinel 2 Imageries. *Int. J. Remote Sens.* **2023**, *44*, 3194–3208. [CrossRef]
140. Tan, Y.; Zhang, W.; Yang, X.; Liu, Q.; Mi, X.; Li, J.; Yang, J.; Gu, X. Cloud and Cloud Shadow Detection of GF-1 Images Based on the Swin-UNet Method. *Atmosphere* **2023**, *14*, 1669. [CrossRef]
141. Fan, S.; Song, T.; Jin, G.; Jin, J.; Li, Q.; Xia, X. A Lightweight Cloud and Cloud Shadow Detection Transformer With Prior-Knowledge Guidance. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
142. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In Proceedings of the Advances in Neural Information Processing Systems; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 12077–12090.

143. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
144. Luc, P.; Couprie, C.; Chintala, S.; Verbeek, J. Semantic Segmentation Using Adversarial Networks. *arXiv* **2016**, arXiv:1611.08408.
145. Souly, N.; Spampinato, C.; Shah, M. Semi Supervised Semantic Segmentation Using Generative Adversarial Network. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5689–5697.
146. 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.
147. Zou, Z.; Li, W.; Shi, T.; Shi, Z.; Ye, J. Generative Adversarial Training for Weakly Supervised Cloud Matting. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 201–210.
148. Nyborg, J.; Assent, I. Weakly-Supervised Cloud Detection with Fixed-Point GANs. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; pp. 4191–4198.
149. Yang, J.; Li, W.; Chen, K.; Liu, Z.; Shi, Z.; Zou, Z. Weakly Supervised Adversarial Training for Remote Sensing Image Cloud and Snow Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 15206–15221. [CrossRef]
150. 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.* **2020**, *17*, 1792–1796. [CrossRef]
151. Liu, Y.; Li, Q.; Yao, Z.; Jiang, J.; Qiu, Z.; Wang, W. CLDiff: Weakly Supervised Cloud Detection With Denoising Diffusion Probabilistic Models. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–18. [CrossRef]
152. Ji, Y.; Chen, Z.; Xie, E.; Hong, L.; Liu, X.; Liu, Z.; Lu, T.; Li, Z.; Luo, P. DDP: Diffusion Model for Dense Visual Prediction. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 21741–21752.
153. 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]
154. Zhai, H.; Zhang, H.; Zhang, L.; Li, P. Cloud/Shadow Detection Based on Spectral Indices for Multi/Hyperspectral Optical Remote Sensing Imagery. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 235–253. [CrossRef]
155. Simpson, J.; Stitt, J. A Procedure for the Detection and Removal of Cloud Shadow from AVHRR Data over Land. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 880–897. [CrossRef]
156. Fraser, A.D.; Massom, R.A.; Michael, K.J. A Method for Compositing MODIS Satellite Images to Remove Cloud Cover. In Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009; Volume 3, pp. III-639–III-641.
157. Li, S.; Sun, D.; Yu, Y. Automatic Cloud-Shadow Removal from Flood/Standing Water Maps Using MSG/SEVIRI Imagery. *Int. J. Remote Sens.* **2013**, *34*, 5487–5502. [CrossRef]
158. Ibrahim, E.; Jiang, J.; Lema, L.; Barnabé, P.; Giuliani, G.; Lacroix, P.; Pirard, E. Cloud and Cloud-Shadow Detection for Applications in Mapping Small-Scale Mining in Colombia Using Sentinel-2 Imagery. *Remote Sens.* **2021**, *13*, 736. [CrossRef]
159. Yan, Z.; Yan, M.; Sun, H.; Fu, K.; Hong, J.; Sun, J.; Zhang, Y.; Sun, X. Cloud and Cloud Shadow Detection Using Multilevel Feature Fused Segmentation Network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1600–1604. [CrossRef]
160. Li, X.; Wang, L.; Cheng, Q.; Wu, P.; Gan, W.; Fang, L. Cloud Removal in Remote Sensing Images Using Nonnegative Matrix Factorization and Error Correction. *ISPRS J. Photogramm. Remote Sens.* **2019**, *148*, 103–113. [CrossRef]
161. Zhang, J.; Wang, H.; Wang, Y.; Zhou, Q.; Li, Y. Deep Network Based on up and down Blocks Using Wavelet Transform and Successive Multi-Scale Spatial Attention for Cloud Detection. *Remote Sens. Environ.* **2021**, *261*, 112483. [CrossRef]
162. Xu, M.; Deng, F.; Jia, S.; Jia, X.; Plaza, A.J. Attention Mechanism-Based Generative Adversarial Networks for Cloud Removal in Landsat Images. *Remote Sens. Environ.* **2022**, *271*, 112902. [CrossRef]
163. Huang, G.-L.; Wu, P.-Y. CTGAN: Cloud Transformer Generative Adversarial Network. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 511–515.
164. Han, S.; Wang, J.; Zhang, S. Former-CR: A Transformer-Based Thick Cloud Removal Method with Optical and SAR Imagery. *Remote Sens.* **2023**, *15*, 1196. [CrossRef]
165. Shen, H.; Li, X.; Cheng, Q.; Zeng, C.; Yang, G.; Li, H.; Zhang, L. Missing Information Reconstruction of Remote Sensing Data: A Technical Review. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 61–85. [CrossRef]
166. Chai, D.; Huang, J.; Wu, M.; Yang, X.; Wang, R. Remote Sensing Image Cloud Detection Using a Shallow Convolutional Neural Network. *ISPRS J. Photogramm. Remote Sens.* **2024**, *209*, 66–84. [CrossRef]
167. Ding, L.; Xia, M.; Lin, H.; Hu, K. Multi-Level Attention Interactive Network for Cloud and Snow Detection Segmentation. *Remote Sens.* **2024**, *16*, 112. [CrossRef]
168. Huang, L.; Jiang, B.; Lv, S.; Liu, Y.; Fu, Y. Deep-Learning-Based Semantic Segmentation of Remote Sensing Images: A Survey. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 8370–8396. [CrossRef]
169. Luo, C.; Feng, S.; Yang, X.; Ye, Y.; Li, X.; Zhang, B.; Chen, Z.; Quan, Y. LWCDnet: A Lightweight Network for Efficient Cloud Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
170. Wei, F.; Wang, S.; Sun, Y.; Yin, B. A Dual Attentional Skip Connection Based Swin-UNet for Real-Time Cloud Segmentation. *IET Image Process.* **2024**, *18*, 3460–3479. [CrossRef]

171. Fan, N.; Li, D.; Pan, J.; Huang, S.; Wang, X. ANNet: Asymmetric Nested Network for Real-Time Cloud Detection in Remote Sensing. *IEEE Trans. Geosci. Remote Sens.* **2024**, 1. [CrossRef]
172. Aybar, C.; Mateo-García, G.; Acciarini, G.; Ruzicka, V.; Meoni, G.; Longépé, N.; Gómez-Chova, L. Onboard Cloud Detection and Atmospheric Correction With Efficient Deep Learning Models. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 19518–19529. [CrossRef]
173. Wang, W.; Kai, X.; Wang, A.; Chen, Y.; Deng, X.; Wang, T. ADAC: An Active Domain Adaptive Network with Progressive Learning Strategy for Cloud Detection of Remote Sensing Imagery. *Geo-Spat. Inf. Sci.* **2024**, 1–18. [CrossRef]
174. Zhu, S.; Li, Z.; Shen, H. Transferring Deep Models for Cloud Detection in Multisensor Images via Weakly Supervised Learning. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–18. [CrossRef]
175. Du, A.; Doan, A.; Law, Y.; Chin, T. Domain Adaptation for Satellite-Borne Multispectral Cloud Detection. *Remote Sens.* **2024**, *16*, 3469. [CrossRef]
176. Mateo-Garcia, G.; Laparra, V.; Lopez-Puigdollers, D.; Gomez-Chova, L. Transferring Deep Learning Models for Cloud Detection between Landsat-8 and Proba-V. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 1–17. [CrossRef]
177. Mateo-Garcia, G.; Laparra, V.; Lopez-Puigdollers, D.; Gomez-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]
178. Sofiiuk, K.; Petrov, I.A.; Konushin, A. Reviving Iterative Training with Mask Guidance for Interactive Segmentation. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 3141–3145.
179. Hao, Y.; Liu, Y.; Chen, Y.; Han, L.; Peng, J.; Tang, S.; Chen, G.; Wu, Z.; Chen, Z.; Lai, B. EISeg: An Efficient Interactive Segmentation Tool Based on PaddlePaddle. *arXiv* **2022**, arXiv:2210.08788.