



Article

# Radiation Feature Fusion Dual-Attention Cloud Segmentation Network

Mingyuan He and Jie Zhang \*

College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China; hemingyuan17@nudt.edu.cn

\* Correspondence: zhangjie19@nudt.edu.cn

**Abstract:** In the field of remote sensing image analysis, the issue of cloud interference in high-resolution images has always been a challenging problem, with traditional methods often facing limitations in addressing this challenge. To this end, this study proposes an innovative solution by integrating radiative feature analysis with cutting-edge deep learning technologies, developing a refined cloud segmentation method. The core innovation lies in the development of FFASPPDANet (Feature Fusion Atrous Spatial Pyramid Pooling Dual Attention Network), a feature fusion dual attention network improved through atrous spatial convolution pooling to enhance the model's ability to recognize cloud features. Moreover, we introduce a probabilistic thresholding method based on pixel radiation spectrum fusion, further improving the accuracy and reliability of cloud segmentation, resulting in the "FFASPPDANet+" algorithm. Experimental validation shows that FFASPPDANet+ performs exceptionally well in various complex scenarios, achieving a 99.27% accuracy rate in water bodies, a 96.79% accuracy rate in complex urban settings, and a 95.82% accuracy rate in a random test set. This research not only enhances the efficiency and accuracy of cloud segmentation in high-resolution remote sensing images but also provides a new direction and application example for the integration of deep learning with radiative algorithms.

**Keywords:** high-resolution remote sensing images; cloud segmentation; fusion of radiative features and deep learning; FFASPPDANet+; adaptability to diverse scenarios; stable performance



**Citation:** He, M.; Zhang, J. Radiation Feature Fusion Dual-Attention Cloud Segmentation Network. *Remote Sens.* **2024**, *16*, 2025. <https://doi.org/10.3390/rs16112025>

Academic Editors: M. Saquib Sarfraz and Muhammad Adnan Siddique

Received: 16 April 2024

Revised: 28 May 2024

Accepted: 1 June 2024

Published: 5 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the rapid advancement of remote sensing technology, satellite remote sensing images have become a crucial means of obtaining information about the Earth's surface. These images contain rich information about the Earth's features, which are essential for environmental monitoring, disaster early warning, and resource management. However, the issue of cloud interference in high-resolution remote sensing images greatly limits the effectiveness and application range of these data. Currently, remote sensing image cloud segmentation methods are primarily categorized into four types [1–5]: multispectral physical property methods, texture and spatial characteristic methods of cloud layers, pattern recognition methods, and deep learning-based cloud segmentation.

Early methods primarily used physical approaches, analyzing the multispectral physical properties of images and applying the characteristics of visible or infrared spectra to individual pixels for cloud identification. For instance, Liu Xinyan and colleagues used spectral difference analysis based on cloud and surface spectral features to perform cloud segmentation on GF-4 satellite data, achieving accuracies above 82% [6]. The challenge of these methods lies in selecting appropriate physical property thresholds and reducing computational loads for hardware implementation. Initially, fixed-threshold detection methods showed good performance for specific sensors. In the U.S., the second stage of the Earth Observing System (EOS) included the MODIS sensor with 36 spectral bands [7,8], which enhances the range and resolution of remote sensing data, leading researchers to

favor MODIS images. In 2018, Xiang used the K-means clustering method and Otsu method to initially extract clouds from MODIS data, then further differentiated cloud pixels from those that were not cloud pixels [9]. Xie et al. introduced a semi-supervised method for cloud detection in multispectral images, utilizing spatial and spectral features [10]. It particularly focuses on reducing mis-detection in complex backgrounds and thicker cloud layers, which aligns with the use of specific channels for robust segmentation across various surface environments. Although simple and effective for on-orbit real-time cloud segmentation, physical methods have limitations, particularly in threshold selection, which requires extensive statistical analysis and iterative experiments. Additionally, in areas covered by ice, snow, or desert, the reflectance similarity between clouds and the surface leads to the same spectrum phenomenon, making precise differentiation challenging. Some scholars suggest using multi-temporal methods to address this issue, which introduces the need for cloud-free images of the scene. Moreover, physical methods depend heavily on the number of spectral bands and are initially designed for specific sensors like MODIS, which are not suitable for satellites with fewer spectral bands.

Cloud classification based on physical features includes texture, shape, and grayscale color. Despite the variability of cloud texture elements with time, temperature, and wind, they exhibit unique features compared to underlying sea surfaces, such as sharp changes near cloud edges. To better extract these features, Hégarat-Masclé used a Markov random framework [11], R. Rossi applied singular value decomposition [12], and SVM techniques were used for cloud cover area determination. Li, Pengfei, and others trained an SVM classifier using cloud brightness temperature and texture features [13]. Başeski, E., and Ç. Cenaras determined cloud presence based on color, while parallel algorithms were proposed for physical feature extraction [14]. Based on satellite data differences, multispectral detection [15], the Bag-of-Words model [16], Bayesian spatiotemporal algorithms [17], and progressive refinement algorithms [18] have been successfully applied to cloud segmentation. As the spatial resolution of satellite remote sensing images increases, cloud texture and spatial characteristics become more prominent, making methods based on these features viable for cloud segmentation. However, these methods, based on shallow features, often miss complex clouds like thin clouds or cumulus, leading to detection inaccuracies.

To overcome the limitations of spectral methods, machine learning algorithms have been introduced for cloud classification in remote sensing images. Li and others used SVM to classify reflectance and co-occurrence matrix, achieving over 90% accuracy for small datasets [13]. Pattern recognition methods—by combining deep features with clustering, SVM, or deep learning—achieve more precise results than threshold or texture methods. McKay extracted SIFT features from sub-images of remote sensing cloud images, combining sparse reconstruction-based classification (SRC) and Localized Pose Management (LPM) algorithms for precise target recognition [19]. Yu proposed a clustering-based pattern recognition method to differentiate cloud layers and glacial snow [20]. SVM and other algorithms enhance cloud segmentation by utilizing texture information, offering broader applicability than spectral methods but requiring manual feature selection. Traditional machine learning algorithms treat image segmentation as a per-pixel classification, with training and prediction complexity dependent on image resolution. For large-scale remote sensing satellite images, these methods often cannot meet real-time operational needs.

With the proliferation of GPUs and advancements in deep learning, techniques using deep neural networks for image object recognition are widely applied in safety monitoring, healthcare, transportation, and industrial production. Employing AI technologies like deep learning to address meteorological issues has been a major research direction in recent years. Recent advancements in deep learning have provided new solutions for cloud segmentation in remote sensing imagery [21–24]. Deep neural networks [25] (DNNs) have been widely applied in fields like image recognition, object detection, and segmentation. Models such as AlexNet [26], YOLO [27], and U-Net [28] have shown effective cloud segmentation capabilities by learning features directly from data.

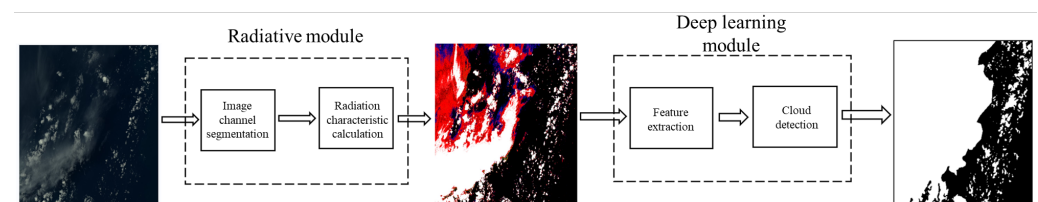
However, current cloud segmentation networks are often modified from models in other domains, leading to limitations. For example, U-Net [29], originally designed for medical image processing, learns overall image features well but struggles with details, affecting segmentation accuracy. To address these issues, dedicated deep learning models incorporating advanced techniques are needed. Dual attention networks [30] improve sensitivity to cloud features through spatial and channel attention mechanisms. Atrous spatial pyramid pooling convolutions [31] expand receptive fields without losing resolution, enhancing detail capture. Probabilistic thresholding methods [32] further refine segmentation by considering pixel-level features and global information.

Combining these three technologies for cloud recognition addresses different challenges in the cloud recognition process, and their comprehensive application can leverage greater advantages. Dual attention networks improve the model's focus and discriminative power by focusing on key areas and feature channels; atrous spatial pyramid pooling convolutions expand the model's receptive field, enhancing the capture of the widespread nature of clouds in remote sensing images; probabilistic thresholding provides the model with prior information, further improving classification accuracy. The combination of these three technologies aims to create a powerful cloud recognition framework capable of effectively handling complex cloud segmentation tasks in high-resolution remote sensing images, achieving more accurate and robust cloud recognition. This comprehensive approach not only improves recognition efficiency and accuracy but also brings new research directions and application prospects to the field of cloud recognition in high-resolution remote sensing images.

In light of this, this study proposes an innovative cloud segmentation framework, FFASPPDANet+, which combines the advantages of atrous spatial pyramid pooling convolutions, dual attention networks, and probabilistic thresholding. This framework aims to address the challenges of cloud segmentation in high-resolution remote sensing images, thus improving the effectiveness of remote sensing images in various application scenarios through refined cloud recognition and segmentation. Through testing in scenarios of different complexities, FFASPPDANet+ has demonstrated outstanding segmentation performance, marking a significant step forward in the application of deep learning technology in the field of remote sensing image processing and providing new directions and ideas for future research.

## 2. Materials and Methods

This study introduces a comprehensive cloud segmentation framework, FFASPPDANet+, which enhances the accuracy and robustness of cloud identification using advanced neural network architectures by combining a deep learning module and a radiative module (Figure 1).



**Figure 1.** “FFASPPDANet+” Algorithm Framework.

The deep learning module is the original Feature Fusion Dual Attention Network (FFASPPDANet), incorporating dual attention mechanisms, atrous spatial pyramid pooling convolutions, and probabilistic thresholding to improve cloud feature recognition and spatial distribution capture. Initially, we developed the FFASPPDANet, utilizing spatial and channel attention mechanisms. These mechanisms enhance the model's capability to recognize cloud features by focusing on relevant spatial and channel information selectively. Concurrently, atrous spatial pyramid pooling convolutions are employed to expand the

model's receptive field. This expansion allows the model to better capture the complex spatial distribution characteristics of clouds, which are essential for accurate segmentation.

The radiative module calculates the probability of each pixel being part of a cloud and sets a dynamic threshold to differentiate between cloud and non-cloud areas effectively. Therefore, by utilizing this cloud segmentation approach that integrates radiative characteristics with deep learning, we can achieve a binary segmentation of the surface and clouds. The algorithm identifies all areas with clouds, with the advantage of mitigating the misidentification of bright surface features by radiative algorithms while also addressing the issue of low recognition rates for thin clouds by deep learning methods. This probability-based approach introduces greater flexibility and precision in cloud segmentation, adapting the threshold based on the local context of the image, which significantly improves the model's robustness.

The FFASPPDANet+ algorithm then combines features from radiative feature-based and deep learning-based cloud segmentation methods. It incorporates the high-confidence portions of the results from a fusion probability threshold method, which is based on pixel spectral lines. The cloud segmentation confidence obtained through the radiative feature cloud judgment algorithm is used as prior information. This prior, along with remote sensing imagery, is input into the neural network for training, forming the comprehensive FFASPPDANet+ cloud segmentation algorithm.

Specifically, a fusion probability threshold method based on pixel spectral lines is used to output cloud judgment results and confidence levels for each pixel. The original optical imagery, along with these results and confidence levels, is fed into the neural network FFASPPDANet. The high-confidence portions of the optical image and result map are stacked in the feature dimension. This stacked input is then fed into the neural network for final cloud segmentation. This approach allows FFASPPDANet+ to adaptively segment clouds from varied scenes effectively, as demonstrated in the figure below. By combining these advanced techniques, FFASPPDANet+ achieves superior performance in cloud segmentation tasks, adapting to a broad range of scenes with varying cloud distributions and characteristics.

The FFASPPDANet+ framework introduces several key innovations to cloud segmentation in satellite imagery, significantly enhancing the performance and applicability of this process in meteorological and climate studies. Firstly, the integration of spatial and channel attention mechanisms within the Feature Fusion Dual Attention Network (FFASPPDANet) allows for selective emphasis on the most relevant features, improving the network's ability to discern intricate cloud patterns against varied backgrounds. Secondly, the use of atrous spatial pyramid pooling expands the model's receptive field, enabling it to capture broader spatial distributions and finer details of cloud formations. Thirdly, the incorporation of a probabilistic thresholding method allows for dynamic adjustment of segmentation thresholds based on the local pixel-wise probability of clouds, thus increasing the robustness and flexibility of the segmentation process. Lastly, by fusing high-confidence segmentation results with prior information derived from radiative feature-based cloud judgment, FFASPPDANet+ achieves superior accuracy and scene adaptability, marking a significant advancement in the field of remote sensing and cloud analysis.

### *2.1. Basic Principles of the Radiative Module*

To further optimize the accuracy of the thresholding method, a targeted utilization of the spectral lines of pixels across six channels is employed. As illustrated in the schematic diagram of cloud segmentation technology based on visible light images below, a thresholding method based on pixel spectral lines is designed. This method takes into account the different sensitivity levels of each channel's pixel spectral lines to various types of clouds within different grayscale intervals, thereby enhancing the precision of the thresholding method. The radiation module can be divided into the following steps: First, the six-channel features to be tested for impact are extracted. Based on the channel probability judgment criteria, the probability of each pixel being a cloud in each channel is calculated.

Then, using the channel fusion equation, the cloud probability maps of these six channels are fused to obtain the total score of cloud possibility for each pixel. In accordance with the channel fusion equation probability judgment criteria, the final cloud segmentation result is obtained.

Initially, experiments are conducted on the spectral lines of pixels in the prediction results of various channels. The spectral differences across the channels are too significant, rendering a uniform threshold inadequate for detection requirements. Consequently, it becomes necessary to establish a threshold correction method based on pixel spectral lines, thereby further enhancing the accuracy of the threshold method. Through extensive experimental statistics, the accuracy criteria for the threshold method have been established. The threshold distributions and judgment criteria used in Equations (1) and (3) are presented in Tables 1–3.

**Table 1.** Grayscale characteristics of six channels, comprising the visible RGB channels and the mutual grayscale difference channels B-R, R-G, and B-G, in the threshold correction method based on pixel spectral lines.

Channel	White (High Probability)	Red (Medium–High Probability)	Green (Medium Probability)	Blue (Medium–Low Probability)	Black (Low Probability)
R	[32, 255]	[30, 32]	[27, 30]	[24, 27]	[0, 24]
G	[60, 255]	[50, 60]	[40, 50]	[25, 40]	[0, 25]
B	[42, 255]	[39, 42]	[37, 39]	[34.5, 37]	[0, 34.5]
B-G	[0, 5]	[5, 5.8]	[5.8, 7.2]	[7.2, 10], [253, 255]	[10, 253]
B-R	[0, 23]	[23, 23.5]	[23.5, 24]	[24, 24.5]	[24.5, 255]
R-G	[245, 255]	[244.4, 245]	[244.3, 244.4]	[244.2, 244.3]	[0, 244.2]

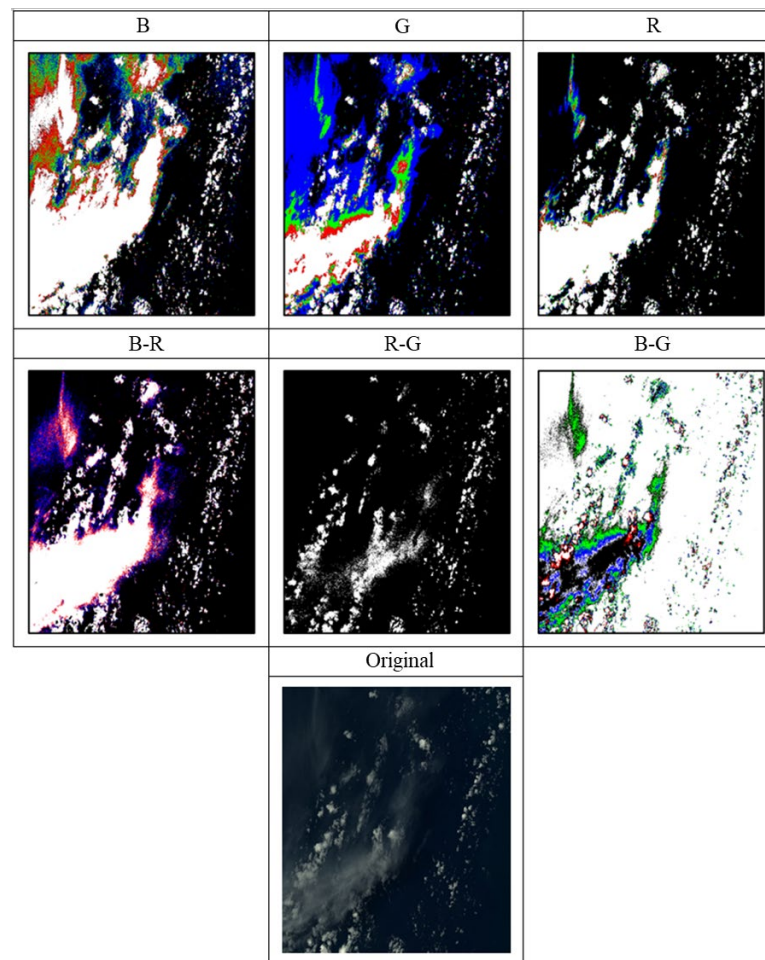
**Table 2.** Confidence table for pixels being clouds.

Confidence of Pixel Being a Cloud	High	Medium–High	Medium	Medium–Low	Low
Pixel Color $P_i$	White 1	Red 0.8	Green 0.6	Blue 0.4	Black 0.2

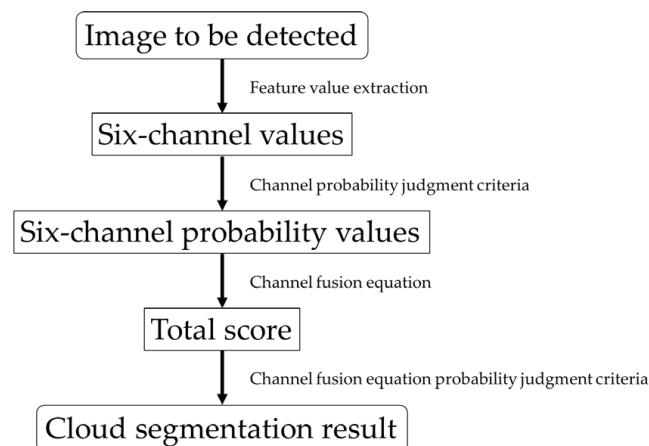
**Table 3.** Thresholds of the channel fusion equation across various probability intervals.

Probability	High	Medium–High	Medium	Medium–Low	Low
$F_u$	1.2	0.8	0.7	0.6	0

The example results of cloud segmentation based on the threshold method using pixel spectral lines are shown in Figure 2. It is evident that the prediction results for the same image differ significantly across channels, a discrepancy attributed to the differences in pixel spectra among the prediction results. Similarly, the sensitivity of different channels to various types of clouds also varies. Some channels are sensitive to convective clouds, while others are sensitive to thin clouds, resulting in varied cloud segmentation performance. For instance, if only the high-confidence parts (white) are selected as the cloud segmentation result, the B channel exhibits a low precision but a high recall, the G, R, and B-R channels show relatively balanced performance in terms of precision and recall, while the R-G channel has high precision but low recall, and the B-G channel's cloud segmentation results are opposite to those of the other channels. The primary principle of the cloud segmentation radiative module in this study is to integrate the advantageous capabilities of these channels for cloud segmentation. The process of cloud segmentation using radiative feature algorithms is illustrated in Figure 3.



**Figure 2.** Cloud probability maps for six channels, comprising the visible RGB channels and the mutual grayscale difference channels B-R, R-G, and B-G. The correspondence between probability and color is shown in Table 2, while the original image is placed at the bottom and the original image.



**Figure 3.** Schematic diagram of the visible light image cloud segmentation technology approach.

To maximize the cloud segmentation capabilities of all channels, a channel fusion equation based on weighted optimization is established here. By weighting and fusing the probabilities identified by each channel of pixels being clouds, a channel fusion equation that integrates the cloud segmentation capabilities of all channels is obtained:

$$Fu = Ch_{weight} * Ch_{score} \quad (1)$$

where  $Fu$  represents the result of the channel fusion equation, i.e., the total score of a pixel across six channels, which is used as the basis for cloud determination.  $Ch_{weight}$  and  $Ch_{score}$  represent the weights and scores of each channel (CH: R, G, B, BG, BR, RG) respectively. The scores for each channel are derived from the grayscale intervals of each channel as determined by the threshold correction method based on pixel spectral lines (Table 1). Through the analysis of channel characteristics, the probability distribution standards for pixel points in different grayscale intervals are established (Table 2). For a single pixel, its score  $Pixel_{score}$  is calculated based on its grayscale in accordance with Tables 1 and 2, while the score  $Ch_{score}$  of a pixel in a specific channel  $Pixel_{score}$  is the pixel score of that channel.

Based on the final results of the channel fusion equation at each pixel point  $Fu$ , each pixel is assigned a final probability of being identified as a cloud  $Pi$ , with the specific calculation method as follows:

$$Pi = \begin{cases} Pi_{\text{high possibility}} & , Fu > Fu_{\text{high possibility}} \\ Pi_{\text{mid-high possibility}} & , Fu_{\text{high possibility}} > Fu > Fu_{\text{mid-high possibility}} \\ Pi_{\text{mid possibility}} & , Fu_{\text{mid-high possibility}} > Fu > Fu_{\text{mid possibility}} \\ Pi_{\text{mid-low possibility}} & , Fu_{\text{mid possibility}} > Fu > Fu_{\text{mid-low possibility}} \\ Pi_{\text{low possibility}} & , Fu_{\text{mid-low possibility}} > Fu > Fu_{\text{low possibility}} \end{cases} \quad (2)$$

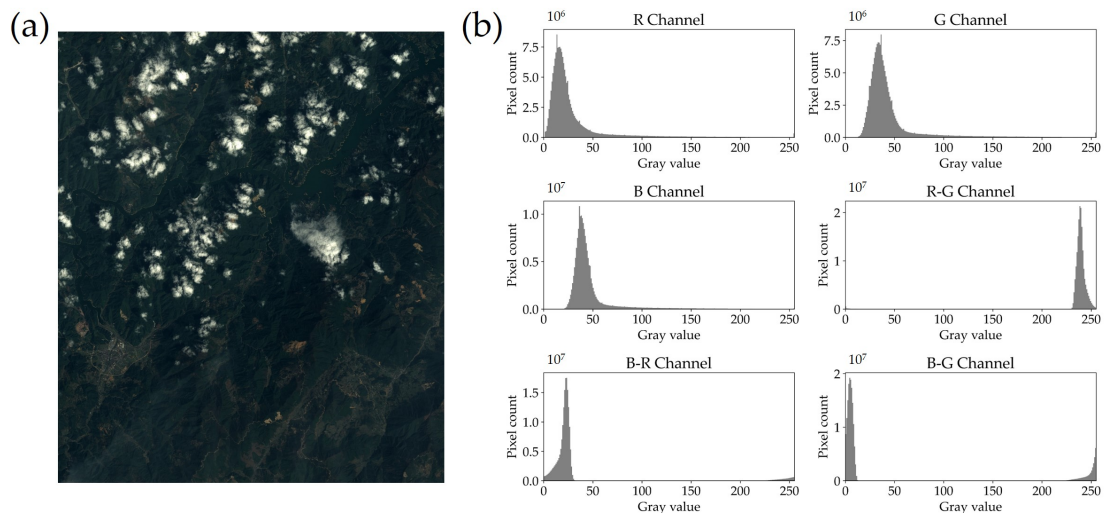
The optimal weights for each channel in the channel fusion equation are as follows.

$$\begin{cases} R_{weight} = 0.15 \\ G_{weight} = 0.25 \\ B_{weight} = 0.4 \\ BG_{weight} = 0.075 \\ BR_{weight} = 0.05 \\ RG_{weight} = 0.075 \end{cases} \quad (3)$$

To further optimize the accuracy of the threshold method, the six-channel pixel spectra were utilized in a targeted manner. As shown in Figure 3, different grayscale ranges of the pixel spectra in each channel have different sensitivities to different cloud types. Based on this feature, a threshold method using the pixel spectra was designed to further improve the accuracy.

The cloud segmentation process begins by inputting the satellite imagery to be analyzed. Feature value extraction is performed to obtain the values for the six channels. These channel values are then compared against the channel probability determination criteria (Table 2) to derive the probability values for each of the six channels. The obtained six-channel probability values are subsequently input into the channel fusion equation (Equation (1)) to calculate the total score. Finally, this total score is assessed against the probability determination criteria for the channel fusion equation (Table 3) to determine the cloud segmentation probability result.

We first conducted experiments on the pixel spectra of the predicted results from each channel (Figure 4) and found that the pixel spectral differences between channels were too large; thus, a uniform threshold could no longer meet the detection requirements. Therefore, it was necessary to establish a threshold correction method based on the pixel spectra to further enhance the accuracy of the threshold method. After multiple experimental statistics, the accuracy determination criteria for the threshold method in Tables 1–3 were derived.



**Figure 4.** Pixel Spectral Profiles: (a) Example satellite image; (b) Pixel spectral lines in different channels.

## 2.2. Basic Principles of the Deep Learning Module

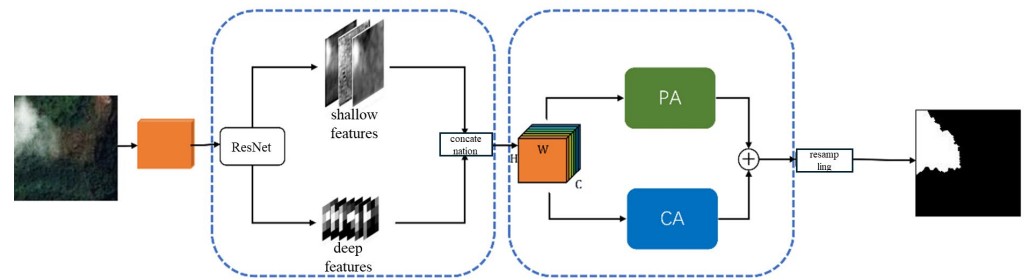
Experimental results for current mainstream neural networks indicate that the stability of cloud segmentation results across different underlying surface environments is poor, necessitating optimization for the multi-scenario characteristics of cloud segmentation tasks to improve adaptability. The main reason for the excessive influence of scene types on cloud segmentation results is that neural networks fail to extract sufficiently key features. Therefore, this paper constructs the Feature Fusion Atrous Spatial Pyramid Pooling Dual Attention Network (FFASPPDANet), which builds on the existing Dual Attention Network (DANet). In this network, a U-shaped feature fusion module (FF) designed in-house is introduced, and the atrous spatial pyramid pooling module (ASPP) proposed in DeepLabv3 replaces the convolution kernel, enhancing the multi-scale information interaction capability of the U-shaped feature fusion module to obtain critical high-discrimination features, solving the problem of poor scene stability in cloud segmentation results.

To fully utilize the semantic information extracted by the backbone network and the adaptive capacity of the attention mechanism, we propose the Feature Fusion Dual Attention Network (FFDANet). The network, when using resnet50 as the feature fusion network, is referred to as FFDANet0, and when using resnet101, it is referred to as FFDANet. Its architecture is shown in the following Figure 5. FFDANet comprises two primary components: (1) The U-shaped feature fusion module; (2) The dual attention mechanism module [30]. The semantic features extracted by (1) are fed into (2), which utilizes attention mechanisms to achieve cloud target segmentation. The proposed U-shaped feature fusion module extracts both deep and shallow semantic features of images and can be implemented based on various feature extraction networks. In this study, the ResNet network was employed as the feature extraction network (see Figure 6) due to its deep architecture, effective residual learning, and ability to capture and reuse hierarchical features efficiently. Although the U-shaped feature fusion module can effectively extract and fuse semantic information from different levels, it still contains some redundant information, resulting in unclear cloud segmentation boundaries. Therefore, we further employed a dual attention network to extract key spatial and channel information.

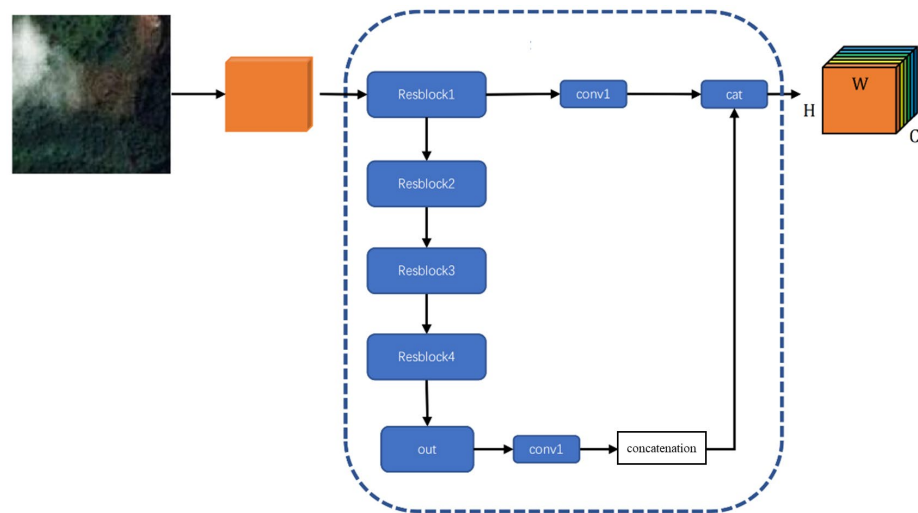
Building upon FFDANet, we further developed FFASPPDANet (Figure 7). In cloud segmentation, neural networks require not only excellent encoding modules and effective attention mechanisms but also multi-scale information interaction. The U-shaped feature fusion module of FFDANet, in the process of information extraction and fusion, uses only simple residual structures and convolutions to continuously extract the original information from remote sensing images. This structure limits the receptive field, failing to perceive more cloud-related information, leading to the omission of critical information (such as thin



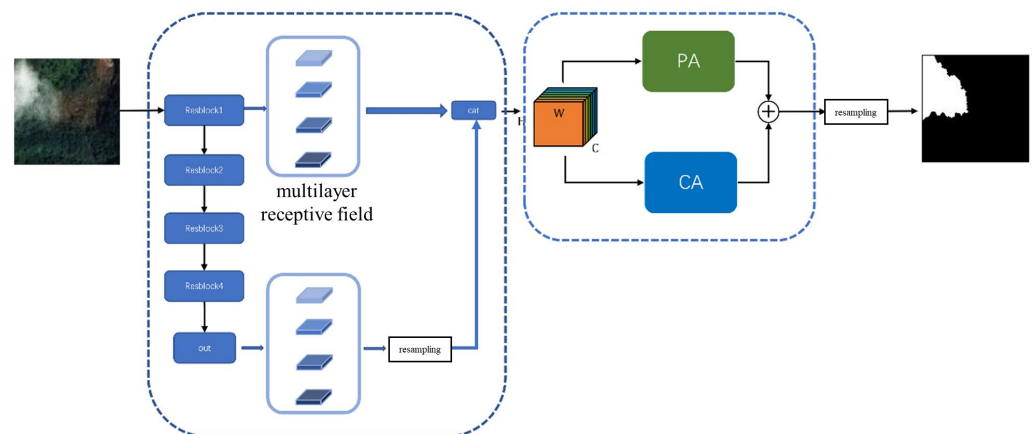
clouds and fragmented clouds). Therefore, we introduced the ASPP module [31] to replace the original  $1 \times 1$  convolution kernel, enhancing the multi-scale information interaction capability of the U-shaped feature fusion module.



**Figure 5.** FFDANet Network Architecture (for specific mechanisms of spatial attention PA and channel attention CA, see Figure 3 in the paper by Jun Fu [30]).



**Figure 6.** Structure of the U-shaped Feature Fusion Module.

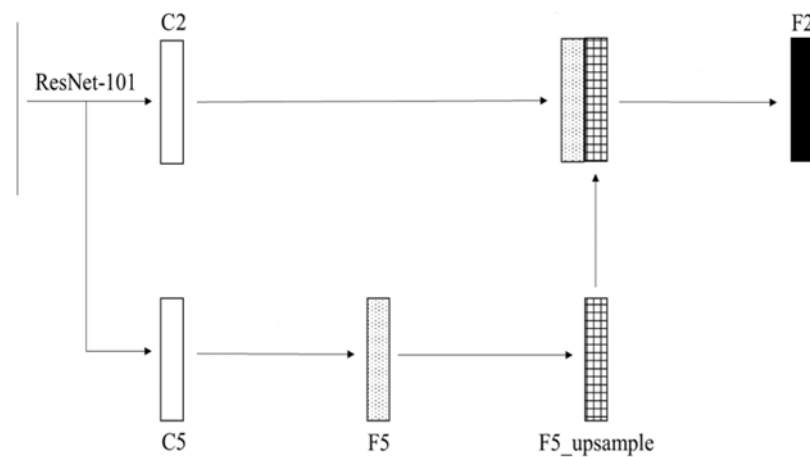


**Figure 7.** FFASPPDANet Network Structure.

In the method illustrated, we use ResNet101 as the backbone network to extract the main features. Then, we process the previously extracted feature maps using a parallel atrous spatial pyramid pooling structure. In the high-resolution feature processing at the top, we use atrous convolution kernels with dilation rates of (1, 6, 12, 18); in the low-resolution feature processing at the bottom, we use atrous convolution kernels with dilation rates of (1, 2, 3, 4). The parallel learning of atrous convolutions with different dilation

rates can further enhance multi-scale features and expand the receptive field during the information extraction process. Through the global average pooling layer, we can capture contextual information. Finally, we use reassembled sampling to fuse the features processed by the atrous spatial pyramid pooling structure.

Specifically, the detailed structure in ResNet-101 was shown in Figure 8. C1 to C5 are the convolutions from the first to fifth layers of ResNet-101. Considering the overly coarse semantic information of C1 and the insufficient semantic information dimensions of C3 and C4, FFDANet only fuses C5 with C2. C2 and C5 are the outputs of the second and fifth layers of ResNet-101, respectively, with output sizes of  $1/4$  and  $1/32$  of the original image, respectively. FFDANet uses a feature stacking method for feature fusion: C2 is first upsampled to 512 dimensions through a  $1 \times 1$  convolution, and C5 is downsampled to form F5. F5, after upsampling to double its spatial dimensions, is stacked with the upsampled C2 in the feature dimension, and finally, a  $3 \times 3$  convolution is used to fuse the stacked features. By ensuring the feature dimensions of C2 and F5 are consistent, semantic and spatial detail information imbalance is prevented. In the upsampling process, FFDANet employs sub-pixel convolution [33], allowing the network to learn the appropriate upsampling information during training.



**Figure 8.** U-shaped Feature Fusion Network with ResNet-101 as the Feature Extraction Network.

### 2.3. Model Parameter Settings

This experiment was conducted using the PyTorch2.2.0+cu121 (Python 3.10.14) deep learning framework [34] on an NVIDIA RTX A5000 GPU equipped with 24GB of memory. The adaptive moment estimation (Adam) optimizer [35] was employed, configured with the exponential decay rate for the first moment estimates  $\beta_1 = 0.9$ , the exponential decay rate for the second-moment estimates  $\beta_2 = 0.999$ , and a very small constant  $\epsilon = 10^{-8}$  to prevent any division by zero in the implementation.

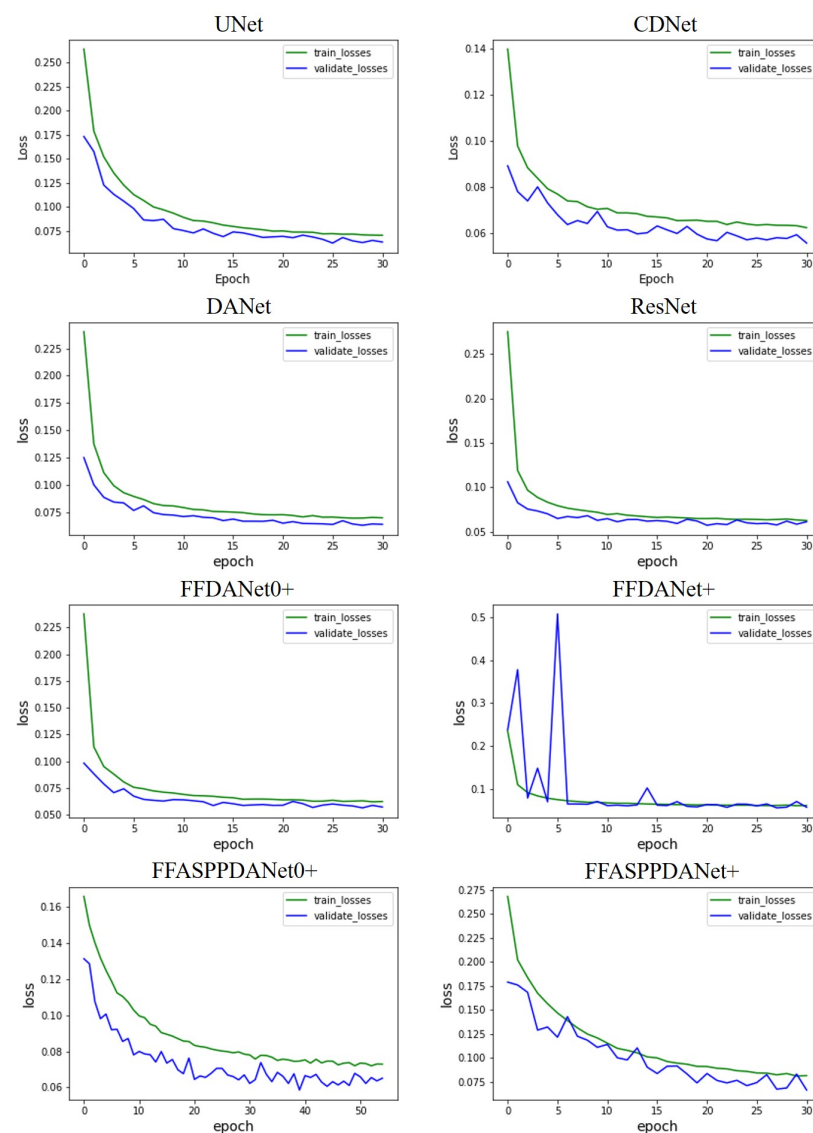
To verify the effectiveness of each module within the neural network, we conducted ablation experiments. All network models employed the same training strategy, undergoing a total of 31 training rounds. The training details are shown in Table 4. Here, we leveraged the Microsoft Neural Network Intelligence (NNI) toolkit [36] for hyperparameter optimization. NNI is a robust open-source toolkit that facilitates automatic feature engineering, hyperparameter tuning, neural architecture search, and model compression. Our hyperparameter tuning process employed several well-established algorithms available within NNI. Specifically, we mainly used the techniques in parentheses to determine the optimal values of the following parameters: the learning rate (grid search) and batch size (random search). This systematic approach to hyperparameter tuning, which was meticulously adapted to the unique characteristics of each algorithm, enabled us to identify the optimal hyperparameters, thereby enhancing the overall performance of our models.

The model training process is illustrated in the following Figure 9. During training, the training and validation set loss functions of all models generally stabilized after

20 rounds of training, dropping below 0.1. Some neural networks (such as FFDANet+) experienced significant fluctuations in the validation set loss function in the early stages but stabilized quickly. Experimentally, the models trained for 31 rounds exhibited the best cloud segmentation performance.

**Table 4.** Training details.

Model	Train Time (s/epoch)	Val Time (s/epoch)	Batch Size	Initial Learning Rate
UNet [37]	62.21	17.36	8	$5 \times 10^{-6}$
CDNet [38]	126.85	35.17	16	$1 \times 10^{-6}$
DANet [30]	108.32	29.79	4	$1 \times 10^{-5}$
ResNet [39]	102.38	28.31	16	$5 \times 10^{-6}$
FFDANet0+	112.95	31.88	16	$1 \times 10^{-6}$
FFDANet+	118.33	32.73	16	$1 \times 10^{-6}$
FFASPPDANet0+	115.54	32.25	16	$1 \times 10^{-6}$
FFASPPDANet+	120.23	33.43	16	$1 \times 10^{-6}$



**Figure 9.** Training Loss Graphs for Various Networks.

Model performance was evaluated using accuracy, precision, recall, and the F1 score as the primary indicators. Accuracy measures the overall correctness of the model in

classifying cloud and non-cloud areas. Precision and recall evaluate the model's accuracy and completeness in identifying cloud areas, respectively, while the F1 score provides a balanced measure of precision and recall, serving as a key indicator of the model's comprehensive performance.

#### 2.4. Dataset

The remote sensing image data used in this study are sourced from the SPOT6 and SPOT7 satellites, covering a variety of land surface types, including but not limited to urban areas, farmland, forests, and bodies of water. These high-resolution images provide a rich source of land surface and atmospheric cloud features, forming an important foundation for the construction and validation of the cloud segmentation model. SPOT6 and SPOT7 are part of the French SPOT (Satellite Pour l'Observation de la Terre, or Earth Observation Satellite) series, which are two civilian Earth observation satellites. These satellites continue the mission of the SPOT series, providing high-resolution optical remote sensing data to support various applications such as map making, urban planning, agricultural monitoring, environmental protection, disaster management, and military uses. SPOT6 and SPOT7 are designed as a pair to offer higher revisit frequency and more flexible data acquisition capabilities. They provide a 1.5-m panchromatic resolution and a 6-m multispectral resolution. The panchromatic images capture very detailed ground features, while the multispectral images capture the characteristics of different materials on the Earth's surface reflecting and absorbing light, which helps in analyzing different types of land surfaces such as vegetation, water bodies, and urban areas. Both satellites are in sun-synchronous orbits at an altitude of approximately 830 km, ensuring they pass over at the same time to acquire imagery under comparable lighting conditions.

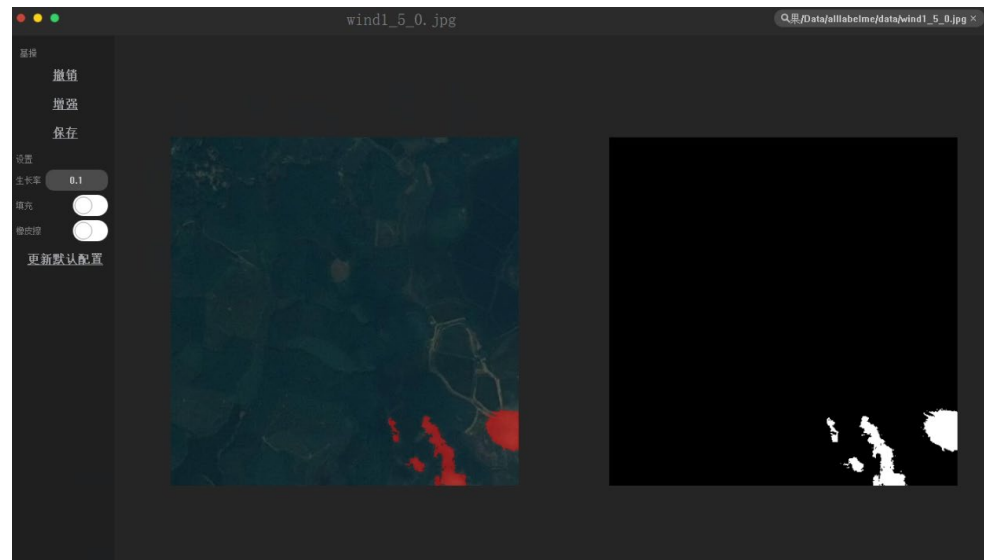
In this study, the annotated dataset was created, consisting of 32,065 satellite cloud images sized  $512 \times 512$ , randomly derived from SPOT-6 and SPOT-7 satellite images [36]. The dataset was generated using an expert-supervised semi-automatic CloudLabel annotation method, incorporating region growing, flood fill, connected components, and guided filter algorithms to optimize cloud segmentation.

Firstly, segmented images are inputted into the CloudLabel UI (Figure 10) for expert assessment. Experts select seed points for region growing and set thresholds based on cloud region characteristics. Typically, for bright, thick clouds, moderately bright pixels are chosen as seed points with a threshold of  $T = 0.4$ ; for darker, thinner clouds, the threshold is  $T = 0.2$ . Annotation details can be refined using tools like erasers and magnifiers, ensuring reliability by comparing the masked image with the original. CloudLabel software v1.0 integrates techniques including region growing, flood fill, connected components, and guided filtering. Users open the cloud image, set an appropriate growth rate, and click within the cloud area to annotate. After basic annotation, the "Enhance" button applies morphological processing to optimize results. Thus, experts only need to select the starting point and set the growth rate to semi-automatically annotate high-resolution satellite images.

Secondly, CloudLabel employs region growing [40,41], enhanced with morphological processing, to semi-automatically annotate cloud regions in high-resolution remote sensing images through human-computer interaction. Region-growing segments are images based on pixel similarity, starting from an initial seed point and iteratively adding neighboring pixels that meet intensity or color criteria. Key components include seed point selection, growth criteria definition, and threshold setting. The process continues until the region ceases to expand, as determined by a gray-level threshold supervised by experts. Proper parameter selection allows the effective partitioning of an image into coherent regions with similar properties.

Then, the flood fill algorithm was employed. High spatial resolution remote sensing images reveal detailed terrain features with significant grayscale variations within cloud regions. Some pixels may have much higher or lower values than their surroundings, leading to improper identification and the formation of holes when using region growing alone. To address this, we employ the flood fill algorithm from morphological image

processing to refine the coarse results from region growing. Although coarse annotations provide rough outlines, some pixels remain unrecognized. The flood-fill operation fills these holes, ensuring a more accurate and complete cloud annotation.



**Figure 10.** CloudLabel annotation software. In the left toolbar, the seven buttons from top to bottom are ‘Undo’, ‘Enhance’, ‘Save’, ‘Growth Rate’, ‘Fill’, ‘Eraser’, and ‘Update Default Configuration’, respectively.

After filling the holes, the region growing method relies on grayscale values of pixels, which can misclassify man-made objects (e.g., buildings, roads) with high reflectance similar to clouds. However, in high-resolution remote sensing images, clouds typically have natural, random shapes with roundedness and smooth edges, while man-made structures have regular shapes (linear, rectangular, or circular). These can be removed based on their regular shapes using the connected-component method. The annotation results from region growth are partitioned into distinct connected regions. For each region, the number of pixels  $N_0$  and the minimum bounding rectangle containing  $N$  pixels are calculated. The ratio between  $N_0$  and  $N$  is computed as  $R = N_0/N$ . By analyzing  $R$ , man-made objects can be distinguished and removed. If  $R \rightarrow 1$ , the region is rectangular; if  $R \rightarrow \pi/4$ , the region is circular; if  $R \rightarrow 0$ , the region is linear. By selecting appropriate ranges and conditions, we could effectively remove man-made objects with high reflectance.

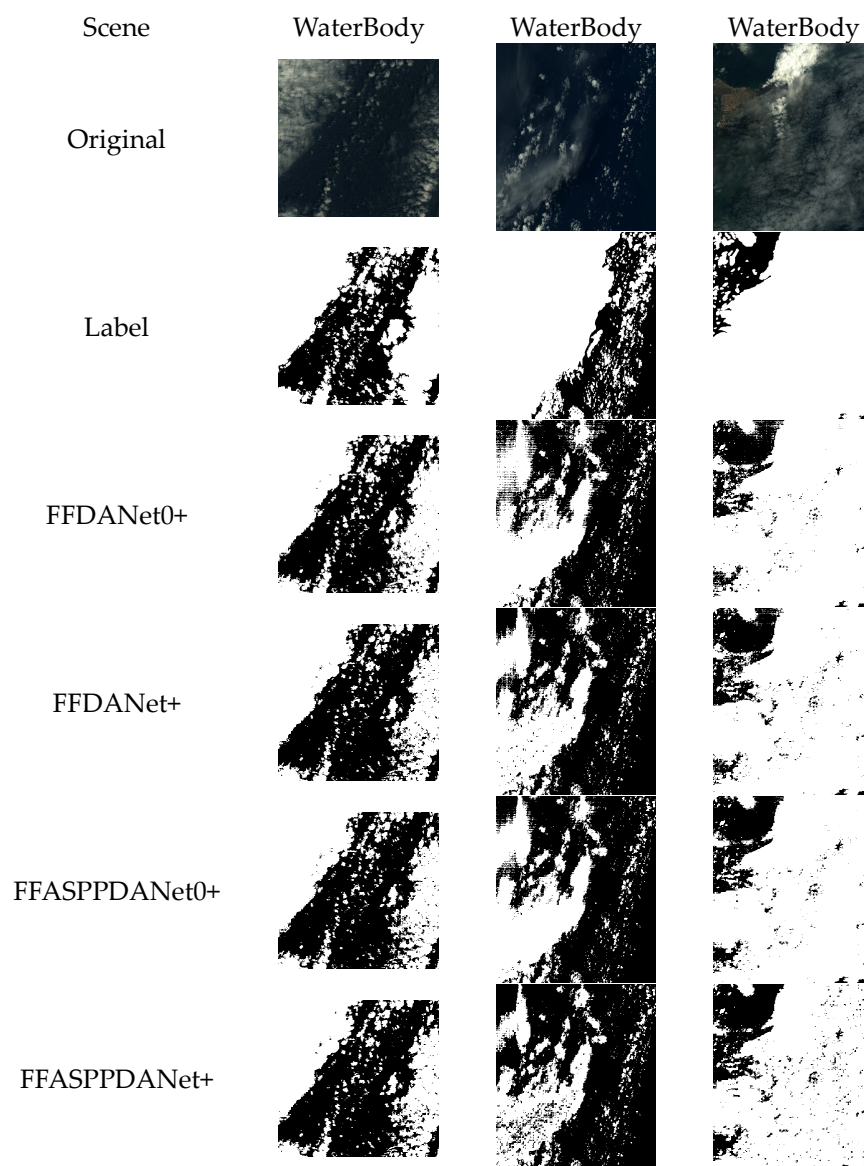
Finally, the images were processed by a guided filter. In high-resolution remote sensing images, cloud region boundaries often appear as blurry, semi-transparent pixels with reflectance similar to the terrain, unlike the high-reflectance pixels at the cloud center. This leads to misclassification. To address this, guided filtering is used for fine segmentation of the coarse annotation results, enhancing cloud region boundary characteristics. The formula is  $C_{guided} = F_{guided}(C_{rough}, C, r, \epsilon)$ , where  $F_{guided}$  is the guided filter,  $C_{rough}$  is the coarse annotation result,  $C$  is the guidance satellite image,  $r$  is the window radius, and  $\epsilon$  is the regularization parameter.

After data augmentation, the dataset was expanded to include the final set of 160,000 annotated images. The dataset was divided by 10:3:3 for training, validation, and testing. Specifically, the training set was expanded to 100,000 images, with the validation and test sets expanded to 30,000 images each. Such data preprocessing not only increased the diversity of the data but also simulated a variety of scene changes that might be encountered in practical applications, providing sufficient training and testing conditions for the deep learning model.

### 3. Results

#### 3.1. “FFASPPDANet+” Algorithm Experiment in Simple Scenarios (Underlying Surfaces of Water Bodies)

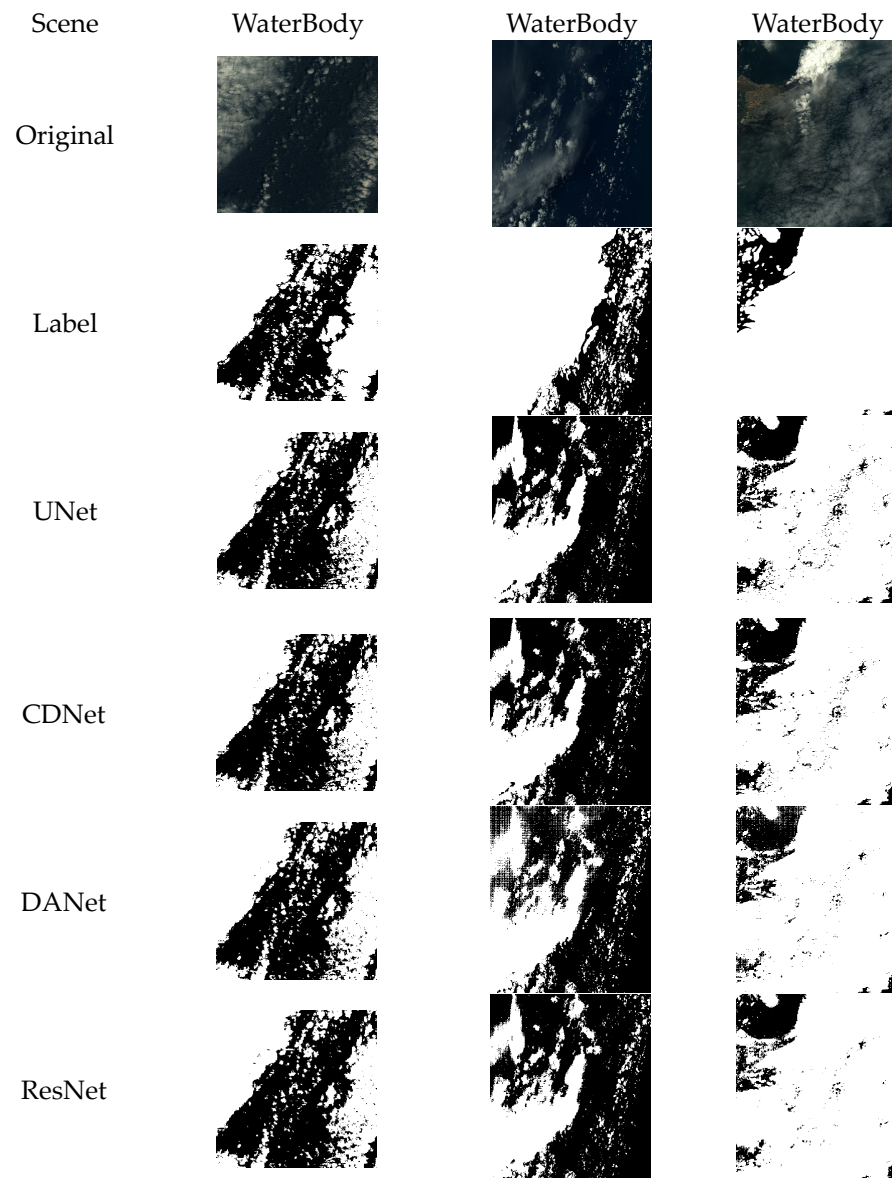
The below Figures 11 and 12 show that in remote sensing satellite images with water bodies as the underlying surface, the underlying conditions are relatively simple and filled with dark ocean geographical information. In water body scenarios, FFDANet0+, FFDANet+, FFASPPDANet0+, and FFASPPDANet+ overall perform well, effectively segmenting the main body of thick clouds.



**Figure 11.** Cloud Segmentation Results of Different Neural Networks in Water Body Scenarios.

In water body scenarios, as shown in Table 5, FFDANet0+, FFDANet+, FFASPPDANet0+, and FFASPPDANet+ demonstrate good performance. The missed detection of thin cloud parts affects the overall metrics, with the IoU around 76%, OA around 83%, and precision exceeding 98%.

It is evident that the metrics for UNet, CDNet, DANet, and ResNet are close to those of the method proposed in this paper, indicating that the room for accuracy improvement in simple scenes is not significant. This contrasts with the substantial accuracy enhancement observed in complex scenes, as discussed in the following subsection.



**Figure 12.** Cloud Segmentation Results of Different Neural Networks in Water Body Scenarios (comparison results with other state-of-the-art methods).

**Table 5.** Cloud Segmentation Results of Different Models in Water Body Scenarios.

Model	IoU	OA	Precision	F1	Recall
FFDANet0+	0.7890	0.8460	0.9837	0.8821	0.7994
FFDANet+	0.7563	0.8228	0.9879	0.8612	0.7634
FFASPPDANet0+	0.7438	0.8114	0.9865	0.8531	0.7478
FFASPPDANet+	0.7235	0.7998	0.9927	0.8396	0.7274
UNet	0.7455	0.8149	0.9872	0.8542	0.7528
CDNet	0.7474	0.8165	0.9890	0.8554	0.7536
DANet	0.7975	0.8519	0.9813	0.8873	0.8097
ResNet	0.8012	0.8530	0.9683	0.8896	0.8228

### 3.2. “FFASPPDANet+” Algorithm Experiment in Complex Scenarios (Underlying Surfaces of Urban Areas)

The results below in Figures 13 and 14 show that FFDANet+ and FFASPPDANet+ effectively utilize the fusion process of shallow and deep features, making good use of both shallow and deep semantic information. Under the influence of the attention mechanism,

which fully considers the comprehensive relationship between pixels, the process can better distinguish between surface highlights and clouds during cloud segmentation. This results in improved segmentation outcomes, with cloud edges having more detailed textures and more accurate identification of fragmented and thin clouds.

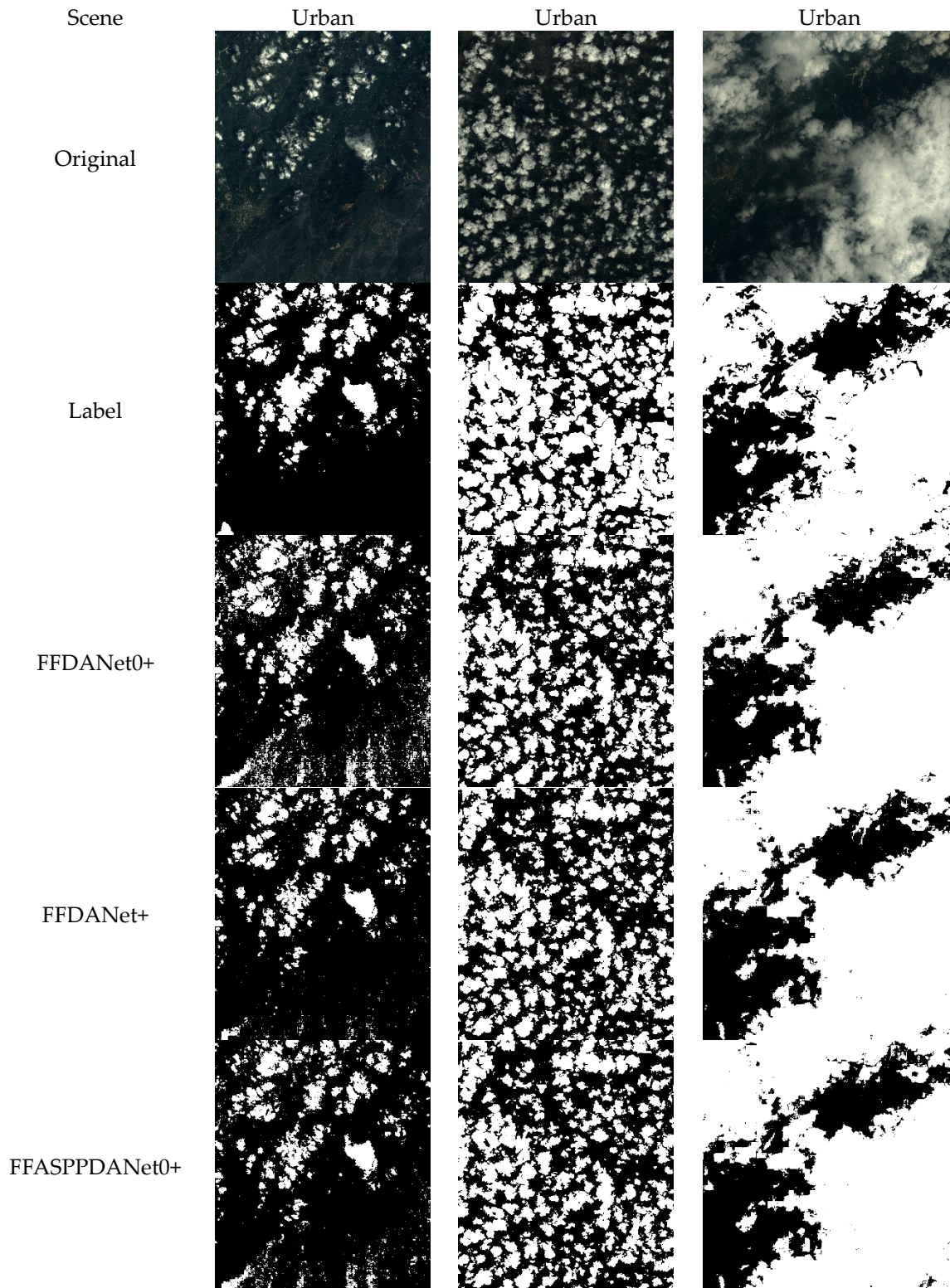


Figure 13. Cont.



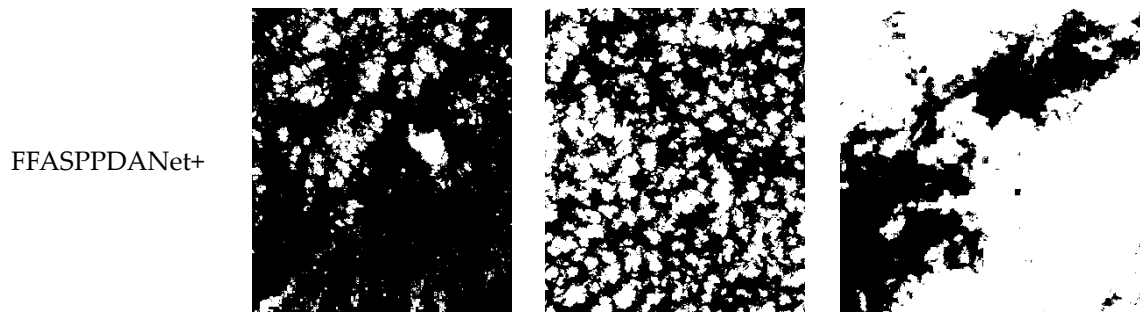


Figure 13. Cloud Segmentation Results of Different Neural Networks in Urban Scenarios.

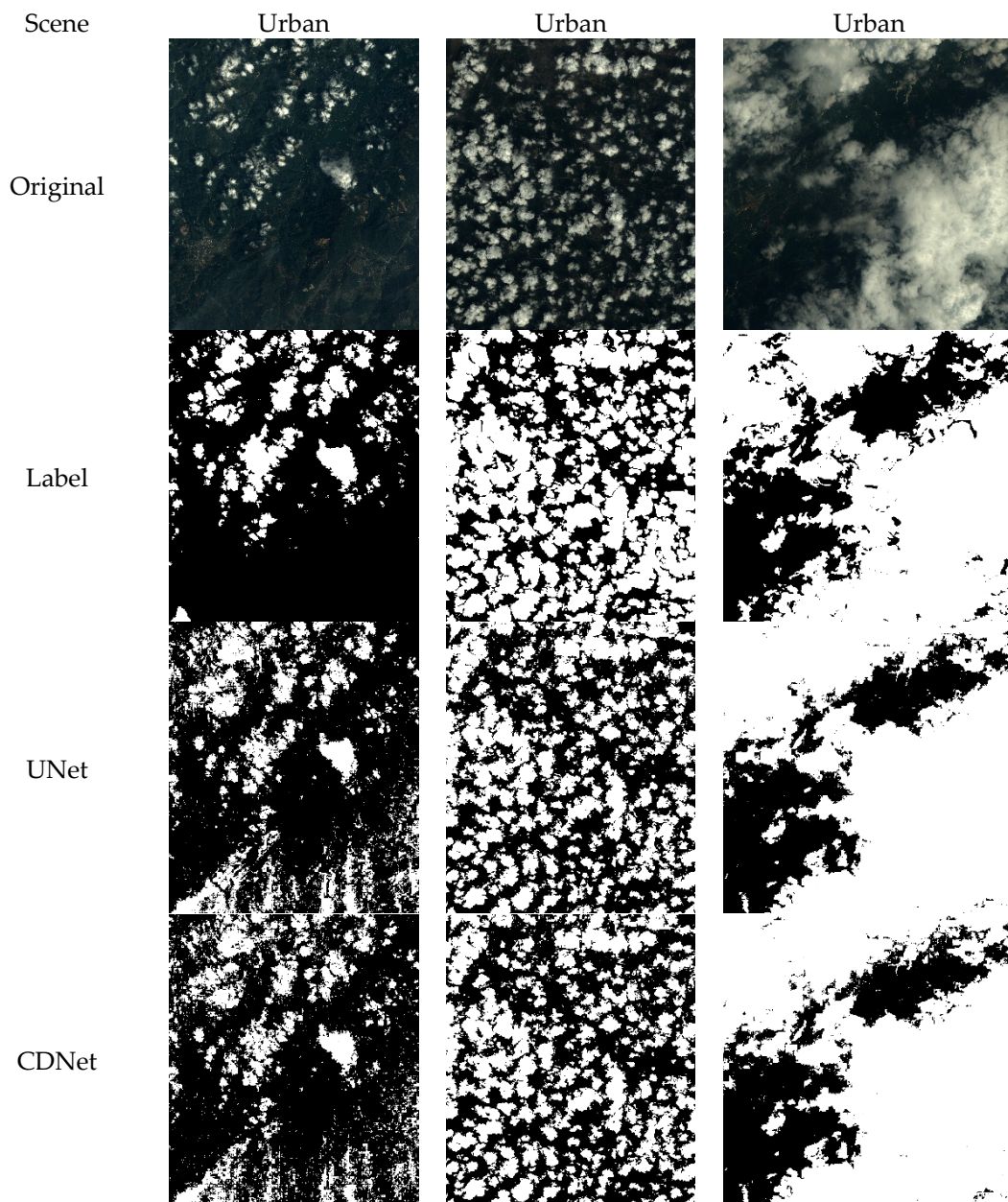
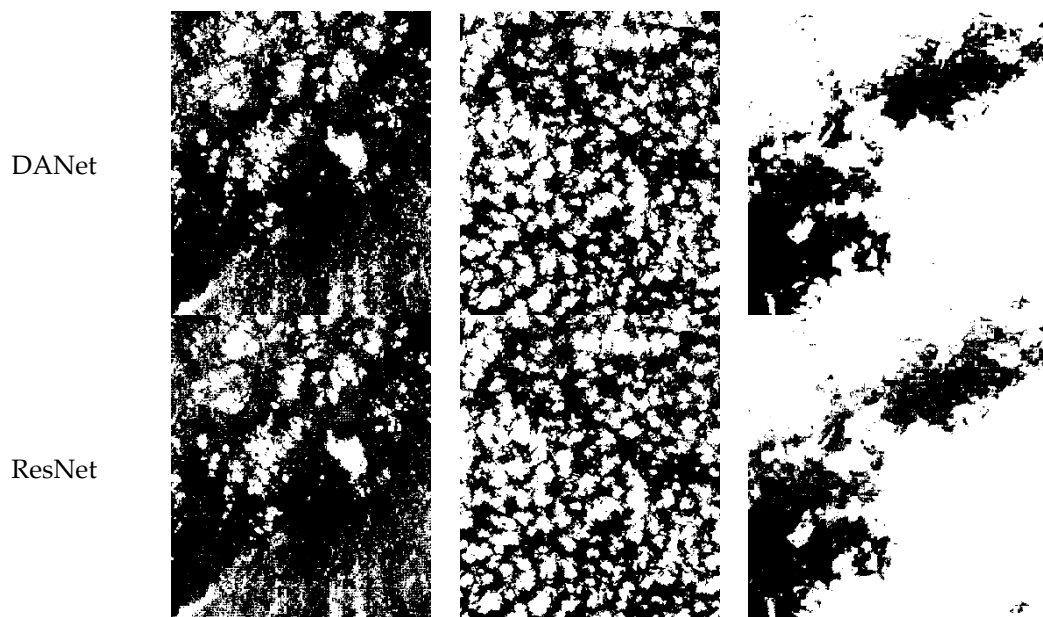


Figure 14. Cont.



**Figure 14.** Cloud Segmentation Results of Different Neural Networks in Urban Scenarios (comparison results with other state-of-the-art methods).

The results indicate that in urban scenarios, FFASPPDANet0+ performs generally well, being able to identify the main body of clouds but with some instances of misidentification. FFDANet+ and FFASPPDANet+ exhibit the best performance, with an IoU of around 74%, OA of around 89%, and F1 scores of around 85% (Table 6). With precision exceeding 96%, they are able to segment clouds effectively, producing clear boundary textures.

**Table 6.** Cloud Segmentation Results of Different Models in Urban Scenarios.

Model	IoU	OA	Precision	F1	Recall
FFDANet0+	0.7184	0.8752	0.8924	0.8361	0.7866
FFDANet+	0.7383	0.8917	0.9723	0.8494	0.7541
FFASPPDANet0+	0.7254	0.8855	0.9623	0.8408	0.7466
FFASPPDANet+	0.7304	0.8881	0.9679	0.8442	0.7485
UNet	0.6665	0.8423	0.8226	0.7999	0.7784
CDNet	0.6901	0.8574	0.8518	0.8166	0.7842
DANet	0.7124	0.8673	0.8534	0.8320	0.8117
ResNet	0.7026	0.8615	0.8436	0.8253	0.8078

In comparison, the UNet network performs the worst, with all evaluation metrics showing deficiencies. The Intersection over Union (IoU) is only 66.65%, the Overall Accuracy (OA) is 84.23%, and the F1 score is 79.99%, with significant missed and false detections. CDNet, DANet, and ResNet exhibit moderate performance in urban scenes, with IoU around 71% and OA around 85%. These networks generally identify the main parts of the clouds but are prone to false detections.

### 3.3. “FFASPPDANet+” Algorithm Experiment on a Random Test Set

To explore the cloud segmentation capabilities of each algorithm on real remote sensing images, during the final five epochs of the training process, the algorithm was used to segment clouds in remote sensing satellite images from the test set. The performance metrics of each algorithm were calculated, and the best-performing algorithm model was retained as the final training outcome. The cloud segmentation capabilities on the test set are shown in the Table 7 below:

**Table 7.** Cloud Segmentation Results of Different Models in Random Test Set.

Model	IoU	OA	Precision	F1	Recall
FFDANet0+	0.9280	0.9732	0.9569	0.9627	0.9686
FFDANet+	0.9291	0.9733	0.9483	0.9633	0.9787
FFASPPDANet0+	0.9278	0.9733	0.9629	0.9626	0.9622
FFASPPDANet+	0.9308	0.9742	0.9582	0.9642	0.9702
UNet	0.9061	0.9325	0.9386	0.9316	0.9346
CDNet	0.9084	0.9437	0.9300	0.9228	0.9258
DANet	0.9156	0.9682	0.9448	0.9359	0.9173
ResNet	0.9060	0.9421	0.9469	0.9216	0.9467

It can be observed that FFASPPDANet0+ and FFASPPDANet+ overall perform well, with an IoU of around 80%, an OA of around 95%, and a precision of around 96%. They are capable of effectively segmenting clouds with richer boundary texture information, and their comprehensive performance is significantly superior to other cloud segmentation methods. Although FFASPPDANet0+ has a slightly higher precision than FFASPPDANet+, its other evaluation metrics are not as good as those of FFASPPDANet+. FFDANet0+ and FFDANet+ are the next best, with a precision of around 95%, able to effectively segment thick and thin clouds, rich in boundary texture information, and also capable of segmenting small fragmented clouds well. It can be observed from the results in the test set that the method proposed in this paper shows improvement across all metrics compared to the four methods mentioned.

### 3.4. Performance in State-of-the-Art Datasets

Table 8 and Figure 15 present the performance of our model compared to other methods over the 38-Cloud dataset [42,43]. A comprehensive analysis of the cloud segmentation results of different models on the 38-Cloud dataset reveals that FFASPPDANet+ excels in several key metrics in the cloud segmentation task. According to the average parameter results of 20 test images of the 38-Cloud dataset (Table 8), FFASPPDANet+ achieves an IoU of 0.9109, an OA of 0.9651, a precision of 0.9540, and an F1 score of 0.8946. These metrics are the highest among all models, indicating the model's relatively superior capability in accurately identifying and detecting cloud regions. Although FFASPPDANet+ was not optimal in the results of the random test set (Table 7), its performance in 38-Cloud (Table 8) indicates the great generalization ability of the model. FFDANet+ and FFASPPDANet0+ also perform robustly, particularly in recall and OA, demonstrating high accuracy and reliability, though they slightly underperform in other metrics. Although ResNet, a widely used segmentation model, shows competitive results in IoU, recall, and precision, its performance in terms of F1 score is relatively lower, suggesting that our proposed models incorporating radiometric feature fusion modules hold an overall performance advantage.

**Table 8.** Cloud Segmentation Results of Different Models in 38-Cloud dataset.

Model	IoU	OA	Precision	F1	Recall
FFDANet0+	0.8431	0.9112	0.9065	0.8708	0.9558
FFDANet+	0.8554	0.9224	0.8920	0.8501	0.9791
FFASPPDANet0+	0.8912	0.9441	0.9336	0.8829	0.9735
FFASPPDANet+	0.9109	0.9651	0.9540	0.8946	0.9763
UNet	0.8455	0.8554	0.8208	0.7274	0.8867
CDNet	0.7739	0.8759	0.8428	0.8658	0.9721
DANet	0.7855	0.8927	0.8662	0.8472	0.9315
ResNet	0.8485	0.9009	0.9212	0.8372	0.9744

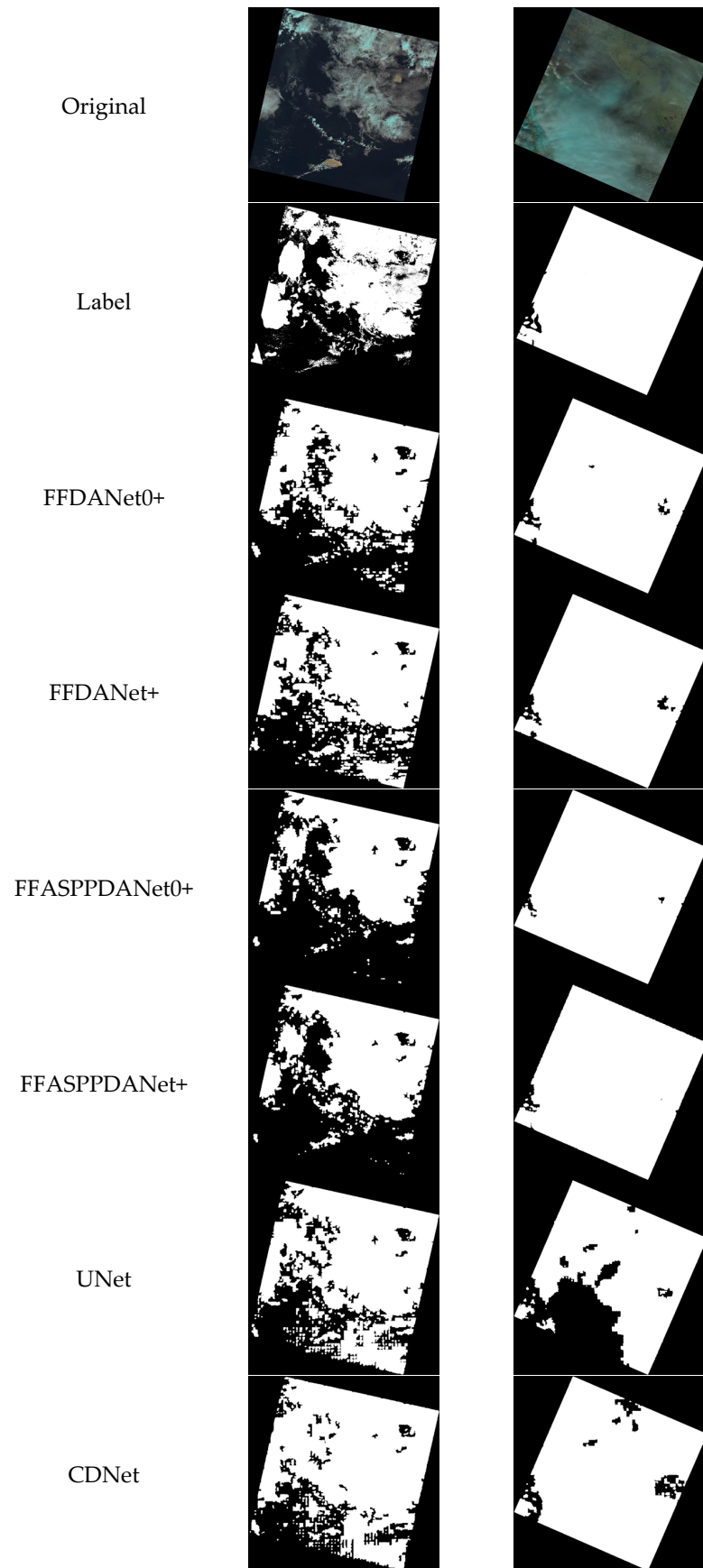
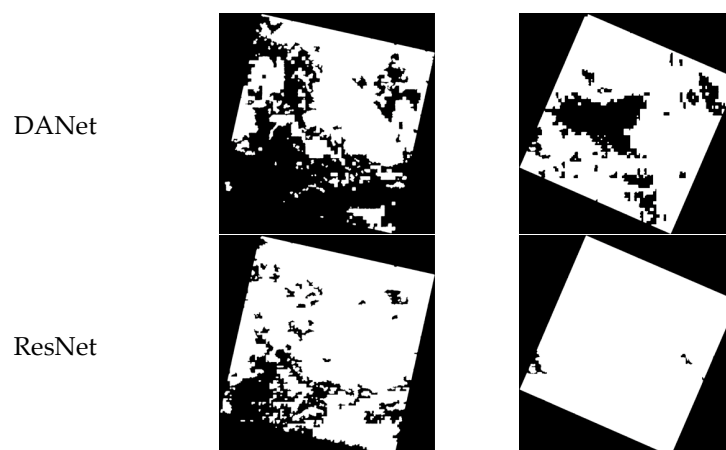


Figure 15. Cont.



**Figure 15.** Cloud Segmentation Results of Different Neural Networks in 38-Cloud dataset (comparison results with other state-of-the-art methods).

It can be observed in Figure 15 that the outcomes of FFASPPDANet0+ and FFASPPDANet+ with the addition of the ASPP mechanism better capture the morphological characteristics of clouds compared to FFDANet0+ and FFDANet+. The boundaries between cloud and non-cloud regions are further clarified, and misclassification is reduced, corresponding to the increases in OA, precision, and F1 score (Table 8). These results indicate that our segmentation model, improved through the use of ASPP and dual attention mechanisms, has made progress in the cloud segmentation task, and this advancement is also evident on the 38-Cloud dataset outside the training set.

#### 4. Discussion

In this study, we propose an innovative remote sensing satellite image cloud segmentation algorithm, FFASPPDANet+, which extensively tests various scenes of different complexities (such as water bodies and urban areas) as well as random test sets through the fusion of dual attention networks, atrous spatial pyramid pooling convolutions, and a probabilistic threshold method.

The integration of the dual attention network with atrous spatial pyramid pooling convolutions and the application of the probabilistic threshold method are the main innovations. FFASPPDANet+ is the first to combine the dual attention network with atrous spatial pyramid pooling technology, effectively enhancing the accuracy and efficiency of cloud segmentation in remote sensing images. This innovative combination not only improves the model's sensitivity and differentiation capability toward cloud features but also expands the model's receptive field, enhancing the model's ability to capture the widespread distribution characteristics of clouds. Moreover, by introducing the probabilistic threshold method, the algorithm further enhances the accuracy and robustness of cloud segmentation based on precise identification of cloud and non-cloud areas. This approach introduces greater flexibility to the cloud segmentation task, allowing the algorithm to better adapt to different remote sensing image conditions.

#### 5. Conclusions

The breakthrough progress of this paper is demonstrated by its efficient handling of different scenes, as evidenced by its outstanding performance on a random test set. In tests of scenes with varying complexities, FFASPPDANet+ showcased its superior performance. Particularly in water body scenes, the algorithm achieved high precision in cloud segmentation, with a precision rate exceeding 98%. In the complex urban scenes, the algorithm also demonstrated excellent cloud segmentation capability, accurately identifying cloud edges and small fragmented clouds, thus showing the algorithm's wide applicability and efficiency.

The experimental results on the random test set further validate the superiority of the FFASPPDANet+ algorithm. The IoU of the algorithm on the random test set is around 93%, with an overall accuracy (OA) of about 97% and a precision of about 96%; these results fully demonstrate the powerful capability and efficiency of FFASPPDANet+ in the task of cloud segmentation in real remote sensing images.

In summary, the innovation and breakthrough of the FFASPPDANet+ algorithm are manifested in its combination of advanced deep learning techniques and the probabilistic threshold method, thus not only improving the accuracy and efficiency of cloud segmentation but also demonstrating strong adaptability and robustness in handling different complex scenes and real-world conditions. These achievements provide an effective new method for cloud segmentation in remote sensing images, as well as new ideas and directions for future research in related fields.

**Author Contributions:** Conceptualization, J.Z. and M.H.; methodology, J.Z.; software, J.Z.; validation, J.Z. and M.H.; formal analysis, J.Z.; investigation, J.Z.; resources, J.Z.; data curation, M.H.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z.; visualization, J.Z.; supervision, M.H.; project administration, M.H.; funding acquisition, M.H. M.H. and J.Z. are co-first authors. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The SPOT-6 and SPOT-7 data used are proprietary and cannot be fully disclosed for public download at this time. If necessary, a download link can be provided after communication with the authors. The 38-Cloud dataset is freely distributed and can be downloaded at <https://www.kaggle.com/datasets/sorour/38cloud-cloud-segmentation-in-satellite-images> (accessed on 20 May 2024).

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

## References

1. Li, C.; Ma, J.; Yang, P.; Li, Z. Detection of cloud cover using dynamic thresholds and radiative transfer models from the polarization satellite image. *J. Quant. Spectrosc. Radiat. Transf.* **2019**, *222–223*, 196–214. [CrossRef]
2. Zhou, K.; Zheng, Y.; Dong, W.; Wang, T.; Technology, O. A deep learning network for cloud-to-ground lightning nowcasting with multisource data. *J. Atmos. Ocean. Technol.* **2020**, *37*, 927–942. [CrossRef]
3. Panfeng, W.; Jian, A.; Baolin, W.; Hongyang, B.; Yunsen, W.; Zhiyuan, L. A Method of Cloud Detection in Remote Sensing Image Based on FPGA. In Proceedings of the 7th International Symposium of Space Optical Instruments and Applications, Beijing, China, 21–23 October 2022; pp. 92–104.
4. Qian, J.; Ci, J.; Tan, H.; Xu, W.; Jiao, Y.; Chen, P. Cloud Detection Method Based on Improved DeeplabV3+ Remote Sensing Image. *IEEE Access* **2024**, *12*, 9229–9242. [CrossRef]
5. 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]
6. Liu, X.; Sun, L.; Yang, Y.; Zhou, X.; Wang, Q.; Chen, T. Cloud and cloud shadow detection algorithm for gaofen-4 satellite data. *Acta Opt. Sin.* **2019**, *39*, 446–457.
7. Barnes, W.L.; Salomonson, V.V. MODIS: A global imaging spectroradiometer for the Earth Observing System. In Proceedings of the Optical Technologies for Aerospace Sensing: A Critical Review, Boston, MA, USA, 16–17 November 1992; pp. 280–302.
8. Nightingale, J.; Nickeson, J.; Justice, C.; Baret, F.; Garrigues, S.; Wolfe, R.; Masuoka, E. Global validation of EOS land products, lessons learned and future challenges: A MODIS case study. In Proceedings of the 33rd International Symposium on Remote Sensing of Environment: Sustaining the Millennium Development Goals, Stresa, Italy, 4–8 May 2008; p. 4. Available online: [https://landval.gsfc.nasa.gov/pdf/ISRSE\\_Nightingale.pdf](https://landval.gsfc.nasa.gov/pdf/ISRSE_Nightingale.pdf) (accessed on 15 April 2024).
9. Xiang, P.S. A Cloud Detection Algorithm for MODIS Images Combining Kmeans Clustering and Otsu Method. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *392*, 062199. [CrossRef]
10. Xie, W.; Bai, K.; Li, Y.; Lei, J.I.E.; Yang, J. Multispectral Cloud Detection Method Based on Semi-Supervision of Spatial and Spectral Features. CN 110567886 A, 10 September 2019.
11. Le Hégarat-Masclé, S.; Kallel, A.; Descombes, X. Ant colony optimization for image regularization based on a nonstationary Markov modeling. *Trans. Image Process.* **2007**, *16*, 865–878. [CrossRef] [PubMed]
12. Rossi, R.; Basili, R.; Del Frate, F.; Luciani, M.; Mesiano, F. Techniques based on support vector machines for cloud detection on quickbird satellite imagery. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 515–518.

13. Li, P.; Dong, L.; Xiao, H.; Xu, M.J.N. A cloud image detection method based on SVM vector machine. *Neurocomputing* **2015**, *169*, 34–42. [[CrossRef](#)]
14. Başeski, E.; Cenasar, Ç. Texture and color based cloud detection. In Proceedings of the 2015 7th International Conference on Recent Advances in Space Technologies (RAST), İstanbul, Türkiye, 16–19 June 2015; pp. 311–315.
15. Sui, Y.; He, B.; Fu, T. Energy-based cloud detection in multispectral images based on the SVM technique. *Int. J. Remote Sens.* **2019**, *40*, 5530–5543. [[CrossRef](#)]
16. 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](#)]
17. Xu, L.; Wong, A.; Clausi, D.A.; Sensing, R. A novel Bayesian spatial–temporal random field model applied to cloud detection from remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4913–4924. [[CrossRef](#)]
18. Zhang, Q.; Xiao, C.; Sensing, R. Cloud detection of RGB color aerial photographs by progressive refinement scheme. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7264–7275. [[CrossRef](#)]
19. McKay, J.; Monga, V.; Raj, R. Localized dictionary design for geometrically robust sonar ATR. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 991–994.
20. Luo, J.; Pan, Y.; Su, D.; Zhong, J.; Wu, L.; Zhao, W.; Hu, X.; Qi, Z.; Lu, D.; Wang, Y. Innovative Cloud Quantification: Deep Learning Classification and Finite Element Clustering for Ground-Based All Sky Imaging. *Preprints* **2023**. [[CrossRef](#)]
21. Chen, Q.; Yin, X.; Li, Y.; Zheng, P.; Chen, M.; Xu, Q. Recognition of Severe Convective Cloud Based on the Cloud Image Prediction Sequence from FY-4A. *Remote Sens.* **2023**, *15*, 4612. [[CrossRef](#)]
22. 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](#)]
23. Li, T.; Wu, D.; Wang, L.; Yu, X. Recognition algorithm for deep convective clouds based on FY4A. *Neural Comput. Appl.* **2022**, *34*, 21067–21088. [[CrossRef](#)]
24. Tian, Y.; Pang, S.; Qu, Y. Fusion Cloud Detection of Multiple Network Models Based on Hard Voting Strategy. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 6646–6649.
25. Xiao, C.; Sun, J. Deep Neural Networks (DNN). In *Introduction to Deep Learning for Healthcare*; Xiao, C., Sun, J., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 41–61.
26. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
27. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
28. Falk, T.; Mai, D.; Bensch, R.; Çiçek, Ö.; Abdulkadir, A.; Marrakchi, Y.; Böhm, A.; Deubner, J.; Jäckel, Z.; Seiwald, K.; et al. U-Net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* **2019**, *16*, 67–70. [[CrossRef](#)]
29. Ronneberger, O. *Invited Talk: U-Net Convolutional Networks for Biomedical Image Segmentation*; Springer: Berlin/Heidelberg, Germany, 2017; p. 3.
30. 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.
31. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**. [[CrossRef](#)]
32. Váša, F.; Bullmore, E.T.; Patel, A.X. Probabilistic thresholding of functional connectomes: Application to schizophrenia. *NeuroImage* **2018**, *172*, 326–340. [[CrossRef](#)]
33. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
34. Ansel, J.; Yang, E.; He, H.; Gimelshein, N.; Jain, A.; Voznesensky, M.; Bao, B.; Bell, P.; Berard, D.; Burovski, E. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS’24). Association for Computing Machinery, New York, NY, USA, 24 April–1 May 2024; pp. 317–335.
35. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**. [[CrossRef](#)]
36. Gridin, I. *Automated Deep Learning Using Neural Network Intelligence: Develop and Design PyTorch and TensorFlow Models Using Python*; Springer: Berlin/Heidelberg, Germany, 2022.
37. Ronneberger, O.; Fischer, P.; Brox, T.J.A. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**. [[CrossRef](#)]
38. 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](#)]
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

40. Otto, G.P.; Chau, T.K. 'Region-growing' algorithm for matching of terrain images. *Image Vis. Comput.* **1989**, *7*, 83–94. [[CrossRef](#)]
41. Tang, J. A color image segmentation algorithm based on region growing. In Proceedings of the 2010 2nd International Conference on Computer Engineering and Technology, Chengdu, China, 16–18 April 2010; pp. V6-634–V636-637.
42. Mohajerani, S.; Krammer, T.A.; Saeedi, P. A Cloud Detection Algorithm for Remote Sensing Images Using Fully Convolutional Neural Networks. In Proceedings of the 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), Vancouver, BC, Canada, 29–31 August 2018; pp. 1–5.
43. 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.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.