Review

Effect of Attention Mechanism in Deep Learning-Based Remote Sensing Image Processing: A Systematic Literature Review

Saman Ghaffarian [1,2,*] **, João Valente** [1] **, Mariska van der Voort** [2] **and Bedir Tekinerdogan** [1]

1   Information Technology Group, Wageningen University & Research, 6707 KN Wageningen, The Netherlands; joao.valente@wur.nl (J.V.); bedir.tekinerdogan@wur.nl (B.T.)
2   Business Economics Group, Wageningen University & Research, 6700 EW Wageningen, The Netherlands; mariska.vandervoort@wur.nl
*   Correspondence: saman.ghaffarian@wur.nl

Abstract: Machine learning, particularly deep learning (DL), has become a central and state-of-the-art method for several computer vision applications and remote sensing (RS) image processing. Researchers are continually trying to improve the performance of the DL methods by developing new architectural designs of the networks and/or developing new techniques, such as attention mechanisms. Since the attention mechanism has been proposed, regardless of its type, it has been increasingly used for diverse RS applications to improve the performances of the existing DL methods. However, these methods are scattered over different studies impeding the selection and application of the feasible approaches. This study provides an overview of the developed attention mechanisms and how to integrate them with different deep learning neural network architectures. In addition, it aims to investigate the effect of the attention mechanism on deep learning-based RS image processing. We identified and analyzed the advances in the corresponding attention mechanism-based deep learning (At-DL) methods. A systematic literature review was performed to identify the trends in publications, publishers, improved DL methods, data types used, attention types used, overall accuracies achieved using At-DL methods, and extracted the current research directions, weaknesses, and open problems to provide insights and recommendations for future studies. For this, five main research questions were formulated to extract the required data and information from the literature. Furthermore, we categorized the papers regarding the addressed RS image processing tasks (e.g., image classification, object detection, and change detection) and discussed the results within each group. In total, 270 papers were retrieved, of which 176 papers were selected according to the defined exclusion criteria for further analysis and detailed review. The results reveal that most of the papers reported an increase in overall accuracy when using the attention mechanism within the DL methods for image classification, image segmentation, change detection, and object detection using remote sensing images.

1. Introduction

Remotely sensed images have been employed as the main data sources in many fields such as agriculture [1–4], urban planning [5–7] and disaster risk management [8–10], and have been shown as an effective and critical tool to provide information. Accordingly, processing remote sensing (RS) images is crucial to extract the useful information from them for such applications. RS image processing tasks include image classification, object detection, change detection, and image fusion [11]. Different processing methods were developed to address them, and they aimed to improve the performance and accuracy of the methods to address RS image processing. Machine learning methods such as support vector machines and ensemble classifiers (e.g., random forest and gradient boosting)

obtained fairly high accuracies for different RS processing tasks [12,13]. In particular, deep learning (DL) methods have recently become state-of-the-art methods in RS image processing and automatically extracting the required information from RS images [14,15]. Since DL has entered this field, researchers try to improve the performance and increase its accuracy by developing new techniques and different architectural designs, e.g., various convolutional neural networks (CNN) [16,17], generative adversarial networks (GAN) [18], graph neural networks (GNN) [19]. Recently, the attention mechanism was proposed by Bahdanau, et al. [20] initially for machine translation application, which aims to guide deep neural network methods by providing focus points and highlighting the important features while minimizing the others. Thereafter, it was used in different applications, including computer vision [21] and RS image processing [22–24]. Accordingly, most of the studies reported an increase in the performance of the DL methods when guided with attention mechanism [25–27].

In recent years, researchers reviewed the developed/used DL methods in RS literature mostly from a general perspective [11,28] or focusing on one application, e.g., image classification [15]. Zhang, et al. [14] reviewed the DL methods in RS big data processing and provided a technical tutorial on the state-of-the-art methods. Zhu, et al. [28] reviewed the DL methods applied to RS data analysis and investigated the challenges of DL in RS applications. They also provided a comprehensive list of resources for DL-RS data analysis. Li, et al. [15] conducted a survey study on the developed DL methods for RS image classification. They also analyzed and compared the performance of the different DL methods. In addition, the recent advances in DL for pixel-level image fusion were reviewed by Li, et al. [29]. Ma, et al. [11] conducted a systematic literature review on applications of the DL on RS and they comprehensively reviewed and categorized DL methods. In addition, Niu, et al. [21] reviewed the different architectural designs of the attention mechanism used in conjunction with DL from a general perspective and provided some application domains. However, the effect of such a mechanism for DL methods in RS image processing has not yet been reviewed and investigated. Accordingly, a systematic literature review is conducted in this study by following a structured review on the DL methods with an embedded attention mechanism for RS image processing applications. Thus, the literature is reviewed systematically to respond to the predefined research questions rather than summarizing the papers. The main objective of this study is to extract the effect of attention mechanism in the performance of deep learning-based RS (DL-RS) image processing. In addition, the current trends, achievements and applications in publications, using attention mechanism-based DL (At-DL) methods and RS image processing applications are extracted to provide insights and guidelines for future studies.

The rest of the paper is organized as follows. Background information regarding the attention mechanism, its different types, and how it is being used in DL methods are provided in Section 2. Section 3 presents and describes the integration of attention mechanisms with different deep neural network architectures to address RS image processing tasks. The steps of the executed systematic literature review are explained in Section 4. Then, Section 5 presents and visualizes the quantitative results and discusses them according to the defined research questions, and reveals the effect of attention mechanism in the performance of the DL-RS image processing. Finally, Section 6 concludes the paper.

2. Attention Mechanism in Deep Learning

The attention mechanism, like other neural network-based methods, tries to mimic the human brain/vision to process data. Human vision does not process the entire image at once; however, it only focuses on the specific parts. With this, the focused parts of the human view space are perceived in "high-resolution" while the surroundings are in "low-resolution". In other words, it gives higher weight to the relevant parts while minimizing the irrelevant ones, giving them lower weights. This allows the brain to process and focus on the most important parts precisely and efficiently, rather than processing the entire view space. This characteristic of human vision inspired researchers to develop the

attention mechanism. It was initially developed in 2014 for natural language processing applications [20], since then it has been widely used for different applications [30], in particular, computer vision tasks [21,31]. Its potential to enhance mostly CNN-based methods has been reported [32]. In addition, it has been used in conjunction with recurrent neural network models [33–36], and graph neural networks [37,38]. The main idea behind the attention mechanism is to give different weights to different information. Thus, giving higher weights to relevant information attracts the attention of the DL model to them [39]. Attention mechanism approaches can be grouped based on four criteria (Figure 1) [21]:
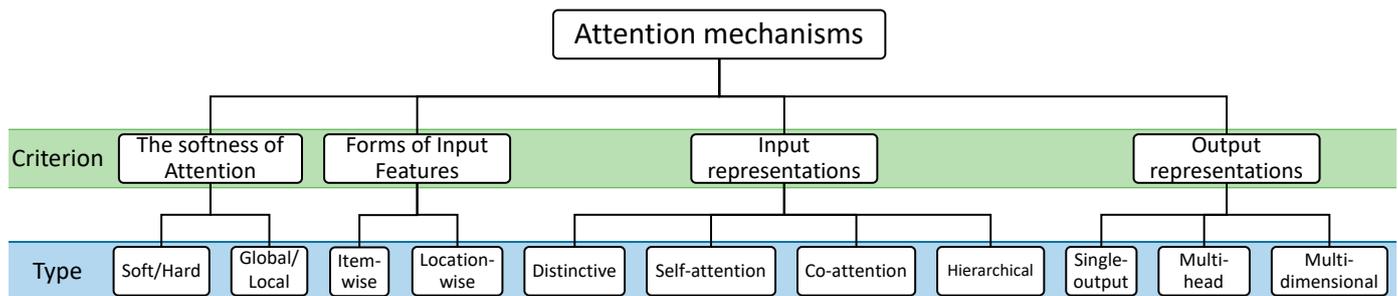

Figure 1. An overview of typical attention mechanism approaches [21].

(i) The softness of attention: the initial attention mechanism proposed by [20] is a soft version, which is also known as deterministic attention. This network considers all input elements (computes the average for each weight) to compute the final context vector. The context vector is the high-dimensional vector representation of the input elements or sequences of the input elements and in general the attention mechanism aims to add more contextual information to compute the final context vector. However, hard attention, which is also known as stochastic attention, randomly selects from the sample elements to compute the final context vector [40]. This, therefore, reduces the computational time. Furthermore, there is another categorization that is frequently used in computer vision tasks and RS image processing, i.e., global and local attentions [41,42]. Global attention is similar to soft attention since it also considers all input elements. However, global attention simplifies soft attention by using the output of the current time step rather than the prior one, while local attention is a combination of soft and hard attentions. This approach considers a subset of input elements at a time, and thus, overcomes the limitation of hard attention, i.e., being nondifferentiable, and in the meantime is less computationally expensive.

(ii) Forms of input features: attention mechanisms can be grouped based on their input requirements: item-wise and location-wise. Item-wise attention requires inputs that are known to the model explicitly or produced with a preprocess [43–45]. However, location-wise attention does not necessarily require known inputs, in this case, the model needs to deal with input items that are difficult to distinguish. Due to the characteristics and features of the RS images and targeted tasks, location-wise attention is commonly used for RS image processing [42,46–48].

(iii) Input representations: there are single-input and multi-input attention models [49,50]. In addition, the general processing procedure of the inputs also varies between the developed models. Most of the current attention networks work with single-input, and the model processes them in two independent sequences (i.e., distinctive model). The co-attention model is a multi-input attention network that parallelly implements the attention mechanism on two different sources but finally merges them [50]. This makes it suitable for change detection from RS images [51]. A self-attention network computes attentions only based on the model inputs, and thus, it decreases the dependence on external information [52–54]. This allows the model to perform better in images with complex background by focusing more on targeted areas [55]. Hierarchical attention mechanism computes weights from the original input and

different levels/scales of the inputs [56]. This attention mechanism is also known as fine-grained attention for image classification [57].

(iv) Output representations: single-output is the commonly used output representation in attention mechanisms. It processes a single feature at a time and computes weight scores. There are also two other multidimensional and multi-head attention mechanisms [21]. Multi-head attention processes the inputs linearly in multiple subsets, and finally merges them to compute the final attention weights [58], and is especially useful when employing the attention mechanism in conjunction with CNN methods [59–61]. Multidimensional attention, which is mostly employed for natural language processing, computes weights based on matrix representation of the features instead of vectors [62,63].

The above-explained attention mechanisms are the same in principle and are developed by researchers to adopt or improve the basic attention mechanism for their tasks. In addition, not all of them have been used for computer vision, and thus, RS image processing. In DL-based image processing, this mechanism is usually used to focus on specific features (feature layers) or a certain location or aspect of an image [64–67]. Accordingly, it can be classified into two major types: channel and spatial attentions.

Figure 2 illustrates simple channel and spatial attention types: (a) The channel attention network aims to boost the feature layers (channel) in the feature map that convey more important information and silence the other feature layers (channels); (b) the spatial attention network highlights regions of interest in the feature space and covers up the background regions. These two attention mechanisms can be used solely or combined within DL methods to provide attention to both important feature layers and the location of the region of interest. Papers in this review were classified according to these two types.
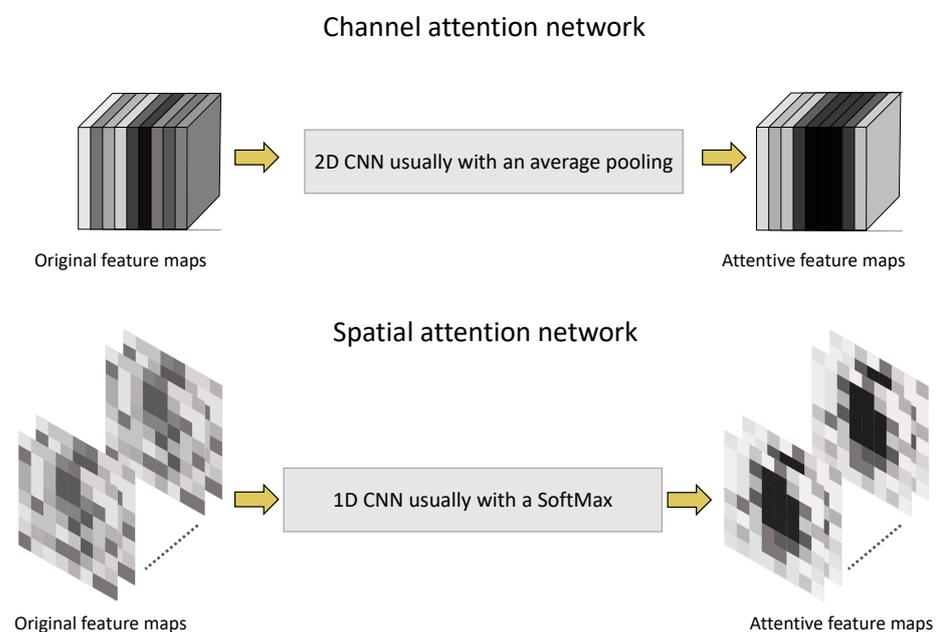

Figure 2. A simple illustration of the channel and spatial attention types/networks, and their effects on the feature maps.

3. Deep Neural Network Architectures with Attention for RS Image Processing

In this section, we describe and provide examples of the four different deep neural network architectures (i.e., CNN, GAN, RNN, and GNN) that are improved using the attention mechanism to address RS image processing. CNN is the main method that has been used for image processing in general, as well as RS applications. Both spatial and channel attentions are embedded in CNN with different attention network designs. For CNNs the channel attention is typically implemented after each convolution but the spatial

attention is mostly added to the end of the network [68–71]. However, in UNet-based networks, spatial attention is usually added to each layer of a decoding/upsampling section [72–74]. Figure 3 shows an example of using spatial and channel attentions, in particular co-attention network, in a Siamese model for building-based change detection [51]. The proposed co-attention network is based on an initial correlation process with a final attention module. For GAN networks which are based on encoding and decoding modules, the process of adding attention networks is the same as of CNNs that can be used in both adversarial and/or discrimination networks depending on the targeted tasks [75] (Figure 4).
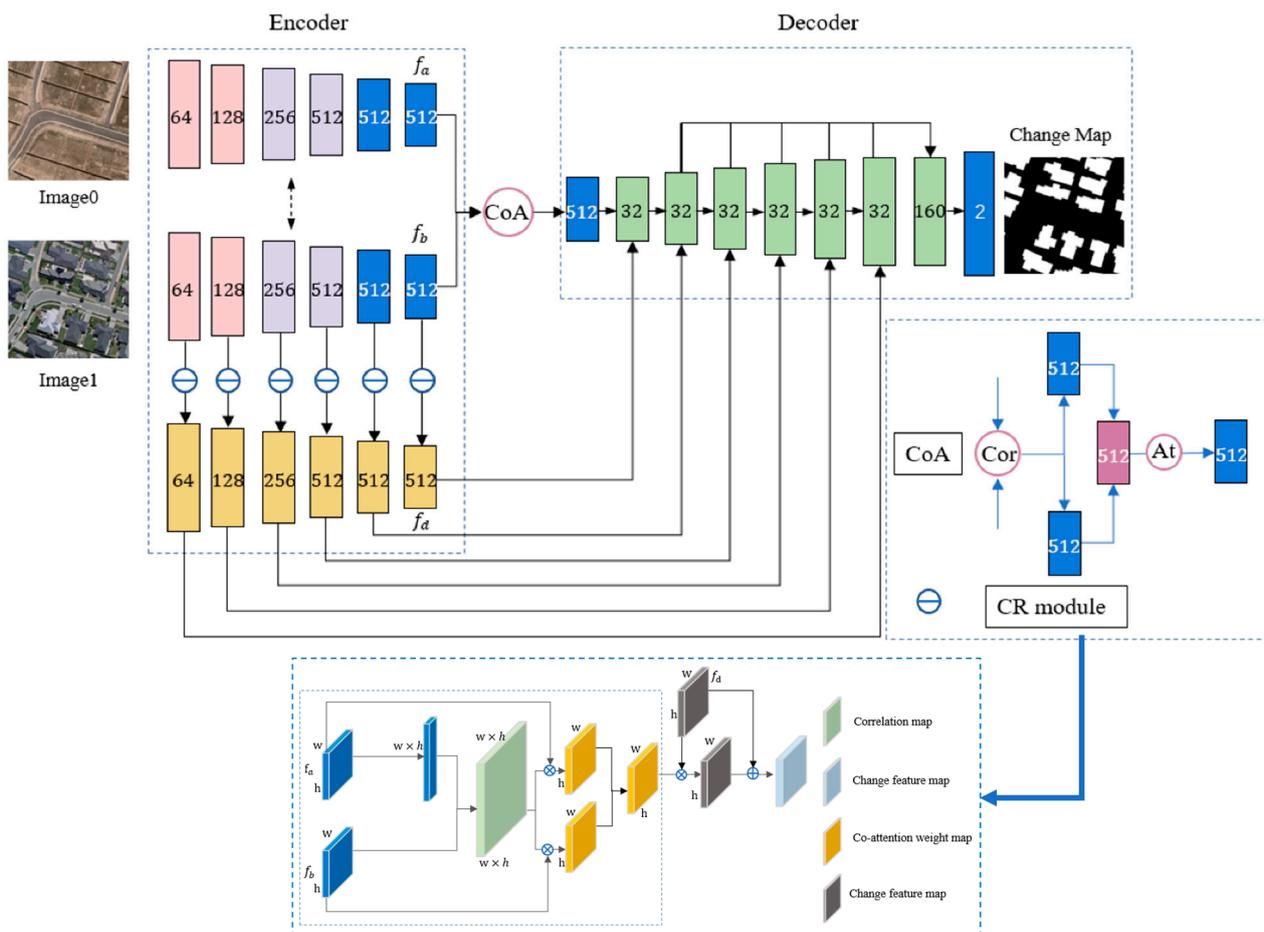

Figure 3. An example of adding attention network (i.e., co-attention) to a CNN module (i.e., Siamese network) for building-based change detection [51]. CoA—co-attention module, At—attention network, CR—change residual module.

RNN is the first deep learning network that is improved by attention mechanism [20] for natural language processing tasks. RNNs are not as popular as CNNs for image processing due to the inherent characteristics of the images. However, RNN has been frequently used in conjunction with CNN for RS image processing [34,76–78]. This also allows the integration of the attention mechanism with RNN for RS applications. For example, Ref. [79] developed a bidirectional RNN module to provide channel attention and add the outcome weights to the CNN-based module which is supported with a spatial attention network for hyperspectral image classification (Figure 5).
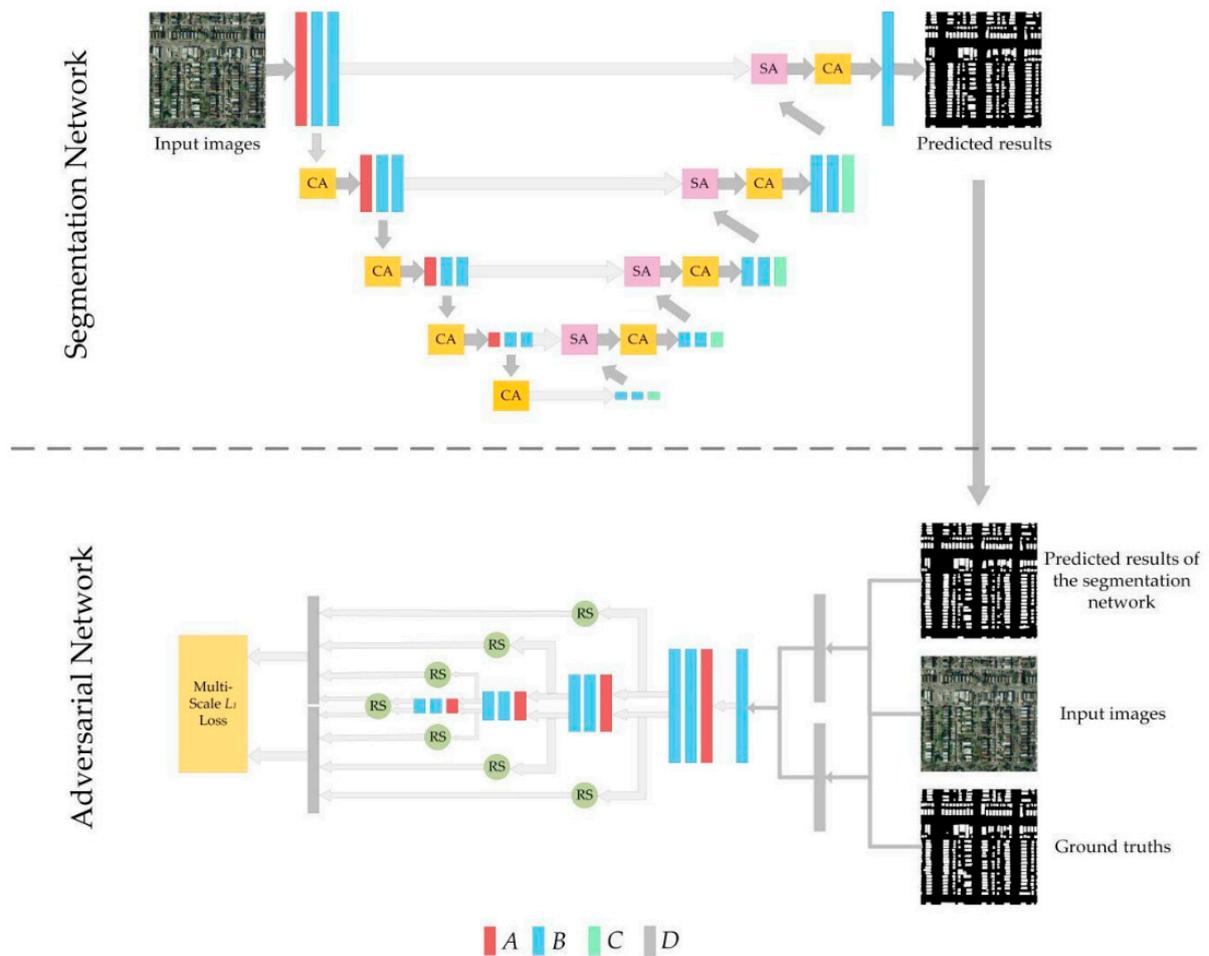
Figure 4. An example of adding spatial and channel attentions to a GAN module for building detection from aerial images [75]. A—max pooling layer; B—convolutional + batch normalization + rectified linear unit (ReLU) layers; C—upsampling layer; D—concatenation operation; SA—spatial attention mechanism; CA—channel attention mechanism; RS—reshape operation.
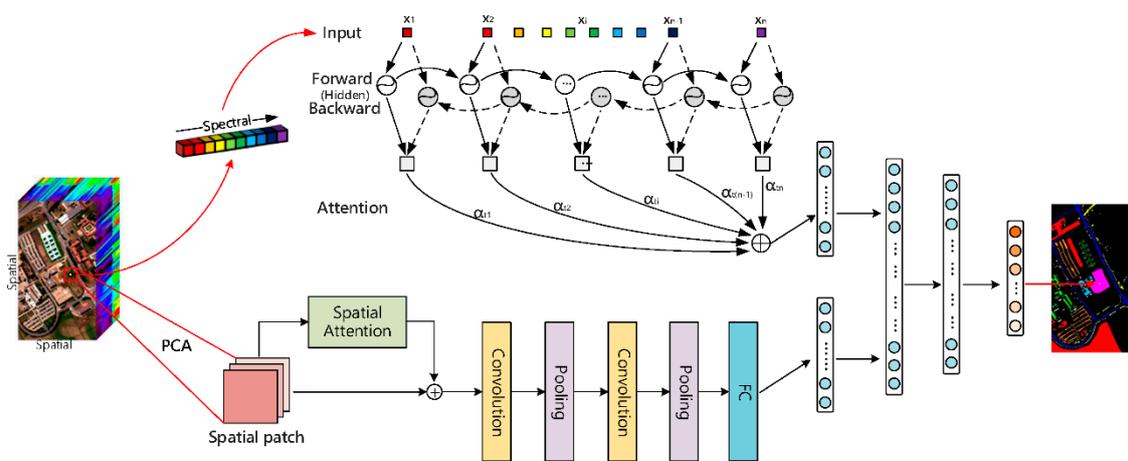

Figure 5. An example of adding attention networks (i.e., spatial and channel attentions) to a RNN + CNN module for hyperspectral image classification [79]. PCA—principal component analysis.

GNN is another network architecture that has been employed in conjunction with CNN for RS image processing. Hence, this mechanism is used to focus on the most

important graph nodes of the network. A typical integration of GNN with CNN is to implement a GNN after a CNN-based image segmentation to produce the final RS image classification results [80,81]. Accordingly, the attention network adjusts the weight for each graph node through the graph convolutional layers (Figure 6) [82].
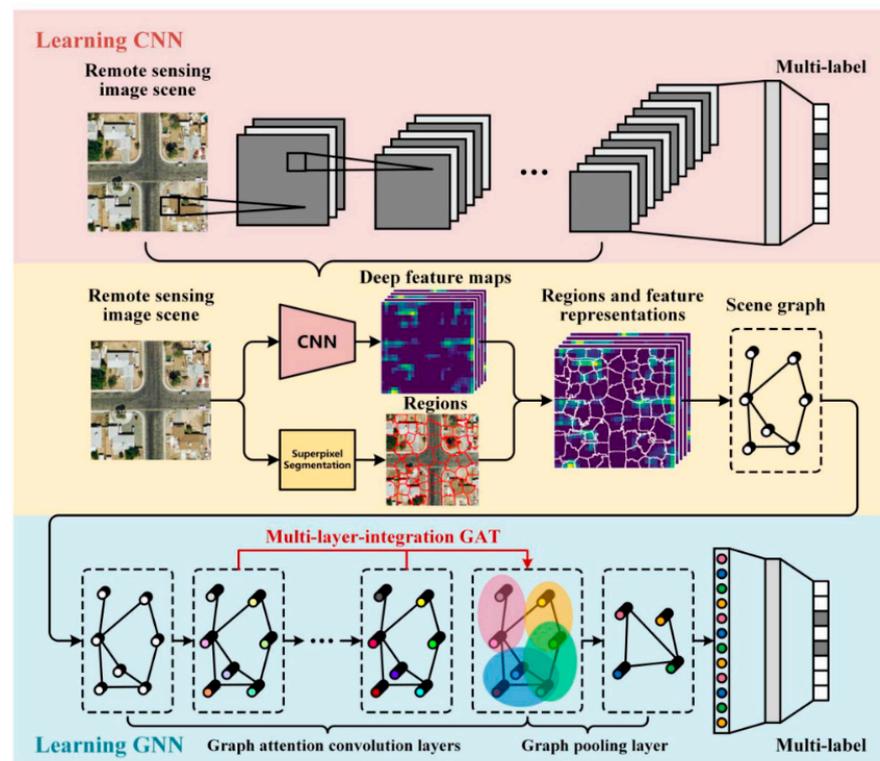

Figure 6. An example of adding an attention network to a GNN module for multi-label RS image classification [82].

4. Methodology

We followed the guidelines provided by Kitchenham, et al. [83] to systematically review the literature and report the results. Accordingly, we developed a review protocol at the start of the study and before conducting the review to reduce the biases. As the first step of the developed protocol, a set of research questions was defined (Section 4.1) according to the objective of this review study (i.e., reviewing and investigating attention-based deep learning methods for remote-sensing image-processing applications). Thereafter, the search strategy including search databases, strings, and a time-period was formulated to automatically find the relevant publications (Section 4.2). The final set of papers for the systematic review were selected by manually screening the papers according to the predefined exclusion criteria (Section 4.3). Then, a data extractions strategy (Section 4.4) and a form (Appendix A—Table A1) were developed to extract the required information from the papers. The extracted data and information were synthesized and the associated results are presented and discussed to answer the research questions.

4.1. Research Questions

A total of five main research questions (RQs) were defined to address the objective of this study. The RQs were specifically selected to extract state-of-the-art and interesting aspects of the developed DL methods with attention mechanism applied to RS image processing, including the effect of such mechanisms in their performance. The review and further structured analysis were built on these RQs.

RQ1. What are the specific objectives in remote sensing image processing that were addressed with attention-based deep learning?

RQ2. What are the deep learning algorithms that were improved with attention mechanism for remote sensing image processing?

RQ3. Which types of attention mechanisms were used in deep learning methods for remote sensing image processing?

RQ4. What are the used data sets/types in attention-based deep learning methods for remote sensing image processing?

RQ4.1. What kind of remote sensing images are used?

RQ4.2. What is the spatial resolution of the used remote sensing images?

RQ5. What are the effects of the attention mechanism in the performance of the deep learning methods in remote sensing image processing?

RQ5.1. What is the level of accuracy achieved with attention-based deep learning methods?

RQ5.2. What is the effect of the attention mechanism on the accuracy level of the deep learning methods?

4.2. Search Strategy

Two main attributes are usually employed to define the search scope of a systemic literature review: publication date and platform. We executed the search with no limit for the published data on the well-known and widely accepted platforms, i.e., ISI Web of Knowledge and Scopus. We formulated the following search string and executed it on the search engine of the selected publication platforms automatically to search in title, abstract, and keywords of the papers.

Search string:

(("attention mechanism" OR "attention guid" OR "attention embed*" OR "attention contain*" OR "attention based" OR "with attention" OR "attention aid*" OR "attention net*" OR "attentive") AND ("remote sensing" OR "satellite image*" OR "UAV image*" OR "hyperspectral image*" OR "aerial image*" OR "SAR") AND ("CNN" OR "deep learning"))*

The defined search query consisted of three main parts that were separated by the term "AND". The first part aimed to find the publications that used attention mechanisms (e.g., attentive). The second part aimed to find the relevant publications concerning their used remote sensing images (e.g., satellite images) and the third part aimed to find the papers that used deep learning methods (e.g., CNN).

4.3. Study Selection Criteria

After automated extraction of the publications from the selected platforms using the defined search query, we manually filtered the papers to select the final list of the most suitable ones. For this, we screened the publications mainly by reading their abstract and introduction sections and based on a set of exclusion criteria (Table 1) that were particularly defined according to the objectives of this review.

Table 1. Exclusion criteria.

ID	Criterion
EC1.	Papers in which the full text is unavailable
EC2.	Papers are not written in English
EC3.	Papers are not aiming to directly contribute to remote sensing image processing
EC4.	Papers do not directly use attention mechanism within DL methods
EC5.	Papers do not validate the proposed study
EC6.	Papers that provide a general summary without a clear contribution
EC7.	Review, conference, and editorial papers

4.4. Data Extraction

To properly answer the defined research questions, first, we needed to extract the necessary data and information from the retrieved papers. For this, a data extraction form was designed and created (Appendix A—Table A1). This form consists of a set of attributes to extract general information from the papers (e.g., publication year and publisher), as well as detailed ones including the study target of the papers, developed DL methods, attention mechanism type used, and the accuracy rates of the employed/developed DL methods with and without attention mechanism. Here, we used only the papers that did this analysis as explained above or compared their produced At-DL results with state-of-the-art DL methods in which no attention mechanism was used. In addition, only the overall accuracy metric was used to compare the papers since this was the only performance measurement used in most of the papers. The general data were extracted with the initial screening of the papers while the more detailed ones were extracted by carefully reading and reviewing of the papers.

4.5. Data Synthesis

The data synthesizing step is to answer the research questions, synthesize the extracted data and present the results. Thus, it is the most important step of a systematic literature review. In this step, the papers were grouped based on the extracted data into defined groups to answer corresponding research questions, and accordingly, the results were summarized and visualized. The detailed discussions over the presented results are provided to elicit and highlight the important points for each research question. Furthermore, the main findings such as current research directions, achievements on the use of attention mechanism to increase the performance of the DL methods for RS image processing applications, open problems, and recommendations for future studies are provided.

5. Results and Discussion

A final number of 176 papers were selected for the detailed review. The main statistics and an overview of the papers are provided in the following subsection. In addition, the detailed results are presented and corresponding discussions are provided for each research question in the next subsections.

5.1. Overview of the Reviewed Papers

At-DL methods entered RS image processing in 2018, while attention mechanism was developed in 2014 [20]. However, only since 2020, have most studies (i.e., 141 papers) employed this technique for different RS image processing applications, which reveals a significant interest in the technique in recent years (Figure 7). Just in 2021, 47 papers were published, knowing that the searches from the online databases were conducted in March 2021.
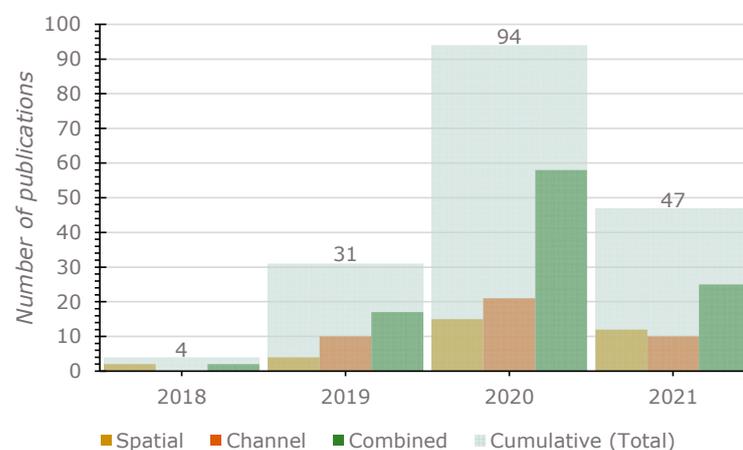

Figure 7. Year-wise classification of the papers and classified based on the attention mechanism type used.

Table 2 shows the journal names with at least two papers, and the rest with only one paper are aggregated in the "other" category. The papers are published in 30 different journals, which shows the usefulness of the At-DL for a wide range of RS image processing applications from water management [84,85] to urban studies [86]. The most popular journal is the "Remote Sensing" journal with 44 papers, and the second one is "IEEE Transactions on Geoscience and Remote Sensing" journal with 33 papers (Table 2). Furthermore, 17 journals only have one paper ("other" category in Table 2). These statistics show that most of the papers are published in technical RS journals rather than subject-specific journals.

Table 2. Journal names and their corresponding number of papers in attention mechanism-based DL for RSIP.

Journal Name	Number of Papers
Remote Sensing	44
IEEE Transactions on Geoscience and Remote Sensing	33
IEEE Access	27
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	17
IEEE Geoscience and Remote Sensing Letters	14
Sensors	6
ISPRS Journal of Photogrammetry and Remote Sensing	5
International Journal of Remote Sensing	3
IET Image Processing	2
ISPRS International Journal of Geo-Information	2
Journal of Applied Remote Sensing	2
Remote Sensing of Environment	2
Symmetry	2
Other	17

5.2. RQ1. What Are the Specific Objectives in Remote Sensing Image Processing That Are Addressed with Attention-Based Deep Learning?

The papers are grouped with regard to their study target similar to the classes used in [11]: image classification, image segmentation, image fusion, object detection, change detection, and other (Figure 8).

(i) Image classification: refers to labeling a group of pixels (objects or patches) in the RS images using training samples (e.g., land cover and land use classification). This is one of the most frequently used RS image processing tasks in various application domains as the starting point of the process [87–89]. Image classification is also called scene classification [88] or land cover and land use classifications [90] in the literature, depending on the aim and the data used in the studies. About half of the papers in At-DL addressed the image classification tasks for images acquired from different sensors such as multispectral satellites [67,91,92], hyperspectral [71,93], and unmanned aerial vehicles (UAV) [34,94] images. The large amount of the freely available benchmark data sets and organized competitions in this regard attracts researchers to develop DL methods in this subject area.

(ii) Object detection: refers to the detection of different objects in an image. It is the second most popular task that is addressed using At-DL including general object/target detection from RS images [46,60,95] or detection of the specific objects and features such as buildings [74,96], ships [97,98], landslides [99], clouds [53,100], airports [101], roads [72] and trees [102].

(iii)　Image segmentation: also known as semantic segmentation refers to labeling each pixel in the image, usually using end-to-end At-DL methods. From the At-DL papers, 17 papers addressed image segmentation [103–105].

(iv)　Image fusion: is mostly known as a fundamental preprocess in the RS field, and aims to produce higher spectral and spatial resolutions. There are two main image fusion tasks that were addressed using At-DL in 13 papers. One is pan-sharpening that aims to fuse a coarse resolution multispectral image with a correspondingly high-resolution panchromatic image to produce a high-resolution multispectral image [106–108]. Another one is image super-resolution which refers to enhancing the resolution of the original image using At-DL methods [106,107,109].

(v)　Change detection: refers to detecting and quantifying the changes in multi-temporal RS images. This is one of the challenging tasks and with the increasing amount of multi-temporal RS images has become more popular. At-DL was used in 7 papers to detect changes in general [110,111], in buildings [51], or any other objects [81,112].

(vi)　Other tasks, such as image dehazing [113], digital elevation model (DEM) void filling [114], and SAR image despeckling [115] were addressed with At-DL in 9 papers.

5.3. RQ2. What Are the Deep Learning Algorithms That Are Improved with Attention Mechanism for Remote Sensing Image Processing?

Figure 9 shows the number of papers that employed the attention mechanism for each DL algorithm. Accordingly, the convolutional neural networks (CNN) algorithm is the predominant DL method that was enhanced with an attention mechanism to address RS image processing, which applied in 154 out of 176 reviewed papers [69,116–120]. This is an expected result since CNN is the most frequently used DL method in general computer vision and image processing. Recurrent neural networks (RNN), such as long-short term memories (LSTM) methods, were the second most frequently used DL method supported by attention mechanism for RS image processing with 18 papers [121–123], this algorithm is also the first DL method that was improved with attention mechanism [20]. In addition, it was observed that most of the RNN methods were used in combination with CNN methods [76,78,124]. Generative adversarial networks (GAN) [53,125,126], Graph Neural Network (GNN) [80,82], and other DL methods including capsule network [72] and autoencoders [61] were the other DL algorithms used in 12, 5, and 4 papers, respectively.
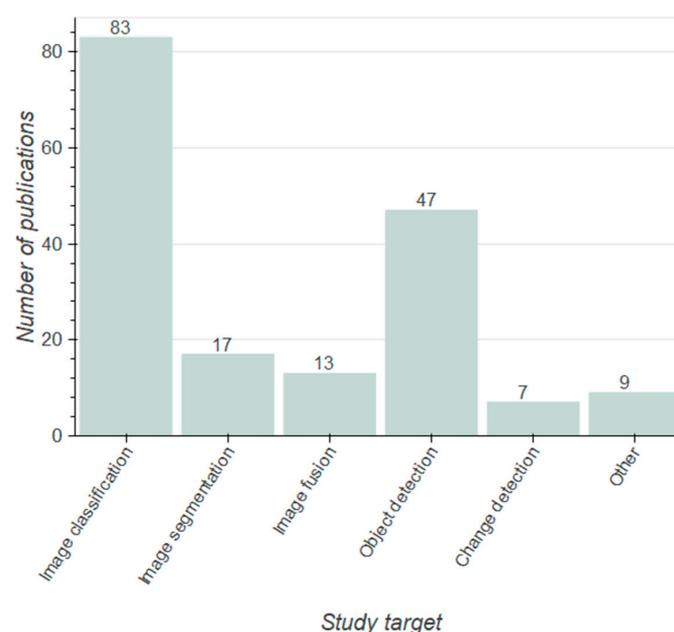

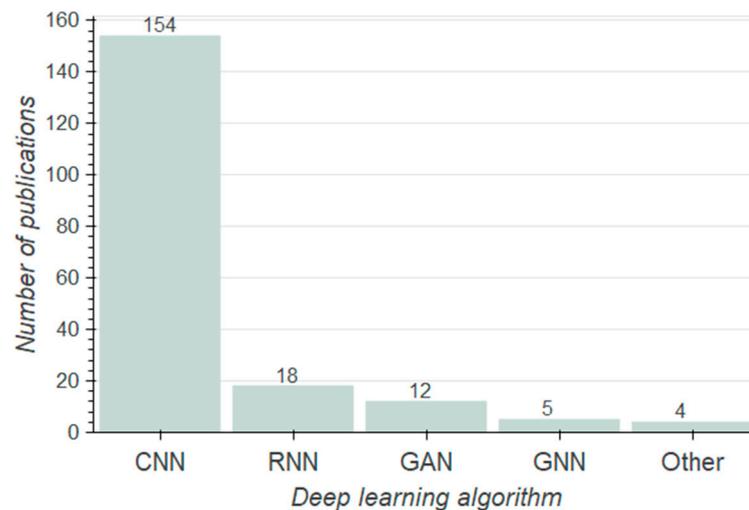**Figure 8.** The number of publications for different study targets.

Figure 9. The improved DL algorithms with attention mechanism in the papers.

5.4. RQ3. Which Types of Attention Mechanisms Were Used in Deep Learning Methods for Remote Sensing Image Processing?

At-DL methods can be classified based on the used attention types (i.e., channel and spatial attention networks) as explained in Section 2 (Figure 10). The combined use of the channel and spatial attention mechanisms were the most frequently used types in the papers [59,127,128]. In addition, the channel type, which is mostly used in hyperspectral image processing [129–131], and the spatial type [47,132,133] were also solely used in 41 and 33 papers, respectively. Depending on the aim of the study, the attention type can be selected; however, because in RS images, the features/channels and spatial location of the objects/features are both important, using a combined type was the predominant choice of the researchers in the papers.
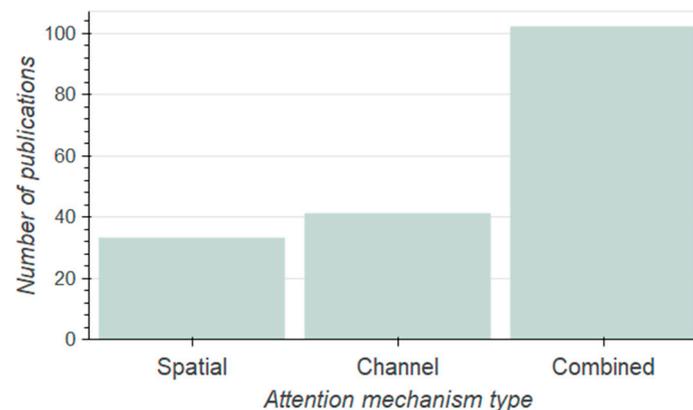

Figure 10. The attention mechanism type used in the papers.

5.5. RQ4. What Are the Used Data Sets/Types in Attention-Based Deep Learning Methods for Remote Sensing Image Processing?

Multispectral satellite images are the most popular images that are processed with At-DL methods (81 papers) [91,92,134] (Figure 11). This is mostly due to the free availability of some MS satellite images and their wide range of applications. Aerial images [54,135,136], hyperspectral images [137–139], and SAR images [97,140,141] were also processed with At-DL methods in 55, 43, and 24 papers, respectively. However, UAV images were used in only three papers [34,94,142]. This is a surprisingly low number; however, due to the very high resolution of the UAV images, the attention mechanism could significantly increase the performance of the DL methods.
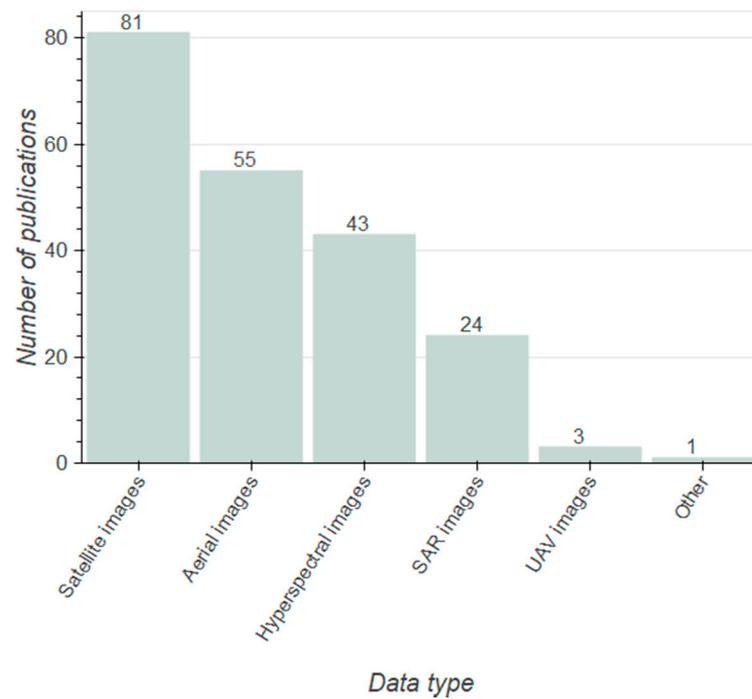
Figure 11. The data sets used in the papers.

The processed RS images were also grouped based on the spatial resolution of the processed images (Figure 12). High- and medium-resolution images were the main processed RS images in 157, and 58 papers, respectively. Low-resolution images (with spatial resolution over 30 m) were only used in four papers.
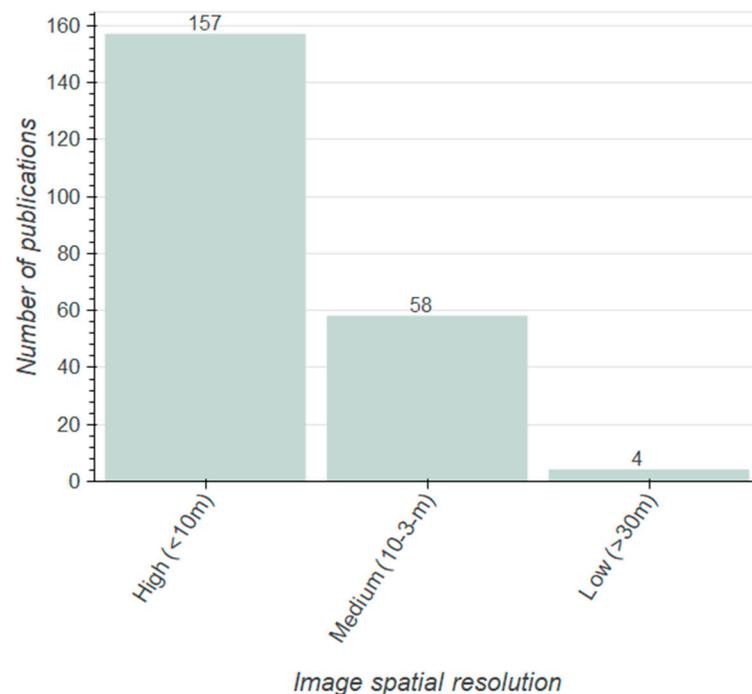

Figure 12. The spatial resolution of the used RS images in the papers.

5.6. RQ5. What Are the Effects of the Attention Mechanism in the Performance of the Deep Learning Methods in Remote Sensing Image Processing?

We investigated the performance of attention mechanism in DL methods for RS image processing in two manners; (i) by extracting the overall accuracies of the used At-DL methods for RS image processing tasks (Figure 13), and (ii) comparing the overall accuracies of the produced results with and without attention mechanism in the papers (Figure 14).

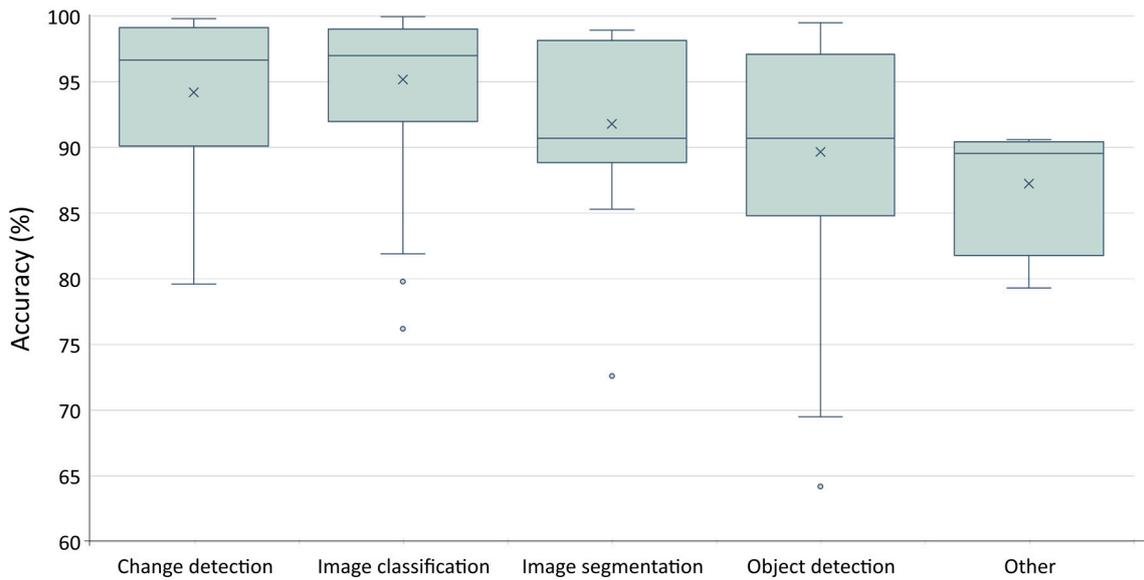**Figure 13.** The produced accuracy of the developed At-DL methods for different tasks in the papers.
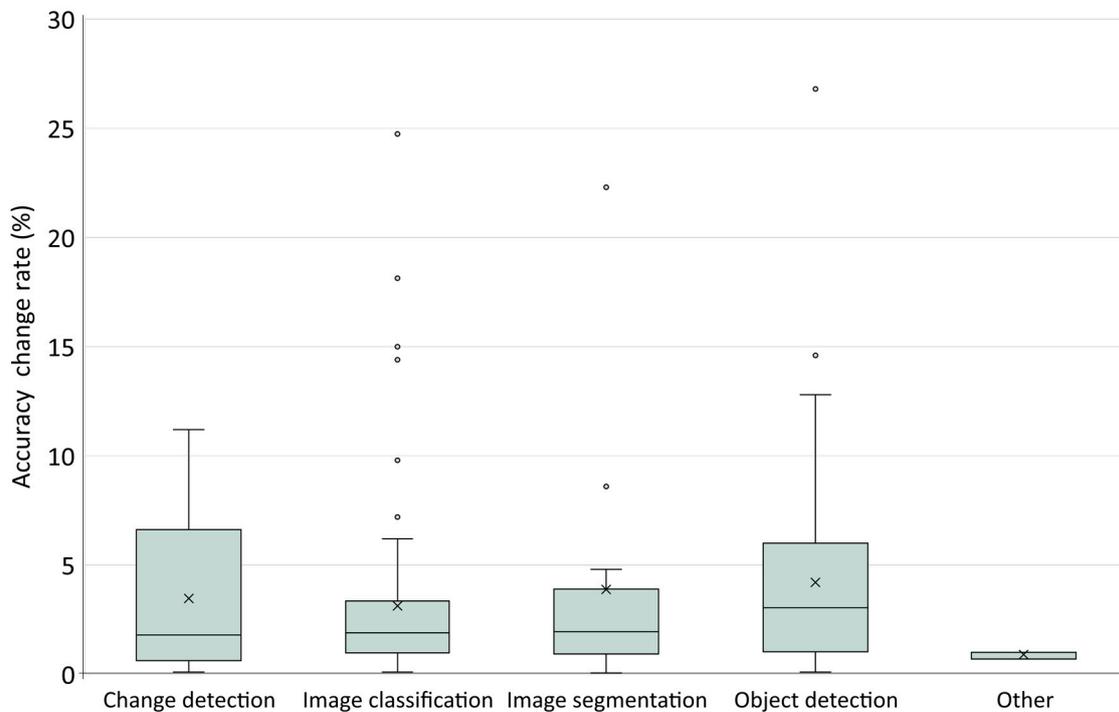

Figure 14. The effect of the use of the attention mechanism within the DL algorithms in terms of accuracy rate for different tasks in the papers.

Figure 13 illustrates a box plot graph of the overall accuracies of the produced results in the papers for change detection, image classification, image segmentation, object detection, and other tasks. Image classification and change detection had the highest median accuracies (~97%). One of the reasons is the availability of the benchmark datasets for such applications that encourage researchers to test their proposed methods on such datasets and tasks. Nevertheless, image classification is one of the fundamental and valuable applications in RS image processing that can be used as the basis in other science fields including agriculture, natural hazards, and thus, the already reached high accuracy levels is a good sign of using At-DL. Change detection with the increasing availability of multi-temporal RS images has become important in different fields [143–145]. Although the results revealed a high performance of At-DL in conducting change detection, only seven papers is not a robust number of papers to conclude a general statement that the At-DL produces above 95% accuracy rate, and thus, more work is needed on this. Image segmentation and object detection had a median accuracy value of about 91%, which is about 5% less than the first two image processing tasks. In addition, other tasks such as digital elevation model (DEM) void filling with At-DL papers had the median of below 90% accuracy values. Providing benchmark RS images and training samples for applications such as object detection would help to attract the attention of the researchers and develop more advanced methods. However, most of the used At-DL methods in image classification can be adopted for other tasks, including object detection.

Figure 14 shows a box plot graph of the effect of the attention mechanism in overall accuracies of the produced results in the papers for change detection, image classification image segmentation, object detection, and other tasks. Most of the papers reported an increase when using the attention mechanism within the DL methods. Only one paper stated that using the attention mechanism did not positively impact the performance of the DL method [146]. The median of the increase rates for all the classes was less than 5%. This increased rate was a remarkable enhancement of overall accuracies, given that the overall accuracy rates for most of the classes were already above 90%. The highest median rate which also showed the highest accuracy increase belonged to the object detection class with ~5%. One of the reasons for the highest increase rate by using attention mechanism in DL methods for object detection class when compared with the others is the inherent characteristics of these methods which need to localize the objects and attention mechanism, in particular, the spatial type, has the same aim by providing a focus on the spatial location of the important features. Image classification, image segmentation, and change detection classes had almost the same increase rate of overall accuracies with ~3–4%. The "other" class with ~1% increase had the lowest increase.

5.7. Threats to Validity of This Review

Every systematic literature review may be biased due to some limitations such as publication bias, data extraction, and classification. The main threats to the validity of our review are discussed as follows:

Construct validity: This study aimed to examine the effect of the attention mechanism on deep learning algorithms for RS image processing through the review of the existing literature that used At-DL methods for RS image processing and accordingly provide insights and recommendations for future studies. We employed automated search queries applied to the ISI Web of Knowledge [147] and Scopus websites. As a result, using these databases as the only sources of publications may lead to missing other relevant publications that are not included in this study. However, this study aimed to provide an overview of high-quality publications. Hence, indexing in ISI and Scopus is an accepted and widely used way to find the corresponding high-quality papers. In addition, there might be missing terms that may affect the final results. However, we tried to keep the search broad (the initial number of papers was 270) and revised the search query several times to reduce such impacts on our results.

Internal validity: In a systematic literature review, systematic errors may occur in the data extraction phase and lead to an incomplete relationship between the extracted data and

findings. In the current study, we precisely defined the research questions to investigate and extract all the required data and necessary information from At-DL studies. Hence, the findings of this study are properly explained and linked to the extracted and presented results.

External validity: This study reviewed the publications which employed At-DL methods for RS image processing applications. However, all of the existing DL methods have not been improved with attention mechanism or have not yet been used for RS image processing applications, and all the possible RS image processing applications were not addressed with At-DL and thus not included or discussed in this study. In addition, we only reviewed the publications that used At-DL for RS image processing applications, and thus, we cannot make judgments about the use and the effect of the At-DL in a broader scope or other applications.

Conclusion validity: We conducted the review based on the accepted structure and protocol for systematic literature review studies [83]. In addition, the steps of the structure review process are comprehensively explained in Section 4 of the paper, and the used search string, data extraction form (Appendix A) and the extracted papers as supplementary materials are provided in the paper. Therefore, the results of this study are reproducible using the given information.

6. Conclusions

This study reviewed the remote sensing (RS) literature that used attention mechanism-based deep learning (At-DL) methods for processing RS imagery. We investigated the advances in the use of At-DL methods and also the effect of the attention mechanism considering its different types on the performance of the DL methods in RS image processing. Accordingly, the current research directions and challenges are presented, and insights and recommendations for future studies are provided. Using a systematic literature review, which is not a well-known and used strategy in RS review papers, led us to a comprehensive review and to precisely answering the predefined research questions and contributing to the objective of this study. The results clearly demonstrate the positive impact of the attention mechanism on the performance of the DL methods in RS image processing, therefore, it is one of the powerful approaches that can be used to improve DL methods for such applications. In addition, the review results show an increasing trend in the use of At-DL methods in RS image processing. However, while image classification attracted most of the attention, other RS image processing tasks, such as object detection and change detection still need more studies to fully understand the effect of the attention mechanism on the performance of the DL methods. There are even important tasks that have not yet been addressed using this mechanism, including object-oriented image analysis. Results also revealed that the CNN methods are the algorithms that are the most frequently improved by the attention mechanism, which is largely due to its general usefulness; and it is a popular method for different computer vision tasks, in general. However, recently generative adversarial networks (GANs) have become state-of-the-art methods in different computer vision tasks when combined with attention mechanisms such as StarGAN [148] and AttentionGAN [149]. Hence, they can be adopted for RS image processing applications in future studies. Moreover, we investigated the performance of the At-DL methods based on the overall accuracy metric, which is widely used for RS applications and provided in the papers. However, the accuracy of the DL methods depends on the dataset used and the aimed tasks. In addition, the performance of the At-DL methods should be studied using other important metrics (e.g., computational time).

Appendix A. Data Extraction Form

Table A1. Data extraction form.

#	Extraction element	Contents
General information		
1	ID	Unique ID for the study
2	Title	Full title of the article
3	Authors	The authors of the article
4	Year	The publication year
5	Journal name	The journal name (e.g., Journal of Dairy Science)
Study description		
6	Study target	☐Image classification ☐Image segmentation ☐Object detection ☐Image fusion ☐Change detection ☐Other
7	Details about the study	E.g., any interesting findings or problems
8	Directly address RS image processing	☐Yes ☐No
9	Deep learning algorithm	☐CNN ☐RNN ☐GAN ☐GNN ☐Other
10	Attention type	☐Spatial ☐Channel ☐Combined
11	Remote sensing image type	☐MS Satellite ☐Aerial ☐Hyperspectral ☐SAR ☐UAV ☐Other
12	Remote sensing image spatial resolution	☐High (<10 m) ☐Medium (10–30 m) ☐Low (>30m)
13	Overall accuracy (%)	The overall accuracy of the produced results using At-DL method
14	Effect of attention mechanism (%)	The increased rate of the overall accuracy when used attention mechanism.
15	Additional notes	E.g., the opinions of the reviewer about the study

References

1. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
2. Ghaffarian, S.; Turker, M. An improved cluster-based snake model for automatic agricultural field boundary extraction from high spatial resolution imagery. *Int. J. Remote Sens.* **2019**, *40*, 1217–1247. [CrossRef]
3. Valente, J.; Sari, B.; Kooistra, L.; Kramer, H.; Mücher, S. Automated crop plant counting from very high-resolution aerial imagery. *Precis. Agric.* **2020**, *21*, 1366–1384. [CrossRef]
4. Zhang, C.; Valente, J.; Kooistra, L.; Guo, L.; Wang, W. Orchard management with small unmanned aerial vehicles: A survey of sensing and analysis approaches. *Precis. Agric.* **2021**. [CrossRef]
5. Nielsen, M.M. Remote sensing for urban planning and management: The use of window-independent context segmentation to extract urban features in Stockholm. *Comput. Environ. Urban Syst.* **2015**, *52*, 1–9. [CrossRef]
6. Kadhim, N.; Mourshed, M.; Bray, M. Advances in remote sensing applications for urban sustainability. *Euro-Mediterr. J. Environ. Integr.* **2016**, *1*, 7. [CrossRef]
7. Ghaffarian, S.; Ghaffarian, S. Automatic building detection based on Purposive FastICA (PFICA) algorithm using monocular high resolution Google Earth images. *ISPRS J. Photogramm. Remote Sens.* **2014**, *97*, 152–159. [CrossRef]
8. Ghaffarian, S.; Kerle, N.; Filatova, T. Remote Sensing-Based Proxies for Urban Disaster Risk Management and Resilience: A Review. *Remote Sens.* **2018**, *10*, 1760. [CrossRef]
9. Ghaffarian, S.; Rezaie Farhadabad, A.; Kerle, N. Post-Disaster Recovery Monitoring with Google Earth Engine. *Appl. Sci.* **2020**, *10*, 4574. [CrossRef]
10. Ghaffarian, S.; Emtehani, S. Monitoring Urban Deprived Areas with Remote Sensing and Machine Learning in Case of Disaster Recovery. *Climate* **2021**, *9*, 58. [CrossRef]

11. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]

12. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

13. Sheykhmousa, M.; Mahdianpari, M.; Ghanbari, H.; Mohammadimanesh, F.; Ghamisi, P.; Homayouni, S. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [CrossRef]

14. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]

15. Li, Y.; Zhang, H.; Xue, X.; Jiang, Y.; Shen, Q. Deep learning for remote sensing image classification: A survey. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1264. [CrossRef]

16. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]

17. Ghanbari, H.; Mahdianpari, M.; Homayouni, S.; Mohammadimanesh, F. A Meta-Analysis of Convolutional Neural Networks for Remote Sensing Applications. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3602–3613. [CrossRef]

18. Liu, X.; Wang, Y.; Liu, Q. Psgan: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpening. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 873–877.

19. Yan, X.; Ai, T.; Yang, M.; Yin, H. A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 259–273. [CrossRef]

20. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.

21. Niu, Z.; Zhong, G.; Yu, H. A Review on the Attention Mechanism of Deep Learning. *Neurocomputing* **2021**. [CrossRef]

22. Zhang, J.; Zhou, Q.; Wu, J.; Wang, Y.C.; Wang, H.; Li, Y.S.; Chai, Y.Z.; 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]

23. Zeng, Y.L.; Ritz, C.; Zhao, J.H.; Lan, J.H. Attention-Based Residual Network with Scattering Transform Features for Hyperspectral Unmixing with Limited Training Samples. *Remote Sens.* **2020**, *12*, 400. [CrossRef]

24. Yu, Y.; Li, X.; Liu, F. Attention GANs: Unsupervised Deep Feature Learning for Aerial Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 519–531. [CrossRef]

25. Gao, F.; He, Y.S.; Wang, J.; Hussain, A.; Zhou, H.Y. Anchor-free Convolutional Network with Dense Attention Feature Aggregation for Ship Detection in SAR Images. *Remote Sens.* **2020**, *12*, 2619. [CrossRef]

26. Li, F.; Feng, R.; Han, W.; Wang, L. High-Resolution Remote Sensing Image Scene Classification via Key Filter Bank Based on Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8077–8092. [CrossRef]

27. Yang, H.; Wu, P.H.; Yao, X.D.; Wu, Y.L.; Wang, B.; Xu, Y.Y. Building Extraction in Very High Resolution Imagery by Dense-Attention Networks. *Remote Sens.* **2018**, *10*, 1768. [CrossRef]

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

29. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [CrossRef]

30. Galassi, A.; Lippi, M.; Torroni, P. Attention in Natural Language Processing. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, 1–18. [CrossRef] [PubMed]

31. Koščević, K.; Subašić, M.; Lončarić, S. Attention-based Convolutional Neural Network for Computer Vision Color Constancy. In Proceedings of the 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 23–25 September 2019; pp. 372–377.

32. Li, W.; Liu, K.; Zhang, L.; Cheng, F. Object detection based on an adaptive attention mechanism. *Sci. Rep.* **2020**, *10*, 11307. [CrossRef]

33. Cui, W.; Wang, F.; He, X.; Zhang, D.Y.; Xu, X.X.; Yao, M.; Wang, Z.W.; Huang, J.J. Multi-Scale Semantic Segmentation and Spatial Relationship Recognition of Remote Sensing Images Based on an Attention Model. *Remote Sens.* **2019**, *11*, 1044. [CrossRef]

34. Alshehri, A.; Bazi, Y.; Ammour, N.; Almubarak, H.; Alajlan, N. Deep Attention Neural Network for Multi-Label Classification in Unmanned Aerial Vehicle Imagery. *IEEE Access* **2019**, *7*, 119873–119880. [CrossRef]

35. Sun, H.; Zheng, X.; Lu, X.; Wu, S. Spectral-Spatial Attention Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3232–3245. [CrossRef]

36. Feng, J.; Wu, X.; Shang, R.; Sui, C.; Li, J.; Jiao, L.; Zhang, X. Attention Multibranch Convolutional Neural Network for Hyperspectral Image Classification Based on Adaptive Region Search. *IEEE Trans. Geosci. Remote Sens.* **2020**. [CrossRef]

37. Zhao, Z.; Wang, H.; Yu, X. Spectral-Spatial Graph Attention Network for Semisupervised Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2021**. [CrossRef]

38. Censi, A.M.; Ienco, D.; Gbodjo, Y.J.E.; Pensa, R.G.; Interdonato, R.; Gaetano, R. Attentive Spatial Temporal Graph CNN for Land Cover Mapping from Multi Temporal Remote Sensing Data. *IEEE Access* **2021**, *9*, 23070–23082. [CrossRef]

39. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

40. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research, Lille, France, 7–9 July 2015; pp. 2048–2057.
41. Guo, Y.; Ji, J.; Lu, X.; Huo, H.; Fang, T.; Li, D. Global-Local Attention Network for Aerial Scene Classification. *IEEE Access* **2019**, *7*, 67200–67212. [CrossRef]
42. Ma, J.; Ma, Q.; Tang, X.; Zhang, X.; Zhu, C.; Peng, Q.; Jiao, L. Remote Sensing Scene Classification Based on Global and Local Consistent Network. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Brussels, Belgium, 11–16 July 2020; pp. 537–540.
43. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef]
44. Wang, L.; Peng, J.T.; Sun, W.W. Spatial-Spectral Squeeze-and-Excitation Residual Network for Hyperspectral Image Classification. *Remote Sens.* **2019**, *11*, 884. [CrossRef]
45. Alswayed, A.S.; Alhichri, H.S.; Bazi, Y. SqueezeNet with Attention for Remote Sensing Scene Classification. In Proceedings of the ICCAIS 2020—3rd International Conference on Computer Applications and Information Security, Riyadh, Saudi Arabia, 19–21 March 2020.
46. Li, C.Y.; Luo, B.; Hong, H.L.; Su, X.; Wang, Y.J.; Liu, J.; Wang, C.J.; Zhang, J.; Wei, L.H. Object Detection Based on Global-Local Saliency Constraint in Aerial Images. *Remote Sens.* **2020**, *12*, 1435. [CrossRef]
47. Zhou, M.; Zou, Z.; Shi, Z.; Zeng, W.J.; Gui, J. Local Attention Networks for Occluded Airplane Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 381–385. [CrossRef]
48. Ding, L.; Tang, H.; Bruzzone, L. LANet: Local Attention Embedding to Improve the Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 426–435. [CrossRef]
49. Chaudhari, S.; Mithal, V.; Polatkan, G.; Ramanath, R. An attentive survey of attention models. *arXiv* **2019**, arXiv:1904.02874.
50. Lu, J.; Yang, J.; Batra, D.; Parikh, D. Hierarchical question-image co-attention for visual question answering. In Proceedings of the NIPS, Barcelona, Spain, 5–10 December 2016; pp. 289–297.
51. Jiang, H.W.; Hu, X.Y.; Li, K.; Zhang, J.M.; Gong, J.Q.; Zhang, M. PGA-SiamNet: Pyramid Feature-Based Attention-Guided Siamese Network for Remote Sensing Orthoimagery Building Change Detection. *Remote Sens.* **2020**, *12*, 484. [CrossRef]
52. He, N.; Fang, L.; Li, Y.; Plaza, A. High-Order Self-Attention Network for Remote Sensing Scene Classification. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Yokohama, Japan, 28 July–2 August 2019; pp. 3013–3016.
53. Wu, Z.C.; Li, J.; Wang, Y.S.; Hu, Z.W.; 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]
54. Cao, R.; Fang, L.; Lu, T.; He, N. Self-Attention-Based Deep Feature Fusion for Remote Sensing Scene Classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 43–47. [CrossRef]
55. Wu, H.L.; Zhao, S.Z.; Li, L.; Lu, C.Q.; Chen, W. Self-Attention Network With Joint Loss for Remote Sensing Image Scene Classification. *IEEE Access* **2020**, *8*, 210347–210359. [CrossRef]
56. Xiao, T.; Xu, Y.; Yang, K.; Zhang, J.; Peng, Y.; Zhang, Z. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In Proceedings of the CVPR, IEEE Computer Society, Boston, MA, USA, 7–12 June 2015; pp. 842–850.
57. Sumbul, G.; Cinbis, R.G.; Aksoy, S. Multisource Region Attention Network for Fine-Grained Object Recognition in Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4929–4937. [CrossRef]
58. Li, J.; Tu, Z.; Yang, B.; Lyu, M.R.; Zhang, T. Multi-Head Attention with Disagreement Regularization. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2897–2903.
59. Zhang, S.Y.; Li, C.R.; Qiu, S.; Gao, C.X.; Zhang, F.; Du, Z.H.; Liu, R.Y. EMMCNN: An ETPS-Based Multi-Scale and Multi-Feature Method Using CNN for High Spatial Resolution Image Land-Cover Classification. *Remote Sens.* **2020**, *12*, 66. [CrossRef]
60. Cheng, B.; Li, Z.Z.; Xu, B.T.; Yao, X.; Ding, Z.Q.; Qin, T.Q. Structured Object-Level Relational Reasoning CNN-Based Target Detection Algorithm in a Remote Sensing Image. *Remote Sens.* **2021**, *13*, 281. [CrossRef]
61. Wu, Z.; Hou, B.; Jiao, L. Multiscale CNN with Autoencoder Regularization Joint Contextual Attention Network for SAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1200–1213. [CrossRef]
62. Shen, T.; Zhou, T.; Long, G.; Jiang, J.; Pan, S.; Zhang, C. DiSAN: Directional Self-Attention Network for RNN/CNN-free Language Understanding. In Proceedings of the AAAI, New Orleans, LA, USA, 2–7 February 2018.
63. Du, J.; Han, J.; Way, A.; Wan, D. Multi-Level Structured Self-Attentions for Distantly Supervised Relation Extraction. In Proceedings of the EMNLP, Brussels, Belgium, 31 October–4 November 2018.
64. Carrasco, M. Visual attention: The past 25 years. *Vis. Res.* **2011**, *51*, 1484–1525. [CrossRef]
65. Beuth, F.; Hamker, F.H. A mechanistic cortical microcircuit of attention for amplification, normalization and suppression. *Vis. Res.* **2015**, *116*, 241–257. [CrossRef]
66. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]
67. Ma, W.P.; Zhao, J.L.; Zhu, H.; Shen, J.C.; Jiao, L.C.; Wu, Y.; Hou, B.A. A Spatial-Channel Collaborative Attention Network for Enhancement of Multiresolution Classification. *Remote Sens.* **2021**, *13*, 106. [CrossRef]

68. Tong, W.; Chen, W.; Han, W.; Li, X.; Wang, L. Channel-Attention-Based DenseNet Network for Remote Sensing Image Scene Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4121–4132. [CrossRef]

69. Guo, D.; Xia, Y.; Luo, X. Scene Classification of Remote Sensing Images Based on Saliency Dual Attention Residual Network. *IEEE Access* **2020**, *8*, 6344–6357. [CrossRef]

70. Hang, R.L.; Li, Z.; Liu, Q.S.; Ghamisi, P.; Bhattacharyya, S.S. Hyperspectral Image Classification With Attention-Aided CNNs. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 2281–2293. [CrossRef]

71. Zhu, M.; Jiao, L.; Liu, F.; Yang, S.; Wang, J. Residual Spectral-Spatial Attention Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 449–462. [CrossRef]

72. Ren, Y.F.; Yu, Y.T.; Guan, H.Y. DA-CapsUNet: A Dual-Attention Capsule U-Net for Road Extraction from Remote Sensing Imagery. *Remote Sens.* **2020**, *12*, 2866. [CrossRef]

73. Ren, Y.; Li, X.; Yang, X.; Xu, H. Development of a Dual-Attention U-Net Model for Sea Ice and Open Water Classification on SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2021**. [CrossRef]

74. He, N.; Fang, L.; Plaza, A. Hybrid first and second order attention Unet for building segmentation in remote sensing images. *Sci. China Inf. Sci.* **2020**, *63*. [CrossRef]

75. Pan, X.; Yang, F.; Gao, L.; Chen, Z.; Zhang, B.; Fan, H.; Ren, J. Building extraction from high-resolution aerial imagery using a generative adversarial network with spatial and channel attention mechanisms. *Remote Sens.* **2019**, *11*, 966. [CrossRef]

76. Liu, R.C.; Cheng, Z.H.; Zhang, L.L.; Li, J.X. Remote Sensing Image Change Detection Based on Information Transmission and Attention Mechanism. *IEEE Access* **2019**, *7*, 156349–156359. [CrossRef]

77. Wang, Q.; Liu, S.T.; Chanussot, J.; Li, X.L. Scene Classification With Recurrent Attention of VHR Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1155–1167. [CrossRef]

78. Li, Z.T.; Chen, G.K.; Zhang, T.X. Temporal Attention Networks for Multitemporal Multisensor Crop Classification. *IEEE Access* **2019**, *7*, 134677–134690. [CrossRef]

79. Mei, X.G.; Pan, E.T.; Ma, Y.; Dai, X.B.; Huang, J.; Fan, F.; Du, Q.L.; Zheng, H.; Ma, J.Y. Spectral-Spatial Attention Networks for Hyperspectral Image Classification. *Remote Sens.* **2019**, *11*, 963. [CrossRef]

80. Ma, F.; Gao, F.; Sun, J.P.; Zhou, H.Y.; Hussain, A. Attention Graph Convolution Network for Image Segmentation in Big SAR Imagery Data. *Remote Sens.* **2019**, *11*, 2586. [CrossRef]

81. Luo, X.; Li, X.; Wu, Y.; Hou, W.; Wang, M.; Jin, Y.; Xu, W. Research on Change Detection Method of High-Resolution Remote Sensing Images Based on Subpixel Convolution. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1447–1457. [CrossRef]

82. Li, Y.; Chen, R.; Zhang, Y.; Zhang, M.; Chen, L. Multi-label remote sensing image scene classification by combining a convolutional neural network and a graph neural network. *Remote Sens.* **2020**, *12*, 4003. [CrossRef]

83. Kitchenham, B.; Pearl Brereton, O.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. Systematic literature reviews in software engineering—A systematic literature review. *Inf. Softw. Technol.* **2009**, *51*, 7–15. [CrossRef]

84. Chen, L.F.; Zhang, P.; Xing, J.; Li, Z.H.; Xing, X.M.; Yuan, Z.H. A Multi-Scale Deep Neural Network for Water Detection from SAR Images in the Mountainous Areas. *Remote Sens.* **2020**, *12*, 3205. [CrossRef]

85. Yang, Q.; Wang, C.; Zeng, T. A method of water change monitoring in remote image time series based on long short time memory. *Remote Sens. Lett.* **2021**, *12*, 67–76. [CrossRef]

86. Zhang, Y.D.; Chen, G.; Vukomanovic, J.; Singh, K.K.; Liu, Y.; Holden, S.; Meentemeyer, R.K. Recurrent Shadow Attention Model (RSAM) for shadow removal in high-resolution urban land-cover mapping. *Remote Sens. Environ.* **2020**, *247*, 111945. [CrossRef]

87. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A review of supervised object-based land-cover image classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 277–293. [CrossRef]

88. Cheng, G.; Han, J.; Lu, X. Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proc. IEEE* **2017**, *105*, 1865–1883. [CrossRef]

89. Li, M.; Zang, S.; Zhang, B.; Li, S.; Wu, C. A Review of Remote Sensing Image Classification Techniques: The Role of Spatio-contextual Information. *Eur. J. Remote Sens.* **2014**, *47*, 389–411. [CrossRef]

90. Alem, A.; Kumar, S. Deep Learning Methods for Land Cover and Land Use Classification in Remote Sensing: A Review. In Proceedings of the 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 4–5 June 2020; pp. 903–908.

91. Sang, Q.; Zhuang, Y.; Dong, S.; Wang, G.; Chen, H. FRF-Net: Land Cover Classification from Large-Scale VHR Optical Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1057–1061. [CrossRef]

92. Ienco, D.; Gbodjo, Y.J.E.; Gaetano, R.; Interdonato, R. Weakly Supervised Learning for Land Cover Mapping of Satellite Image Time Series via Attention-Based CNN. *IEEE Access* **2020**, *8*, 179547–179560. [CrossRef]

93. Tang, X.; Meng, F.; Zhang, X.; Cheung, Y.M.; Ma, J.; Liu, F.; Jiao, L. Hyperspectral Image Classification Based on 3-D Octave Convolution with Spatial-Spectral Attention Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 2430–2447. [CrossRef]

94. Feng, Q.L.; Yang, J.Y.; Liu, Y.M.; Ou, C.; Zhu, D.H.; Niu, B.W.; Liu, J.T.; Li, B.G. Multi-Temporal Unmanned Aerial Vehicle Remote Sensing for Vegetable Mapping Using an Attention-Based Recurrent Convolutional Neural Network. *Remote Sens.* **2020**, *12*, 1668. [CrossRef]

95. Li, Y.Y.; Huang, Q.; Pei, X.; Jiao, L.C.; Shang, R.H. RADet: Refine Feature Pyramid Network and Multi-Layer Attention Network for Arbitrary-Oriented Object Detection of Remote Sensing Images. *Remote Sens.* **2020**, *12*, 389. [CrossRef]

96.	Zhou, D.; Wang, G.; He, G.; Long, T.; Yin, R.; Zhang, Z.; Chen, S.; Luo, B. Robust building extraction for high spatial resolution remote sensing images with self-attention network. *Sensors* **2020**, *20*, 7241. [CrossRef]

97.	Zhao, Y.; Zhao, L.; Xiong, B.; Kuang, G. Attention receptive pyramid network for ship detection in SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2738–2756. [CrossRef]

98.	Fu, J.; Sun, X.; Wang, Z.; Fu, K. An Anchor-Free Method Based on Feature Balancing and Refinement Network for Multiscale Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1331–1344. [CrossRef]

99.	Ji, S.P.; Yu, D.W.; Shen, C.Y.; Li, W.L.; Xu, Q. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* **2020**, *17*, 1337–1352. [CrossRef]

100.	Yao, Z.; Jia, J.; Qian, Y. Mcnet: Multi-scale feature extraction and content-aware reassembly cloud detection model for remote sensing images. *Symmetry* **2021**, *13*, 28. [CrossRef]

101.	Tan, S.Y.; Chen, L.F.; Pan, Z.H.; Xing, J.; Li, Z.H.; Yuan, Z.H. Geospatial Contextual Attention Mechanism for Automatic and Fast Airport Detection in SAR Imagery. *IEEE Access* **2020**, *8*, 173627–173640. [CrossRef]

102.	Zheng, J.; Fu, H.; Li, W.; Wu, W.; Zhao, Y.; Dong, R.; Yu, L. Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 154–177. [CrossRef]

103.	Qi, X.; Li, K.; Liu, P.; Zhou, X.; Sun, M. Deep Attention and Multi-Scale Networks for Accurate Remote Sensing Image Segmentation. *IEEE Access* **2020**, *8*, 146627–146639. [CrossRef]

104.	Xiao, D.; Wang, Z.; Wu, Y.; Gao, X.; Sun, X. Terrain Segmentation in Polarimetric SAR Images Using Dual-Attention Fusion Network. *IEEE Geosci. Remote Sens. Lett.* **2020**. [CrossRef]

105.	Li, J.L.; Xiu, J.P.; Yang, Z.Q.; Liu, C. Dual Path Attention Net for Remote Sensing Semantic Image Segmentation. *Isprs Int. J. Geo-Inf.* **2020**, *9*, 571. [CrossRef]

106.	Dong, X.; Sun, X.; Jia, X.; Xi, Z.; Gao, L.; Zhang, B. Remote Sensing Image Super-Resolution Using Novel Dense-Sampling Networks. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1618–1633. [CrossRef]

107.	Wang, H.; Hu, Q.; Wu, C.D.; Chi, J.N.; Yu, X.S. Non-Locally up-Down Convolutional Attention Network for Remote Sensing Image Super-Resolution. *IEEE Access* **2020**, *8*, 166304–166319. [CrossRef]

108.	Li, X.; Xu, F.; Lyu, X.; Tong, Y.; Chen, Z.; Li, S.; Liu, D. A Remote-Sensing Image Pan-Sharpening Method Based on Multi-Scale Channel Attention Residual Network. *IEEE Access* **2020**, *8*, 27163–27177. [CrossRef]

109.	Li, J.J.; Cui, R.X.; Li, B.; Song, R.; Li, Y.S.; Du, Q. Hyperspectral Image Super-Resolution with 1D-2D Attentional Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 2859. [CrossRef]

110.	Chen, L.; Zhang, D.Z.; Li, P.; Lv, P. Change Detection of Remote Sensing Images Based on Attention Mechanism. *Comput. Intell. Neurosci.* **2020**, *2020*, 6430627. [CrossRef] [PubMed]

111.	Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Liu, Y.; Li, H. DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1194–1206. [CrossRef]

112.	Zhang, Y.; Zhang, S.Z.; Li, Y.; Zhang, Y.N. Coarse-to-Fine Satellite Images Change Detection Framework via Boundary-Aware Attentive Network. *Sensors* **2020**, *20*, 6735. [CrossRef]

113.	Gu, Z.Q.; Zhan, Z.Q.; Yuan, Q.Q.; Yan, L. Single Remote Sensing Image Dehazing Using a Prior-Based Dense Attentive Network. *Remote Sens.* **2019**, *11*, 3008. [CrossRef]

114.	Gavriil, K.; Muntingh, G.; Barrowclough, O.J.D. Void Filling of Digital Elevation Models with Deep Generative Models. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1645–1649. [CrossRef]

115.	Shen, H.; Zhou, C.; Li, J.; Yuan, Q. SAR Image Despeckling Employing a Recursive Deep CNN Prior. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 273–286. [CrossRef]

116.	Li, J.; Lin, D.Y.; Wang, Y.; Xu, G.L.; Zhang, Y.Y.; Ding, C.B.; Zhou, Y.H. Deep Discriminative Representation Learning with Attention Map for Scene Classification. *Remote Sens.* **2020**, *12*, 1366. [CrossRef]

117.	Bahri, A.; Majelan, S.G.; Mohammadi, S.; Noori, M.; Mohammadi, K. Remote Sensing Image Classification via Improved Cross-Entropy Loss and Transfer Learning Strategy Based on Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1087–1091. [CrossRef]

118.	Zhang, C.; Yue, J.; Qin, Q. Global prototypical network for few-shot hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4748–4759. [CrossRef]

119.	Lei, P.C.; Liu, C. Inception residual attention network for remote sensing image super-resolution. *Int. J. Remote Sens.* **2020**, *41*, 9565–9587. [CrossRef]

120.	Cheng, W.S.; Yang, W.; Wang, M.; Wang, G.; Chen, J.Y. Context Aggregation Network for Semantic Labeling in Aerial Images. *Remote Sens.* **2019**, *11*, 1158. [CrossRef]

121.	Gbodjo, Y.J.E.; Ienco, D.; Leroux, L.; Interdonato, R.; Gaetano, R.; Ndao, B. Object-based multi-temporal and multi-source land cover mapping leveraging hierarchical class relationships. *Remote Sens.* **2020**, *12*, 2814. [CrossRef]

122.	Liang, L.; Wang, G. Efficient recurrent attention network for remote sensing scene classification. *IET Image Process.* **2021**. [CrossRef]

123.	Wang, Z.S.; Zou, C.; Cai, W.W. Small Sample Classification of Hyperspectral Remote Sensing Images Based on Sequential Joint Deeping Learning Model. *IEEE Access* **2020**, *8*, 71353–71363. [CrossRef]

124. You, H.; Tian, S.; Yu, L.; Lv, Y. Pixel-Level Remote Sensing Image Recognition Based on Bidirectional Word Vectors. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1281–1293. [CrossRef]

125. Zhang, X.K.; Pun, M.O.; Liu, M. Semi-Supervised Multi-Temporal Deep Representation Fusion Network for Landslide Mapping from Aerial Orthophotos. *Remote Sens.* **2021**, *13*, 548. [CrossRef]

126. Wong, R.; Zhang, Z.J.; Wang, Y.M.; Chen, F.S.; Zeng, D. HSI-IPNet: Hyperspectral Imagery Inpainting by Deep Learning With Adaptive Spectral Extraction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4369–4380. [CrossRef]

127. Xu, R.D.; Tao, Y.T.; Lu, Z.Y.; Zhong, Y.F. Attention-Mechanism-Containing Neural Networks for High-Resolution Remote Sensing Image Classification. *Remote Sens.* **2018**, *10*, 1602. [CrossRef]

128. Gao, H.; Cao, L.; Yu, D.; Xiong, X.; Cao, M. Semantic Segmentation of Marine Remote Sensing Based on a Cross Direction Attention Mechanism. *IEEE Access* **2020**, *8*, 142483–142494. [CrossRef]

129. Zheng, J.; Feng, Y.; Bai, C.; Zhang, J. Hyperspectral Image Classification Using Mixed Convolutions and Covariance Pooling. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 522–534. [CrossRef]

130. Zhao, L.; Yi, J.; Li, X.; Hu, W.; Wu, J.; Zhang, G. Compact Band Weighting Module Based on Attention-Driven for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**. [CrossRef]

131. He, X.; Chen, Y.; Ghamisi, P. Heterogeneous Transfer Learning for Hyperspectral Image Classification Based on Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3246–3263. [CrossRef]

132. Chen, H.; Chen, R.; Li, N.N. Attentive generative adversarial network for removing thin cloud from a single remote sensing image. *IET Image Process.* **2021**, *15*, 856–867. [CrossRef]

133. Wang, J.; Xiao, H.; Chen, L.; Xing, J.; Pan, Z.; Luo, R.; Cai, X. Integrating weighted feature fusion and the spatial attention module with convolutional neural networks for automatic aircraft detection from sar images. *Remote Sens.* **2021**, *13*, 910. [CrossRef]

134. Zhang, H.; Ma, J.; Chen, C.; Tian, X. NDVI-Net: A fusion network for generating high-resolution normalized difference vegetation index in remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 182–196. [CrossRef]

135. Haut, J.M.; Fernandez-Beltran, R.; Paoletti, M.E.; Plaza, J.; Plaza, A. Remote Sensing Image Superresolution Using Deep Residual Channel Attention. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9277–9289. [CrossRef]

136. Dong, X.; Xi, Z.; Sun, X.; Yang, L. Remote Sensing Image Super-Resolution via Enhanced Back-Projection Networks. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Waikoloa, HI, USA, 26 September–2 October 2020; pp. 1480–1483.

137. Guo, H.; Liu, J.; Yang, J.; Xiao, Z.; Wu, Z. Deep Collaborative Attention Network for Hyperspectral Image Classification by Combining 2-D CNN and 3-D CNN. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4789–4802. [CrossRef]

138. Li, R.; Zheng, S.Y.; Duan, C.X.; Yang, Y.; Wang, X.Q. Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network. *Remote Sens.* **2020**, *12*, 582. [CrossRef]

139. Wang, J.; Huang, R.; Guo, S.; Li, L.; Zhu, M.; Yang, S.; Jiao, L. NAS-Guided Lightweight Multiscale Attention Fusion Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**. [CrossRef]

140. Chen, S.; Zhan, R.; Wang, W.; Zhang, J. Learning Slimming SAR Ship Object Detector through Network Pruning and Knowledge Distillation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1267–1282. [CrossRef]

141. Li, R.; Wang, X.; Wang, J.; Song, Y.; Lei, L. SAR Target Recognition Based on Efficient Fully Convolutional Attention Block CNN. *IEEE Geosci. Remote Sens. Lett.* **2020**. [CrossRef]

142. Qin, J.; Wang, B.; Wu, Y.L.; Lu, Q.; Zhu, H.C. Identifying Pine Wood Nematode Disease Using UAV Images and Deep Learning Algorithms. *Remote Sens.* **2021**, *13*, 162. [CrossRef]

143. de Alwis Pitts, D.A.; So, E. Enhanced change detection index for disaster response, recovery assessment and monitoring of accessibility and open spaces (camp sites). *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *57*, 49–60. [CrossRef]

144. Ghaffarian, S.; Kerle, N.; Pasolli, E.; Jokar Arsanjani, J. Post-Disaster Building Database Updating Using Automated Deep Learning: An Integration of Pre-Disaster OpenStreetMap and Multi-Temporal Satellite Data. *Remote Sens.* **2019**, *11*, 2427. [CrossRef]

145. Kumar, S.; Anouncia, M.; Johnson, S.; Agarwal, A.; Dwivedi, P. Agriculture change detection model using remote sensing images and GIS: Study area Vellore. In Proceedings of the 2012 International Conference on Radar, Communication and Computing (ICRCC), Tiruvannamalai, India, 21–22 December 2012; pp. 54–57.

146. Liu, Q.; Zhou, H.; Xu, Q.; Liu, X.; Wang, Y. PSGAN: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpening. *IEEE Trans. Geosci. Remote Sens.* **2020**. [CrossRef]

147. Web of Science. Available online: www.isiwebofknowledge.com (accessed on 17 March 2021).

148. Choi, Y.; Uh, Y.; Yoo, J.; Ha, J.W. StarGAN v2: Diverse Image Synthesis for Multiple Domains. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 8188–8197.

149. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-Attention Generative Adversarial Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 7354–7363.