



Review

SAR-to-Optical Image Translation and Cloud Removal Based on Conditional Generative Adversarial Networks: Literature Survey, Taxonomy, Evaluation Indicators, Limits and Future Directions

Quan Xiong ¹, Guoqing Li ^{1,*}, Xiaochuang Yao ² and Xiaodong Zhang ²

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

² College of Land Science and Technology, China Agricultural University, Beijing 100083, China

* Correspondence: ligq@aircas.ac.cn

Abstract: Due to the limitation of optical images that their waves cannot penetrate clouds, such images always suffer from cloud contamination, which causes missing information and limitations for subsequent agricultural applications, among others. Synthetic aperture radar (SAR) is able to provide surface information for all times and all weather. Therefore, translating SAR or fusing SAR and optical images to obtain cloud-free optical-like images are ideal ways to solve the cloud contamination issue. In this paper, we investigate the existing literature and provides two kinds of taxonomies, one based on the type of input and the other on the method used. Meanwhile, in this paper, we analyze the advantages and disadvantages while using different data as input. In the last section, we discuss the limitations of these current methods and propose several possible directions for future studies in this field.

Keywords: synthetic aperture radar (SAR); optical images; translation; cloud removal; data fusion; conditional generative adversarial networks



Citation: Xiong, Q.; Li, G.; Yao, X.; Zhang, X. SAR-to-Optical Image Translation and Cloud Removal Based on Conditional Generative Adversarial Networks: Literature Survey, Taxonomy, Evaluation Indicators, Limits and Future Directions. *Remote Sens.* **2023**, *15*, 1137. <https://doi.org/10.3390/rs15041137>

Academic Editor: Alfonso Vitti

Received: 25 November 2022

Revised: 15 February 2023

Accepted: 17 February 2023

Published: 19 February 2023



Copyright: © 2023 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

Free optical imageries such as Sentinel-2 [1], GaoFen 1 (GF-1, GaoFen means high resolution in Chinese) [2], Landsat 8 [3] and Moderate Resolution Imaging Spectroradiometer (MODIS) [4] play essential roles on current agricultural applications, e.g., crop classification [5], cropland monitoring [6], analysis of historical samples [7], and research on arable land-use intensity [8]. However, due to the limitation of wavelength, these optical imageries inevitably suffer from cloud and shadow contamination, which could lead to inefficient earth surface reflectance data needed to conduct subsequent research [5,9].

In order to solve the above problems, researchers have begun to carry out research from two different ideas.

One idea is to fuse multisource [10], multitemporal [11], or multispectral [12] optical imageries. The rationale of this kind of idea is to use cloud-free imageries or bands on the reference phase to recover the missing information on the target phase [13]. The advantage of this method is that we have many multisource optical image data from which to choose. At the same time, each band range of these data is relatively similar, as is the reflectance of ground objects captured, which is more conducive to the collaborative fusion among the data. However, the disadvantage is that it is not effective at capturing reference data if cloud conditions occur continuously [14]. Thus, such methods cannot fundamentally solve the problem of missing information caused by the existence of cloud.

The other idea is to fuse the synthetic aperture radar (SAR) [15] and target optical imageries or use the SAR data only. SAR has a penetration capacity sufficient to capture surface features through cloud [16]. Therefore, in theory, the fusion of SAR and optical

data can solve the deficiency of the first method. However, this approach has its limitations. SAR and optical remote sensing are fundamentally different in imaging principles, resulting in different features of some ground features (e.g., roads, playgrounds and airport runways) in spectral reflectance and the SAR backscattering coefficient. Therefore, it is still difficult to establish a mapping relationship between the two types of data on these ground features [17]. In addition, according to these authors' original goal, we found that this part of the literature can be roughly divided into two categories: the first category is cloud removal on optical remote sensing data, and the second is SAR-to-optical translation. The second type of target was originally intended to facilitate the interpretation of obscure SAR images, but we believe that this type of target can be extended to the production of cloud-free optical remote sensing data. Although, for the second type, cloud removal is not performed on the basis of the original cloud contaminated data, it is similar to the results of the first type, where the cloud-free optical remote sensing data is obtained in the end. The research on cloud removal using SAR data discussed below actually includes these two types.

In addition, in most cases, these two methods will try to select the same space and time according to the target data when selecting reference data [18,19]. In terms of space, the closer the distance, the greater the correlation between ground objects, the greater the distance, and the greater the difference (the First Law of Geography [20]). In terms of time, the closer the interval is, the smaller the change between ground features will be. However, a few studies did not follow the above rules when selecting data. Zhang et al. used the similarity correlation measurement index to measure the correlation of all remote sensing images in the region within a period of time, sorted them according to index level, and selected the image with a high score as the input [21]. The unsupervised Cycle-Consistent Adversarial Networks (CycleGAN) [22] does not need the remote sensing image pairs in the same region. Although it is difficult to obtain the mapping function of styles between two images like supervised learning with this method, it can obtain a mapping relationship between styles through a series of images of two different styles, which is more convenient for the expansion of the dataset [23,24].

We used "TI = "cloud removal" OR AK = "cloud removal"" as retrieving conditions and retrieved all the relevant literature through the Web of Science. At the same time, we used "(TI = "SAR" OR AK = "SAR")AND(TI = "cloud removal" OR AK = "cloud removal")" as retrieving conditions and retrieved the relevant literature through the Web of Science. On this basis, we further searched for the relevant papers through the citations obtained from this body of literature to expand the number of papers. Finally, the literature on optical data cloud removal based on SAR data were collected and sorted (Figure 1). The figure shows that, in studies on optical remote sensing data cloud removal, the first method of integrating optical remote sensing data starts earlier than the second method, and there are more papers on the first method. It can be seen in Figure 1 that the development of cloud removal methods used for optical data based on SAR started in recent years. Once generative adversarial networks (GANs) were constructed, many scholars tried to use this kind of neural network to study SAR-to-optical translation.

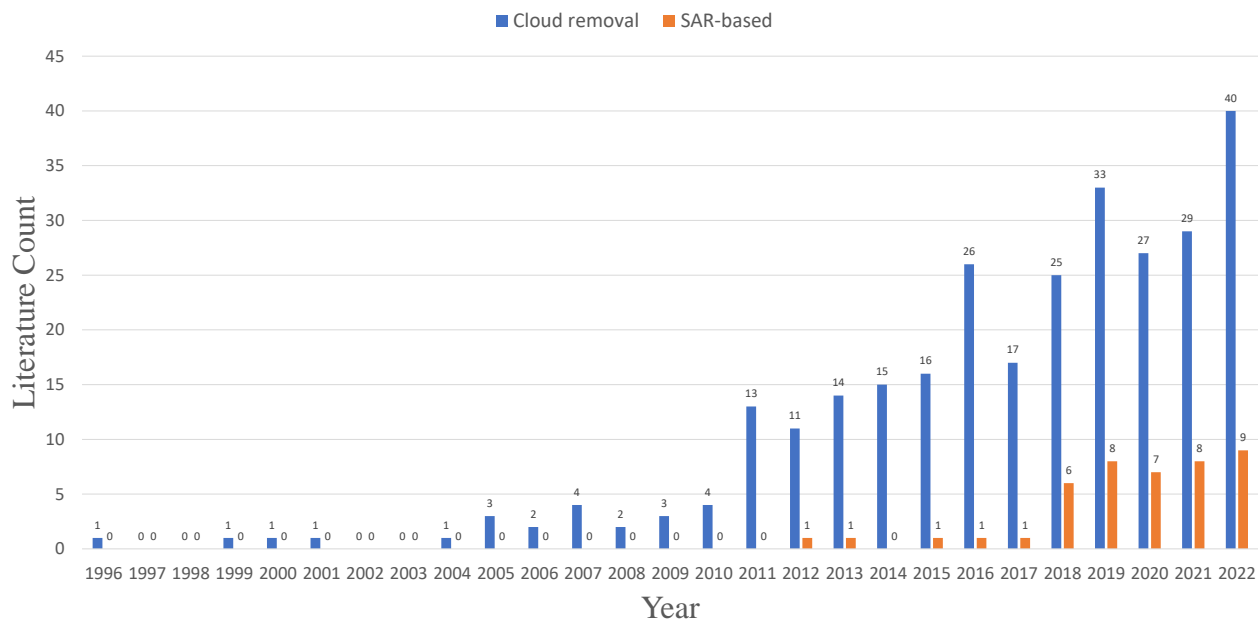


Figure 1. Yearly literature counts of journal papers introducing cloud removal.

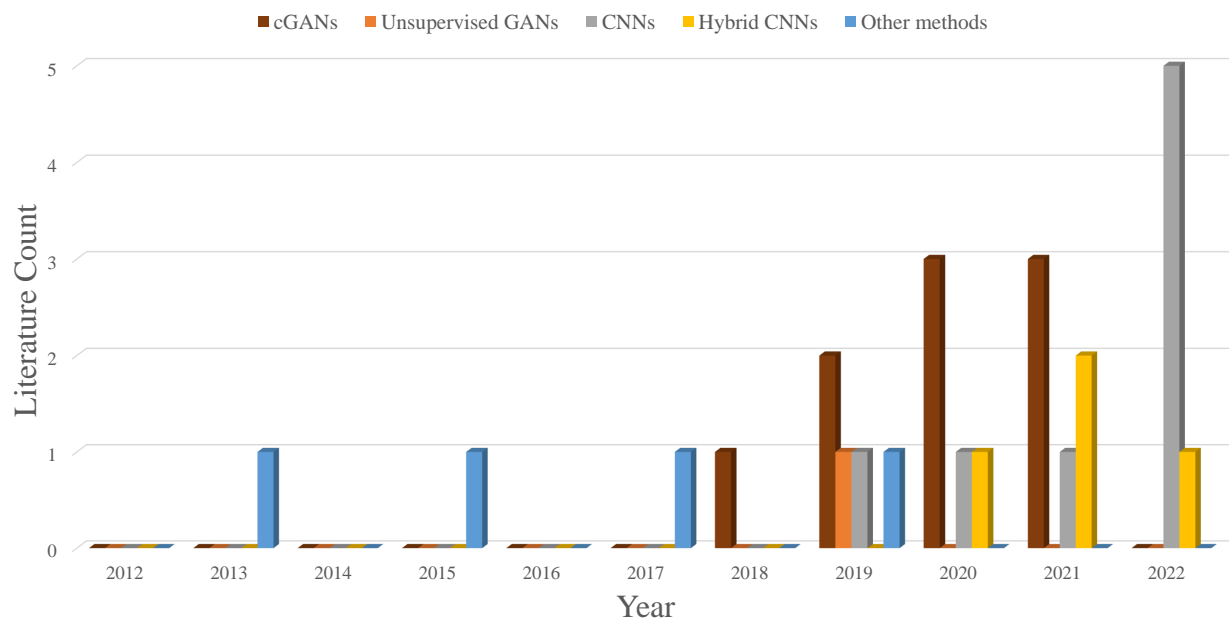
The study of cloud removal using optical remote sensing data alone has taken a long time to develop. So far, there have been a number of relevant literature reviews [25–27], which supports the further development of cloud removal research. However, due to the short development time of the research on cloud removal integrating SAR data, no relevant review can be found. The academic paper published by Fuentes Reyes et al. [28] could be regarded as a relevant review, but its content is more focused on the effect of the CycleGAN network in SAR-to-optical translation. Therefore, we hope that this review can indicate the current cloud removal development of optical remote sensing data using SAR data, especially the emerging methods based on deep learning from 2018.

2. Literature Survey

We researched cloud removal methods for optical imageries using SAR published in journals and conferences. We found 26 papers published in 9 international journals (Table 1), 13 conferences papers [29–41] and 4 papers [42–45] published in arXiv (pronounced “archive”—the X represents the Greek letter chi [χ]), which is an open-access repository of electronic preprints (known as e-prints) approved for posting after moderation but not peer review. Among these 26 journal papers, 14 were published in journals including Remote Sensing and IEEE Transactions on Geoscience and Remote Sensing. 84% of the papers were published after 2019, and the total number of papers is not large, which indicates that the research on cloud removal from optical remote sensing data using SAR data is still at a new stage and still has great potential. We divided these papers into five types of methods, as shown in Section 3.1, and the development of these methods, which are published in journals, is shown in Figure 2.

Table 1. Number of papers published in 9 journals using SAR for cloud removal. The search was conducted on 20 November 2022.

Source Titles	Records	Percentage	Literature
Remote Sensing	9	34.61%	[28,46–53]
IEEE Transactions on Geoscience and Remote Sensing	5	19.23%	[54–58]
IEEE Access	3	11.54%	[17,59,60]
IEEE Geoscience and Remote Sensing Letters	2	7.69%	[61,62]
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	2	7.69%	[63,64]
ISPRS Journal of Photogrammetry and Remote Sensing	2	7.69%	[65,66]
ISPRS International Journal of Geo-Information	1	3.85%	[67]
Journal of the Indian Society of Remote Sensing	1	3.85%	[68]
International Journal of Remote Sensing	1	3.85%	[69]
Total	26	100%	

**Figure 2.** Yearly journal literature counts of different methods for removing clouds from optical data using SAR.

At the same time, in order to help scholars to search for papers with keywords, we carried out text analysis on the 26 papers and extracted the words commonly used in the titles (Figure 3). It can be seen in the figure that different scholars have different named methods for the title of the paper, but the words with the highest frequency are still about cloud removal of optical remote sensing data using SAR data, such as 'SAR', 'cloud', 'optical', 'image' and 'removal'. There are also the keywords about deep learning, especially generative adversarial networks, such as 'learning' and 'adversarial', which indicates that the generative adversarial networks is widely applied in this field. In the following, we will focus on the current situation, opportunities and challenges of using SAR data to decloud optical remote sensing data based on GANs.

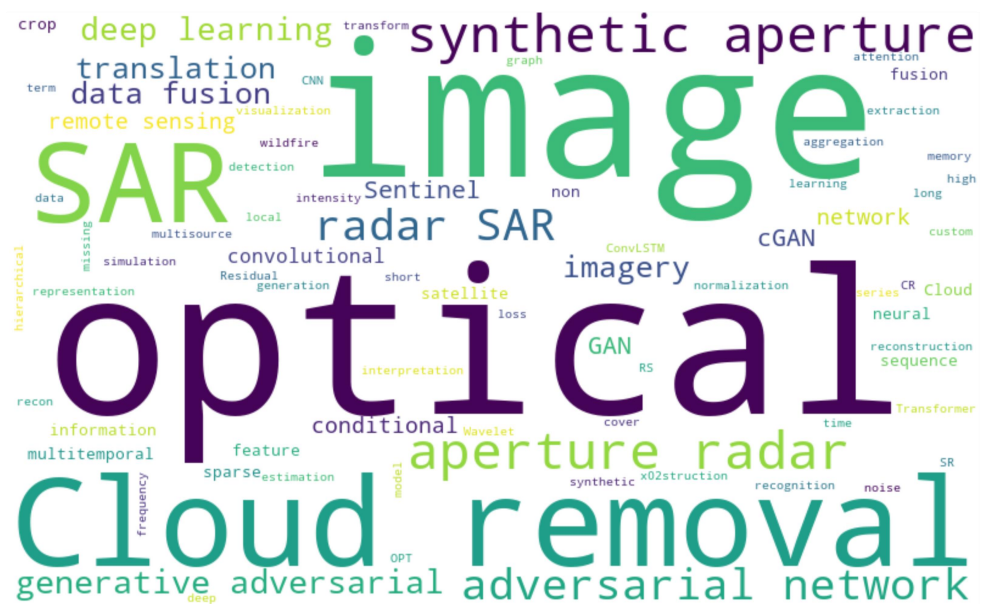


Figure 3. A tag cloud of the literature titles of the 26 journal papers about the cloud removal from optical remote sensing imageries using SAR, where a larger font and darker color denote a higher frequency.

We also plotted a literature citation map of journal papers with VOSViewer, as Figure 4 shows. The total number of papers in the figure is 25, because [66] cannot be searched in the Web of Science Core Collection (WOSCC). The most cited paper is [62], which presents a cloud removal method for reconstructing the missing information in cloud-contaminated regions of a high-resolution optical satellite image using two types of auxiliary images, i.e., a low-resolution optical satellite composite image and a SAR image.

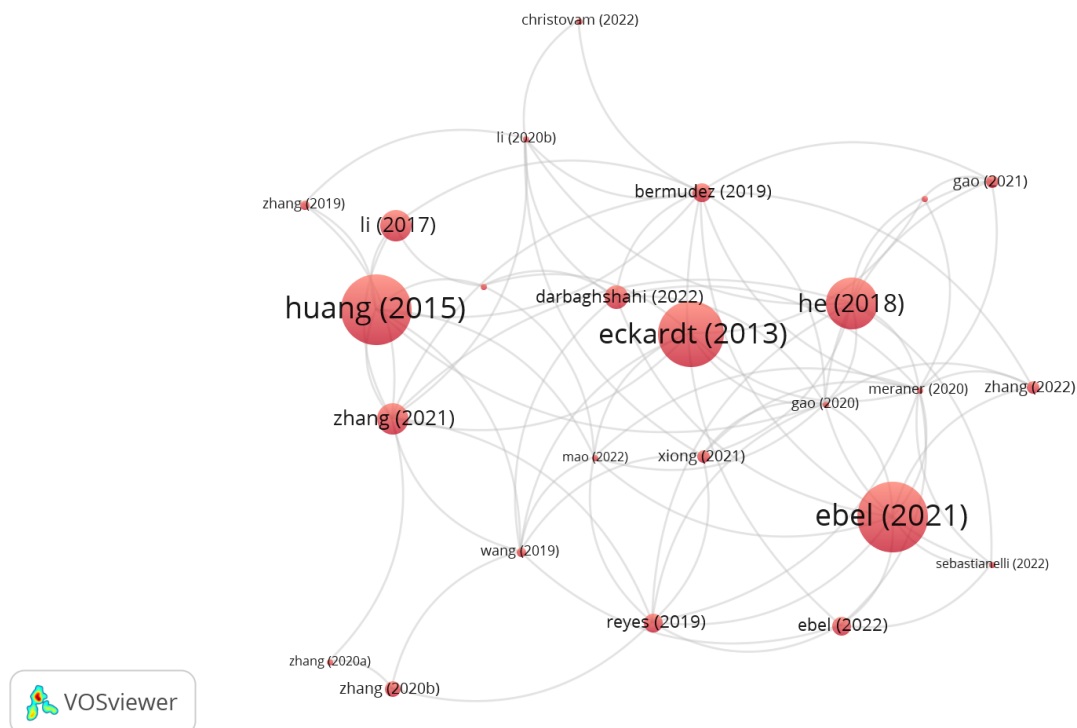


Figure 4. Citation map of the journal literature in Table 1. The size of nodes indicates the count of citations. The citation relationship was derived from Web of Science.

3. Taxonomy

This section classifies a part of the papers based on the methods used and the input data used. It is hoped that this multiangle classification method will enable scholars to have a more diversified understanding of this field.

3.1. The Categories of the Methods

It has been described in the introduction and literature survey that GANs have been widely used in optical remote sensing data decloud research based on SAR data, so we classified the methods into five categories according to GANs: the conditional generative adversarial networks (cGANs), the unsupervised GANs, the convolutional neural networks (CNNs, not GANs), the hybrid CNNs and other methods, as Table 2 shows.

Table 2. Categories of the existing cloud removal methods of optical remote sensing imagery using SAR. (¹ Names of these methods were extracted from the literature. If the methods were not named in the literature, the words describing the methods' major characteristics were used to name them here. cGANs: conditional generative adversarial networks. CNNs: convolutional neural networks.)

Category	Methods ¹ and References
cGANs	modified cGAN-based SAR-to-optical translation method [60]; feature-guided method [59]; MTcGAN [67]; S-CycleGAN [17]; Synthesis of Multispectral Optical Images method [61]
Unsupervised GANs	CycleGAN [28]
CNNs (not GANs)	DSen2-CR [65]; CMD [54]; SEN12MS-CR-TS [55]; G-FAN [48]
Hybrid CNNs	SFGAN [46]; PLFM [56]
Other methods	Sparse Representation [62]; cloud removal based on multifrequency SAR [47]; IHSW [68]; DGL-INJSC [63]

3.1.1. cGANs

The cGAN [70] is an extension of the original GAN [71], which is made of two adversarial neural network modules, the Generator (G) and the Discriminator (D). G attempts to generate the simulated images in domain Y from the realistic images in domain X to fool D. D is to be trained from realistic images in domain X and Y to discriminate the input data between synthetic and realistic. When training the cGAN model, the input has to be paired SAR/Optical data.

The flowchart of the cGAN is shown in Figure 5. The loss of the G comes from two parts, GANLoss and L1Loss. GANLoss expresses whether the simulated images y' are able to fool the D (cause the D to judge the simulated images as real images). L1Loss presents the distance between real images y and simulated images y' . The G tries to make these two loss functions' values close to zero, which means the simulated results are more realistic. Contrarily, the D is in charge of discriminating the real images y as true and the simulated images y' as false, i.e., DLoss. Similar to the G, the D also tries to make the DLoss function's value close to zero to improve its own distinguishing capacity.

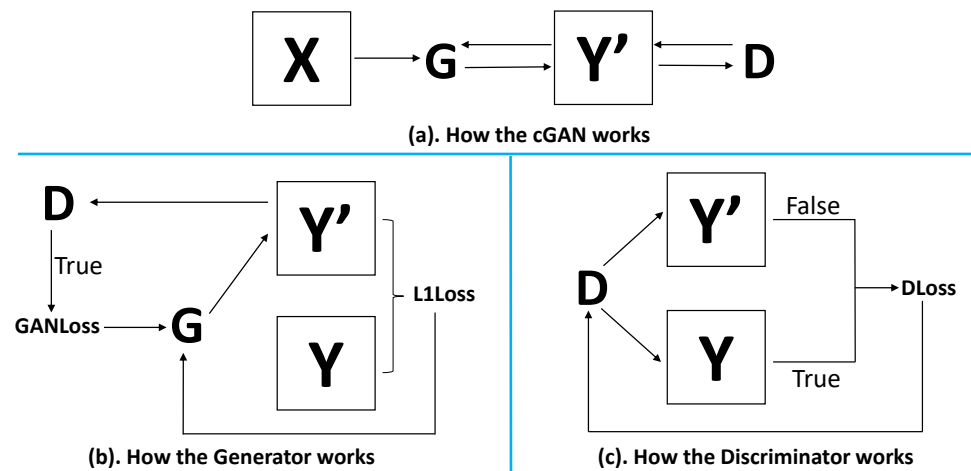


Figure 5. The flowchart of the cGAN (D: Discriminator; G: Generator) [50].

GANLoss is a loss function that digitizes the process of the Generator and the Discriminator, which is literally the most fundamental loss function in GANs. The objective function for cGAN ($\mathcal{L}_{cGAN}(G,D)$) could be expressed in Equation (1):

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y \sim p_{data}(x,y)} [\log D(x,y)] + \mathbb{E}_{x \sim p(x), z \sim p(z)} [\log(1 - D(x, G(x,z)))] \quad (1)$$

where \mathbb{E} and \log are expectation and logarithmic operators, respectively; p is the distribution of images; z is a random noise vector that follows a prior known distribution $p(z)$, typically uniform or Gaussian. Ordinarily, a L1 norm distance loss (L1Loss) would be added to the objective function of the cGAN to cause the simulated images to be more similar to the realistic images, as shown in Equation (2):

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_{L1}(G) \quad (2)$$

where λ is a regularization weight, and \mathcal{L}_{L1} is defined as

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y \sim p_{data}(x,y), z \sim p(z)} [\|y - G(x,z)\|_1]. \quad (3)$$

Some authors [60] have proposed a modified cGAN-based SAR-to-optical translation method that uses the structural similarity index measurement (SSIM) as a loss function to strengthen the connection between the generated image and the target image and produce the generated image with better details. Others [29,33,61,67] have added the channels of cGANs to use the information of the multitemporal paired SAR/Optical data, which improves the quality of the generated image. Some authors [59] have paid more attention to the high-dimension feature extraction with VGG19 [72], which leads to better results. Although S-CycleGAN [17] uses supervised cycle-consistent adversarial networks, the paired SAR/Optical data were utilized as input, and a MSELoss was added to measure the similarity between the generated image and the real image, so we considered S-CycleGAN a cGAN method.

3.1.2. Unsupervised GANs

CycleGAN [22] is a typical unsupervised GAN, and its flowchart is shown in Figure 6. In the absence of truth value data corresponding to generated images, it is impossible for us to use these two data to obtain L1 loss and measure the model effect. Zhu believes that, when there are two Generators (G and F), adversarial training can learn mappings G and F that produce outputs identically distributed as target domains Y and X, respectively. However, the premise is that the direction of the mapping must be restricted. CycleGAN mainly

includes two Generators (G and F) and two Discriminators (D_X and D_Y). As shown in Figure 6a, Generator G is mainly responsible for converting image X to image Y , Generator F is mainly responsible for converting image Y to image X , and the two Discriminators are respectively responsible for judging the generated image as true or false. The whole model operation process can be divided into two directions. The first direction is from image X to image Y , as shown in Figure 6b. Generator G will generate the image Y' based on the image X , and the Discriminator D_Y will then determine the true or false of the generated image Y' . This process is consistent with the conditional generative adversarial network. Generator F will then continue to generate the image X' based on the generated image Y' and finally calculate the loss function using the image X and generated image X' . The second direction is from image Y to image X , as shown in Figure 6c. The process is similar to the first direction. In Figure 6b,c, it can be found that the calculation of the cycle-consistency loss function only needs image X , generated image X' , image Y , and generated image Y' and does not need the paired of image X and image Y , so the whole network is an unsupervised model. The formula of cycle consistency loss is shown in Equation (4). The full objective is shown in Equation (5):

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \quad (4)$$

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F) \quad (5)$$

where D_X and D_Y are two adversarial discriminators, D_X aims to distinguish between images X and generated images $F(y)$, D_Y aims to distinguish between images Y and generated images $G(x)$, and λ controls the relative importance of the two objectives. The final aim is shown in Equation (6):

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y). \quad (6)$$

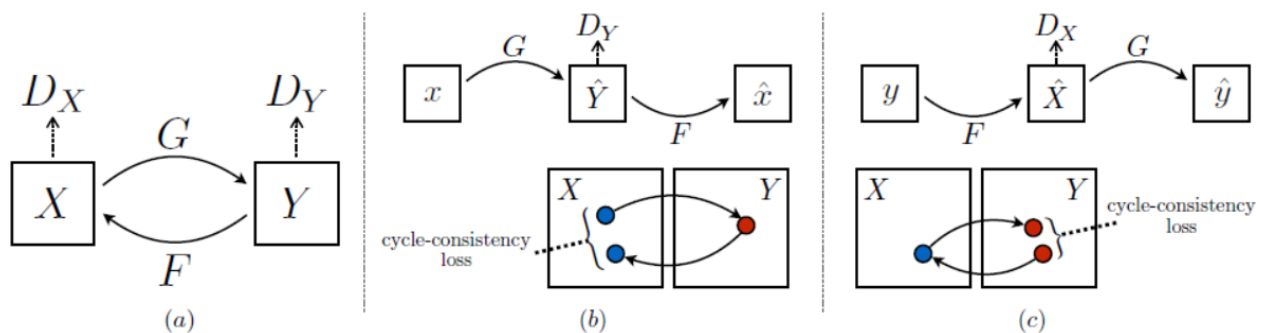


Figure 6. The flowchart of CycleGAN [22]. (a) The whole model operation process. (b) Direction from image X to image Y . (c) Direction from image Y to image X .

In [28], the author tried to use CycleGAN for SAR-to-optical translation (TerraSAR-X satellite to ALOS PRISM satellite and Sentinel-1 to Sentinel-2). All images were from 13 European cities spread over seven counties. The results showed that CycleGAN is able to take advantage of the large datasets acquired by satellite and has potential for SAR-to-optical image translation. Compared to urban areas, CycleGAN is best used for rural or semiurban areas, as it can recover missing information in mountain or water areas well. It is difficult to isolate walls and buildings, which can be blurred or merged into larger elements.

3.1.3. CNNs

The biggest difference between CNNs and the two methods mentioned above lies in whether there is an adversarial network. Of course, the main network structure is also

different. To some extent, the adversarial network is actually a way of extracting high-dimensional features, and it ultimately acts upon the network model training through a loss function [54]. However, the ordinary loss function is a fixed formula that will not change, while the adversarial neural network can constantly change the model formula through continuous learning. That is to say, as long as the structure of the main network is designed reasonably, and the design of the loss function can measure the training effect of the model, it is unnecessary to use an adversarial neural network, as was done in [54,65].

In [65], the author used a residual neural network as the main network structure to recover the missing information of Sentinel-2 images by fusing Sentinel-1 images and all bands of the cloud-contaminated Sentinel-2 images. Meanwhile, the author proposed a new loss function, Cloud-Adaptive Regularized Loss (\mathcal{L}_{CARL}), to estimate the recovery accuracy of the original cloud area and the original cloud-free area in the generated image. The formulation is shown in Equation (7):

$$\mathcal{L}_{CARL} = \frac{\|\mathbf{CSM} \odot (\mathbf{P} - \mathbf{T}) + (\mathbf{1} - \mathbf{CSM}) \odot (\mathbf{P} - \mathbf{I})\|_1}{N_{tot}} + \lambda \frac{\|\mathbf{P} - \mathbf{T}\|_1}{N_{tot}} \quad (7)$$

where \mathbf{P} , \mathbf{T} and \mathbf{I} denote the predicted, target and input optical images, respectively. The \mathbf{CSM} (cloud-shadow mask) has the same spatial property of the optical images. The pixel value 1 represents the cloud or cloud shadow and pixel value 0 for the cloud-free area. $\mathbf{1}$ is a matrix with values of 1 and has the same spatial dimensions as the optical images. N_{tot} is the total number of pixels in all bands of the optical images. Compared with the commonly used loss function to calculate the effect of cloud removal, this loss function takes into account two different regional attributes of the image: cloud and cloudless.

3.1.4. Hybrid CNNs

Hybrid CNNs can mainly be divided into two stages, and a typical flowchart is shown in Figure 7. The first stage is the cloudless image simulation stage, and the second stage is to fuse the output of the first stage with the data contaminated by the cloud to obtain the final cloud-free optical remote sensing image data. In this process, adversarial neural network can be used at one stage [46] or both stages [35] or none at all. The advantage of the hybrid network is that it can not only save the cloud-free area information of the original cloud-contaminated data as much as possible but also use the multisource data to recover the information of the cloud area.

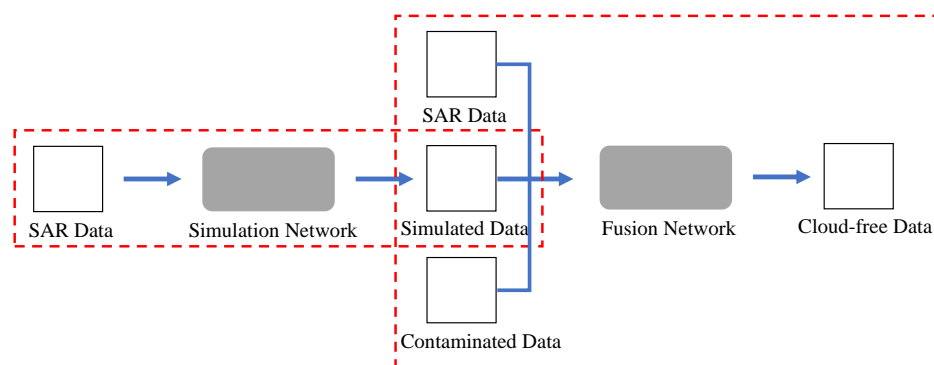


Figure 7. The flowchart of typical Hybrid CNNs.

In order to train the network model more specifically, one author used two new loss functions [46]. The first is the perceptual loss function, which is used to judge the accuracy of the model output from the high-dimensional feature level. The second is the local loss function, which is used to dynamically judge the data recovery accuracy of the cloud-polluted area. For the perceptual loss function, the author tried to extract feature maps of

the fusion result and the ground truth from the 8th layer of VGG16 [73]. The formula is shown as Equation (8):

$$\mathcal{L}_{perc} = \|v_{gg8}(F) - v_{gg8}(GT)\|_1 \quad (8)$$

where F stands for the fusion result, and GT stands for the ground truth data. The formula of the local loss function is shown as Equation (9):

$$\mathcal{L}_{local} = \|M \cdot F - M \cdot GT\|_1 \quad (9)$$

where M is the cloud mask, the author adapted MSCFF (multiscale convolutional feature fusion) [74], which is based on deep learning, to dynamically obtain the cloud mask of each training image of each batch.

3.1.5. Other Methods

In addition to the above four types of methods, there are also studies on SAR-to-optical or optical image cloud removal, which are classified as other methods. For example, Sparse Dictionary Learning was used to remove cloud pollution in optical data by using SAR data [62,63]. Eckardt et al. [47] combined the Closest Spectral Fit (CSF) [75] algorithm with multispectral images and multifrequency SAR data to remove the thick cloud of optical images. Fu et al. [34] tried to build a new concept of remote sensing image translation based on image understanding, target transform and representation. Zhang et al. [68] adapted the intensity–hue–saturation and wavelet transform integrated fusion for SAR-to-optical image translation.

3.2. The Categories of the Input

In addition to the method and the loss function used, we consider the input data to be important in the modeling. Therefore, we classified the above-mentioned literature according to the type of input data combination, as shown in Table 3. We hope that readers will be able to see the advantages and disadvantages of different input data in this subsection. Moreover, we found several papers [17,55,57,76] on the dataset for the fusion of SAR and optical image data, which are widely cited and used by other papers. In these papers, Sentinel-1 is the most popular source of SAR data. Ground-range-detected (GRD) products acquired in the most frequently available interferometric wide swath (IW) mode are used. They contain VV and VH polarized channels, and the resolution is 10 m.

Table 3. Different inputs of the existing cloud removal methods of optical remote sensing imagery using SAR. (¹ Names of these methods were extracted from the literature. If the methods were not named in the literature, the words describing the methods' major characteristics were used to name them here. cGANs: conditional generative adversarial networks. CNNs: convolutional neural networks.)

Input	Methods ¹ and References
monotemporal SAR	modified cGAN-based SAR-to-optical translation method [60]; feature-guided method [59]; DGL-INJSC [63]; CycleGAN [28]; S-CycleGAN [17]; IHSW [68]
monotemporal SAR & corrupted Optical	Sparse Representation [62]; DSen2-CR [65]; SFGAN [46]; CMD [54]; PLFM [56]; cloud removal based on multifrequency SAR [47]; SEN12MS-CR-TS [55]; G-FAN [48]
multi-temporal SAR & Optical	MTcGAN [67]; Synthesis of Multispectral Optical Images method [61]

3.2.1. Mono-Temporal SAR

The first type of input data in the test stage is monotemporal SAR data, which is the commonly used input data in the traditional SAR-to-optical image translation field [39,60]. In the training stage, we often need to provide SAR data and optical remote sensing data in pairs on a single time phase [31]. SAR data is mainly used to provide feature information, while optical remote sensing data is mainly used to calculate the loss function together with the model output. Moreover, although CycleGAN also needs to input SAR and optical remote sensing data, it does not need to have the same spatiotemporal properties [28]. The advantages of using monotemporal SAR data to generate cloud-free optical remote sensing images are that there are fewer restrictions on the acquisition of input data and that the production of the training dataset is relatively simple. At the same time, this method can also increase the amount of S2 data in the time dimension. However, due to the essential imaging differences between radar data and optical data, the reflection information of some ground objects on these two types of data is different, so the accuracy of using SAR data alone to acquire the optical image on some ground objects cannot be guaranteed [17]. At the same time, this method requires a large number of samples for the model to fully learn the conversion relationship between SAR and optical remote sensing data.

3.2.2. Monotemporal SAR and Corrupted Optical Images

The second type of input data is the combination of monotemporal SAR data and cloud contaminated optical remote sensing data [30]. Unlike the first method, corrupted optical data is added in this method, which helps the model to assist the recovery of the target area according to the feature of the adjacent areas of the cloud area; at the same time, the output results can be more consistent with the real situation [65]. However, since the recovery of the target region (especially in the case of thick clouds) mainly relies on the SAR data, the accuracy of the SAR-to-optical translation process still has a great impact on the results, which is similar to the first method [46].

3.2.3. Multitemporal SAR and Optical Images

The last type is multitemporal SAR and optical image data pairs. In this way, the model learns the conversion relationship between SAR and optical remote sensing data at the reference time and then generates the target optical remote sensing image according to the feature of the SAR data at the target time [67]. The advantage of this method is that the main object distribution of the target image can be roughly obtained from the optical image reference time data, and the reflectance of the target image can then be inferred by using the change of the target time and reference time of the SAR data [33]. However, the problem is that the features of ground objects in reference time and target time cannot change much; otherwise, the accuracy cannot be guaranteed.

4. Evaluation Indicators

How the accuracy of a generated cloud-free image is evaluated is also an important issue for the relevant research. In this section, we counted the existing literature one by one and obtained several evaluation indexes that scholars like to use, as shown in Table 4.

Table 4. Different input of the existing cloud removal methods of optical remote sensing imagery using SAR.

Indicators	References	Frequency
Structural Similarity Index Measurement (SSIM)	[17,46,48,54–56,59,60,62,65,67,68]	12
Peak Signal-to-Noise Ratio (PSNR)	[17,48,54–56,59,60,62,65,67]	10
Spectral Angle Mapper (SAM)	[46–48,55,56,65,68]	7
Root Mean Square Error (RMSE)	[46,55,56,62,65,68]	6
Correlation Coefficient (CC)	[46,48,56,68]	4
Feature Similarity Index Measurement (FSIM)	[17,59]	2
Universal Image Quality Index (UIQI)	[56,68]	2
Mean Absolute Error (MAE)	[48,65]	2
Mean Square Error (MSE)	[59]	1
Degree of Distortion (DD)	[56]	1

The most frequently used indicator is Structural Similarity Index Measurement (SSIM) [77]. In terms of image composition, SSIM defines structural information as independent of brightness and contrast, reflecting the object structure property in the scene, and modeling distortion as a combination of brightness, contrast and structure. The mean value, standard deviation and covariance were used to estimate the luminance, contrast, and structural similarity. The formula is shown as Equation (10).

$$SSIM = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)} \quad (10)$$

where $\sigma_{y\hat{y}}$ is the covariance between the realistic and simulated values, and C_1 and C_2 are the constants used to enhance the stability of SSIM. SSIM ranges from -1 to 1 ; the closer the SSIM is to 1 , the more accurate the simulated image is.

The second indicator is the Peak Signal-to-Noise Ratio (PSNR) [78], which is a traditional image quality assessment (IQA) index. A higher PSNR generally indicates that the image is of higher quality. The formula can be defined as

$$PSNR = 10 \lg\left(\frac{255^2}{MSE}\right) \quad (11)$$

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N} \quad (12)$$

where y_i and \hat{y}_i are the realistic and simulated values for the i th pixel, respectively, N is the number of pixels, and MSE is the mean square error.

The third indicator is Spectral Angle Mapper (SAM), which was proposed by Kruse et al. [79] in 1993, which regards the spectrum of each pixel in an image as a high-dimensional vector and measures the similarity between the spectra by calculating the included angle between two vectors. The smaller the included angle, the more similar the two spectra are.

The fourth indicator is the Root Mean Square Error (RMSE) [80], which is used to measure the difference between the actual and simulated values. The formula is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (13)$$

where y_i and \hat{y}_i are the realistic and simulated values for the i th pixel, respectively, and N is the number of pixels. The smaller the RMSE is, the more similar the two images are.

In addition to the above four evaluation indicators, scholars also use the following indicators to evaluate accuracy: Correlation Coefficient (CC) [81], Feature Similarity Index

Measurement (FSIM) [82], Universal Image Quality Index (UIQI) [83], Mean Absolute Error (MAE) [80], Mean Square Error (MSE) [84] and Degree of Distortion (DD) [56].

5. Limits and Future Directions

On the basis of this literature review, we analyze the problems of existing methods and propose the following six directions for future research.

5.1. Image Pixel Depth

Remote sensing data present the reflectivity information of surface objects, which are recorded continuously from 0 to 1. Considering that the storage of float data will consume a high amount of storage resources, we usually multiply the reflectivity by 10,000 and retain an integer from 1 to 10,000. The storage format of remote sensing data is usually GeoTIFF, but the image input format used for deep learning is usually PNG format, an 8-bit unsigned integer with a pixel depth of 0–255. Therefore, in order to ingest remote sensing data into the deep learning model smoothly, most scholars adopt the method of saving remote sensing images in PNG format.

However, this approach is similar to a lossy compression, with a loss rate of spectral information reaching 97.44%, and most of the spectral information is lost, which will lead to two problems. First, the amount of information obtained by the model is far lower than the amount of information that the data itself can provide, which is not conducive to the model's feature extraction of spectral information, and a large amount of information can easily be misclassified, which is not conducive to finding the relationship between the backscattering coefficient and spectral reflectance. This may lead to poor cloud removal effect. Second, the final output results of the model can only be images with pixel values of 0–255. These results are acceptable for simple visual inspection, but cannot be used for subsequent quantitative calculations of the model, because the information loss is too large.

In fact, this problem is not limited to the field of cloud removal studied in this paper, but also exists in other fields that use deep learning to study remote sensing data. We do not recommend choosing to lose the accuracy of remote sensing data itself in order to blindly put the data into a deep learning model, which may cause substantial interference to our research. We suggest that the GDAL library be used to modify the data loading module and data output module of the deep learning model, instead of the usual image format of deep learning, which will enable the deep learning model to directly read and generate data with a higher pixel depth.

5.2. Number of Image Bands

The essential reason for this problem is the same as that in Section 5.1, which is caused by the inconsistency between the professional storage format used for remote sensing data and the data reading and writing format commonly used for deep learning. Optical remote sensing data capture the reflectivity information of different electromagnetic bands: one channel of a panchromatic image, multiple channels of multispectral data and dozens or hundreds of hyperspectral data channels. SAR data also have different combinations according to different polarization modes. The PNG data format commonly used for deep learning can only accommodate three channels, so sometimes optical image data and SAR data are forced to delete the remaining channels or repeat an additional channel, so that the channel number of the image data can reach the three channels required by the PNG data format.

When the remaining channels are deleted, some spectral information will be lost, which causes the features that can be learned by deep learning to be deleted artificially during data processing from the beginning, which is not conducive to the training of cloud removal models. In particular, the impact of clouds is different on different bands. Using the fusion of different bands is an effective way to remove thin clouds from images. However, in order to meet the requirements of PNG format, the remaining bands can only be deleted. We suggest that the GDAL library be used to modify the data loading module

and data output module of the deep learning model, instead of the usual image format of deep learning with three channels, which will enable the deep learning model to directly read and generate data with proper channel amounts.

5.3. Global Training Datasets

At present, most research involves the use of a small amount of data in the local area for model training. However, for deep learning models such as generative adversarial networks, many training samples are required for the model to learn knowledge, and sufficient verification samples are also required to evaluate the model effect. However, due to the large repetitive workload and low innovation in dataset production, there are still few publicly available and downloadable datasets for SAR and optical remote sensing image fusion. Therefore, the global regional dataset for optical image cloud removal based on SAR data will be important for research in this field. This dataset needs to have the following characteristics:

1. it involves multiple regions of interests in the world,
2. it has paired SAR and optical data,
3. it is a multi-temporal series,
4. it is in GeoTIFF format, and
5. the target image which is corrupted by clouds and its cloud free image.

First, different regions of the world must be covered as much as possible in the dataset, so that the model has the opportunity to learn features in different terrains and regions. In this way, the generalization ability of the trained model can be stronger. Second, SAR and optical image data must be fused, so the paired datasets are essential. Third, the model must have a chance to learn the characteristics of the time dimension, which is conducive to further improving the accuracy of the model. Fourth, the feature is to ensure that the dataset has a relatively complete information feature, which was mentioned in the previous two sections. Last, there must be available verification data in the process of cloud removal research, so that the model can be continuously trained and optimized.

5.4. Accuracy Verification for Cloud Regions

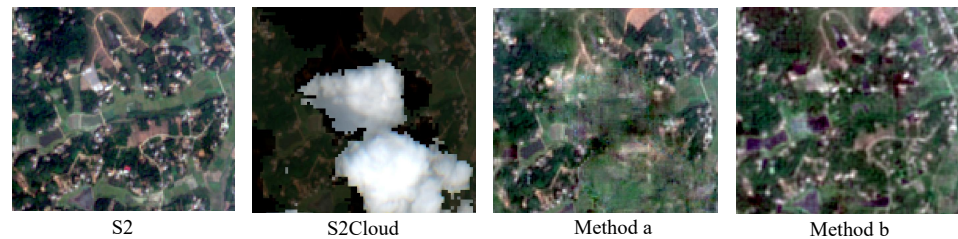
At present, the vast majority of research is to evaluate the effect of cloud removal for whole scene data, and the evaluation result is the average result of the whole scene image. However, for cloud removal applications, it is key to recover the areas polluted by clouds or cloud shadows, and the original pixel information in the cloud image to be removed can be used in other nonpolluted areas. Therefore, for cloud removal research, there are certain limitations in judging model accuracy from the whole scene image range, and it is necessary to try to conduct targeted accuracy evaluations in the areas polluted by clouds or cloud shadows.

For example, Table 5 records the precision comparison of whole scene results of two SAR-based cloud removal methods. Method A has obvious advantages, but when we view the local cloud removal effect, as shown in Figure 8, it is obvious that the cloud removal effect of Method B is better, so it is difficult to evaluate the quality of cloud removal only from the evaluation indicators of the whole scene results. Therefore, we need to extract the area to be cloud-removed separately, and we need to calculate the indicators to evaluate the accuracy of the cloud removal effect, which is more objective.

Table 5. Evaluation of cloud removal results of the whole scene image.

	Method A	Method B
RMSE	0.0067	0.0094
R ²	0.9596	0.9256
KGE	0.9224	0.9256
SSIM	0.9948	0.9943
PSNR	54.4722	54.1575
SAM	0.0869	0.0940

The bold result is the best among the different metrics.

**Figure 8.** Contrast in local details.

5.5. Auxiliary Data

The input data of most research are SAR and optical image data, but few studies have paid attention to the value of cloud mask data in this field. At present, many optical data are provided with corresponding cloud mask data when downloading, and there are also many open-source cloud detection model algorithms, making cloud mask data easier to obtain. It is a worthwhile idea to input cloud mask data into the model, which also allows the model to obtain cloud distribution in advance, so as to conduct more targeted cloud removal operations.

Furthermore, we can enrich the types of cloud mask data. In addition to cloud and cloudless, we can further distinguish thin clouds, thick clouds, cloud shadow and other areas and cooperate with the loss function so that the results of the model are more satisfactory. Compared to the \mathcal{L}_{CARL} (7), the new loss function is able to measure and feed back more information to improve the accuracy. The equation can be defined as

$$\mathcal{L}_{NEW} = \left(\frac{\|\mathbf{TCM1} \odot (\mathbf{P} - \mathbf{T})\|_1}{N_{tot}} + \lambda_1 \frac{\|\mathbf{TCM2} \odot (\mathbf{P} - \mathbf{T})\|_1}{N_{tot}} + \lambda_2 \frac{\|\mathbf{SM} \odot (\mathbf{P} - \mathbf{T})\|_1}{N_{tot}} + \lambda_3 \frac{\|\mathbf{CSM} \odot (\mathbf{P} - \mathbf{T}) + (\mathbf{1} - \mathbf{CSM}) \odot (\mathbf{P} - \mathbf{I})\|_1}{N_{tot}} + \lambda_4 \frac{\|\mathbf{P} - \mathbf{T}\|_1}{N_{tot}} \right) \quad (14)$$

where \mathbf{P} , \mathbf{T} and \mathbf{I} denote the predicted, target and input optical images, respectively. The **TCM1** (thick cloud mask) has the same spatial property of the optical images. The pixel value 1 represents thick cloud, and the pixel value 0 represents other areas. The **TCM2** (thin cloud mask) has the same spatial property of the optical images. The pixel value 1 represents thin cloud, and the pixel value 0 represents other areas. The **SM** (shadow mask) has the same spatial property of the optical images. The pixel value 1 represents shadow, and the pixel value 0 represents other areas. The **CSM** (cloud-shadow mask) has the same spatial property of the optical images. The pixel value 1 represents cloud or cloud shadow, and the pixel value 0 represents cloud-free areas. $\mathbf{1}$ is a matrix with values of 1 and has the same spatial dimensions as the optical images. N_{tot} is the total number of pixels in all bands of the optical images.

5.6. Loss Functions

The current studies use a variety of precision indicators to verify the results from multiple dimensions, as shown in Table 4. However, this can only be used for the final quantitative evaluation of the results and cannot greatly help the optimization of the model.

At the same time, most studies use relatively simple L1 loss or L2 loss for the entire space to calculate the loss function. Next, we can try to put these multidimensional indicators, such as SAM, into the model optimization stage (i.e., the loss function calculation) to see whether they can help the model to optimize training from different dimensions.

It is worth noting that, due to the matrix calculation used in the loss function calculation, we need to modify these commonly used precision index calculation formulas so that they conform to the matrix calculation method, which will involve some mathematical knowledge and programming knowledge. Of course, in theory, we can also directly put the precision index calculation formula into the loss function without modification for calculation, but this will greatly increase the training time, and the initial estimation time consumption may be 1000 times higher, so it is not recommended.

6. Conclusions

Through literature retrieval and analysis, in this paper, we find that the research of optical image cloud removal based on SAR data is a valuable research direction. Compared to traditional methods using optical remote sensing data alone to remove clouds, the research on cloud removal integrating SAR data has risen in recent years. After the emergence of deep learning, more scholars began studying cloud removal by SAR and optical data fusion. Therefore, there is a lack of review papers on the research of optical image cloud removal using SAR data. We hope that, through this paper, more relevant research will be carried out to understand the development of this study field and to communicate the advances with more scholars.

In this paper, we present the main contributing journals, keywords, fundamental literature and other research in the field of optical image cloud removal using SAR data. It will be helpful for scholars to search and study the relevant literature in this field. Scholars can read nearly 54% of the literature in this field from two sources: Remote Sensing and IEEE Transactions on Geoscience and Remote Sensing.

In this paper, we summarize the relevant literature from two dimensions: research methods and data input. We classify the research methods into five categories according to GANs: conditional generative adversarial networks (cGANs), unsupervised GANs, convolutional neural networks (CNNs, not GANs), hybrid CNNs and other methods, as Table 2 shows. We provide the general principles and loss functions of these methods to help scholars understand them. At the same time, we describe the input data used by these methods and summarize three types of data input, as shown in Table 3, to help scholars understand the data and provide advance preparation for studying these methods.

Moreover, this paper documents the accuracy verification indicators used in the current mainstream literature, which can help subsequent scholars make appropriate choices. The results show that the most popular indicators are SSIM, PSNR, SAM and RMSE. We believe that these four indicators are good choices for relevant scholars to quantitatively verify the accuracy of their models.

Finally, this paper discusses the key points of the future development of this field in terms of six aspects: image pixel depth, the number of image bands, global training datasets, accuracy verification for cloud regions, auxiliary data and loss functions. We discuss the current problems in these six areas and provide our solutions. We hope these problems and solutions can inspire scholars and promote the development of optical image cloud removal using SAR data.

Author Contributions: Q.X., G.L. and X.Z. conceived and designed the paper; Q.X. and X.Y. wrote the paper; Q.X. and G.L. realized the visualization; Q.X. collected data and processed them. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [Special Research Assistant Project of CAS].

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

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

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Synthetic Aperture Radar
GAN	Generative Adversarial Network
cGAN	Conditional Generative Adversarial Network

References

1. Desnos, Y.L.; Borgeaud, M.; Doherty, M.; Rast, M.; Liebig, V. The European Space Agency's Earth Observation Program. *IEEE Geosci. Remote Sens. Mag.* **2014**, *2*, 37–46. [\[CrossRef\]](#)
2. Xiong, Q.; Wang, Y.; Liu, D.; Ye, S.; Du, Z.; Liu, W.; Huang, J.; Su, W.; Zhu, D.; Yao, X.; et al. A Cloud Detection Approach Based on Hybrid Multispectral Features with Dynamic Thresholds for GF-1 Remote Sensing Images. *Remote Sens.* **2020**, *12*, 450. [\[CrossRef\]](#)
3. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [\[CrossRef\]](#)
4. Friedl, M.A.; McIver, D.K.; Hodges, J.C.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A.; et al. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens. Environ.* **2002**, *83*, 287–302. [\[CrossRef\]](#)
5. Yang, N.; Liu, D.; Feng, Q.; Xiong, Q.; Zhang, L.; Ren, T.; Zhao, Y.; Zhu, D.; Huang, J. Large-scale crop mapping based on machine learning and parallel computation with grids. *Remote Sens.* **2019**, *11*, 1500. [\[CrossRef\]](#)
6. Bontemps, S.; Arias, M.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Matton, N.; Morin, D.; Popescu, R.; et al. Building a data set over 12 globally distributed sites to support the development of agriculture monitoring applications with Sentinel-2. *Remote Sens.* **2015**, *7*, 16062–16090. [\[CrossRef\]](#)
7. Zhang, L.; Liu, Z.; Liu, D.; Xiong, Q.; Yang, N.; Ren, T.; Zhang, C.; Zhang, X.; Li, S. Crop Mapping Based on Historical Samples and New Training Samples Generation in Heilongjiang Province, China. *Sustainability* **2019**, *11*, 5052. [\[CrossRef\]](#)
8. Ye, S.; Song, C.; Shen, S.; Gao, P.; Cheng, C.; Cheng, F.; Wan, C.; Zhu, D. Spatial pattern of arable land-use intensity in China. *Land Use Policy* **2020**, *99*, 104845. [\[CrossRef\]](#)
9. Cao, R.; Chen, Y.; Chen, J.; Zhu, X.; Shen, M. Thick cloud removal in Landsat images based on autoregression of Landsat time-series data. *Remote Sens. Environ.* **2020**, *249*, 112001. [\[CrossRef\]](#)
10. Zhang, Q.; Yuan, Q.; Shen, H.; Zhang, L. A Unified Spatial-Temporal-Spectral Learning Framework for Reconstructing Missing Data in Remote Sensing Images. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4981–4984.
11. Lin, C.H.; Tsai, P.H.; Lai, K.H.; Chen, J.Y. Cloud removal from multitemporal satellite images using information cloning. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, 232–241. [\[CrossRef\]](#)
12. Enomoto, K.; Sakurada, K.; Wang, W.; Fukui, H.; Matsuoka, M.; Nakamura, R.; Kawaguchi, N. Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 48–56.
13. Das, S.; Das, P.; Roy, B.R. Cloud detection and cloud removal of satellite image—A case study. In *Trends in Communication, Cloud, and Big Data*; Springer: Singapore, 2020; pp. 53–63.
14. Li, J.; Wu, Z.; Hu, Z.; Zhang, J.; Li, M.; Mo, L.; Molinier, M. Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 373–389. [\[CrossRef\]](#)
15. Feng, Q.; Yang, J.; Zhu, D.; Liu, J.; Guo, H.; Bayartungalag, B.; Li, B. Integrating multitemporal sentinel-1/2 data for coastal land cover classification using a multibranch convolutional neural network: A case of the Yellow River Delta. *Remote Sens.* **2019**, *11*, 1006. [\[CrossRef\]](#)
16. Wang, J.; Yang, X.; Yang, X.; Jia, L.; Fang, S. Unsupervised change detection between SAR images based on hypergraphs. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 61–72. [\[CrossRef\]](#)
17. Wang, L.; Xu, X.; Yu, Y.; Yang, R.; Gui, R.; Xu, Z.; Pu, F. SAR-to-optical image translation using supervised cycle-consistent adversarial networks. *IEEE Access* **2019**, *7*, 129136–129149. [\[CrossRef\]](#)
18. Singh, P.; Komodakis, N. Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1772–1775.
19. Xu, M.; Jia, X.; Pickering, M.; Plaza, A.J. Cloud removal based on sparse representation via multitemporal dictionary learning. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2998–3006. [\[CrossRef\]](#)
20. Tobler, W.R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **1970**, *46*, 234–240. [\[CrossRef\]](#)
21. Zhang, Q.; Yuan, Q.; Li, J.; Li, Z.; Shen, H.; Zhang, L. Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 148–160. [\[CrossRef\]](#)

22. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
23. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
24. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797.
25. Mostafa, Y. A review on various shadow detection and compensation techniques in remote sensing images. *Can. J. Remote Sens.* **2017**, *43*, 545–562. [[CrossRef](#)]
26. Shen, H.; Li, X.; Cheng, Q.; Zeng, C.; Yang, G.; Li, H.; Zhang, L. Missing information reconstruction of remote sensing data: A technical review. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 61–85. [[CrossRef](#)]
27. Shahtahmassebi, A.; Yang, N.; Wang, K.; Moore, N.; Shen, Z. Review of shadow detection and de-shadowing methods in remote sensing. *Chin. Geogr. Sci.* **2013**, *23*, 403–420. [[CrossRef](#)]
28. Fuentes Reyes, M.; Auer, S.; Merkle, N.; Henry, C.; Schmitt, M. Sar-to-optical image translation based on conditional generative adversarial networks—Optimization, opportunities and limits. *Remote Sens.* **2019**, *11*, 2067. [[CrossRef](#)]
29. Xia, Y.; Zhang, H.; Zhang, L.; Fan, Z. Cloud Removal of Optical Remote Sensing Imagery with Multitemporal Sar-Optical Data Using X-Mtgan. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3396–3399.
30. Grohnfeldt, C.; Schmitt, M.; Zhu, X. A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1726–1729.
31. Enomoto, K.; Sakurada, K.; Wang, W.; Kawaguchi, N.; Matsuoka, M.; Nakamura, R. Image translation between SAR and optical imagery with generative adversarial nets. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1752–1755.
32. Ebel, P.; Schmitt, M.; Zhu, X.X. Cloud removal in unpaired Sentinel-2 imagery using cycle-consistent GAN and SAR-optical data fusion. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 2065–2068.
33. Bermudez, J.; Happ, P.; Oliveira, D.; Feitosa, R. SAR TO OPTICAL IMAGE SYNTHESIS FOR CLOUD REMOVAL WITH GENERATIVE ADVERSARIAL NETWORKS. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *4*, 5–11. [[CrossRef](#)]
34. Fu, Z.; Zhang, W. Research on image translation between SAR and optical imagery. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *1*, 273–278. [[CrossRef](#)]
35. Gao, J.; Zhang, H.; Yuan, Q. Cloud removal with fusion of SAR and Optical Images by Deep Learning. In Proceedings of the 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Shanghai, China, 5–7 August 2019; pp. 1–3.
36. Zhu, C.; Zhao, Z.; Zhu, X.; Nie, Z.; Liu, Q.H. Cloud removal for optical images using SAR structure data. In Proceedings of the 2016 IEEE 13th International Conference on Signal Processing (ICSP), Chengdu, China, 6–10 November 2016; pp. 1872–1875.
37. Wang, P.; Patel, V.M. Generating high quality visible images from SAR images using CNNs. In Proceedings of the 2018 IEEE Radar Conference (RadarConf18), Oklahoma City, OK, USA, 23–27 April 2018; pp. 0570–0575.
38. Xiao, X.; Lu, Y. Cloud Removal of Optical Remote Sensing Imageries using SAR Data and Deep Learning. In Proceedings of the 2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Virtual, 1–3 November 2021; pp. 1–5.
39. Ley, A.; Dhondt, O.; Valade, S.; Haensch, R.; Hellwich, O. Exploiting GAN-based SAR to optical image transcoding for improved classification via deep learning. In Proceedings of the EUSAR 2018, 12th European Conference on Synthetic Aperture Radar, VDE, Aachen, Germany, 4–7 June 2018; pp. 1–6.
40. Martinez, J.; Adarme, M.; Turnes, J.; Costa, G.; De Almeida, C.; Feitosa, R. A Comparison of Cloud Removal Methods for Deforestation Monitoring in Amazon Rainforest. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *43*, 665–671. [[CrossRef](#)]
41. Hwang, J.; Yu, C.; Shin, Y. SAR-to-optical image translation using SSIM and perceptual loss based cycle-consistent GAN. In Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 21–23 October 2020; pp. 191–194.
42. Fu, S.; Xu, F.; Jin, Y.Q. Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks. *arXiv* **2019**, arXiv:1901.08236.
43. Cresson, R.; Narçon, N.; Gaetano, R.; Dupuis, A.; Tanguy, Y.; May, S.; Commandre, B. Comparison of convolutional neural networks for cloudy optical images reconstruction from single or multitemporal joint SAR and optical images. *arXiv* **2022**, arXiv:2204.00424.
44. Xu, F.; Shi, Y.; Ebel, P.; Yu, L.; Xia, G.S.; Yang, W.; Zhu, X.X. Exploring the Potential of SAR Data for Cloud Removal in Optical Satellite Imagery. *arXiv* **2022**, arXiv:2206.02850.
45. Sebastianelli, A.; Nowakowski, A.; Puglisi, E.; Del Rosso, M.P.; Mifdal, J.; Pirri, F.; Mathieu, P.P.; Ullo, S.L. Spatio-Temporal SAR-Optical Data Fusion for Cloud Removal via a Deep Hierarchical Model. *arXiv* **2021**, arXiv:2106.12226.
46. Gao, J.; Yuan, Q.; Li, J.; Zhang, H.; Su, X. Cloud Removal with Fusion of High Resolution Optical and SAR Images Using Generative Adversarial Networks. *Remote Sens.* **2020**, *12*, 191. [[CrossRef](#)]

47. Eckardt, R.; Berger, C.; Thiel, C.; Schmullius, C. Removal of optically thick clouds from multi-spectral satellite images using multi-frequency SAR data. *Remote Sens.* **2013**, *5*, 2973–3006. [\[CrossRef\]](#)
48. Chen, S.; Zhang, W.; Li, Z.; Wang, Y.; Zhang, B. Cloud Removal with SAR-Optical Data Fusion and Graph-Based Feature Aggregation Network. *Remote Sens.* **2022**, *14*, 3374. [\[CrossRef\]](#)
49. Gao, J.; Yi, Y.; Wei, T.; Zhang, G. Sentinel-2 cloud removal considering ground changes by fusing multitemporal SAR and optical images. *Remote Sens.* **2021**, *13*, 3998. [\[CrossRef\]](#)
50. Xiong, Q.; Di, L.; Feng, Q.; Liu, D.; Liu, W.; Zan, X.; Zhang, L.; Zhu, D.; Liu, Z.; Yao, X.; et al. Deriving non-cloud contaminated sentinel-2 images with RGB and near-infrared bands from sentinel-1 images based on a conditional generative adversarial network. *Remote Sens.* **2021**, *13*, 1512. [\[CrossRef\]](#)
51. Zhang, J.; Zhou, J.; Li, M.; Zhou, H.; Yu, T. Quality assessment of SAR-to-optical image translation. *Remote Sens.* **2020**, *12*, 3472. [\[CrossRef\]](#)
52. Zhang, Q.; Liu, X.; Liu, M.; Zou, X.; Zhu, L.; Ruan, X. Comparative analysis of edge information and polarization on sar-to-optical translation based on conditional generative adversarial networks. *Remote Sens.* **2021**, *13*, 128. [\[CrossRef\]](#)
53. Christovam, L.E.; Shimabukuro, M.H.; Galo, M.d.L.B.; Honkavaara, E. Pix2pix Conditional Generative Adversarial Network with MLP Loss Function for Cloud Removal in a Cropland Time Series. *Remote Sens.* **2021**, *14*, 144. [\[CrossRef\]](#)
54. Li, W.; Li, Y.; Chan, J.C.W. Thick Cloud Removal With Optical and SAR Imagery via Convolutional-Mapping-Deconvolutional Network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2865–2879. [\[CrossRef\]](#)
55. Ebel, P.; Xu, Y.; Schmitt, M.; Zhu, X.X. SEN12MS-CR-TS: A Remote-Sensing Data Set for Multimodal Multitemporal Cloud Removal. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [\[CrossRef\]](#)
56. Sebastianelli, A.; Puglisi, E.; Del Rosso, M.P.; Mifdal, J.; Nowakowski, A.; Mathieu, P.P.; Pirri, F.; Ullo, S.L. PLFM: Pixel-Level Merging of Intermediate Feature Maps by Disentangling and Fusing Spatial and Temporal Data for Cloud Removal. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [\[CrossRef\]](#)
57. Ebel, P.; Meraner, A.; Schmitt, M.; Zhu, X.X. Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5866–5878. [\[CrossRef\]](#)
58. Darbaghshahi, F.N.; Mohammadi, M.R.; Soryani, M. Cloud removal in remote sensing images using generative adversarial networks and SAR-to-optical image translation. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–9. [\[CrossRef\]](#)
59. Zhang, J.; Zhou, J.; Lu, X. Feature-Guided SAR-to-Optical Image Translation. *IEEE Access* **2020**, *8*, 70925–70937. [\[CrossRef\]](#)
60. Li, Y.; Fu, R.; Meng, X.; Jin, W.; Shao, F. A SAR-to-Optical Image Translation Method Based on Conditional Generation Adversarial Network (cGAN). *IEEE Access* **2020**, *8*, 60338–60343. [\[CrossRef\]](#)
61. Bermudez, J.D.; Happ, P.N.; Feitosa, R.Q.; Oliveira, D.A. Synthesis of multispectral optical images from sar/optical multitemporal data using conditional generative adversarial networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1220–1224. [\[CrossRef\]](#)
62. Huang, B.; Li, Y.; Han, X.; Cui, Y.; Li, W.; Li, R. Cloud removal from optical satellite imagery with SAR imagery using sparse representation. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1046–1050. [\[CrossRef\]](#)
63. Li, Y.; Li, W.; Shen, C. Removal of optically thick clouds from high-resolution satellite imagery using dictionary group learning and interdictionary nonlocal joint sparse coding. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1870–1882. [\[CrossRef\]](#)
64. Mao, R.; Li, H.; Ren, G.; Yin, Z. Cloud Removal Based on SAR-Optical Remote Sensing Data Fusion via a Two-Flow Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7677–7686. [\[CrossRef\]](#)
65. Meraner, A.; Ebel, P.; Zhu, X.X.; Schmitt, M. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 333–346. [\[CrossRef\]](#)
66. Xu, F.; Shi, Y.; Ebel, P.; Yu, L.; Xia, G.S.; Yang, W.; Zhu, X.X. GLF-CR: SAR-enhanced cloud removal with global–local fusion. *ISPRS J. Photogramm. Remote Sens.* **2022**, *192*, 268–278. [\[CrossRef\]](#)
67. He, W.; Yokoya, N. Multi-temporal sentinel-1 and-2 data fusion for optical image simulation. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 389. [\[CrossRef\]](#)
68. Zhang, W.; Xu, M. Translate SAR data into optical image using IHS and wavelet transform integrated fusion. *J. Indian Soc. Remote Sens.* **2019**, *47*, 125–137. [\[CrossRef\]](#)
69. Zhang, X.; Qiu, Z.; Peng, C.; Ye, P. Removing Cloud Cover Interference from Sentinel-2 imagery in Google Earth Engine by fusing Sentinel-1 SAR data with a CNN model. *Int. J. Remote Sens.* **2022**, *43*, 132–147. [\[CrossRef\]](#)
70. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
71. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Proceedings of the Advances in Neural Information Processing Systems*, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
72. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
73. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, 2016; pp. 694–711.
74. Li, Z.; Shen, H.; Cheng, Q.; Liu, Y.; You, S.; He, Z. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 197–212. [\[CrossRef\]](#)
75. Meng, Q.; Borders, B.E.; Cieszewski, C.J.; Madden, M. Closest spectral fit for removing clouds and cloud shadows. *Photogramm. Eng. Remote Sens.* **2009**, *75*, 569–576. [\[CrossRef\]](#)

76. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS—A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *arXiv* **2019**, arXiv:1906.07789.
77. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
78. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. ssim. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
79. Kruse, F.A.; Lefkoff, A.; Boardman, J.; Heidebrecht, K.; Shapiro, A.; Barloon, P.; Goetz, A. The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* **1993**, *44*, 145–163. [[CrossRef](#)]
80. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE). *Geosci. Model Dev. Discuss.* **2014**, *7*, 1525–1534.
81. Asuero, A.G.; Sayago, A.; Gonzalez, A. The correlation coefficient: An overview. *Crit. Rev. Anal. Chem.* **2006**, *36*, 41–59. [[CrossRef](#)]
82. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [[CrossRef](#)]
83. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [[CrossRef](#)]
84. Allen, D.M. Mean square error of prediction as a criterion for selecting variables. *Technometrics* **1971**, *13*, 469–475. [[CrossRef](#)]

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.