remote sensing

MDPI

Technical Note

A Survey on SAR and Optical Satellite Image Registration

Oscar Sommervold, Michele Gazzea * and Reza Arghandeh

Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, 5020 Bergen, Norway

* Correspondence: mgaz@hvl.no

Abstract: After decades of research, automatic synthetic aperture radar (SAR)-optical registration remains an unsolved problem. SAR and optical satellites utilize different imaging mechanisms, resulting in imagery with dissimilar heterogeneous characteristics. Transforming and translating these characteristics into a shared domain has been the main challenge in SAR-optical matching for many years. Combining the two sensors will improve the quality of existing and future remote sensing applications across multiple industries. Several approaches have emerged as promising candidates in the search for combining SAR and optical imagery. In addition, recent research has indicated that machine learning-based approaches have great potential for filling the information gap posed by utilizing only one sensor type in Earth observation applications. However, several challenges remain, and combining them is a multi-step process where no one-size-fits-all approach is available. This article reviews traditional, state-of-the-art, and recent development trends in SAR-optical co-registration methods.

Keywords: remote sensing; SAR and optical registration; deep learning

1. Introduction

In recent years, the number of Earth observation satellites orbiting the Earth has increased rapidly [1]. As a result, remote sensing is the go-to tool for examining the Earth at various scales due to the vast amount of data from different sensors in orbit. Satellite sensors observe specific parts of the electromagnetic spectrum and are divided into mainly two categories: active and passive. In short, passive sensors rely on reflected sunlight as their illumination source, while active sensors have their own illumination source. The distinction is important because the two sensor types produce different characteristics with complementary information.

Synthetic aperture radar (SAR) satellites are active sensors with an onboard microwave energy source. The long wavelengths of microwave radiation are not affected by weather conditions and can continuously provide images. SAR is excellent at characterizing the structural properties of surface objects. However, the backscatter from the sensor is prone to speckle noise, reducing clarity and detail. In addition, since the sensor only collects the intensity of the backscatter, it provides no spectral information, resulting in noisy black-and-white imagery. Many Earth observation applications rely on spectral data, which renders SAR imagery at a significant disadvantage in real-world applications despite its high-quality surface detail and robustness [2].

On the other hand, optical satellites rely on solar illumination as their energy source and capture the reflected sunlight from objects. This reliance makes optical imagery prone to weather conditions and time of day. In addition, optical satellites cover specific spectral bands of the electromagnetic spectrum. The most common bands are red, green, blue, and near-infrared (NIR). Information from different bands can be combined to create various imagery and land cover indices. Land cover indices are widely used in terrain analysis applications such as vegetation and forest monitoring [3], and power line monitoring [4]. These properties have made optical satellite imagery the most commonly used sensor in remote sensing applications.

check for updates

Citation: Sommervold, O.; Gazzea, M.; Arghandeh, R. A Survey on SAR and Optical Satellite Image Registration. *Remote Sens.* **2023**, *15*, 850. https://doi.org/10.3390/rs15030850

Academic Editor: Andrea Garzelli

Received: 16 December 2022
Revised: 20 January 2023
Accepted: 1 February 2023
Published: 3 February 2023

The favorable aspects of SAR images, such as being "weatherproof" and having high surface detail, when combined with the advantages of optical images, such as spectral characteristics and undeniability for human eyes, would enhance the performance of remote sensing solutions.

The SAR and optical registration process are one of the most fundamental and challenging operations in remote sensing [5]. In other words, image registration is the process of aligning two or more images (the reference andsensed images). The SAR-optical alignment is especially tricky due to the radiometric and geometric differences. SAR and optical images from the same area differ in spatial resolution, spatial alignment, satellite type, and temporal dimensions. Moreover, imagery captured by different satellites introduces several inconsistencies that need to be corrected in the image registration process. Irregularities such as varying positioning of the sensors, object deformation, object movement, and viewpoint mismatch emerge when superimposing two images from different satellites. Temporal differences introduce further complexity to the image registration process. Seasonal and urban changes drastically alter the Earth's surface, resulting in scenarios where satellite observations of the same area appear dissimilar, particularly in optical imagery. Furthermore, imagery captured at various intervals may not share features due to changes in the period between the sensed images.

The SAR-optical image registration process has four major steps: Feature detection and extraction, feature matching, affine transformation, and image alignment. Figure 1 shows a step-by-step example of optical-optical image matching using the classical scale-invariant-feature-transform (SIFT) algorithm [6]. In general, single-sensor image alignment is a much simpler procedure than SAR-optical alignment and is here used to showcase the different steps in the image registration process. Figure 1a shows an optical image from the Sentinel-2 satellite with a 10 m resolution. A random affine transformation has been applied to the sensed image to give a complete example of the image-matching problem. The reference and sensed image contains 50% shared information. Figure 1b shows features detected by the SIFT descriptor denoted by yellow circles in the reference and sensed image, respectively. Figure 1c illustrates the remaining matched features after evaluating feature correspondence and distance thresholding. Finally, Figure 1d presents the affine correction and image alignment. Affine correction is performed manually by selecting three non-collinear matched features as control points. Note the blurred appearance of the overlapping regions of the reference and sensed image due to the slight misalignment.

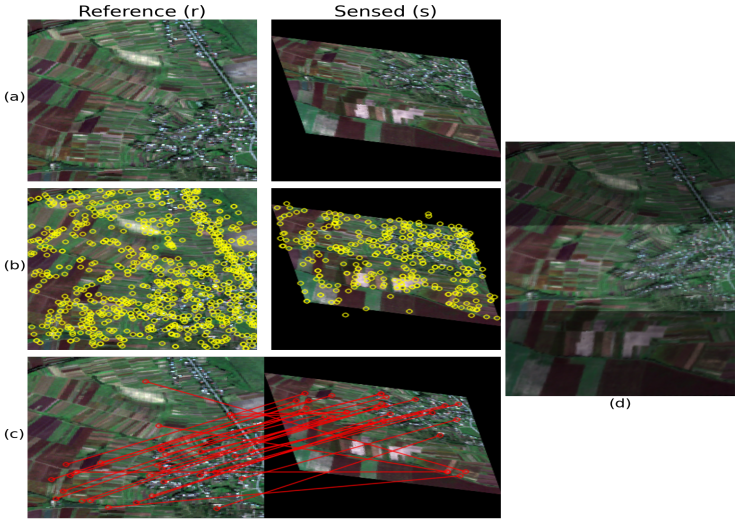

Figure 1. Step-by-step example of optical-optical image matching. (**a**) two optical images; (**b**) features detected by the SIFT descriptor on each image separately; (**c**) features are matched between the two images after evaluating the feature correspondence; (**d**) affine correction and image alignment. The image is taken from the SEN1-2 dataset [7], which is a benchmark dataset for SAR-optimal co-registration, sampled around the world and in all meteorological seasons. Images are acquired from Sentinel-1 (SAR) and Sentinel-2 (optical) with a 10 m/pixel resolution.

The paper structure is as follows. First, we overview the state-of-the-art approaches for SAR-optical image registration. Then, we introduce four common approaches that have been used to address this task more comprehensively. Finally, we discuss the practical challenges.

2. State-of-the-Art Approaches

Available approaches in the literature for SAR-optical image registration problems are classified into two major categories: area-based approaches and feature-based approaches. Features, in particular, can be extracted and calculated using handcrafted methods (which are non-learning methods), or extracted automatically using a neural network (thus via deep learning techniques). In each sub-section, we overview each category mentioned above.

2.1. Area-Based Methods

Area-based methods are often called template matching. The idea is to find the location of a smaller image (the template) in a larger image (the reference image). Here, the template image can be thought of as the sensed image in the image registration process. Template matching is commonly used in object detection, face recognition, motion tracking, medical image analysis, and image registration [8]. As shown in Figure 2, matching the template to the reference image is traditionally done by sliding the template across the reference image and calculating the image similarity using a suitable similarity metric. This approach is especially suited for image alignment. However, it is susceptible to scale and rotation differences, so it is a prerequisite that the template and reference image have the same affine relationship.

Several metrics are used to evaluate the similarity between SAR and optical images. These metrics are mostly intensity-based metrics that exploit pixel-wise intensity patterns present in both types of imagery. A widely used metric is the Normalized Cross Correlation (NCC). The main advantages of NCC are its speed and invariance to linear deviations in brightness and contrast. It is relatively straightforward to implement and is especially suited for template matching scenarios [9]. However, NCC is sensitive to scale, rotation, radiometric, and geometric differences in SAR-optical images. The Sum of Squared Differences (SSD) is another fast but unstable metric. SSD directly compares the image intensity values between the images; consequently, it is susceptible to image intensity variations. Structural Similarity Index Measure (SSIM) [10] is traditionally used to assess image quality and degradation after passing an image through a transformer or filter. SSIM is a weighted measurement of the difference between two images' structure, contrast, and luminance discrepancies. SSIM considers image degradation as changes in the structural information of the image. Hence, in the context of SAR-optical matching, the difference between the two sensors can be simplified to changes in structure, contrast, and luminance, where the SSIM would be a suitable metric. Mutual Information (MI) is a notable exception to the intensity-based approach. MI is a traditional measure of the shared information between two random variables. In the 90s, the concept of MI was extended to perform multi-modal medical image registration [11,12]. This inspired the adaptation of MI-based SAR-optical image registration, cumulating in the Compressed And Segmented Mutual Information technique published in 2009, achieving state-of-the-art performance for its time [13]. The main drawback of such a method is the amount of computing required to perform the registration.

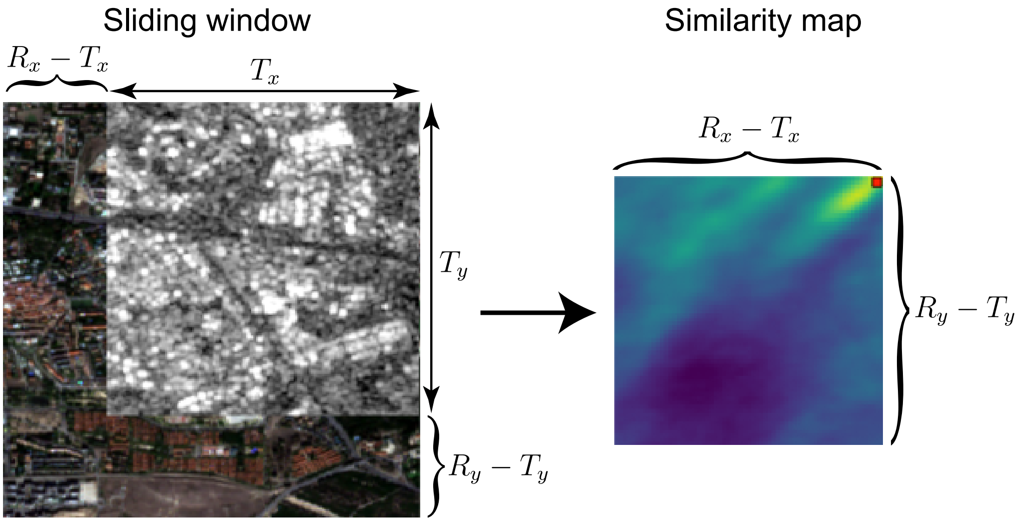

Figure 2. Computing image similarity between the smaller SAR template (T) and the larger optical reference image (R) by iteratively moving the template across the reference image. Although precise, the sliding window technique is a computationally expensive method. With a 256 × 256 reference image and a 192 × 192 template, the resulting similarity map requires 4225 comparisons of 192 × 192 images. The image example is taken from the SEN1-2 dataset.

2.2. Features Extraction via Hand-Crafted Methods

Feature-based approaches are traditionally used in computer vision areas such as object recognition, camera calibration, and image registration [14]. This type of approach aims to exploit the distinctive features of an image. In this context, a feature is a geometric or contrasting characteristic easily detected in both the reference and sensed images. It is vital that the selected features are salient in both images, resulting in features typically being either closed boundary regions in the form of polygons (lakes, forests) [15] or salient points (corners, intersections, edges, roads) [16,17]. The selected features are then translated into vector representations by a feature descriptor before their correspondence is evaluated in the feature-matching step. Finally, features are matched using various similarity metrics and feature descriptors, combined with the spatial relationship of the selected features [18].

A complete step-by-step example of this process is shown in Figure 1, and as illustrated, the mismatching of features is inevitable. Therefore, effective removal of mismatched features is essential to achieve an accurate alignment [19]. In SAR-optical matching, alignment is generally achieved by affine transformation of the sensed image by selecting three candidate feature points. Eventual differences in spatial resolution are corrected using interpolation, where bilinear interpolation is among the most commonly used methods due to its balance between computational complexity and performance [18].

Non-learning feature-based methods for SAR-optical matching are mostly deterministic algorithms. The SIFT algorithm is among the most influential methods in traditional feature matching. It works by creating a Gaussian scale space where feature points are extracted by calculating the Difference of Gaussian between two consecutive scales. Orientation and magnitude gradients are then calculated for pixels surrounding the extracted feature points. These gradient directions are placed in a histogram with 36 bins representing 360 degrees, where the peak determines the rotation angle. Yet, SIFT could perform better on the SAR-optical matching problem due to the noisy nature of the SAR imagery and the radiometric differences between the two sensors [20,21]. As a result, numerous extensions of SIFT, such as speeded-up robust features (SURF) [22], principal component analysis SIFT (PCA-SIFT) [23], affine SIFT (ASIFT) [24], ORB (Oriented FAST and Rotated BRIEF) [25], uniform robust SIFT (UR-SIFT) [26], etc., have been devised to extend the functionality and improve performance of the original SIFT algorithm.

Notable variations relevant to SAR-optical image matching are the SAR-SIFT [27] and radiation-variation insensitive feature transform (RIFT) algorithms [21]. SAR-SIFT introduced domain-specific gradients based on the noisy characteristic of SAR imagery, making it more robust to speckle noise. RIFT addressed the challenge of radiometric differences in multi-modal imagery, such as SAR-optical, by exploiting a phase congruency-based feature descriptor in 2020. Since then, non-learning feature-based SAR-optical image matching has been a hot research topic.

Yu et al. computed optical flow fields from SIFT and phase congruency descriptors to achieve dense feature-based matching at the cost of increased time complexity [28]. Xiong et al. introduced the Adjacent Self-Similarity (ASS) feature using offset mean filtering, yielding robust state-of-the-art performance [29]. Computational cost is the Achilles heel of non-learning methods. Most state-of-the-art techniques take anywhere from roughly 15 to 200 s to perform a single instance of SAR-optical matching, depending on the image dimensions and algorithm used [28,29]. That is to say, only some non-learning methods are slow. The Channel Features of Orientated Gradients (CFOG) is a fast state-of-the-art feature descriptor that uses oriented gradients and a gaussian kernel to build multiple feature vectors for each pixel [30]. Ye et al. proposed an extension of the CFOG descriptor by extracting CFOG feature maps across multiple resolutions before employing a SAR image masking technique to create an edge map and thus achieve more fine-grained matching results [31].

Nonetheless, handcrafted feature descriptors have one major inherent disadvantage. They assume the presence of salient and distinct features. This is not a given in remote sensing imagery, especially in SAR imagery. For example, imagery of forests and rural areas contain few features, and thus the presented non-learning methods can only produce stable results in more urban areas. As a result, current non-learning feature-based methods are not viable solutions for big data SAR-optical image matching or real-time applications, where efficiency and robustness are of utmost importance.

2.3. Features Extraction via Deep Learning

The current state of non-learning handcrafted feature methods cannot meet the requirements of efficient, automatic SAR-optical registration. For many years, the need for publicly available remote sensing data limited the usability of machine learning methods in the SAR-optical domain. This changed with the launch of the SpaceNet program in 2016 [32]. A year later, Google made the Google Earth Engine and its extensive remote sensing data catalog available to the public [33]. Subsequently, in 2019 Schmitt et al. [7] published the SEN1-2 dataset for SAR-optical data fusion, combining imagery from the Sentinel-1 and Sentinel-2 satellites. As such, machine-learning methods suddenly became viable, and a surge of research in machine-learning-based SAR-optical image matching followed in the coming years. Neural networks have been used to extract features automatically from the SAR-optical image pair, as opposed to the handcrafted feature methods (e.g., SIFT, SURF, etc.).

Earlier research on single-modal image matching by Fischer et al. [34] showed that Convolutional Neural Networks (CNNs) outperformed traditional non-learning feature descriptors such as SIFT. This outcome inspired the use of CNNs with shared weights (Siamese architecture) to extract features in SAR and optical imagery. A Siamese Neural Network is a particular architecture that contains two or more identical subnetworks, mirrored with respect to the others. Merkle et al. [35] proposed the first notable example of a Siamese machine learning architecture to perform SAR-optical image registration. The advantage of such an architecture with shared weights is that the features are mapped into the same latent space. This leads to an efficient calculation of the similarity between feature vectors. The main limitation is that the overall architecture complexity is doubled, resulting in more trainable parameters and a longer training time than standard networks. The machine learning method proposed by Merkle extracted features using stacked dilated convolutions. The similarity was assessed by convolving the two feature maps from each sensor and computing the dot product. An overview of the proposed architecture is outlined in Figure 3. In Figure 3, the blue CNN modules detect and extract features, and the green similarity module assesses the similarity of the feature maps, that is, feature matching. Finally, the peak of the resulting similarity heatmap determines the predicted offset and alignment of the template within the reference image. Due to their high performance, Siamese architecture has become the standard scheme for image registration. While the architecture remains similar, more models have been implemented, mainly to increase performance and optimize the computational burden. The SFcNet, described in Ref. [36], introduced a novel loss function that maximizes the distance between positive and negative samples. Traditionally, negative sampling is used in classification tasks to reduce significant false positive rates between similar classes. SFcNet utilizes this strategy by selecting the negative sample as the area with the largest mismatch and the positive sample as the correct matching point. The proposed negative sampling strategy in SFcNet increases the discriminability of the model. Hughes et al. proposed a novel component-based framework using three separate networks [37]. The framework consists of a goodness network, a correspondence network, and an outlier removal network. The main advantage of this approach is that the framework can determine the matching quality via its outlier removal network, acknowledging the problem of high false positive rates. A major drawback of the traditional Siamese CNNs is that they rely on the pixel-wise sliding window technique to compute similarity heatmaps. This operation is computationally intensive, especially for large images. To solve this limitation, Zhang et al. bypassed the sliding window approach by accelerating the SSD computation in the frequency domain in Ref. [30] and introducing the Deep Dense Feature Network (DDFN) in 2022 [38]. Skipping the sliding window method is possible by exploiting the Fast Fourier Transform (FFT) and the convolution theorem. The inverse FFT of pointwise multiplication in the frequency domain is equivalent to the convolution with a sliding window while being magnitudes faster to compute for large images. Fang et al. [39] used the same FFT method to implement the NCC similarity metric in the frequency domain and used the popular image segmentation model U-Net [40] as a feature extractor. It is worth mentioning that a similar Siamese U-Net architecture combined with NCC was also proposed in Ref. [41], but without the FFT accelerated implementation. Thus practically, the FFT U-Net is preferred. Cui et al. [42] proposed an architecture (MAP-net) augmented with an attention mechanism and spatial pyramid aggregated pooling (SPAP). According to the authors, the adoption of the SPAP module makes the network more capable of integrating global and local contextual information. The attention block weights the dense features generated from the network to extract more invariant and distinguishable key features.

Differently from the Siamese models early presented, Zhou et al. [43] extract multi-orientated gradient features using the CFOG descriptors (initially tested in Ref. [30]) to depict the structure properties of images. Then, they implement a shallow pseudo-Siamese network to convolve the gradient feature maps in a multiscale manner, which produces the Multiscale Convolutional Gradient Features (MCGFs). In addition, MCGF employed both negative samples and an FFT-accelerated cross-correlation similarity metric to achieve satisfactory matching performance and computational efficiency.

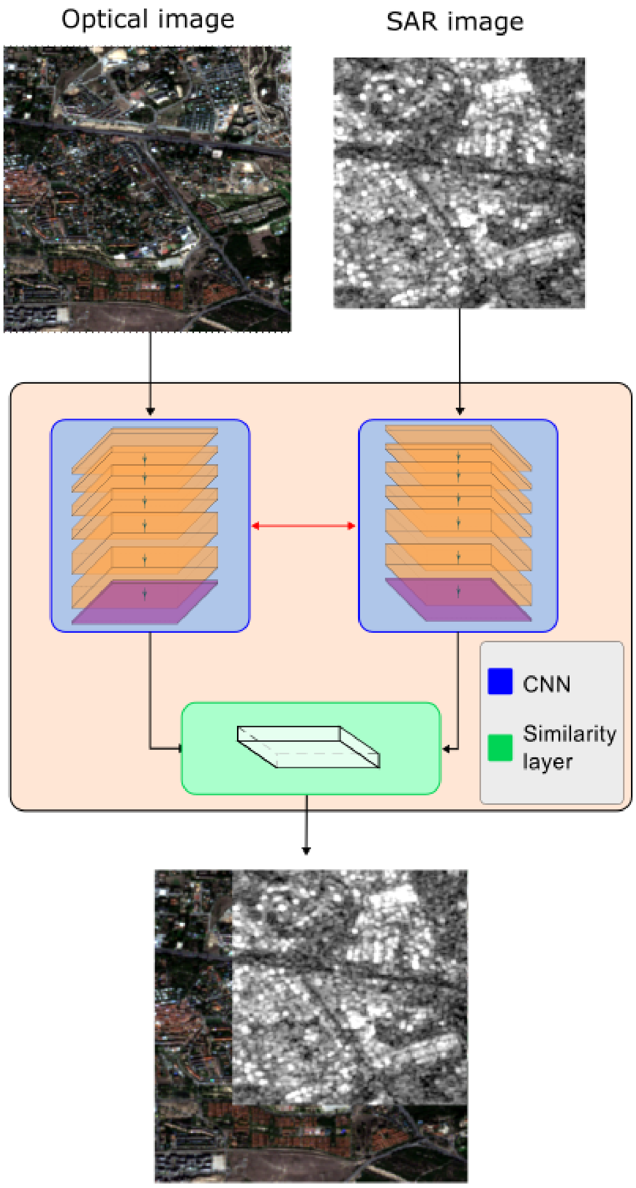

Figure 3. A general Siamese template matches architecture. The red arrow indicates the possible weight-sharing between the two CNN-based feature extractors. Since the CNN translates the images into a homogeneous space, similarity metrics previously demonstrated to be inadequate, such as SSIM and NCC, can now be used to assess the similarity between the CNN feature maps. The satellite images are taken from the SEN1-2 dataset.

2.4. Recent Trends

Generative Adversarial Networks (GANs) [44] is a relatively recent branch of machine learning that generates synthetic images. As a result, researchers began exploring the potential use cases of generating SAR images from optical imagery [45]. The same study noted that generated imagery could help improve existing non-learning SAR-optical image-matching techniques. Ref. [46] used GAN to improve SAR-Optical matching accuracy. More recently, using GANs to translate SAR images into pseudo-optical directly has been explored in Ref. [47,48], showing great promise. Combining unsupervised and supervised machine learning models is a natural next step in the field.

A primary shortcoming of all the models mentioned above is that they are not robust to scale and rotation differences. As such, recent research has focused on making models robust to scale and rotation changes. In Ref. [49] Markiewicz et al. present a new approach to the estimation of shift and rotation between two images from different sensors using the

ASIFT feature descriptor and Structure from Motion (SfM) technique. Hybrid approaches such as in Ref. [50] have previously been proposed, but recently machine learning models based on semantic segmentation methods have appeared. In particular, Li et al. [50] proposed a possible two-step framework for rotation-invariant matching, consisting of a machine-learning module and a novel non-learning module. The machine learning module, called RotNET, is trained to classify the rotation relationship between the reference- and template images. Furthermore, Li et al. [51] stated that semantic-dependent self-attention layers could effectively handle minor affine irregularities present in SAR-optical imagery while achieving state-of-the-art matching accuracy. Nevertheless, more research is needed into developing independent scale and rotation invariant machine learning modules.

3. Survey of the Most Common Methods

The main categories of methods for the SAR-optical matching problem are described in the previous section. This section provides a more extensive explanation of four powerful methods for SAR-optical matching.

3.1. Mutual Information

The Mutual Information (*MI*) similarity-based method is a non-learning approach. *MI* is able to handle the non-linear intensity variations between SAR and optical imagery. This is because *MI* does not rely on pixel intensity. Instead, *MI* assumes a statistical relationship between the two sensors that can be captured by exploiting the entropy and joint entropy of the images. To compute entropy on an image, the pixel values are binned into histograms. The joint entropy, given SAR image *S* and optical image *O*, can be derived from the following equation:

$$H(S,O) = -\sum_{s,o} p_{SO}(s,o)\, log\, p_{SO}(s,o) \tag{1}$$

where p_{SO} denotes the joint probability mass function of *S* and *O*. The entropy of a single image is given by

$$H(X) = -\sum_{x} p_X(x)\, log\, p_X(x) \tag{2}$$

MI can then be calculated as:

$$MI(S,O) = H(S) + H(O) - H(S,O) \tag{3}$$

Using *MI* to perform template matching, the most overlapping region is interpreted as the point maximizing the *MI* value. Specifically, the point where the individual entropies are maximized and the joint entropy is minimized.

3.2. Siamese CNN

The Siamese Convolutional Neural Network (Siamese-CNN) proposed by Merkle [35] was one of the first notable methods published in the context of SAR-optical image matching. Merkle proposed the general architecture depicted in Figure 3 and evaluated the performance of both Siamese and pseudo-Siamese CNNs. The results deemed the Siamese model superior and, consequently, is the primary strategy adopted by later methods. The similarity heatmap is generated using the time-consuming sliding window technique shown in Figure 2, and similarity is evaluated from the dot product of the output feature vectors. The Siamese-CNN consists of several stacked layers of dilated convolutions with 5 × 5 filters to achieve a feature extractor with the desired receptive field size. The complete architecture of the feature extraction network is outlined in Figure 4. The dilated convolutions allow for the exponential growth of the receptive without reducing the resolution of the images [35]. The authors also utilized a soft ground truth distribution by using the discrete approximation of the Gaussian function to blur the region surrounding the correct matching position.

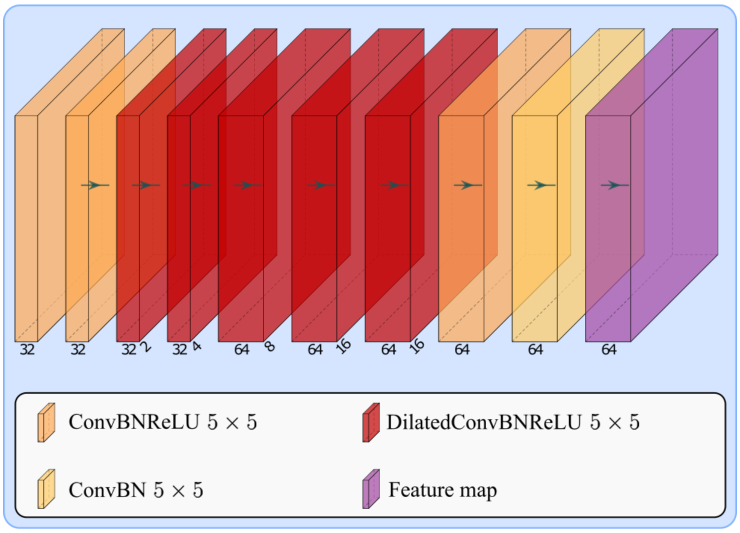

Figure 4. Siamese-CNN architecture. The dilation rate of the dilated convolutions is denoted along the z-axis. Abbreviations: Convolution (Conv), Batch Normalization (BN), Rectified Linear Unit (ReLU).

3.3. Deep Dense Feature Network

Zhang et al. propose several improvements, such as pixel-wise deep dense features, FFT accelerated SSD computation, negative mining, and stacked convolutional layers with small 3 × 3 filters [38]. For each pixel, the network produces a 9-D feature vector, an approach inspired by the CFOG feature descriptor [30]. The structure of the network is shown in Figure 5. Different from other state-of-the-art machine learning approaches, the DDFN use convolutional layers with no padding and instead apply a symmetric pad of 1 for each convolutional layer. Compared to the traditional sliding window SSD computation, the FFT accelerated approach is 14 times faster [38]. Similar to MCGF [43] and SFcNet [36], DDFN employs negative mining to increase the discriminability of the network. Negative mining is used because the machine learning approach essentially transforms the SAR-optical image-matching problem into a classification problem. Individual pixels in the ground truth offset space are interpreted as categories. Following the example of Figure 2, this produces 4225 categories where only one is interpreted as the correct matching position. As a result, negative mining can be utilized to accelerate training and help guide the network in the direction of the correct matching position. Zhang et al. proposed a novel loss function aimed at maximizing the disparity between the correct matching point and the hardest negative sample. The study showed that the proposed DDFN outperformed previous state-of-the-art models, such as SFcNet.

3.4. FFT U-Net

The final selected method is the FFT NCC U-Net proposed by Fang et al. [39]. They argue that the U-Net, a classic image segmentation model, is well-suited for the SAR-optical image-matching problem. The encoder-decoder architecture of the U-Net enables the extraction of both high- and low-level features in the image. The details of the adopted U-Net are shown in Figure 6. Similar to many other state-of-the-art approaches, an FFT accelerated similarity metric is adopted. The authors evaluate the performance of cross correlation (CC) and NCC in the frequency domain, showing that CC yields better pixel-level accuracy while NCC produces the best average precision at the cost of slightly increased computational complexity. Similar to the Siamese-CNN [35], the FFT U-Net utilizes the cross-entropy as the loss function, but here without a soft ground truth distribution. The authors also compared the matching performance against the Siamese-CNN, where the FFT U-Net yielded improved matching accuracy, precision, and time complexity. Compared to the other selected methods, the FFT U-Net is a considerably deeper model with a larger parameter size.

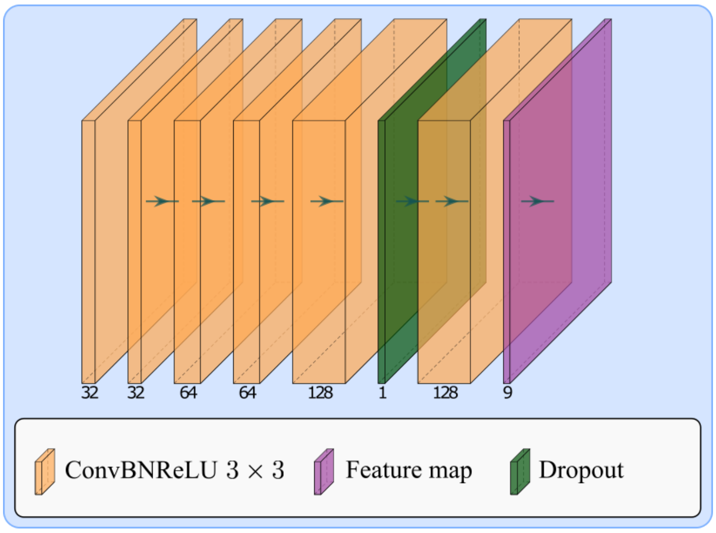

Figure 5. CNN architecture of the Deep Dense Feature Network.

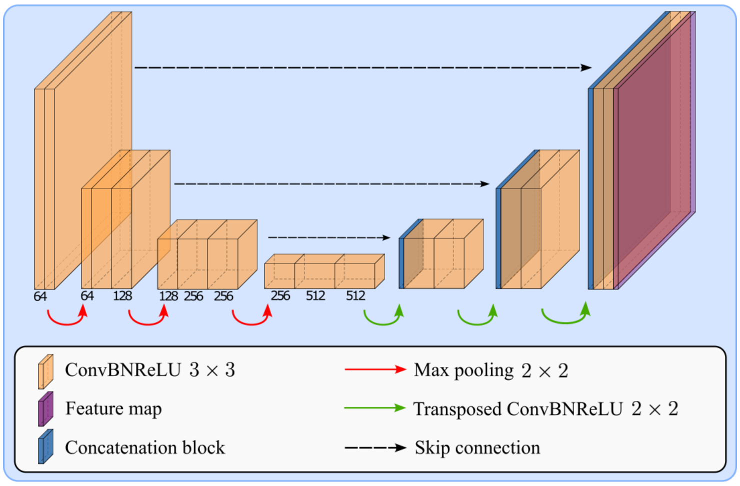

Figure 6. The U-Net architecture adopted in the FFT U-Net paper, here shown in a more shallow configuration for the sake of illustration. The actual network has four encoder-decoder stages.

4. Practical Challenges

Although significant advances have been made toward automatic SAR-optical registration, several challenges remain. Handcrafted state-of-the-art feature descriptors such as SIFT and CFOG cannot reliably handle satellite imagery of homogeneous areas with few salient features. Recent approaches such as SAR masking proposed in Ref. [31] attempt to solve the problem. However, methods able to extract salient features across both sensors are a prerequisite for stable and robust performance. Thus, developing robust feature extractors across a wide class of remote sensing imagery remains a challenge to solve.

Even though machine learning approaches have emerged as the preferred method, they have some shortcomings. In contrast to their non-learning counterparts, state-of-the-art machine learning models often need to be more robust to affine deformations in remote sensing imagery. Making models scale- and rotation-invariant is still a challenge. In addition, there is room for improvement regarding the accuracy and precision of the matching. Another remaining challenge is the reliance on large amounts of preprocessed data. As a result, machine learning models risk overfitting the specific data. This is particularly unwanted in the context of SAR-optical image matching because sensors of the same type can differ in spatial resolution and, consequently, image characteristics. As a result, developing robust machine learning models that are able to generalize well on unseen data from unseen sensors is an imminent challenge in automatic SAR-optical image matching.

Finally, there are intrinsic challenges due to the different radiometric and geometric properties in both optical and SAR imagery. The methods presented in this review have been implemented and tested on various benchmarking datasets, spanning different spatial resolutions: 10 m/px from the SEN1-2 dataset ([7]), the 3 m/px used in Ref. [43], and 1 m/px (as in Ref. [42]). The optical images are generally ortho-photos consisting of three channels RGB. However, one direction of research can be to perform a sensitivity analysis investigating the effects of having multiple spectral channels in the performances of the different methods. Although SAR consists of complex signals (having both amplitude and phase), most of the works deal with only amplitude. Because of the radar speckle noise, the multi-looking technique is used [52]. Multi-look is a technique to reduce speckle noise by processing the image in sections (looks) and later combining them back together. More looks reduce more speckle but at the same time lead to a decrease in the resolution and loss of information in the process. Differently from optical images, which work best with a down-looking view, SAR is intrinsically a side-looking sensor capturing information in varied, rugged terrain. This leads to geometric distortions in slant range, foreshortening, layover, and shadowing. To remove or reduce the distortions arising from these effects, the application of image preprocessing procedures is necessary. Procedures for preprocessing SAR images in terms of the radiometric corrections and calibration are presented in Ref. [53]. In rugged terrain, the changing local imaging geometry may result in backscatter changes up to ±5 dB [54]. Radiometric terrain correction corrects the backscatter intensity of pixels that are distorted by the local incidence angle. Terrain correction can be performed using available tools such as the SNAP toolbox [55] with high-resolution Digital Terrain Models (DTMs) available for the considered area of interest.

5. Conclusions

SAR-optical image registration is a rapidly evolving topic in the remote sensing industry. Early research focused on a purely algorithmic approach with feature-based methods. Feature-based methods developed from the traditional SIFT algorithm to SAR-optical specific techniques, including SAR-SIFT and RIFT. The lack of robustness was addressed with the introduction of state-of-the-art methods such as CFOG and the ASS features. Nevertheless, further research is required to overcome the inherent shortcomings of handcrafted feature descriptors, namely the inability to detect and extract high-level and more abstract features present in SAR-optical imagery.

The public release of large quantities of remote sensing data in the late 2010s enabled the development of machine learning-based SAR-optical matching. Since then, research in the field has been dominated by machine learning techniques. Employing the innate feature extraction capabilities of CNNs, the proposed models build on principles from the traditional feature and template matching methods discussed in this article. State-of-the-art models such as DDFN and FFT U-Net show that machine-learning models perform fast and accurate SAR-optical matching. Nevertheless, this branch of SAR-optical matching is still in its early phase, and several shortcomings highlighted in this article need to be addressed in further work.

Author Contributions: Conceptualization, M.G. and R.A.; investigation, O.S. and M.G.; writing—original draft preparation, O.S. and M.G.; writing—review and editing, M.G. and R.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: This research didn't use any data.

Acknowledgments: The authors would like to thank Andreas Hay Kaljord and Charlotte Bishop from Kongsberg Satellite Services AS (KSAT), Norway, for their insight and discussion.

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

References

1. Liu, P. A survey of remote-sensing big data. *Front. Environ. Sci.* **2015**, *3*, 45. [CrossRef]
2. 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]
3. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* **2017**, *2017*, 1353691. [CrossRef]
4. Gazzea, M.; Pacevicius, M.; Dammann, D.O.; Sapronova, A.; Lunde, T.M.; Arghandeh, R. Automated Power Lines Vegetation Monitoring Using High-Resolution Satellite Imagery. *IEEE Trans. Power Deliv.* **2022**, *37*, 308–316. [CrossRef]
5. Wu, Y.; Liu, J.W.; Zhu, C.Z.; Bai, Z.F.; Miao, Q.G.; Ma, W.P.; Gong, M.G. Computational Intelligence in Remote Sensing Image Registration: A survey. *Int. J. Autom. Comput.* **2021**, *18*, 1–17. [CrossRef]
6. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
7. Schmitt, M.; Hughes, L.H.; Zhu, X.X. The SEN1-2 Dataset for Deep Learning in SAR-Optical Data Fusion. *arXiv* **2018**, arXiv:1807.01569.
8. Hashemi, N.S.; Aghdam, R.B.; Ghiasi, A.S.B.; Fatemi, P. Template Matching Advances and Applications in Image Analysis. *arXiv* **2016**, arXiv:1610.07231.
9. Sarvaiya, J.; Patnaik, S.; Bombaywala, S. Image Registration by Template Matching Using Normalized Cross-Correlation. In Proceedings of the 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, Bangalore, India, 28–29 December 2009; pp. 819–822. [CrossRef]
10. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
11. Collignon, A.; Maes, F.; Delaere, D.; Vandermeulen, D.; Suetens, P.; Marchal, G. Automated multi-modality image registration based on information theory. In *Proceedings of the Information Processing in Medical Imaging*; Citeseer: Princeton, NJ, USA, 1995; Volume 3, pp. 263–274.
12. Wells, W.M., III; Viola, P.; Atsumi, H.; Nakajima, S.; Kikinis, R. Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.* **1996**, *1*, 35–51. [CrossRef]
13. Suri, S.; Reinartz, P. Mutual-Information-Based Registration of TerraSAR-X and Ikonos Imagery in Urban Areas. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 939–949. [CrossRef]
14. Wang, Z.; Kieu, H.; Nguyen, H.; Le, M. Digital image correlation in experimental mechanics and image registration in computer vision: Similarities, differences and complements. *Opt. Lasers Eng.* **2015**, *65*, 18–27. [CrossRef]
15. Goncalves, H.; Corte-Real, L.; Goncalves, J.A. Automatic Image Registration Through Image Segmentation and SIFT. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2589–2600. [CrossRef]
16. Huo, C.; Pan, C.; Huo, L.; Zhou, Z. Multilevel SIFT Matching for Large-Size VHR Image Registration. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 171–175. [CrossRef]
17. Yu, L.; Zhang, D.; Holden, E.J. A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. *Comput. Geosci.* **2008**, *34*, 838–848. [CrossRef]
18. Zitová, B.; Flusser, J. Image registration methods: A survey. *Image Vis. Comput.* **2003**, *21*, 977–1000. [CrossRef]
19. Feng, R.; Shen, H.; Bai, J.; Li, X. Advances and Opportunities in Remote Sensing Image Geometric Registration: A systematic review of state-of-the-art approaches and future research directions. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 120–142. [CrossRef]
20. Hughes, L.H.; Merkle, N.; Bürgmann, T.; Auer, S.; Schmitt, M. Deep Learning for SAR-Optical Image Matching. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 4877–4880. [CrossRef]
21. Li, J.; Hu, Q.; Ai, M. RIFT: Multi-Modal Image Matching Based on Radiation-Variation Insensitive Feature Transform. *IEEE Trans. Image Process.* **2020**, *29*, 3296–3310. [CrossRef] [PubMed]
22. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
23. Ke, Y.; Sukthankar, R. PCA-SIFT: A more distinctive representation for local image descriptors. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), Washington, DC, USA, 27 June–2 July 2004; Volume 2, p. II. [CrossRef]
24. Morel, J.M.; Yu, G. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM J. Imaging Sci.* **2009**, *2*, 438–469.
25. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571. [CrossRef]
26. Sedaghat, A.; Mokhtarzade, M.; Ebadi, H. Uniform Robust Scale-Invariant Feature Matching for Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4516–4527. [CrossRef]
27. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-Like Algorithm for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 453–466. [CrossRef]
28. Yu, Q.; Jiang, Y.; Zhao, W.; Sun, T. High-Precision Pixelwise SAR–Optical Image Registration via Flow Fusion Estimation Based on an Attention Mechanism. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3958–3971. [CrossRef]
29. Xiong, X.; Jin, G.; Xu, Q.; Zhang, H.; Wang, L.; Wu, K. Robust Registration Algorithm for Optical and SAR Images Based on Adjacent Self-Similarity Feature. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 1–17. [CrossRef]

30. Ye, Y.; Bruzzone, L.; Shan, J.; Bovolo, F.; Zhu, Q. Fast and Robust Matching for Multimodal Remote Sensing Image Registration. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9059–9070. [CrossRef]
31. Ye, Y.; Yang, C.; Zhang, J.; Fan, J.; Feng, R.; Qin, Y. Optical-to-SAR Image Matching Using Multiscale Masked Structure Features. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
32. Van Etten, A.; Lindenbaum, D.; Bacastow, T.M. SpaceNet: A Remote Sensing Dataset and Challenge Series. *arXiv* **2018**, arXiv:1807.01232.
33. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. Big Remotely Sensed Data: Tools, applications and experiences. [CrossRef]
34. Fischer, P.; Dosovitskiy, A.; Brox, T. Descriptor Matching with Convolutional Neural Networks: A Comparison to SIFT. *arXiv* **2014**, arXiv:1405.5769,
35. Merkle, N.; Luo, W.; Auer, S.; Müller, R.; Urtasun, R. Exploiting Deep Matching and SAR Data for the Geo-Localization Accuracy Improvement of Optical Satellite Images. *Remote Sens.* **2017**, *9*, 586. [CrossRef]
36. Zhang, H.; Ni, W.; Yan, W.; Xiang, D.; Wu, J.; Yang, X.; Bian, H. Registration of Multimodal Remote Sensing Image Based on Deep Fully Convolutional Neural Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3028–3042. [CrossRef]
37. Hughes, L.H.; Marcos, D.; Lobry, S.; Tuia, D.; Schmitt, M. A deep learning framework for matching of SAR and optical imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 166–179. [CrossRef]
38. Zhang, H.; Lei, L.; Ni, W.; Tang, T.; Wu, J.; Xiang, D.; Kuang, G. Optical and SAR Image Matching Using Pixelwise Deep Dense Features. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
39. Fang, Y.; Hu, J.; Du, C.; Liu, Z.; Zhang, L. SAR-Optical Image Matching by Integrating Siamese U-Net With FFT Correlation. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
40. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
41. Wu, W.; Xian, Y.; Su, J.; Ren, L. A Siamese Template Matching Method for SAR and Optical Image. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
42. Cui, S.; Ma, A.; Zhang, L.; Xu, M.; Zhong, Y. MAP-Net: SAR and Optical Image Matching via Image-Based Convolutional Network With Attention Mechanism and Spatial Pyramid Aggregated Pooling. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]
43. Zhou, L.; Ye, Y.; Tang, T.; Nan, K.; Qin, Y. Robust Matching for SAR and Optical Images Using Multiscale Convolutional Gradient Features. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
44. 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, USA, 8–13 December 2014; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; Volume 27.
45. Merkle, N.; Auer, S.; Müller, R.; Reinartz, P. Exploring the Potential of Conditional Adversarial Networks for Optical and SAR Image Matching. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1811–1820. [CrossRef]
46. Hughes, L.H.; Schmitt, M.; Zhu, X.X. Mining Hard Negative Samples for SAR-Optical Image Matching Using Generative Adversarial Networks. *Remote Sens.* **2018**, *10*, 1552. [CrossRef]
47. Yang, X.; Wang, Z.; Zhao, J.; Yang, D. FG-GAN: A Fine-Grained Generative Adversarial Network for Unsupervised SAR-to-Optical Image Translation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]
48. Nie, H.; Fu, Z.; Tang, B.H.; Li, Z.; Chen, S.; Wang, L. A Dual-Generator Translation Network Fusing Texture and Structure Features for SAR and Optical Image Matching. *Remote Sens.* **2022**, *14*, 2946. [CrossRef]
49. Markiewicz, J.; Abratkiewicz, K.; Gromek, A.; Ostrowski, W.; Samczyński, P.; Gromek, D. Geometrical Matching of SAR and Optical Images Utilizing ASIFT Features for SAR-based Navigation Aided Systems. *Sensors* **2019**, *19*, 5500. [CrossRef]
50. Li, Z.; Zhang, H.; Huang, Y. A Rotation-Invariant Optical and SAR Image Registration Algorithm Based on Deep and Gaussian Features. *Remote Sens.* **2021**, *13*, 2628. [CrossRef]
51. Li, L.; Han, L.; Cao, H.; Hu, H. Joint Self-Attention for Remote Sensing Image Matching. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
52. Vavriv, D.M.; Bezvesilniy, O.O. Advantages of multi-look SAR processing. In Proceedings of the 2013 IX Internatioal Conference on Antenna Theory and Techniques, Odessa, Ukraine, 16–20 September 2013; pp. 217–219. [CrossRef]
53. Frulla, L.; Milovich, J.; Karszenbaum, H.; Gagliardini, D. Radiometric corrections and calibration of SAR images. In Proceedings of the IGARSS '98. Sensing and Managing the Environment. 1998 IEEE International Geoscience and Remote Sensing. Symposium Proceedings. (Cat. No.98CH36174), Seattle, WA, USA, 6–10 July 1998; Volume 2, pp. 1147–1149. [CrossRef]
54. Loew, A.; Mauser, W. Generation of geometrically and radiometrically terrain corrected SAR image products. *Remote Sens. Environ.* **2007**, *106*, 337–349. [CrossRef]
55. Sentinel Application Platform (SNAP). Available online: https://step.esa.int/main/download/snap-download/ (accessed on 28 October 2022).

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.