



# Article CycleGAN-Based SAR-Optical Image Fusion for Target Recognition

Yuchuang Sun <sup>1,2</sup>, Kaijia Yan <sup>1,2</sup> and Wangzhe Li <sup>1,2,\*</sup>

- <sup>1</sup> National Key Laboratory of Microwave Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China; sunyuchuang19@mails.ucas.ac.cn (Y.S.); yankaijia21@mails.ucas.edu.cn (K.Y.)
- <sup>2</sup> School of Electronics, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
- \* Correspondence: wzli@mail.ie.ac.cn

Abstract: The efficiency and accuracy of target recognition in synthetic aperture radar (SAR) imagery have seen significant progress lately, stemming from the encouraging advancements of automatic target recognition (ATR) technology based on deep learning. However, the development of a deep learning-based SAR ATR algorithm still faces two critical challenges: the difficulty of feature extraction caused by the unique nature of SAR imagery and the scarcity of datasets caused by the high acquisition cost. Due to its desirable image nature and extremely low acquisition cost, the simulated optical target imagery obtained through computer simulation is considered a valuable complement to SAR imagery. In this study, a CycleGAN-based SAR and simulated optical image fusion network (SOIF-CycleGAN) is designed and demonstrated to mitigate the adverse effects of both challenges simultaneously through SAR-optical image bidirectional translation. SAR-to-optical (S2O) image translation produces artificial optical images that are high-quality and rich in details, which are used as supplementary information for SAR images to assist ATR. Conversely, optical-to-SAR (O2S) image translation generates pattern-rich artificial SAR images and provides additional training data for SAR ATR algorithms. Meanwhile, a new dataset of SAR-optical image pairs containing eight different types of aircraft has been created for training and testing SOIF-CycleGAN. By combining image-quality assessment (IQA) methods and human vision, the evaluation verified that the proposed network possesses exceptional bidirectional translation capability. Finally, the results of the S2O and O2S image translations are simultaneously integrated into a SAR ATR network, resulting in an overall accuracy improvement of 6.33%. This demonstrates the effectiveness of SAR-optical image fusion in enhancing the performance of SAR ATR.

**Keywords:** synthetic aperture radar (SAR); automatic target recognition (ATR); image fusion; deep learning; CycleGAN

# 1. Introduction

With strong penetrability, electromagnetic waves can penetrate clouds, vegetation, snow, ice, sand, and other earth surface cover. By actively transmitting electromagnetic waves and receiving echoes for imaging, synthetic aperture radar (SAR) can monitor ground conditions in all weather conditions throughout the day, which plays an irreplaceable role in time-sensitive tasks such as disaster detection, ship monitoring, and traffic investigation. The research of automatic target recognition (ATR) using SAR images has attracted considerable attention in recent years. Numerous attempts at progress have been made around the creation of SAR target datasets and the proposal of SAR ATR algorithms [1]. However, current SAR ATR technologies still have room for improvement in practical applications, which is mainly challenged by two aspects: the difficulty of SAR target feature extraction and the scarcity of training datasets. (1) Due to the unique imaging mechanism of SAR,



Citation: Sun, Y.; Yan, K.; Li, W. CycleGAN-Based SAR-Optical Image Fusion for Target Recognition. *Remote Sens.* 2023, *15*, 5569. https://doi.org/ 10.3390/rs15235569

Academic Editors: Dusan Gleich and Alex Hay-Man Ng

Received: 22 September 2023 Revised: 7 November 2023 Accepted: 28 November 2023 Published: 30 November 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/). the target in SAR images exhibits distinct features that differ from its optical appearance. These characteristics, such as geometric distortion and speckle noise, seriously affect image quality. In addition, target features in SAR images are rarely robust against pose variation and exhibit significant viewpoint-dependent variations, which complicates the tasks of feature extraction, generalization, and classification. (2) The scarcity of SAR datasets is not solely attributed to costly SAR systems and experiments but also due to the inherent difficulty in manually interpreting and labeling SAR images. Accordingly, existing open-source datasets for SAR target recognition are seriously insufficient in terms of data volume, target type diversity, and scene variability. The recognition algorithm developed on this basis has strong limitations and poses challenges when applied to complex real-world scenarios.

The SAR-to-optical (S2O) method, based on deep learning, provides a solution for processing SAR target features by translating them into optical expressions through imageto-image translation. This approach enhances the quality and interpretability of SAR images, thereby reducing the difficulty of both manual interpretation and automatic target recognition [2–4]. The vast majority of S2O methods rely on SAR and optical remotesensing images acquired from satellite platforms to demonstrate large-scale scene S2O translation, which provides limited improvement for enhancing the recognition algorithms of small targets. An earlier work [5] by our team showed the S2O translation of small ground targets for the first time and verified its enhancement for SAR target recognition, both automatic and manual. The translation performance is further enhanced in this paper through improvements to the network architecture, integration of appropriate loss functions, expansion of the dataset, and other enhancements.

In order to alleviate the scarcity of training datasets, a common approach is to utilize the data generation ability of deep networks for data augmentation. Specifically, the deep networks are used to learn the mapping relationship between labels and real data and then generate artificial data with a similar distribution to real data by adjusting the labels. Among these methods, generative adversarial networks (GANs), which were proposed in 2014 [6], have been particularly effective in expanding SAR target datasets due to their excellent generation effect [7–10]. There are few SAR target datasets available for reference, resulting in a single scenario for SAR data generation applications. Meanwhile, because of the complex operating conditions of SAR imaging, the GAN-based generation of SAR target datasets still faces the problems of complex input condition labels [11]. By using the same generative adversarial idea, this paper achieves optical-to-SAR (O2S) translation with the simulated optical images of the targets as labels and SAR images of the targets as real data, which provides a cost-effective approach to augment the existing SAR target datasets.

In order to tackle the aforementioned challenges in SAR ATR, a SAR ATR system enhanced by a SAR-optical image bidirectional translation method is proposed, where a SAR-optical image fusion CycleGAN (SOIF-CycleGAN) improves the performance of SAR ATR algorithms by taking advantage of the high quality and low acquisition cost of optical target images. On the one hand, the S2O translation path of SOIF-CycleGAN is implemented using a combination of supervised and unsupervised learning. A new joint loss function, which integrates GAN losses, cycle-consistency losses, L1 loss, and LPIPS loss, is proposed to significantly enhance the translation performance compared with conventional ones. The SAR target image and its artificial optical image generated by S2O translation form a co-registration image as different channels, which is input into SAR ATR networks to achieve higher accuracy recognition. On the other hand, the O2S translation path of SOIF-CycleGAN uses optical target images from computer simulations in various poses to generate SAR target images through unsupervised learning. The generator of the O2S translation path incorporates random noise injection in its back-end layers to ensure the diversity of details in the generated SAR images. These generated images serve as an augmentation to existing SAR target datasets at a low cost. The proposed system is trained and tested on a SAR-optical target dataset comprising eight types of small aircraft and helicopters (SPH8). Within the SPH8 dataset, SAR images are acquired using a UAV-borne SAR for ground imaging, whereas the optical images are generated through computer simulation referring to the active infrared imaging processes. The evaluations of the SOIF-CycleGAN translation results based on both human vision and image quality assessment (IQA) methods indicate that S2O translation produces optical images with high-quality, rich details, and strong pose robustness. Meanwhile, O2S translation generates SAR images with a feature representation and statistical distribution similar to those of real SAR images. The results of SAR ATR confirm that the artificial optical image can be used as a complement to SAR images, significantly reducing the difficulty in target feature extraction and enhancing the LeNet-based SAR ATR algorithm to obtain more than a 5% accuracy improvement. The artificial SAR image can accurately restore the structural features of the target under the corresponding SAR imaging viewpoints and effectively augment patterns in training data, which effectively improves the adaptability of the SAR ATR algorithm. The main contributions of this work can be summarized as the following three aspects:

- 1. A method for bidirectional translation of SAR-optical images is demonstrated by utilizing the bidirectional generation ability of CycleGAN. The feasibility of this data-fusion method in solving the difficulty of feature extraction and the scarcity of training datasets in SAR ATR is verified.
- 2. A joint loss function that takes into account both the whole and local factors for S2O translation is proposed by comparing the impacts of various supervised and unsupervised losses. Through a combination of human vision and numerical evaluation, it has been validated that the joint loss function improves the translation results.
- 3. A new dataset, SPH8, comprising SAR images and the simulated optical images of eight types of ground aircraft targets, is created. It includes both paired and unpaired SAR-optical target images, making it suitable for supporting SAR-optical data fusion, SAR ATR, SAR data generation, and other research, both supervised and unsupervised.

The rest of the paper is structured as follows. In Section 2, we provide an overview of the existing research on SAR ATR, SAR-to-optical image fusion, and SAR data generation that is relevant to this study. Section 3 introduces SOIF-CycleGAN, a method for the SAR-optical image fusion of targets, where two entry points for SAR ATR enhancement using the SAR-optical image fusion are given. The dataset and experiment configuration are elaborated in Section 4, followed by the presentation of translation and recognition results in Section 5. The impact of loss functions and the sample size on the translation results is discussed in Section 6, along with some special cases, demonstrating the robustness of S2O translation. In the end, we draw a conclusion in Section 7.

# 2. Related Works

In this Section, the previous works related to the paper are introduced, encompassing research in the area of SAR ATR, SAR-optical image fusion, and SAR data generation. By incorporating the literature review here, we hope this makes the argument in the Introduction more concise.

# 2.1. SAR ATR

SAR images have a vast amount of information and are difficult to interpret by untrained people, which makes it necessary to replace manual labor with ATR algorithms for accurate large-scale target recognition [12]. According to the detection, discrimination, and classification three-level flow chart [13] launched by the Lincoln laboratory, the current so-called SAR target recognition is essentially the process of classifying the SAR image that contains a single target according to the type of target. The datasets designed for this task usually contain SAR slices of a single target, among which the most typical one is the Moving and Stationary Target Acquisition Recognition (MSTAR) dataset [14] organized by the Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL). MSTAR contains multi-angle airborne SAR images of 10 types of ground military vehicles, covering 1° to 360°, with a resolution of 0.3 m  $\times$  0.3 m. MSTAR continues to expand as data collection continues. Because it has complete azimuth angles and a considerable amount of data, the majority of representative SAR ATR systems [12,15–19] are developed based on MSTAR. In the task of ship recognition, OpenSARShip [20] and OpenSARShip2.0 [21], collected from Sentinel-1A by Shanghai Jiao Tong University, are widely used, which have a resolution of  $20 \text{ m} \times 22 \text{ m}$  and  $(2.7–3.5 \text{ m}) \times 22 \text{ m}$ , with 45874 SAR ship slices in total. While in OpenSARShip, the number of different types of ships is not balanced, with most of the ship samples being cargo ships. Another commonly used dataset, FUSAR [22], containing 15 types of ship targets published by Fudan University has a higher resolution,  $(1.7–1.754 \text{ m}) \times 1.124 \text{ m}$ , from the GF-3 satellite. At present, there are few works pertaining to aircraft recognition, with some studies relying on the SAR images of aircraft targets acquired from satellite platforms [23,24]. Aiming at the classification task of airborne high-resolution SAR slices of aircraft, this paper introduces the SPH8 dataset, which contains multi-angle SAR slices of eight types of small aircraft, to breathe new life into this field.

#### 2.2. SAR-Optical Image Fusion

Various sensors mounted on the remote-sensing platform possess unique characteristics that complement each other in collecting object information. A comparison between SAR imaging and optical imaging highlights the importance of this complementarity. On one hand, SAR imaging captures texture, geometry, and moisture-sensitive information of the surface under all-day and all-weather conditions, but the SAR image is affected by speckle noise and is difficult to understand. Optical imaging, on the other hand, collects spectral information, which is influenced severely by solar illumination and weather conditions. Therefore, the association, correlation, and combination of SAR and optical data have garnered extensive interest [25]. Recent efforts demonstrate that deep learning models have dominated the SAR-optical image fusion task. At GARSS 2017, Ref. [26] verified that cGAN-based Pix2Pix can translate a despeckled SAR image from TerraSAR-X to grayscale optical representation by referring to optical images from PRISM in farmland areas. Then, the grayscale bidirectional translation between high-resolution airborne SAR images and GoogleEarth optical images with a classic CycleGAN network [2], as well as the S2O translation from ALOS-PALSAR SAR images to Terra-ASTER optical images with three channels: R, G, and NIR [27], were displayed at GARSS 2018. Additionally, in the same year, Ref. [28] published the SEN1-2 dataset, including 282,384 pairs of matched SAR and optical image patches, which were collected from Sentinel-1 and Sentinel-2, respectively. SEN1-2 have greatly fostered the development of SAR-optical translation technology. Ref. [3] is dedicated to unpaired SAR-optical image translation with unsupervised learning, and the evaluation of the translation results based on feedback from experts in SAR remote sensing indicates the superiority of image fusion. Four baseline networks, Pix2pix, CycleGAN, Pix2pixHD, and FGGAN, were tested in the SAR-optical image fusion task in [29], and several IQA methods were adopted in various scenes to select a suitable method for more accurate image translation. The parallel feature fusion generator, multi-scale discriminator, and chromatic aberration loss were exploited for improving the contour sharpness, the texture fine-graininess, and the color fidelity of translated optical images, respectively, in [4]. The introduction of new supervised learning loss, such as SSIM, is used in [30,31] for retaining more structural information and the connection between generated optical images and original SAR images, which brings new ideas of loss function design for improving the performance of SAR-to-optical image translation.

In terms of practical application, Ref. [32] focuses on the analysis of SAR-optical image fusion for removing both thin and thick clouds in optical remote-sensing images, which can generate not only RGB images but also other spectral bands. Aiming at the same goal of cloud removal, Refs. [33–36] modified the network architecture, loss function, and dataset, which greatly enhanced the ability to generate multi-temporal cloud-free optical images. In an effort to achieve high-precision building extraction, Ref. [37] proposed a progressive fusion-learning framework that uses phase as a modal invariant between optical and SAR images and realizes fusion through multi-stage learning. Ref. [38] generated optical images

from SAR images by using cGAN as a way to extract common features for multi-modal image alignment. For change detection in heterogeneous images, Ref. [39] trained an unsupervised CycleGAN for SAR-optical image bidirectional translation and implemented change detection between real and translated images in both the SAR and optical image domains. The final change detection result was obtained from the image fusion of the results of the two domains. Another work [40] used a supervised CD network that utilizes deep context features to carry out pixel change detection after translating images to a unity domain. Other applications also benefit greatly from SAR-optical image fusion, such as crop classification [41], wildfire monitoring [42], vegetation monitoring [43,44], road extraction [3], etc.

However, there are few studies applying SAR-optical image fusion in SAR ATR; there are mainly two reasons for this: firstly, the majority of SAR images used for SAR-optical image fusion are obtained from satellite platforms that show severe information loss for small targets. As a result, existing SAR-optical image fusion works are restricted to the applications of large-scale scenes. Our last work [5] overcomes this limitation by building a new high-resolution SAR-optical image dataset of targets for SAR-optical image fusion. Secondarily, the dissimilarity between SAR and optical images impedes their fusion, particularly in areas where ground features undergo significant changes. In order to illustrate this, the results in S2O translation [3,27,29,45] are better in mountains, rivers, forests, farmland, and other natural scenes, whereas man-made scenes like buildings and vehicles are hard to restore. In this paper, the high-resolution SAR-optical image dataset of targets is further expanded to reduce the difficulty of image fusion, and a carefully designed image fusion network is explored to facilitate high-quality SAR-optical image bidirectional translation of targets.

# 2.3. SAR Data Generation

SAR applications based on deep learning require a substantial amount of data to support network training. However, SAR images are difficult to acquire, understand, and process, resulting in limited datasets that meet the requirements [46]. Some computer simulation methods [47,48] have been implemented to directly calculate target SAR images using ray-tracing algorithms combined with the computer-aided design (CAD) models of the targets. This is due to the fact that high-frequency electromagnetic waves exhibit similar scattering characteristics to light. The simulated images, although having similar radiometric representation to SAR images, are quite different in detail and texture and, thus, are often used to help understand SAR images rather than extend training datasets. Data augmentation is often used to alleviate the problem of data scarcity. While SAR images are mappings of target-scattering characteristics in the Range-Doppler domain, their expression is closely related to imaging bands, the azimuth angles of target orientation, and other factors. Traditional data augmentation methods, such as random rotation, fail to introduce new reference patterns to SAR datasets. Since the amazing results of generative adversarial networks were demonstrated [6], deep learning-based generative models have been used to generate high-quality, pattern-rich image data. Hence, the need to augment SAR datasets for ATR gave rise to a growing corpus of GAN-based SAR target image generation methods.

When exploiting DCGAN to generate MSTAR data, Ref. [7] jointly used two discriminators to recognize the generated results, and the recognition results are backpropagated to the generator to enhance its performance. This method can be well adapted to the proposed task of semi-supervised learning in the case of fewer labeled samples. Ref. [9] adopted a cGAN to generate SAR target images with a noise + latent variable input, and information theory was adopted by adding a CNN, fitting the mutual information item between the generated images and the latent variable. The CNN kept calculating the lower bound and forcing the GAN to optimize itself to reach the lower bound, which can help to increase sample diversity. In [49], which had the same goal, real SAR target images from MSTAR were encoded to feature codes and were mixed with noise and category labels. The mixed feature codes, as input, trained multiconstraint GAN outputs generated SAR target images with high sample diversity, SAR feature similarity, and correct categories. In order to complement the azimuth interpolation of the SAR target datasets, Refs. [10,11,50] improved the generation ability of SAR images under specific azimuth angle labels; among them, Ref. [11] trained a generator, the inputs of which contain four kinds of condition information: azimuth angle, elevation angle, target type, and image resolution for SAR deceptive jamming with high fidelity. Aiming to generate highly realistic simulated SAR images, Ref. [8] fused the simulated SAR target image from OpenSARSim and the background image from MSTAR and fine-tuned the fused image on a CycleGAN-based image-to-image translation network. The recognition results showed that the network is able to preserve the label information of targets and improve the realism of images simultaneously. Based on these examples, it can be found that there are two primary issues in current SAR data generation: firstly, the baseline dataset is single, while most of the works are based on MSTAR. Secondarily, The conditional label information input to a cGAN is complex, and a cGAN is unable to generate valid data beyond labels. In this paper, we build a new SAR dataset containing eight aircraft targets as the baseline dataset, which extends the scope of SAR data generation. Moreover, the image-to-image translation architecture is used to translate the optical image obtained by computer simulation into the corresponding SAR image. By utilizing simulated optical images as conditional information, generated SAR images of those targets that are not included in the training set can be obtained.

## 3. Methods

The proposed SAR ATR system in this study comprises two components: SOIF-CycleGAN for achieving S2O and O2S translation and a deep network for SAR ATR. The processes of image fusion and SAR ATR are illustrated in Figure 1a,b, respectively. Both SOIF-CycleGAN and the SAR ATR network are trained on the SPH8 dataset. In order to improve the accuracy of SAR ATR, we choose two entry points for using the SAR-optical image fusion: SAR-optical co-registration image recognition and SAR data augmentation. In terms of SAR-optical co-registration image recognition, the simulated optical target image possesses the ideal quality and abundant features for recognition, which can be utilized as a complement to the SAR image that is input into the SAR ATR network to achieve more effective feature extraction. As shown in Figure 1b, only real SAR target images are available for regular SAR ATR tasks. The S2O translation path of SOIF-CycleGAN is trained to translate the real SAR image into the optical image domain and generate the corresponding artificial optical image. The final input of the SAR ATR network is the co-registration image, which consists of two channels: a real SAR image and an artificial optical image. With regard to SAR data augmentation, computer simulation can offer optical target images of arbitrary aircraft types from any viewpoint at a very low cost, whereas the type and the viewpoint of the data in SAR target datasets are usually limited. The O2S translation path of SOIF-CycleGAN is trained to translate the real optical image obtained by computer simulation into the corresponding artificial SAR image. These artificial SAR images containing the complete viewpoint interpolation of targets are used as extra training data for the SAR ATR network. The two aforementioned entry points for enhancing SAR ATR accuracy can be concurrently leveraged.



**Figure 1.** Entire recognition system. (a) Processes of image fusion with SOIF-CycleGAN; (b) processes of SAR ATR with a deep network. The artificial optical image and the artificial SAR image output by SOIF-CycleGAN are introduced into the SAR ATR. 0 to 7 in Type represent eight types of aircraft targets, respectively.

# 3.1. Bidirectional Translation Network

With a series of modifications, SOIF-CycleGAN is proposed based on CycleGAN to achieve high-quality bidirectional image translation. These modifications will be covered below in the introduction to the network architecture and loss function.

## 3.1.1. Network Architecture

SOIF-CycleGAN consists of two paths, S2O and O2S, and its architecture is fully illustrated in Figure 2. The S2O path will be introduced first, which contains an S2O generator and an S2O discriminator. The S2O generator has the input of the real SAR image and the output of the corresponding artificial optical image, which adopts the encoder-decoder structure and contains nine residual blocks in the middle of the network. Its network composition is presented in Table 1; first, the input single-channel grayscale image  $256 \times 256$  in size is normalized to the interval [-1, 1] and is encoded by three groups of convolution + instance normalization + ReLU. This process conducts dimension compression and channel expansion for feature maps. The parameter C of the convolutional layers in Table 1 represents the number of output channels, K the size of convolution kernels, S the stride, and P the padding. Instancenormalization normalizes each sample, which is often used in style transfer tasks [51]. Next, the following series of residual blocks help to optimize even large networks with inputs and outputs of the same size [52]. Finally,

the decoding process adopts upsampling layers to expand the dimension and convolution groups to realize the compression of the channel, which restores the feature map to the size of the input image. Note that the image is finally normalized to the range [-1, 1] by the Tanh activation function [53].



Figure 2. Architecture of SOIF-CycleGAN.

Table 1. Architecture of generators.

Layer Information	Output Shape
Cov(C64, K7, S1, P3) + InsNorm + ReLU	(64 × 256 × 256)
Cov(C128, K3, S2, P1) + InsNorm + ReLU	$(128 \times 128 \times 128)$
Cov(C256, K3, S2, P1) + InsNorm + ReLU	$(256 \times 64 \times 64)$
ResBlock(C256)	$(256 \times 64 \times 64)$
Upsample(S2)	$(256 \times 128 \times 128)$
Cov(C128, K3, S1, P1) + InsNorm + ReLU	$(128 \times 128 \times 128)$
Upsample(S2)	$(128 \times 256 \times 256)$
Cov(C64, K3, S1, P1) + InsNorm + ReLU	$(64 \times 256 \times 256)$
Cov(C1, K7, S1, P3) + Tanh	$(1 \times 256 \times 256)$

The artificial optical image output by the S2O generator has three destinations: (1) the artificial optical image is fed into the S2O discriminator, which is trained to try to distinguish artificial optical images from real ones. Table 2 shows the structure of the S2O discriminator. Four groups of convolution + instance normalization + LeakyReLU transform images into feature maps with 512 channels and a size of  $16 \times 16$ . Using LeakyReLU activation in the discriminator follows the guidelines for stable deep convolutional GANs in [53]. Through a final convolutional layer with one output channel, the output is a  $15 \times 15$  matrix, in which each value in the matrix represents the correctness of the corresponding local patch. The MSE error between this matrix and the real or fake labels of the input optical image is the GAN loss of S2O. No sigmoid is added after the final convolutional layer borrowing from Wasserstein GAN [54]. (2) Through supervised learning, the error is calculated for the artificial optical image and its corresponding real optical image, which is named S2O supervised learning loss. This supervised learning loss will also be backpropagated to

the S2O generator to adjust the parameters, making the output closer to the real optical image. The reason for adding supervised learning is elaborated in Section 3.1.2. (3) The artificial optical image is fed into the O2S generator, translating images from the optical image domain into the SAR image domain to output the reconstructed SAR image. The error between the reconstructed SAR image and the real SAR image is named S2O-O2S cycle-consistency loss, the reduction of which represents the improvement of the joint performance of the two generators. The O2S generator has the same structure as the S2O generator, except that noise injection [55] is added after the last two convolutions to increase the diversity of the output artificial SAR images.

Table 2. Architecture of discriminators.

Layer Information	Output Shape
Cov(C64, K4, S2, P1) + InsNorm + LeakyReLU	$(64 \times 128 \times 128)$
Cov(C128, K4, S2, P1) + InsNorm + LeakyReLU	$(128 \times 64 \times 64)$
Cov(C256, K4, S2, P1) + InsNorm + LeakyReLU	$(256 \times 32 \times 32)$
Cov(C512, K4, S2, P1) + InsNorm + LeakyReLU	$(512 \times 16 \times 16)$
Cov(C1, K4, S1, P1)	$(1 \times 15 \times 15)$

As for the O2S path, it is totally the same in the calculation of both GAN loss and cycle-consistency loss using the S2O path, with a discriminator of the same structure. Nevertheless, instead of using supervised learning, the O2S path adopts a histogram discriminator to determine whether the statistic of artificial SAR images conform to the distribution in real SAR. The function that directly computes the histogram of images is not continuously derivable, which will prevent backpropagation. Therefore, we set a sigmoid function as the threshold for counting the number of points in each interval, and the resulting histogram is an approximation of the true histogram of the image. The structure of the histogram discriminator adopts a multi-layer perceptron, which is presented in Table 3. The whole SOIF-CycleGAN exploits the adversarial learning between the generators and the discriminators, the joint learning between the S2O and the O2S paths, and the combination of supervised and unsupervised learning, to improve the performance of both the generators and the discriminators together. Eventually, a powerful SAR-optical image bidirectional translation capability is acquired.

Table 3. Architecture of histogram discriminators.

Layer Information	Output Shape
Linear(C512) + LeakyReLU Linear(C256) + LeakyReLU	(1 × 512) (1 × 256)
Linear(C1)	$(1 \times 250)$ $(1 \times 1)$

# 3.1.2. Loss Function

The S2O and O2S translation need to be discussed and designed separately due to the huge difference in target characteristics caused by the distinct mechanisms of SAR imaging and optical imaging. The optical image obtained by computer simulation contains complete target structure information and rich details without noise. The SAR image from the real SAR system has inevitable speckle noise and clutter, with the target structure information emerging or disappearing with the change in the viewpoint. Meanwhile, because the SAR image is a combination of scattering points, the target details are difficult to identify. Therefore, the SAR image and the optical image cannot be considered equivalent from the perspective of the target structure restoration through image translation. The target structure information contained in SAR images is usually a subset of that found in optical images. In other words, the O2S translation is an overdetermined problem that can be addressed via unsupervised learning with weaker constraints. While the S2O translation is underdetermined, additional constraints need to be incorporated to ensure that the

network is effectively trained, such as supervised learning. Therefore, as shown in Figure 2, in the O2S translation path, the network is trained with the unsupervised GAN loss and the cycle-consistency loss in typical CycleGAN. In contrast, supervised loss functions are added to the S2O translation path to constrain the training of the network. The test results obtained using different loss functions separately in Section 6.1 further support the above inferences. All the losses used in SOIF-CycleGAN will be described as follows.

Think of it in terms of the loss function, where the trainable discriminator evaluating whether the output of the generator is close enough to the real samples is essentially an adaptive loss function, which is known as GAN loss. GAN loss usually achieves better results than a fixed loss because its objective function adjusts as the discriminator becomes more powerful during training, forcing the generator to move towards a higher standard of performance. Equation (1) shows the total GAN loss in SOIF-CycleGAN, where the real optical image O and the real SAR image S follow the distribution probabilities of  $p_{data}(O)$  and  $p_{data}(S)$ , respectively.

$$L_{\text{GAN}} = E_{\text{O}\sim p_{\text{data}}(\text{O})} \left[ D_{\text{S2O}}(\text{O})^2 \right] + E_{\text{S}\sim p_{\text{data}}(\text{S})} \left[ 1 - D_{\text{S2O}}(G_{\text{S2O}}(\text{S}))^2 \right] + E_{\text{S}\sim p_{\text{data}}(\text{S})} \left[ D_{\text{O2S}}(\text{S})^2 \right] + E_{\text{O}\sim p_{\text{data}}(\text{O})} \left[ 1 - D_{\text{O2S}}(G_{\text{O2S}}(\text{O}))^2 \right]$$
(1)

In the O2S translation path of SOIF-CycleGAN, the histogram discriminator for monitoring the statistical parameters of the generated image is also adopted. A derivable histogram calculation function is designed by using a steep Sigmoid function instead of the activation function at the thresholds. The approximate number of points in each histogram interval can be denoted as Equation (2), where *b* represents the width of the intervals, *m* is the sequence number of the intervals, *I* represents the image, *N* represents the number of pixels in the image, *S* represents the Sigmoid function, and *W* is its weight parameter. The histogram GAN loss in the O2S translation path can be denoted as Equation (3):

$$Hist_{m}(I) = \sum_{n=1}^{N} S(I(n) - m \cdot b) - \sum_{n=1}^{N} S(I(n) - (m+1) \cdot b)$$

$$S(x) = \frac{1}{1 + e^{-w \cdot x}}$$
(2)

$$L_{\text{Hist}} = E_{\text{S} \sim p_{\text{data}}(\text{S})} \left[ D_{\text{Hist}}(\text{Hist}(\text{S}))^2 \right] + E_{\text{O} \sim p_{\text{data}}(\text{O})} \left[ 1 - D_{\text{Hist}}(\text{Hist}(G_{\text{O2S}}(\text{O})))^2 \right]$$
(3)

When calculating cycle-consistency loss, the original image passes through a pair of opposite generators in turn, and the output reconstructed image is obtained. Then, the reconstructed image is compared with the original image using a fixed loss function (usually L1). Cycle consistency loss provides constraints on the two generators during training so that their functions remain symmetric. Using it in isolation from other loss functions is not effective. The total cycle-consistency loss in SOIF-CycleGAN is denoted by Equation (4).

$$L_{\text{cycle}} = E_{S \sim p_{\text{data}}(S)} [\|G_{\text{O2S}}(G_{\text{S2O}}(S)) - S\|_1] + E_{O \sim p_{\text{data}}(O)} [\|G_{\text{S2O}}(G_{\text{O2S}}(O)) - O\|_1]$$
(4)

The artificial optical image output in the S2O translation path is also directly compared with the corresponding real optical image through supervised learning. The supervised learning loss combines the L1 loss that focuses on the whole and the LPIPS loss that focuses on the local (demonstrated in Section 6.1), which is denoted as Equation (5). L1 loss is one of the most commonly used loss functions for image translation tasks and is obtained by taking the L1 norm of two images pixel by pixel and then averaging this. Learned perceptual image patch similarity (LPIPS) loss [56] compares the two input images by a neural network that has been fully trained on an optical image dataset for feature extraction. AlexNet trained by ImageNet is used when computing LPIPS in this paper. The outputs of

each layer of AlexNet are activated, normalized, and weighted. Then, the spatial-average L2 norm is calculated and averaged to obtain LPIPS. The calculation process of LPIPS is shown in Figure 3.



Figure 3. The calculation process of LPIPS.

$$L_{\text{super}} = E_{S \sim p_{\text{data}}(S)}[\|G_{\text{S2O}}(S) - O\|_{1}] + E_{S \sim p_{\text{data}}(S)}[L_{\text{LPIPS}}(G_{\text{S2O}}(S), O)]$$
(5)

SOIF-CycleGAN optimizes all the loss functions in each learning epoch. The identity loss commonly used in CycleGAN can help keep the hue of RGB images stable. In this study, the SAR image and the optical image obtained by simulation with reference to active infrared imaging do not contain color information, so this loss is not used.

## 3.2. Recognition Network

A modified LeNet network is adopted as the recognition network. LeNet uses the classical convolutional layer + linear layer architecture, and most mainstream image recognition networks are extended and improved based on this architecture. The objective of this study is to verify that SAR-optical image data fusion can enhance the performance of SAR ATR rather than be used to explore the upper limit of accuracy achievable by various SAR ATR methods in the proposed SPH8 dataset. We finally chose a concise network because an excessively intricate structure runs the risk of masking the benefits derived from the data.

#### 4. Experiments

In this section, the source and composition of the SPH8 dataset are presented in detail first. Next, the training parameters and hardware configuration in the experiments are provided.

# 4.1. Dataset

A new SPH8 dataset containing paired, as well as unpaired, SAR-optical target images is created for the SAR-optical image fusion by using the combination of supervised and unsupervised learning in this paper. The targets in SPH8 cover five types of fixed-wing aircraft (Quest Kodiak 100 Series II, Cessna 208B, Air Tractor 504, PC-12, and Beech King Air 350) and three types of helicopters (Ka-32, AW 139, and AS350). The SAR image in SPH8 contains HH, HV, and VV polarization modes, with a resolution of 0.3 m × 0.3 m, which is obtained from a UAV-borne Ku-band SAR for the multi-angle imaging of ground aircraft targets, with a center frequency of 14.6 GHz, a bandwidth of 600 MHz, a sideways viewing Angle of 45°, and a flight altitude of 150 m. Due to the small sideways viewing angle and the range-based imaging mechanism, the target in the original SAR image is inverted with the layover effect, which is not in accordance with the perspective used by optical imaging and human vision. Thus, the SAR image in SPH8 is the result of flipping the original SAR image upside down; that is, the upper end of the image is proximal, and the lower end is distal. The optical images are obtained through a ray-tracing algorithm referring to active infrared imaging. As active infrared imaging actively emits electromagnetic waves and receives the echoes for imaging just like SAR imaging, they share similar radiation expression. The noncoherent imaging mechanism of active infrared imaging enables it to avoid the speckle noise in an SAR image. The elaborate CAD models of the targets are established based on prior knowledge, with their surfaces programmed to be smooth, referring to the strong specular reflection of the metal shells to microwaves. The creation of the dataset is shown in Figure 4, with the illustration of UAV SAR imaging of the real scene and the simulated active infrared imaging of the CAD scene. The aircraft position during SAR imaging is obtained, and a camera is set at the corresponding position in the CAD scene as the UAV works at a fixed altitude (H = 150 m) and the multi-view routes are known. Analogous to the electromagnetic plane wave in the far field, the light source is set to parallel the infrared rays with the same  $\theta = 45^{\circ}$  as the viewpoint of SAR imaging in the CAD scene. The ray-tracing algorithm tracks the incoming ray backwards at each pixel of the image received by the camera and calculates the reflection and refraction of the ray according to the target and generates optical images, which are then matched with SAR images.



Figure 4. Creation of the SPH8-P dataset.

The paired data, named SPH8-P, are the SAR image and the corresponding optical image under the same viewpoint. Samples from SPH8-P and photos of the corresponding targets are shown in Figure 5. Among them, the SAR images with three polarization modes under the same viewpoint are given a category; they share one simulated optical image as an independent sample to give the network the ability to process SAR images with different polarization modes. There is high consistency between the SAR images and the corresponding simulated optical images. SPH8-P contains 269 categories, with a total of 807 SAR-optical image pairs. Besides, the optical images of the eight targets under the angle views of  $0-355^{\circ}$ , with  $5^{\circ}$  interpolation (some samples are shown in Figure 6) are named SPH8-U, which formed the unpaired data with the SAR images in SPH8-P. As the viewpoints of SAR imaging are not comprehensive due to the limitation of field test conditions, SPH8-U makes a supplement. All images are converted into 8-bit grayscales with a size of  $256 \times 256$  pixels and are labeled according to the type of the single aircraft target they contain.



**Figure 5.** Samples of SPH8-P and photos of the targets. Columns (**a**–**h**) show one of the categories of the Quest Kodiak 100 Series II, Cessna 208B, Air Tractor 504, PC-12, Beech King Air 350, Ka-32, AW 139, and AS350, respectively. Images of each row from top to bottom belong to SAR images in HH, SAR images in HV, SAR images in VV, optical images, and photos of the targets, respectively.



**Figure 6.** Samples of SPH8-U. The sample images of eight types of aircraft are arranged into an ellipse according to their poses, with their backgrounds removed. In fact, the computer simulation can generate optical images of any type of target under any viewpoint.

### 4.2. Implement Details

All the optimizers used for training the networks are Adam. The initial learning rate is 0.0002, which becomes one-tenth of the previous learning rate every 100 epochs. The decreasing learning rate allows the network training to go from drastic changes to fine-tuning and helps the network converge reach a better result.

Using both supervised and unsupervised learning, SOIF-CycleGAN requires paired and unpaired data for training simultaneously. Each training or test requires a pair of SAR-optical images from SPH8-P and an optical image from SPH8-U. Thus, SPH8 is randomly divided into five nearly equal groups, each containing 1/5 SPH8-P and 1/5 SPH8-U. It takes five trials to translate all images of SPH8 in the test, using five sets as test data in turn and using the others as training data. The images of the three different polarization modes contained in each category in SPH8 are similar. Images from one category appearing in the training dataset and the test dataset at the same time should be avoided. Thus, the category is used as the smallest unit to split the groups.

Data augmentation is exploited to expand the number of patterns in the training data, which can avoid overfitting. Random cropping and random horizontal flipping are considered reasonable because data augmentation should simulate what would happen in

a real situation. SAR and optical image pairs from SPH8-P are subjected to exactly the same data augmentation in one round of training to keep their semantics consistent. In addition, the values of the images in both the training and test data are normalized to the interval [-1, 1]. The total number of parameters for both the S2O generator and O2S generator is 43.36 MB. ALL the experiments are implemented based on an Intel(R) Xeon(R) Gold 5218R CPU at 2.10 GHz and an NVIDIA GeForce RTX 3090 GPU with a dedicated GPU memory of 24.0 GB. The training time of the whole SOIF-GAN for 500 epochs is 10.5 h. All the codes are written in Python 3.8.5 (A programming language first released by Guido van Rossum as Python 0.9.0 in Amsterdam) , using the deep learning tools of the Pytorch 1.7.1 (An open-source machine learning framework developed by Meta AI) package in Anaconda.

## 5. Results and Analysis

In this section, the results of data fusion are presented and evaluated through both human vision and IQA methods. Subsequently, the results of the SAR ATR experiments enhanced by image fusion are shown.

## 5.1. Results of Image Fusion

In this study, the artificial optical image and the artificial SAR image output by SOIF-CycleGAN are introduced into the SAR ATR. The quality of the translated images is the key to achieving the improvement of accuracy. By exploiting the grouping method in Section 4.2 to train and test SOIF-CycleGAN, translated images corresponding to all the SAR images and optical images are obtained. Each training lasted 500 epochs. Due to the diversity of image fusion tasks, there is no uniform standard to evaluate the quality of the image fusion results. We chose human vision and IQA methods to evaluate the quality of SAR-optical image fusion combined with practical application and mathematical analysis.

Samples of the S2O translation results are shown in Figure 7. Firstly, by identifying the local features in the SAR image and translating them into the corresponding expression in the optical image, the set of scattered points in the target is transformed into a continuous region with the chiaroscuro added, and the speckle noise in the background is purified. These changes improve the image quality and make the artificial optical image very suitable for human eye observation. Some clutter caused by the SAR imaging process, such as bright stripes in the background in the HV SAR image of Figure 7e and the HH, VV SAR images of Figure 7g, are effectively judged as noise and eliminated. Even some extreme clutter, such as the aircraft propeller in the HH, VV SAR images of Figure 7b, and the aircraft tail in all SAR images of Figure 7c (the formation of the angular reflection structure brings the strong trailing appearing near the strong scattering point), are not misjudged as entities. Secondarily, the main bodies of the aircraft targets, including fuselages, wings, and tail fins (and including the horizontal stabilizer fin and the vertical fin), are successfully restored to their optical counterpart. Thanks to the abundant prior information from the optical image, any missing and distorted details, like undercarriages and wing radars, are also recovered. Structure reconstruction and detail recovery can effectively enhance the recognition of aircraft types. For instance, when distinguishing the target types in Figure 7a,b, it is difficult to judge when using the SAR images because they are both high-wing. However, with the ratio of wings and tails, the aspect ratio, the size of undercarriages, and the position of wing radar in the artificial optical images, we can effectively distinguish them. In fact, these artificial optical images can greatly reduce the difficulty of manual interpretation and recognition. Moreover, there are usually slight pose differences between the targets in the artificial optical images and the optical images, which is particularly obvious in Figure 7b,e. When using computer simulation, due to the difference in mapping geometry and the minor error of viewpoint, the target poses in some generated optical images are different from those in SAR images. Even if there are semantic information errors in the training data, with the help of unsupervised learning, SOIF-CycleGAN fully respects the semantic information in the SAR images to complete the S2O translation task. This indicates the network possesses an excellent local feature reconstruction ability without

overfitting. These positive results verify the effectiveness of the S2O translation path of SOIF-CycleGAN. In practical applications, S2O translation can also effectively facilitate human interpretation and improve the efficiency of annotating SAR images, resulting in a more efficient dataset construction process.



**Figure 7.** Samples of S2O translation results. Columns (**a**–**h**) show the results of different categories, respectively, among which (**a**) belongs to Quest Kodiak 100 Series II, (**b**) Cessna 208B, (**c**) Air Tractor 504, (**d**) PC-12, (**e**) Beech King Air 350, (**f**) Ka-32, (**g**) AW 139, and (**h**) AS350. The first row shows optical images, then SAR images in HH, SAR images in HV, and SAR images in VV, each followed by the corresponding artificial optical images.

In order to numerically evaluate the results of S2O translation, we used the three IQA methods of SSIM, PSNR, and LPIPS. SSIM evaluates an image from the perspective of human visual perception, considering the brightness, contrast, and structure of the image, which is defined as Equation (6).  $\mu_X$  and  $\mu_Y$  represent the mean values of images *X* and *Y*, respectively,  $\sigma_X$  and  $\sigma_Y$  the variances of images *X* and *Y*, respectively,  $\sigma_{XY}$  the covariance of images *X* and *Y*, and *C*1, *C*2, and *C*3 are the constants for avoiding the denominator being 0. A higher SSIM means lower image distortion.

$$SSIM(X,Y) = l(X,Y) \cdot c(X,Y) \cdot s(X,Y)$$

$$l(X,Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1}, c(X,Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2}, s(X,Y) = \frac{2\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}$$
(6)

PSNR, denoted as Equation (7), is defined by the mean square error (MSE), which calculates the difference between the corresponding pixels in the images point by point.

 $MAX_I^2$  means the maximum value of the image pixels, which is 255 in the 8-bit grayscales used in this paper. The higher the PSNR value, the less distortion appears.

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$
(7)

Since most images generated by GAN are too smooth, traditional evaluation methods may lose accuracy in terms of the assessment of image quality [56]. In order to solve this problem, the use of LPIPS [56] is proposed to compare the quality of two images by using a trained deep network, which is more in accordance with the assessment of human vision. The method to compute the LPIPS of two images has been presented in Section 3.1.2.

Table 4 displays a comparison of the S2O results between SOIF-CycleGAN and Pix2Pix in the previous work [5]. In that, the SSIM, PSNR, and LPIPS are obtained by using the corresponding real optical image as references and taking the average result after calculation. As can be seen from the table, when compared to Pix2Pix, which only uses supervised learning, SOIF-CycleGAN combines supervised and unsupervised learning, gives the target clearer main bodies and more realistic details with lower noise, and improves the SSIM, PSNR, and LPIPS significantly. These evaluations further validate the advancement of SOIF-CycleGAN in the S2O translation task.

**Table 4.** Comparison of S2O results between SOIF-CycleGAN and Pix2Pix. The number in bold indicates that one is better in the evaluation of the translation results.

	SAR	Pix2Pix	SOIF-CycleGAN	Optical
SSIM↑	0.4312	0.7420	0.8000	1
PSNR↑	18.1603	21.4738	22.5710	$+\infty$
LPIPS↓	0.4253	0.1236	0.0855	0
Samples		* *	* *	* *

Figure 8 illustrates the results of the O2S translation and the real HH SAR images under similar viewpoints. As shown in Figure 8, the O2S translation path of SOIF-Cycle can accurately predict the distribution of scattering intensity on the target from different viewpoints and successfully restore the point scattering characteristics of SAR images. For example, in the SAR target image in Figure 8c, the body, tail, and wing ends of the target have strong scattering, and these parts are also prominently emphasized in the corresponding artificial SAR image. Meanwhile, the special effects in SAR images, such as the secondary scattering formed by the wing and the ground in Figure 8a and formed by the tail and the ground in Figure 8h, are also accurately restored in the artificial SAR images. The details in the real SAR images are complex, which is expressed differently in the SAR images obtained in polarization modes. The current network is unable to fully restore the speckle noise texture and distinguish between polarization modes, resulting in slight differences in the appearance and the histograms of the artificial SAR images compared to the real SAR images. However, these artificial SAR images do partially capture SAR target features under different viewpoints, thereby increasing the diversity of the patterns in the training data and contributing towards the improved performance of SAR ATR. The specific verification will be detailed in the next subsection.



**Figure 8.** Samples of S2O translation results. Columns (**a**–**h**) show the results of different categories respectively, among which (**a**) belongs to the Quest Kodiak 100 Series II, (**b**) Cessna 208B, (**c**) Air Tractor 504, (**d**) PC-12, (**e**) Beech King Air 350, (**f**) Ka-32, (**g**) AW 139, and (**h**) AS350. The first row shows optical images, then the corresponding artificial SAR images, followed by the real SAR images, and finally, the comparisons between the histograms of the real and the artificial SAR images.

#### 5.2. Results of SAR ATR Enhanced by Image Fusion

SAR-optical co-registration image recognition and SAR data augmentation are the two entry points for using SAR-optical image fusion to enhance SAR ATR. In order to verify the effectiveness of the two entry points for improving SAR ATR accuracy, an ablation study is conducted based on the modified LeNet network. In the baseline experiment labeled as Experiment 0, all the SAR images in SPH8 are randomly divided into equal sets with no overlap in categories. Each group is used as the test set of LeNet and the other as the training set in turn. The average test accuracy obtained from these two (training and test) is used as the final accuracy. On the basis of Experiment 0, Experiment 1 uses SAR-optical co-registration images as the input of the two-channel LeNet. In order to simulate a realistic situation, the two channels of training data are the real SAR image and the real optical image, whereas the two channels of test data are the real SAR image and the artificial optical image. In addition, the test data consist of the real SAR image and the real optical image, and this is also used to test the network as a control. Experiment 2 added artificial SAR images to the training data based on Experiment 0. Experiment 3, which uses both SAR-optical co-registration images and SAR data augmentation, is a synthesis of Experiment 1 and Experiment 2. In order to exclude uncertainties in training, each experiment is repeated five times, and the highest accuracy rates achieved are used as the final results.

The results of the ablation study are presented in Table 5. First of all, by comparing the results of Experiment 1 and Experiment 0, it can be concluded that SAR-optical co-registration images as the input of the recognition network can improve the accuracy outstandingly. Secondly, in Experiment 1 and Experiment 3, the test results using the artificial optical image and the artificial optical image have small differences, which indicates that the artificial optical image is quite close to the real optical image in the view of the recognition network. The high-quality S2O translation ability of the proposed SOIF-CycleGAN is further demonstrated. Next, when comparing the results of Experiment 2 and Experiment 0, the accuracy of the network trained by using SAR data augmentation is improved, although not significantly. The accuracy of Experiment 2 is lower than that of Experiment 0 in a few tests. It was found that although the artificial SAR image restores the structure and radiometric expression of the target in SAR images, the details are still a certain distance from the real SAR image. In order to achieve higher accuracy, the O2S translation of SOIF-CycleGAN has room for improvement. Finally, based on the comprehensive results, two entry points of using SAR-optical image fusion to enhance SAR ATR are both valid and can be used jointly to further improve the performance of SAR ATR.

**Table 5.** Results of ablation study. The data types connected by a | means they compose the SAR-optical co-registration image as different channels, and with a +, this means additional data are added.

Experiment	Input Channel	Training Data	Test Data	Accuracy
0	One	Real SAR	Real SAR	79.92%
1	Тwo	Real SAR   Real optical	Real SAR   Real optical	86.00%
1	100		Real SAR   Artificial optical	85.50%
2	One	Real + Artificial SAR	Real SAR	81.54%
3	Two	Real + Artificial SAR   Real optical	Real SAR   Real optical Real SAR   Artificial optical	87.61% 86.25%

### 6. Discussion

In this section, we first show the effect of different loss functions on the results of S2O translation, which justifies our adoption of the joint loss function. Then, the effect of the differences in the sample sizes among target types on image fusion is discussed. In the end, the robustness of S2O translation is demonstrated by displaying some special cases.

#### 6.1. Effect of Loss Functions on S2O Translation

In order to test the effect of the loss functions, both in supervised learning and unsupervised learning, on the S2O translation results and to select a better loss function combination, a series of experiments were conducted. All training in the experiments lasted 150 epochs using the same training set, and the trained networks were tested on the same test set.

On the one hand, CycleGAN loss (a combination of GAN losses and cycle consistency loss), L1 loss, SSIM loss, and LPIPS loss are used to train SOIF-CycleGAN, respectively. Firstly, as per the average IQA values and result samples shown in Table 6, L1 loss achieves better SSIM and PSNR, and LPIPS loss performs better on the LPIPS. Subsequently, when used alone, CycleGAN loss can restore the local features of the targets well, such as with the continuous regions and realistic chiaroscuro, whereas it is not accurate for the main bodies and contours of the targets, which is caused by the weak constraint of unsupervised learning. In contrast, the three losses using supervised learning: L1 loss, SSIM loss, and LPIPS loss, can restore the main bodies of the targets better. Thirdly, by observing the training process and the corresponding result samples in Table 6, it can be found that the networks using L1 loss and SSIM loss first learn the low-frequency main-body information to obtain the fuzzy target body and then gradually optimize the high-frequency information. Conversely, the network using LPIPS loss first learns the high-frequency detail information and then optimizes the contour. This phenomenon is related to the different optimizing methods of the networks using different loss functions. As L1 loss and SSIM loss optimize the average of all pixel errors, the low-frequency components that account for the vast majority of errors are optimized first. Whereas, as LPIPS loss calculates the average error after decomposing the image through convolution kernels, the learning of the reference image is limited to the range of each convolution kernel, so the image is optimized locally first.

On the other hand, when combining supervised and unsupervised learning, Cycle-GAN loss is combined with L1 loss, SSIM loss, and LPIPS loss, respectively. The three IQA values of CycleGAN + L1 loss, CycleGAN + SSIM loss, and CycleGAN + LPIPS loss comprehensively improve compared to using only supervised losses, which is thanks to CycleGAN loss. In addition, the combination of whole-focused loss and local-focused loss can take into account both the main bodies and details of the target in each round of learning, which can greatly shorten the period required for training. Therefore, two combinations

of CycleGAN + L1 + LPIPS loss and CycleGAN + SSIM + LPIPS loss were tested. Among them, CycleGAN + L1 + LPIPS loss obtains the highest SSIM while maintaining better PSNR and LPIPS, according to Table 6, the result samples of which have targets with clear edges and rich details. Based on the above experimental results and analysis, CycleGAN + L1 + LPIPS loss is determined as the final joint loss in this paper.

**Table 6.** Test result of S2O translation trained with different loss function combinations for 150 epochs. The number in bold indicates that one is better in a set of comparisons in the evaluation of the translation results. CycleGAN + L1 + LPIPS is the final loss combination.

Туре	<b>SSIM</b> ↑	PSNR↑	LPIPS↓	Samples
SAR	0.4312	18.1603	0.4253	
CycleGAN	0.6575	20.0650	0.1673	$\varkappa$ $\varkappa$ $\varkappa$
L1	0.7830	22.2319	0.1262	そ ト ギ ゲ
SSIM	0.7412	21.7073	0.1417	* * * *
LPIPS	0.7765	21.8677	0.0894	<b>米 水 水 水</b>
CycleGAN + L1	0.7917	22.5961	0.1122	* * * *
CycleGAN + SSIM	0.7589	22.3098	0.1169	* * * *
CycleGAN + LPIPS	0.7880	22.0856	0.0857	ネ ふ ぞ が
CycleGAN + L1 + LPIPS	0.7999	22.4452	0.0879	<b>寒水</b> 家 X
CycleGAN + SSIM + LPIPS	0.7975	22.4793	0.0881	メ ネ ギ ギ

# 6.2. Effect of the Unequal Sample Number on Image Fusion

In the translation results shown in Section 5.1, the Quest Kodiak 100 Series II, Cessna 208B, Air Tractor 504, and Ka-32 achieve satisfactory outcomes. However, the PC-12, Beech King Air 350, AW 139, and AS350 only just obtained acceptable results. This problem

is mainly caused by the small sample number in the latter four types of targets. Due to the irresistible factor in field experiments, the number of various types of targets in the available SAR images varies greatly. In order to quantitatively represent the relationship between the sample number and the image translation performance, three IQA methods were used to evaluate the S2O translation results for eight types of targets, respectively. Regarding the relationship between the three IQA values and the sample number shown

between the sample number and the image translation performance, three IQA methods were used to evaluate the S2O translation results for eight types of targets, respectively. Regarding the relationship between the three IQA values and the sample number shown in Figure 9, there is a strong positive correlation between the sample number and the quality of the artificial optical images. Furthermore, in addition to the difference in the number of samples between types, the image translation performance is also affected by the quality of the sample images, the total number of fixed-wing aircraft samples and helicopter samples, and other factors. For example, Air Tractor 504, which does not have a maximum sample number, also achieves the top three average IQA values, as most of its samples belong to successful SAR imaging results. Whereas, because the total sample number of fixed-wing aircraft is significantly larger than that of helicopters, the overall translation results of the three helicopters Ka-32, AS350, and AW 139 are relatively poor. The above analysis also demonstrates that the network, trained on a substantial amount of data, exhibits adaptability to new target types with small sample numbers. Therefore, it is reasonable to anticipate that the image fusion performance of SOIF-CycleGAN will be further improved with the continued implementation of field experiments and expansion of the number of samples.



**Figure 9.** Scatter diagram of the three IQA values versus the number of samples of eight types of targets. (a) SSIM, with the linear correlation coefficient of 0.9195; (b) PSNR, with the linear correlation coefficient of 0.8050; (c) LPIPS, with the linear correlation coefficient of -0.7517.

# 6.3. Spacial Cases

Robustness is crucial for image fusion networks, which denotes the network's ability to maintain a stable output despite disturbances in the input image. Simulating the disturbances encountered by SAR images in practical applications using precise mathematical methods can be challenging. Fortunately, there are some disturbed samples in the SAR images of SPH8, which can aid our understanding of the performance of SOIF-CycleGAN. These special cases of S2O are shown in Figure 10, including the S2O translation results under four different types of disturbances: azimuth ambiguity, azimuth ghosting, extra bright strips, and missing structures.

- 1. Azimuth ambiguity. The results of the samples with azimuth ambiguity are shown in Figure 10a,b. The severe ambiguity makes the target in the SAR image difficult to recognize, which also affects the artificial optical image, resulting in slight geometric distortions and missing structures, such as the tail in Figure 10a,b. Nevertheless, the target can be effectively recovered in artificial optical images, which greatly facilitates target recognition.
- 2. Azimuth ghosting. As shown in Figure 10c,d, azimuth ghosting appears in the SAR images, which is caused by the periodic high-frequency vibration of the platform.

Despite the extra entities in the SAR images causing the translation network confusion, it still adheres to the prior knowledge and avoids generating an aircraft with four wings. However, the extra fuselage still results in an elongated nose on the aircraft.

- 3. Extra bright strips. In Figure 10e–i, bright strips can be classified into two types: periodic and aperiodic. The periodic bright stripes are well eliminated in Figure 10e,f, which hardly affect the results of the transformation. However, the aperiodic bright strips in Figure 10g and the HH SAR image in Figure 10h show the target structure missing in the same place as the artificial optical images. The bright stripes in the tail of the SAR images in Figure 10i arise from the secondary scattering of the rotor and the ground, which are regarded as an entity and added to the tails of the aircraft.
- 4. Missing structures. Due to the edge of the imaging area, the targets in Figure 10e,i have missing wings. With the help of prior information, the translation network restores the wings of the aircraft but with a slight distortion. Proper incomplete information completion is also highly advantageous for target recognition.



**Figure 10.** Special cases of S2O translation results. Columns (a,b) show the results of samples with azimuth ambiguity, columns (c,d) show the results of samples with azimuth ghosting, columns (e-i) show the results of samples with extra bright strips, and columns (j,k) show the results of samples with missing structures. The first row shows optical images, then SAR images in HH, SAR images in HV, and SAR images in VV, each followed by the corresponding artificial optical images.

# 7. Conclusions

The experimental results and analysis presented in this paper verify that SAR-optical image fusion can significantly enhance the performance of SAR ATR. We propose a SAR-optical image fusion network named SOIF-CyleGAN for high-quality image bidirectional translation. New constraints, such as supervised learning and additional discriminators,

are introduced into the training process of SOIF-CyleGAN by analyzing the characteristics of S2O and O2S image translation tasks separately. A joint loss function that addresses both the whole and local considerations was adopted to further improve the training efficiency and translation performance of the network. The network proposed in this paper exhibits significant improvements in terms of both visual effects and IQA values compared to previous works. The artificial optical image and the artificial SAR image obtained through image bidirectional translation can result in 1.62% and 5.58% accuracy improvements, respectively, when introduced into SAR ATR, and this value reaches 6.33% when the two are used simultaneously, as the ablation study results show. Meanwhile, a new approach for constructing SAR-optical image pairs of targets is proposed and validated, with the optical images generated by the computer simulation referring to active infrared imaging, which yields highly consistent semantic information with the corresponding SAR images. Based on this approach, a new multi-view SAR-optical image dataset named SPH8 has been created, which can support various tasks, such as SAR ATR, SAR-optical image fusion, and multi-polarimetric SAR information fusion, whether supervised or unsupervised. Additionally, the exploration of the impact of diverse loss functions and sample number discrepancies among different types on the SAR-optical image fusion holds a high reference value for future studies.

In future research, the focus will be on improving the performance of O2S translation. In addition, the ability to generate SAR images with specific polarization modes will also be implemented with the further expansion of the dataset.

**Author Contributions:** Conceptualization of this study, Methodology, Software, Data processing, Original draft preparation, Y.S.; Review and editing, Supervision, Project administration, W.L.; Literature collection, K.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China, grant numbers: (2018YFA0701900 and 2018YFA0701901).

**Data Availability Statement:** The SPH8 dataset mentioned in this paper was created by our team and will be made available to any potential collaborators upon the publication of this paper. If need be, please email wzli@mail.ie.ac.cn to access the SPH8 dataset.

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

# References

- 1. Li, J.; Yu, Z.; Yu, L.; Cheng, P.; Chen, J.; Chi, C. A Comprehensive Survey on SAR ATR in Deep-Learning Era. *Remote Sens.* 2023, 15, 1454. [CrossRef]
- Liu, L.; Lei, B. Can SAR Images and Optical Images Transfer with Each Other? In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018. [CrossRef]
- 3. 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]
- 4. Yang, X.; Zhao, J.Y.; Wei, Z.Y.; Wang, N.N.; Gao, X.B. SAR-to-optical image translation based on improved CGAN. *Pattern Recognit*. **2022**, 121, 108208. [CrossRef]
- Sun, Y.; Jiang, W.; Yang, J.; Li, W. SAR Target Recognition Using cGAN-Based SAR-to-Optical Image Translation. *Remote Sens.* 2022, 14, 1793. [CrossRef]
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* 2020, 63, 139–144. [CrossRef]
- Gao, F.; Yang, Y.; Wang, J.; Sun, J.; Yang, E.; Zhou, H. A Deep Convolutional Generative Adversarial Networks (DCGANs)-Based Semi-Supervised Method for Object Recognition in Synthetic Aperture Radar (SAR) Images. *Remote Sens.* 2018, 10, 846. [CrossRef]
- Liu, W.; Zhao, Y.; Liu, M.; Dong, L.; Liu, X.; Hui, M. Generating simulated sar images using generative adversarial network. In Proceedings of the Applications of Digital Image Processing XLI. International Society for Optics and Photonics, San Diego, CA, USA, 17 September 2018; Volume 10752, p. 1075205.
- Xie, D.; Ma, J.; Li, Y.; Liu, X. Data Augmentation of Sar Sensor Image via Information Maximizing Generative Adversarial Net. In Proceedings of the 2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT), Xi'an, China, 18–20 August 2021. [CrossRef]
- Song, Q.; Xu, F.; Zhu, X.X.; Jin, Y.Q. Learning to Generate SAR Images With Adversarial Autoencoder. *IEEE Trans. Geosci. Remote Sens.* 2022, 60, 1–15. [CrossRef]

- 11. Fan, W.; Zhou, F.; Zhang, Z.; Bai, X.; Tian, T. Deceptive jamming template synthesis for SAR based on generative adversarial nets. *Signal Process.* **2020**, 172, 107528. [CrossRef]
- Chen, S.; Wang, H.; Xu, F.; Jin, Y.Q. Target Classification Using the Deep Convolutional Networks for SAR Images. *IEEE Trans. Geosci. Remote Sens.* 2016, 54, 4806–4817. [CrossRef]
- 13. Dudgeon, D.E.; Lacoss, R.T. An overview of automatic target recognition. Linc. Lab. J. 1993, 6, 3–10.
- Keydel, E.R.; Lee, S.W.; Moore, J.T. MSTAR extended operating conditions: A tutorial. In Proceedings of the Algorithms for Synthetic Aperture Radar Imagery III. International Society for Optics and Photonics, Orlando, FL, USA, 10 June 1996; Volume 2757, pp. 228–242.
- 15. Zhao, Q.; Principe, J.C. Support vector machines for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* 2001, 37, 643–654. [CrossRef]
- 16. Bhanu, B.; Lin, Y. Genetic algorithm based feature selection for target detection in SAR images. *Image Vis. Comput.* 2003, 21, 591–608. [CrossRef]
- 17. Mishra, A.K.; Motaung, T. Application of linear and nonlinear PCA to SAR ATR. In Proceedings of the 2015 25th International Conference Radioelektronika (RADIOELEKTRONIKA), Pardubice, Czech Republic, 21–22 April 2015. [CrossRef]
- Majumder, U.; Christiansen, E.; Wu, Q.; Inkawhich, N.; Blasch, E.; Nehrbass, J. High-performance computing for automatic target recognition in synthetic aperture radar imagery. In Proceedings of the Cyber Sensing 2017. International Society for Optics and Photonics, Anaheim, CA, USA, 1 May 2017; Volume 10185, p. 1018508.
- 19. Zhang, Y.; Guo, X.; Leung, H.; Li, L. Cross-task and cross-domain SAR target recognition: A meta-transfer learning approach. *Pattern Recognit.* **2023**, *138*, 109402. [CrossRef]
- 20. Huang, L.; Liu, B.; Li, B.; Guo, W.; Yu, W.; Zhang, Z.; Yu, W. OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *11*, 195–208. [CrossRef]
- Li, B.; Liu, B.; Huang, L.; Guo, W.; Zhang, Z.; Yu, W. OpenSARShip 2.0: A large-volume dataset for deeper interpretation of ship targets in Sentinel-1 imagery. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 13–14 November 2017. [CrossRef]
- 22. Hou, X.; Ao, W.; Song, Q.; Lai, J.; Wang, H.; Xu, F. FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition. *Sci. China Inform. Sci.* **2020**, *63*, 140303. [CrossRef]
- Liu, L.; Pan, Z.; Qiu, X.; Peng, L. SAR Target Classification with CycleGAN Transferred Simulated Samples. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018. [CrossRef]
- 24. Sun, X.; Lv, Y.; Wang, Z.; Fu, K. SCAN: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution SAR images. *IEEE Trans. Geosci. Remote Sens.* 2022, 60, 5226517. [CrossRef]
- 25. Pohl, C.; Van Genderen, J. Remote Sensing Image Fusion; CRC Press: Boca Raton, FL, USA, 2016. [CrossRef]
- Merkle, N.; Fischer, P.; Auer, S.; Muller, R. On the possibility of conditional adversarial networks for multi-sensor image matching. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017. [CrossRef]
- 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. [CrossRef]
- 28. 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. [CrossRef]
- Zhang, J.; Zhou, J.; Li, M.; Zhou, H.; Yu, T. Quality Assessment of SAR-to-Optical Image Translation. *Remote Sens.* 2020, 12, 3472. [CrossRef]
- 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.
- 31. 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]
- 32. 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]
- Bermudez, J.D.; Happ, P.N.; Feitosa, R.Q.; Oliveira, D.A.B. 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]
- 34. 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] [PubMed]
- 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]
- Zhao, M.; Olsen, P.; Chandra, R. Seeing Through Clouds in Satellite Images. *IEEE Trans. Geosci. Remote Sens.* 2023, 61, 4704616. [CrossRef]
- Li, X.; Zhang, G.; Cui, H.; Hou, S.; Chen, Y.; Li, Z.; Li, H.; Wang, H. Progressive fusion learning: A multimodal joint segmentation framework for building extraction from optical and SAR images. *ISPRS J. Photogramm. Remote Sens.* 2023, 195, 178–191. [CrossRef]

- Toriya, H.; Dewan, A.; Kitahara, I. SAR2OPT: Image Alignment Between Multi-Modal Images Using Generative Adversarial Networks. In Proceedings of the IGARSS 2019, Yokohama, Japan, 28 July–2 August 2019. [CrossRef]
- Chen, Z.; Liu, J.; Liu, F.; Zhang, W.; Xiao, L.; Shi, J. Learning Transformations between Heterogeneous SAR and Optical Images for Change Detection. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022. [CrossRef]
- Li, X.; Du, Z.; Huang, Y.; Tan, Z. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 2021, 179, 14–34. [CrossRef]
- 41. Adrian, J.; Sagan, V.; Maimaitijiang, M. Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 215–235. [CrossRef]
- 42. Zhang, P.; Ban, Y.; Nascetti, A. Learning U-Net without forgetting for near real-time wildfire monitoring by the fusion of SAR and optical time series. *Remote Sens. Environ.* **2021**, *261*, 112467. [CrossRef]
- Li, J.; Li, C.; Xu, W.; Feng, H.; Zhao, F.; Long, H.; Meng, Y.; Chen, W.; Yang, H.; Yang, G. Fusion of optical and SAR images based on deep learning to reconstruct vegetation NDVI time series in cloud-prone regions. *Int. J. Appl. Earth Obs. Geoinf.* 2022, 112, 102818. [CrossRef]
- 44. Mao, Y.; Van Niel, T.G.; McVicar, T.R. Reconstructing cloud-contaminated NDVI images with SAR-Optical fusion using spatio-temporal partitioning and multiple linear regression. *ISPRS J. Photogramm. Remote Sens.* 2023, 198, 115–139. [CrossRef]
- 45. 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.
- 46. Lewis, B.; Scarnati, T.; Sudkamp, E.; Nehrbass, J.; Rosencrantz, S.; Zelnio, E. A SAR dataset for ATR development: the Synthetic and Measured Paired Labeled Experiment (SAMPLE). In Proceedings of the Algorithms for Synthetic Aperture Radar Imagery XXVI. International Society for Optics and Photonics, Baltimore, MD, USA, 14 May 2019; Volume 10987, p. 109870H.
- Auer, S.; Hinz, S.; Bamler, R. Ray-Tracing Simulation Techniques for Understanding High-Resolution SAR Images. *IEEE Trans. Geosci. Remote Sens.* 2010, 48, 1445–1456. [CrossRef]
- 48. Gartley, M.; Goodenough, A.; Brown, S.; Kauffman, R.P. A comparison of spatial sampling techniques enabling first principles modeling of a synthetic aperture RADAR imaging platform. In Proceedings of the Algorithms for Synthetic Aperture Radar Imagery XVII, International Society for Optics and Photonics, Orlando, FL, USA, 18 April 2010; Volume 7699, p. 76990N.
- Du, S.; Hong, J.; Wang, Y.; Qi, Y. A High-Quality Multicategory SAR Images Generation Method With Multiconstraint GAN for ATR. IEEE Geosci. Remote Sens. Lett. 2022, 19, 4011005. [CrossRef]
- Oh, J.; Kim, M. PeaceGAN: A GAN-Based Multi-Task Learning Method for SAR Target Image Generation with a Pose Estimator and an Auxiliary Classifier. *Remote Sens.* 2021, 13, 3939. [CrossRef]
- 51. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv* 2016, arXiv:1607.08022.
- 52. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 1997; pp. 770–778.
- 53. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* 2015, arXiv:1511.06434.
- Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 214–223.
- 55. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
- Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595. [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.