remote sensing

MDPI

Article

Style Transfer from Sentinel-1 to Sentinel-2 for Fluvial Scenes with Multi-Modal and Multi-Temporal Image Fusion

Patrice E. Carbonneau

Geography Department, Mountjoy Site, Durham University, Durham DH13LE, UK; patrice.carbonneau@durham.ac.uk

Highlights

What are the main findings?

- Fusion of an 8-year cloud-free composite mosaic of Sentinel-2 NIR data (band 8) to Sentinel-1 VV and VH SAR imagery enhances deep learning style transfer and delivers high-quality synthetic Sentinel-2 imagery that is 100% cloud-free.
- Our single model, trained on a global, multi-annual, and multi-seasonal dataset of 1.4 million samples, can synthesize cloud-free Sentinel-2 imagery for 99.2% of the globe.

What are the implications of the main findings?

- When used in an existing semantic classification pipeline designed for native Sentinel-2 imagery, our cloud-free synthetic Sentinel-2 imagery reaches an IoU score of 0.93 against an external benchmark, whereas native Sentinel-2 imagery achieves an IoU of 0.94.
- This method allows for semantic classification approaches focused on water to function in any cloud condition, which enhances our ability to remotely detect and monitor high discharge events at global scales with the native 10 m resolution of Sentinel-2 data.

Abstract

Recently, there has been significant progress in the area of semantic classification of water bodies at global scales with deep learning. For the key purposes of water inventory and change detection, advanced deep learning classifiers such as UNets and Vision Transformers have been shown to be both accurate and flexible when applied to large-scale, or even global, satellite image datasets from optical (e.g., Sentinel-2) and radar sensors (e.g., Sentinel-1). Most of this work is conducted with optical sensors, which usually have better image quality, but their obvious limitation is cloud cover, which is why radar imagery is an important complementary dataset. However, radar imagery is generally more sensitive to soil moisture than optical data. Furthermore, topography and wind-ripple effects can alter the reflected intensity of radar waves, which can induce errors in water classification models that fundamentally rely on the fact that water is darker than the surrounding landscape. In this paper, we develop a solution to the use of Sentinel-1 radar images for the semantic classification of water bodies that uses style transfer with multi-modal and multi-temporal image fusion. Instead of developing new semantic classification models that work directly on Sentinel-1 images, we develop a global style transfer model that produces synthetic Sentinel-2 images from Sentinel-1 input. The resulting synthetic Sentinel-2 imagery can then be classified with existing models. This has the advantage of obviating the need for large volumes of manually labeled Sentinel-1 water masks. Next, we show that fusing an 8-year cloud-free composite of the near-infrared band 8 of Sentinel-2 to the input Sentinel-1 image improves the

check for updates

Academic Editors: Xiaoyan Lu, Yuting Wan, Zhuo Zheng, Ce Zhang and Ji Zhao

Received: 29 August 2025
Revised: 10 October 2025
Accepted: 11 October 2025
Published: 15 October 2025

Citation: Carbonneau, P.E. Style Transfer from Sentinel-1 to Sentinel-2 for Fluvial Scenes with Multi-Modal and Multi-Temporal Image Fusion. *Remote Sens.* **2025**, *17*, 3445. https://doi.org/10.3390/rs17203445

Copyright: © 2025 by the author. 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/).

classification performance. Style transfer models were trained and validated with global scale data covering the years 2017 to 2024, and include every month of the year. When tested against a global independent benchmark, S1S2-Water, the semantic classifications produced from our synthetic imagery show a marked improvement with the use of image fusion. When we use only Sentinel-1 data, we find an overall IoU (Intersection over Union) score of 0.70, but when we add image fusion, the overall IoU score rises to 0.93.

Keywords: fluvial remote sensing; deep learning style transfer; big data; Sentinel-1; Sentinel-2; cloud removal

1. Introduction

Recently, global monitoring of freshwater resources has begun to leverage technical progress in the fields of big data and deep learning [1–9]. We are also seeing the emergence of geomorphology-oriented approaches that are specifically capable of mapping rivers and fluvial sediments in the landscape as distinct semantic entities [2,3,8,10]. However, one of the main challenges associated with freshwater monitoring from Earth Observation is the presence of clouds. Most studies tend to choose optical data such as Landsat-8 [11], Sentinel-2 [3], or commercial multispectral sensors [8]. These sensors have the well-known limitation of cloud obstructions. For example, the data of Carbonneau and Bizzi [3,12] has large gaps in the Amazon and Ganges catchments because cloud-free periods in these catchments are very rare. In the specific case of rivers, this problem is even more significant since floods and the peak discharge events, which perform geomorphic work, are driven by precipitation events and therefore tend to occur under cloud cover. Waiting for the availability of cloud-free images will inevitably result in imagery where river flows are at a lower discharge. This will bias any observations made from remotely sensed data. The classic solution to this problem of cloud occlusion is to use Synthetic Aperture Radar (SAR) data [13–20]. These works generally report excellent performance when delineating water bodies from SAR data in local case studies. However, reports for global applications are less frequent. Arguably, the most robust and direct comparison of SAR and optical data classification performance specific to surface water at global scales comes from Weiland et al. [9]. These authors identified 64 sites and dates where synchronous Sentinel-1 and -2 images are available. They train deep learning models for water classification for both sensors and compare them to high-quality manual ground truth datasets. Their global study found that Sentinel-1 data classified water bodies with an IoU of 0.844, while the model based on Sentinel-2 data reached an IoU of 0.940. Performance drops from Sentinel-2- to Sentinel-1-derived classifications are not unexpected. Twele et al. [18] report that wind ripple on large water bodies can increase backscatter. High levels of soil moisture on bare ground can lead to SAR absorption, which darkens the SAR imagery to the point of creating confusion with water bodies [21]. A possible solution to these issues is to use modern deep learning style transfer and image translation approaches in order to transform Sentinel-1 image data into synthetic Sentinel-2 image data [22,23]. These methods that leverage deep learning approaches and potentially large training datasets have a notable advantage in that they do not require labor-intensive manual labeling of semantic classes. All they need is near-synchronous imagery from two image sources, at which point it becomes possible to train deep convolution networks of various types to synthesize one type of imagery from another type of imagery [24,25]. This, therefore, suggests a workflow where large data gaps in Sentinel-2 data acquisition could be filled by using Sentinel-1 data to create synthetic Sentinel-2 images that simulate an acquisition in the absence of clouds.

1.1. Aim

The core scientific aim of this paper is to improve the performance of global assessments of rivers, their health, and their morphology by developing a data product that is insensitive to cloud and thus capable of semantic classifications of rivers in all weather and discharge conditions. Rather than directly using Sentinel-1 imagery, we develop an image fusion and style transfer approach that will produce synthetic optical imagery from both Sentinel-1 and long-term cloud-free Sentinel-2 inputs for use in an existing deep learning classification model [2]. This has two advantages: (1) the resulting pipeline will benefit from the synergies and complementarity of Sentinel-1 and Sentinel-2 data, as noted by Meraner et al. [23], and (2) the training of a style transfer model requires only synchronous image pairs without the need for labor-intensive manual labeling.

1.2. Specific Objectives

Develop a deep learning pipeline that will:

1. Perform a style transfer (*i.e.*, translation) from Sentinel-1 SAR imagery to Sentinel-2 optical imagery;
2. Function in conditions with as much as 100% cloud;
3. Data-sparse and computationally efficient, thus capable of global processing;
4. Optimized to deliver synthetic optical imagery suited to the specific task of semantic classification of fluvial features with an existing deep learning model [2].

2. Methods

2.1. Key Innovations

Meraner et al. [23] demonstrated a workflow capable of removing clouds in Sentinel-2 data based on a deep learning style transfer from Sentinel-1 data. In order to adapt this concept to our first three objectives, we propose the following innovations: (1) Use a data fusion approach that is both multi-temporal and multi-modal. Instead of using only near-synchronous Sentinel-2 and Sentinel-1 data, we propose to add cloud-free mosaics of Sentinel-2 imagery based on 8 years of data to complement the Sentinel-1 data. This represents the near-totality of Sentinel-2 data made available on Google Earth Engine; (2) to make the process data-sparse and operational in conditions where large areas are obstructed by 100% cloud, we use only band 8 (NIR) from the cloud-free mosaic stored in an unsigned 8-bit integer range for a minimal data footprint.

2.2. Full Archive Cloud-Free Mosaics

The intended function of these cloud-free NIR mosaics is to provide information on the large-scale features of the scene. NIR is less sensitive to high soil moisture values when compared to SAR. It is therefore expected that the NIR cloud-free mosaic will help disambiguate open water from very moist surfaces, which may both appear as very dark in SAR data. Furthermore, NIR is not affected by water surface wind ripple and thus is expected to have a constant dark appearance for all water bodies, in contrast to SAR data, where wind ripple on large water bodies can cause increased backscatter, leading to reflectance values similar to vegetation or bare ground [18].

Cloud-free NIR mosaics are produced in Google Earth Engine (GEE). Simonetti et al. [26] developed a GEE algorithm capable of producing quarterly cloud-free mosaics for Sentinel-2 imagery. They use advanced cloud masking algorithms to effectively derive cloud-free imagery suitable for long-term change detection. However, in our case, the cloud-free mosaics are not needed for change detection. Therefore, they can be constructed from the full archive of Sentinel-2 data, which obviously increases the number of available images. Our assumption is that over the full span of Sentinel-2 data acquisition, the full

globe has been covered. We use GEE via the Python API in the Google Colaboratory cloud computing platform. Our GEE download algorithm selects a bounding box AOI and retrieves all Sentinel-2 images from the 1st April 2017 to the 30th March 2025 (~8 years of data). In the case of areas north of 50° latitude, we only use the months of June, July, and August to remove snow cover. We select band 8 (NIR) and reduce the available imagery to a single raster with a median operator. We then normalize the values by a factor of 0.051 [26]. The final image is cast to an 8-bit unsigned integer data type and then projected to a UTM projected coordinate system. We find that this straightforward process achieves 99.2% coverage of the study area shown in Figure 1.

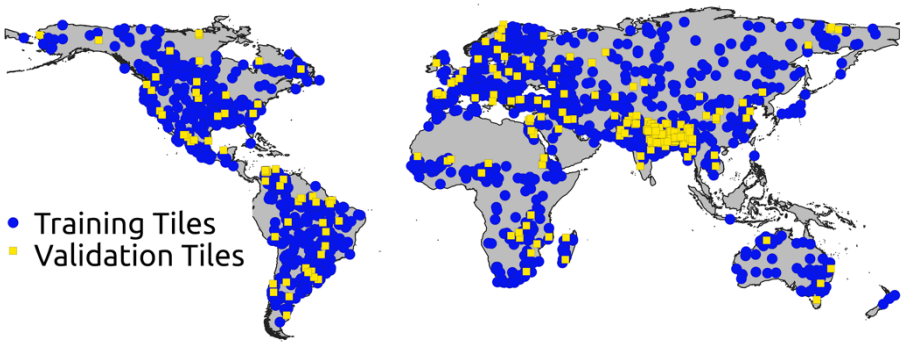


Figure 1. Global map with training and validation sites.

2.3. Study Area, Data, and Sampling Sites

Figure 1 shows the study area for this work. Given our focus on fluvial systems, this spans the globe but excludes the poles, Greenland, Iceland, and both the Canadian and Russian Arctic islands. We used existing global scale classifications of rivers from [3,12] in order to establish an initial database of over 2 million non-overlapping river reaches and 1 million lakes. From this database, we randomly extracted 30,000 river samples and 10,000 lake sample locations. Then, we generated an additional 10,000 sample locations from the Amazon and Ganges catchment in order to ensure that our model was well trained on these 2 major rivers that are often difficult to image with optical data. For these 50,000 sample locations, we wrote a Google Earth Engine download script in Python to download Sentinel-2 and Sentinel-1 data. For Sentinel-2, we only download bands 8, 4, and 3. For Sentinel-1, we download VV and VH polarization. We also downloaded the NIR cloud-free image composite described above. For each location, we sample an area 0.75 × 0.75 degrees (~77 km). The cloud threshold for Sentinel-2 data was set to 5%. As a final constraint, the image acquisitions for Sentinel-1 and Sentinel-2 were set to be within the same randomly generated 6-day period. In order to facilitate the process of data augmentation described below, the data is downloaded as a single 6-channel stack, projected to a UTM (Universal Transverse Mercator) coordinate reference system (CRS) using the WGS84 datum and with a spatial resolution of 10 m. We manage the multiple CRSs in the data by using the EPSG (European Petroleum Survey Group) numbers and use EPSG numbers ranging from 32,601 to 32,660 for UTM zones 1 to 60 in the Northern Hemisphere and 32,701 to 32,760 for UTM zones in the Southern Hemisphere. We found that whilst the Sentinel-2 data performed well when cast to an unsigned 8-bit data type, the Sentinel-1 data performed better if kept in a 16-bit integer data type. The image sample stacks are therefore downloaded in a 16-bit integer format. We run the above process for the initial attempts of 50,000 samples twice (*i.e.*, for two different randomly generated dates). Our objective is to create some duplicate sites where samples for 2 dates are available.

This process had a very high attrition rate. About 80k sample locations and date combinations failed to produce any data due to the unavailability of synchronous, cloud-free imagery from both Sentinel-1 and -2. The remaining images were downloaded and manually inspected. Even at this level, the attrition rate was high once again. We deleted samples with visible clouds, snow, and incomplete data. This left 2678 samples mapped in Figure 1. We kept 2411 samples as training data (blue in Figure 1) and reserved 267 samples for a validation set (yellow in Figure 1). We note that areas of Northern Canada and Russia impacted by winter have a lower density of points, with a lack of samples from winter months, which are excluded from the process. We also note that data acquisition for the Gangetic basin was more successful than anticipated. We then used the classification pipeline of Carbonneau [2] to estimate that, in total, the 2411 tiles of this training dataset has 1.70×10^9 river pixels (170,000 km^2), 1.12×10^{10} lake pixels (1.1×10^6 km^2), 0.51×10^9 gravel pixels (51,000 km^2), and 1.19×10^{12} 'background' pixels (1.19×10^6 km^2) that are not in any of these 3 classes.

We use 267 sites reserved from the main samples to construct our validation set. This is composed of 198 tiles of 0.2×0.2 degrees (~22 km), which are completely distinct from the training set in both space and time. The remaining 69 tiles have a spatial overlap with the training set, but are acquired at different times. This is intended as a test of the seasonal robustness of the method. In total, this validation dataset has 3.14×10^7 river pixels (3144 km^2), 5.99×10^7 lake pixels (5998 km^2), 1.01×10^7 gravel pixels (1010 km^2), and 1.28×10^9 'background' pixels (12,800 km^2). This dataset is smaller than the typical 20% of training that is customary in machine learning, but it has the benefit of spatial and temporal segregation. Overall, the final training and validation sets show a good distribution across time, with samples well distributed across the year and with most data being samples from 2019 to 2024, after the launch of Sentinel-2B in March 2017 (Figure 2). Furthermore, we will supplement this dataset with a large external benchmark dataset.

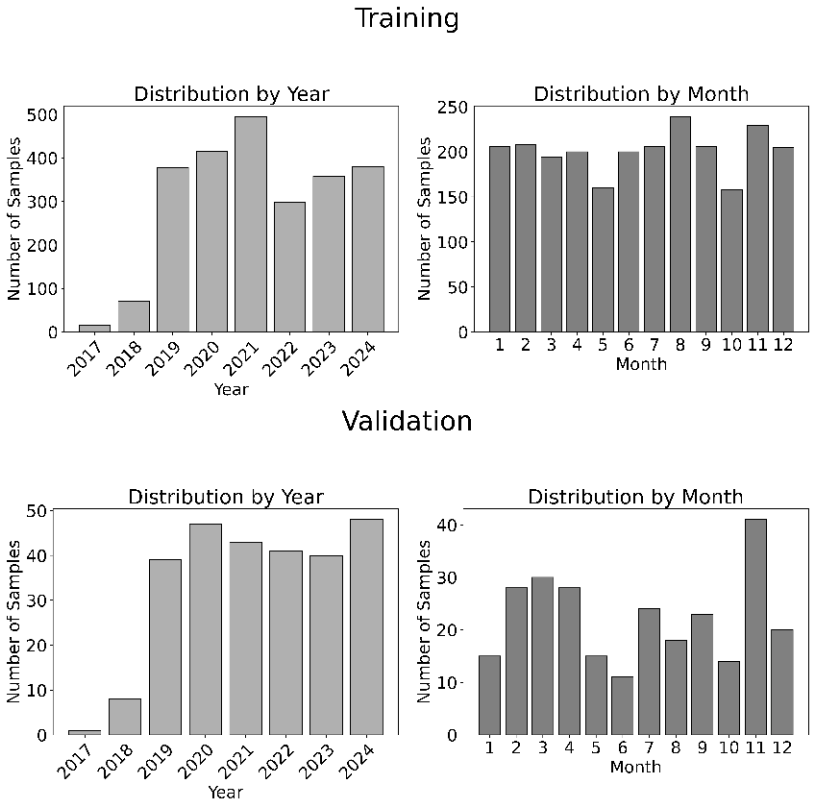


Figure 2. Temporal coverage and distribution with training and validation samples.

2.4. External Testing Data

Wieland et al. [9] present a dataset, S1S2-Water, that was specifically constructed for use with deep learning pipelines designed to map water bodies at global scales. It consists of Sentinel-1 and cloud-free Sentinel-2 imagery for 65 large tiles of 100 × 100 km, well distributed across the world. Images are accompanied by high-quality masks that identify water pixels. The Sentinel-1 and Sentinel-2 data are acquired with an average difference of 1 day with a standard deviation of 4 days. The dataset is intended for the training and validation of deep learning models. But here, we used 63 tile masks for Sentinel-2 data from this dataset purely as a testing dataset (Figure 3). We found that the brightness normalization factors used by Wieland et al. [9] did not give optimal performance with our models. Therefore, we used the dates embedded in the metadata to re-acquire the Sentinel-1 imagery from GEE. We also acquired the Sentinel-2 band 8 cloud-free mosaics for these sites. These will be processed with our models, with results validated against the masks from Wieland et al. [9].

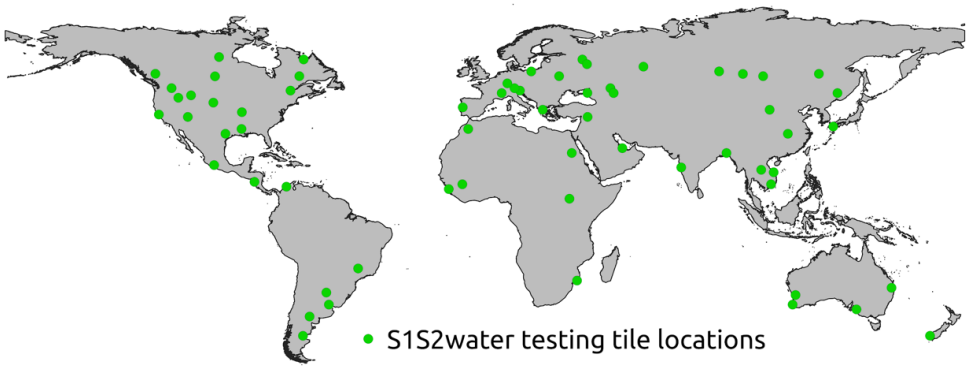


Figure 3. World locations of the S1S2-Water tiles.

2.5. Model Architecture

Style transfers and/or translations from between sensor formats are typically carried out using Generative Adversarial Networks (GANs) [27–30] or Fully Convolutional Networks (e.g., UNets) [31,32]. GANs require the training of multiple networks and are significantly more processing-intensive. For example, Li et al. [33] report that GANs can be 1 or 2 orders more computationally demanding. Given our objective of working at global scales (objective 3), we therefore opted for a UNet type of architecture. The core task of this architecture is to take a 3-channel input composed of Sentinel-1 VV and VH polarization with the cloud-free Sentinel-2 band 8 mosaic and translate this into a 3-channel output matching bands 8, 4, and 3 of a Sentinel-2 target image that is synchronous to the Sentinel-1 VV and VH data. Crucially, this architecture will need to learn small-scale details from the VV and VH channels that are synchronous to the target Sentinel-2 image and actively prevent the model from learning to minimize its training loss by transferring small-scale details from the cloud-free Sentinel-2 band 8 input channel to the Sentinel-2 band 8 output channel. Another desired feature of this model is an upsampling of outputs. Given our objective of a data-sparse model, we decided that the Sentinel-1 inputs would enter the model at a resolution of 20 m and deliver outputs at 10 m of resolution. The first justification for this is data volume. Based on our work in Carbonneau and Bizzi [3] and the need to store the VV and VH data in a 16-bit integer format, coverage of the full study area shown in Figure 1 at a resolution of 10 m would require ~6 Tb of storage. By cutting the input resolution to 20 m, we estimate that global coverage will be achieved in ~1.5 Tb of storage. Furthermore, whilst Sentinel-1 data is available to download at a resolution of 10 m in GEE, the native pixel dimensions of Sentinel-1 SAR data are in fact 5 × 20 m. We argue that downsampling to 20 m removes any complication associated with the direction

of the pixels and thus mitigates the possibility of directionally dependent errors (*i.e.*, rivers running parallel to the native Sentinel-1 pixel direction would be easier to detect).

Figure 4 shows our final attention UNet architecture, similar to the one used for semantic classification in Carbonneau [2]. The architecture takes a 3-channel input with 224 × 224 pixels at the start of the encoder phase. The input pixel dimensions of 224 × 224 are chosen for consistency with the models of Carbonneau [2]. The architecture first uses a lambda layer to extract the cloud-free band 8 channel from the input stack. This is then processed in a blurring module that performs a 5 × 5 Gaussian blur and a downsampling of the B8 data from 224 to 112 pixels. This severely blurred output is then concatenated in the encoder pathway. In practice, and with a 20 m resolution input, this means that any feature smaller than 200 m will, in effect, be blurred out of the cloud-free band 8 input channel. This ensures that the model cannot use small-scale details from this input channel in the outputs. After 3 standard encoder blocks composed of a 2D convolution, a 2 × 2 maxpool layer, and a batch normalization, the encoder pathway reaches a bottleneck convolution of 28 × 28 × 256. In the decoder pathway, we use attention gates to focus the model on small details [34]. We again concatenate the blurred B8 data to the second decoder block to enforce the large-scale structures therein. After 3 decoding blocks linked to the matching encoder blocks with skip connections, we add a 2D transpose convolution that can perform our upsampling with learned parameters. The final convolution therefore outputs 448 × 448 images in simulated channels of NIR, red, and green corresponding to bands 8, 4, and 3 from Sentinel-2. In total, this architecture has 18.1 million trainable parameters.

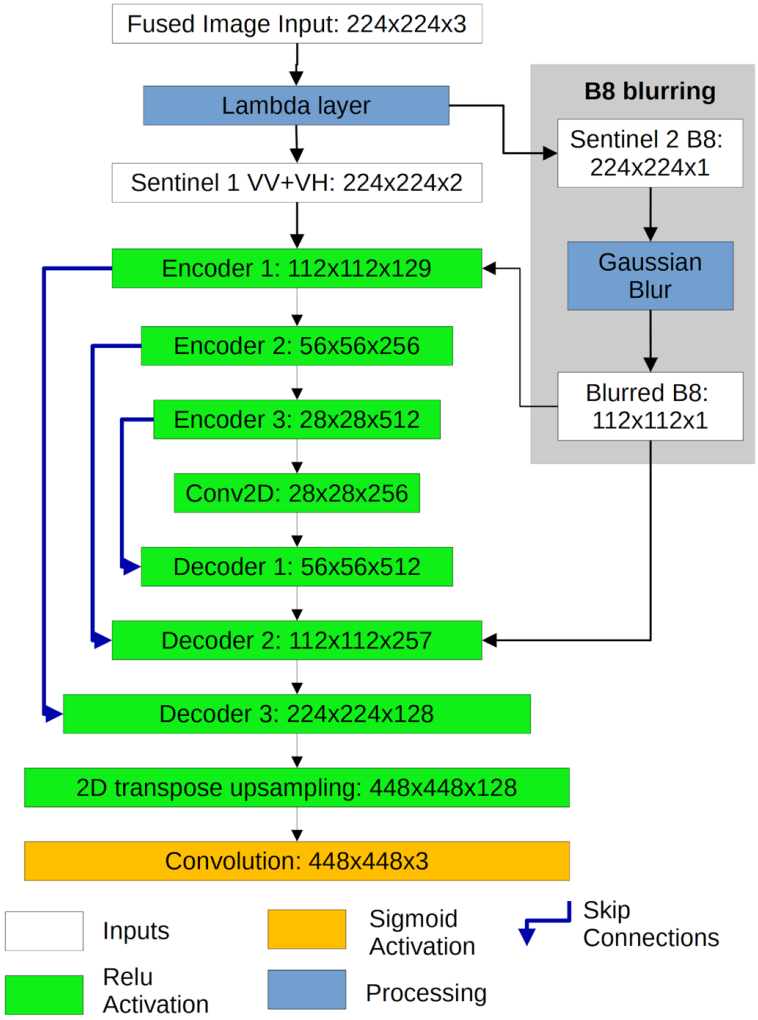


Figure 4. Style transfer UNet model architecture.

2.6. Data Augmentation

The final training dataset is created with the use of data augmentation. We use the albumentations toolbox [35] to extract a total of 1.4 million samples from our training tiles. Training samples are composed of an input image with Sentinel-1 VV, VH, and cloud-free Sentinel-2 band 8 mosaics composited as described above. The input image is downsampled to cover a planimetric extent of 224 × 224 pixels at a resolution of 20 m. The input is matched to a target image from Sentinel-2 with bands 8, 4, and 3 at the native resolution of 10 m and therefore with 448 × 448 pixels in XY. The target image is synchronous with the VV and VH bands of the Sentinel-1 input. The augmentation uses a series of random flips, rotations, and transposes to vary the data. We also add random scaling changes (preserving the aspect ratio) up to a factor of 2. In order to mitigate the natural class imbalance of water features where lakes, rivers, and gravel bars occupy roughly 10%, 1% and 0.1% of the Earth's surface, respectively [3], we use the semantic class rasters described above to check the content of the random samples before saving them to disk. The random sample generation was designed to extract a target of 1×10^7 river samples, 250,000 lake samples, and 250,000 gravel bar samples. After reaching 1.4 million samples, we deemed the process satisfactory. Figure 5 shows sample training data for small river channels, and Figure 6 shows samples with larger water bodies and gravel bars. Readers should note the differences visible in the VV and VH bands when compared to the band 8 mosaics. In the middle row of Figure 5, we can clearly see differences in the channel configuration, with the VV and VH channels being a closer match to the target. The same effect can be seen in the middle row of Figure 6, where the configuration of the large braiding channel is clearly different. Also of note is Figure 6 in the bottom row, where patches of water clearly show higher reflectance in the VV channel but not in the B8 channel.

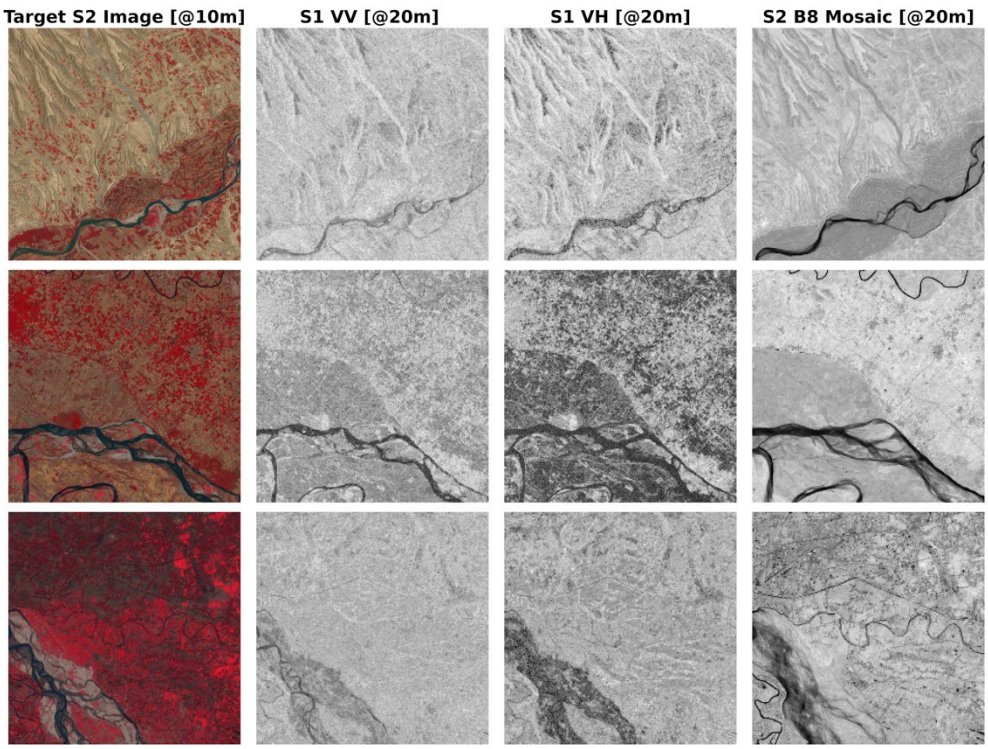


Figure 5. Small water body training samples with the target output image in Band 8, 4, and 3, false color, Sentinel-1 VV input, Sentinel-1 VH input, and Cloud-Free Sentinel-2 band 8 inputs. Readers should note that the cloud-free input channel is generally not identical to the VV and VH inputs; this is most notable in the bottom 2 samples.

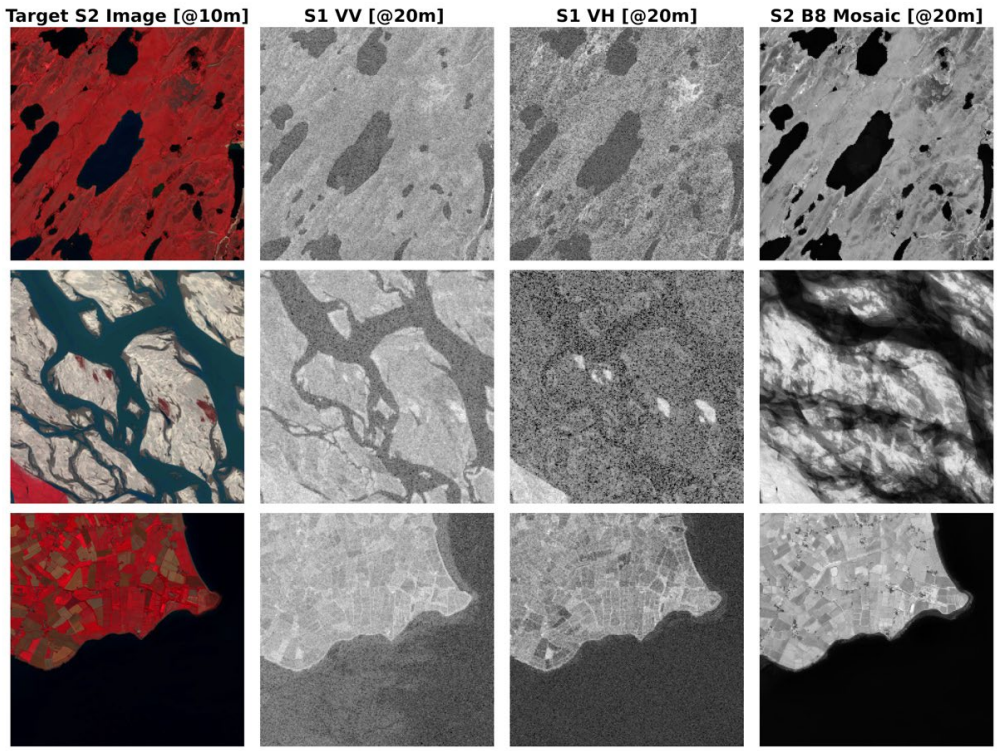


Figure 6. Large water body training samples with the target output image in Band 8, 4, and 3, false color, Sentinel-1 VV input, Sentinel-1 VH input, and Cloud-Free Sentinel-2 band 8 inputs.

2.7. Loss Functions and Training

Style transfer models are very sensitive to the choice of the loss function. In our case, our objective is to produce synthetic Sentinel-2 imagery for the purpose of semantic classification of fluvial landscapes. Specifically, we aim to use the model of Carbonneau [2]. Early experimentation at the start of this work clearly showed that using a standard loss of the mean square error of pixel brightness values gave poor results. We therefore designed a custom loss function with 2 components. First, we use a mean square loss calculation that is inverse weighted with respect to brightness. This gives more weight to dark pixels in the imagery, which are usually water. Second, we use a perceptual loss component. Perceptual loss was developed by [36] in order to use an existing, trained model and evaluate the loss of a training iteration by running the model outputs and the target through this external model and calculating a loss from these external model outputs. The term 'perceptual' comes from the idea that we are checking how the synthetic data will be 'perceived' by an existing model. To this end, we use the model from Carbonneau [2]. We isolate the encoder part of our high-resolution UNet classifier. This model works with a 224 × 224 input and outputs 224 × 224 pixels. Since our style transfer architecture (Figure 4) outputs 448 × 448 pixels, we split outputs and targets into 4 images of 224 × 224, use the encoder from Carbonneau [2] to calculate 4 sets of features, and then use the mean square error of these features to obtain a loss value. After some checks that the magnitudes of losses were comparable, this is added to our inverse weighted mean square value to obtain our final loss value. Ultimately, in accordance with objective 4, the use of this custom loss function produces synthetic images that are predisposed to perform well with our existing classification model since this model has been used in the loss function.

In our experience [2,3], such models trained with millions of samples converge rapidly, and in practice, we have found that they converge in less than 8 iterations. Rather than using our validation data to control the training with learning rate scheduling, early stopping, etc., we opted to systematically train models for 8 epochs and save the weights for each

iteration, irrespective of validation score performance. On an NVIDIA RTX A5000 GPU, this required 10 h per epoch. Each set of weights is then checked against both our own validation and the external training data.

2.8. Experimental Structure

The final objective of this work (see objective 4 above) is to produce synthetic Sentinel-2 imagery that performs well with the semantic classification workflow of Carbonneau [2]. Consequently, we evaluated the success of our image style transfer model with classification metrics and by comparing the classification performance of a synthetic image with that of its native Sentinel-2 equivalent. In essence, we define a 'perfect' synthetic image as one where the semantic class raster output from the pipeline of Carbonneau [2] is identical to the one output from a native Sentinel-2 input. Therefore, we consider the reproduction of exact pixel brightness values and accurate spectral indices (e.g., NDVI) to be beyond the remit of this work. Furthermore, we wished to explicitly test the impact and effectiveness of our data fusion using cloud-free band 8 mosaics. We therefore tested the effect of this band 8 input channel by slightly modifying our architecture (Figure 4). We produced an alternative model where the band 8 channel is stripped from the input via the same lambda layer, but where this data is not blurred and concatenated with other layers, and remains unused. The effect is therefore an alternative model that only uses the VV and VH channels from Sentinel-1 to produce the synthetic image data.

The first part of our experiment is the comparison with our own validation data, which, crucially, has not been used to control model training and remains distinct from the training data. In this case, we use Cohen's kappa to check the similarity of outputs for each class separately (rivers, lakes, and gravel bars). In addition, we merge our river and lake classes to create a single 'water' class. For our approach to be deemed successful, it is essential that the model learns to translate small-scale features based on the SAR channels and not the long-term near-infrared data. We therefore repeat the error analysis on the 22-sample subset of our validation data from Figure 1, which is composed of large meandering and braiding rivers. These rivers would have changed over the 8-year acquisition period of the cloud-free data, and it is therefore expected that the appearance of the cloud-free composite band 8 data will significantly differ from the SAR VV and VH data. In addition to examining classification performance for these 22 sites with and without data fusion, we constructed cloud-free Sentinel-2 images in bands 8, 4, and 3 in the same manner as above. These images are compatible with the classification pipeline of Carbonneau [2]. If our approach successfully learned features from the contemporary Sentinel-1 VV and VH bands, the semantic class rasters from the cloud-free imagery will be different from those generated by our style transfer approach. Finally, we use our external validation data drawn from the S1S2-Water dataset of Wieland et al. [9]. Here, we consider the Sentinel-2 masks of Wieland et al. [9] as ground truth data. We merge our river and lake classes to achieve comparable data and calculate precision, recall, F1, and IoU scores. Error evaluations are repeated for models with and without data fusion of cloud-free Sentinel-2 band 8 mosaics.

3. Results

Figure 7 shows some examples comparing an original Sentinel-2 image, a synthetic image generated with the fusion of the full archive band 8 data, and a synthetic image generated without this long-term data. Generally speaking, all the images appear relatively similar. However, close examination reveals key differences. On the first row (top), we can see the presence of a large falsely generated body of water that is not present in the original Sentinel-2 image. This false water body does not appear in the synthetic image generated with the cloud-free band 8 data fusion. A similar effect can be seen in the second

row of images, where once again the fusion with cloud-free band 8 data has prevented the generation of a non-existent water body in the image. In the bottom two rows of Figure 7, we see a reverse effect where existing water bodies have significant areas that are of lighter colors that almost approach that of senescent vegetation or bare soil. In the third row of Figure 7, we see a complex reservoir lake. Close examination of the synthetic image generated without cloud-free band 8 data shows that the lake has shrunk, with some areas of the border appearing to be vegetated. In the bottom row of Figure 7, we can see a large lake, and here it can be seen that a large portion of the interior of this lake is lighter in color and resembles bare soil or sediment. Once again, the fusion of full archive cloud-free data from band 8 of Sentinel-2 data prevented these artifacts from being generated.

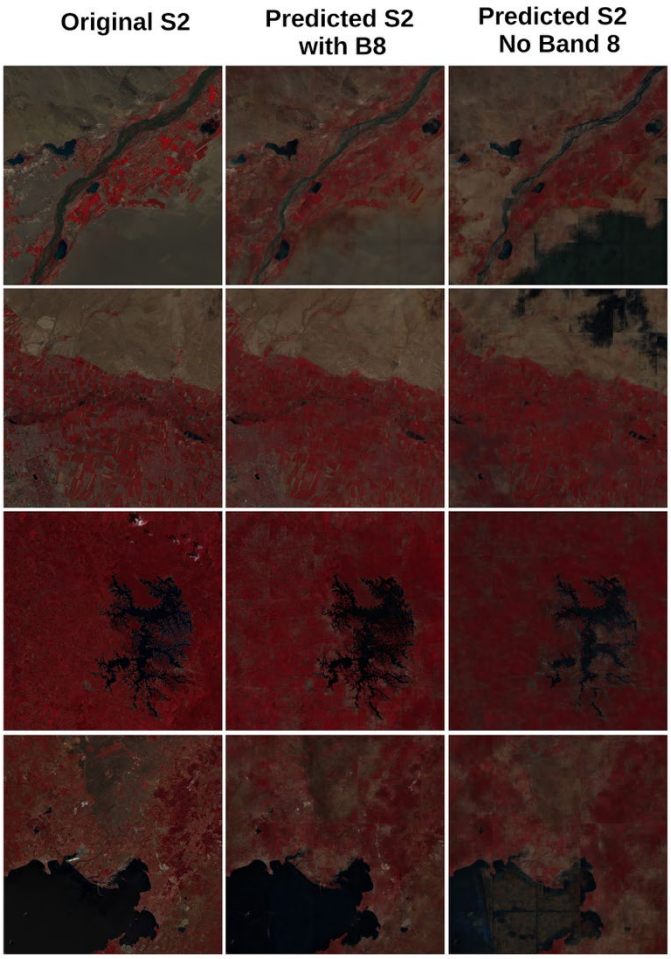


Figure 7. Sample outputs. The left-hand column shows native Sentinel-2 tiles in false color with channels 8, 4, and 3. The middle column shows the outputs of our final model with the inclusion of cloud-free Sentinel-2 mosaics. For comparison, the right-hand column shows the output of the alternative model that does not use cloud-free mosaics in the input and only predicts the synthetic imagery from Sentinel-1 data.

Figure 8 shows the same imagery as in Figure 7, but with the overlay of the semantic class rasters generated by the pipeline of Carbonneau [2]. Here we can see that the best results appear to be from the middle column, where data fusion was used. We can clearly see, in the right column, large water bodies that were falsely generated in the absence of data fusion, which resulted in false lake identification. Interestingly, we also note that in the case of the first row, the river is successfully detected by the classification pipeline even if the water color appears to be slightly brighter in Figure 7.

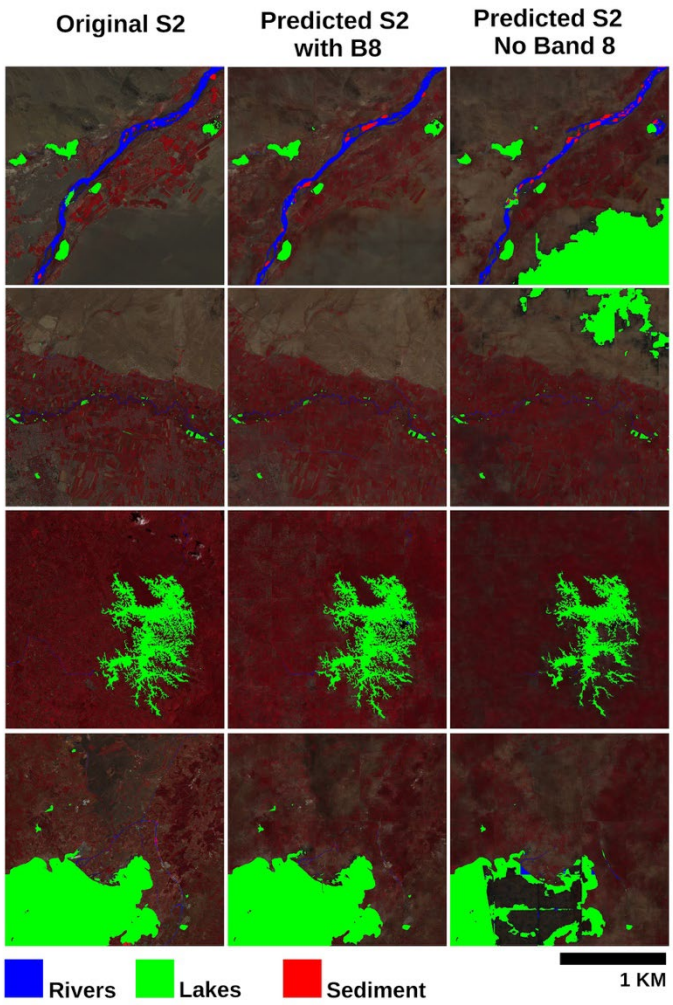


Figure 8. Sample outputs with semantic class overlays. Here we see the same examples as in Figure 7, but with the addition of semantic class overlays obtained when running the images from Figure 7 through the classification pipeline of Carbonneau [2].

Table 1 shows the full results of a validation data comparison (see Figure 1). Given that we are comparing two model outputs, neither of which is a ground truth measure, we use Cohen's kappa as a measure of similarity. In addition to comparing the results for our three modeled semantic classes of rivers, lakes, and sediment bars, we also merge the river and lake classes to create a water class. Here we see very strong agreement for the rivers, lakes, and overall water with kappa scores of 0.87, 0.89, and 0.91, respectively. The similarity of the bars is good, with a kappa score of 0.75. Visual observation of the data in Figure 8 and across our broader validation set confirms this. We also see in Table 1 that the omission of band 8 data leads to systematic performance degradation in all classes.

Table 1. Kappa scores comparing semantic classification outputs of native Sentinel-2 images to synthetically generated imagery, with and without image fusion.

	Water	Rivers	Lakes	Gravel Bars
With cloud-free B8	0.91	0.87	0.89	0.75
Without cloud-free B8	0.85	0.84	0.81	0.71

We use confusion matrices, Figure 9, in order to examine model performance more closely. On the left of Figure 9, we see a classic confusion matrix that compares semantic classifications derived from the optimal synthetic generation process with image fusion,

notated as 'Predicted Labels', to the results obtained from native Sentinel-2 imagery, here assumed to be 'True Labels'. The numbers in the categories represent total pixel counts, with the coloring giving a logarithmic color ramp for these total pixel counts. The main off-diagonal elements can be seen as follows: 1) 6.85 million (11.4% of total lake pixels) true lake pixels being predicted as background pixels; 2) 2.75 million true background pixels predicted as lake pixels (4.5% of total lake pixels), 3) 2.55 million true gravel bar pixels (25% of total gravel bars) predicted as background pixels and 4) 2.96 million true river pixels (4.5% of total river pixels) being predicted as background pixels.

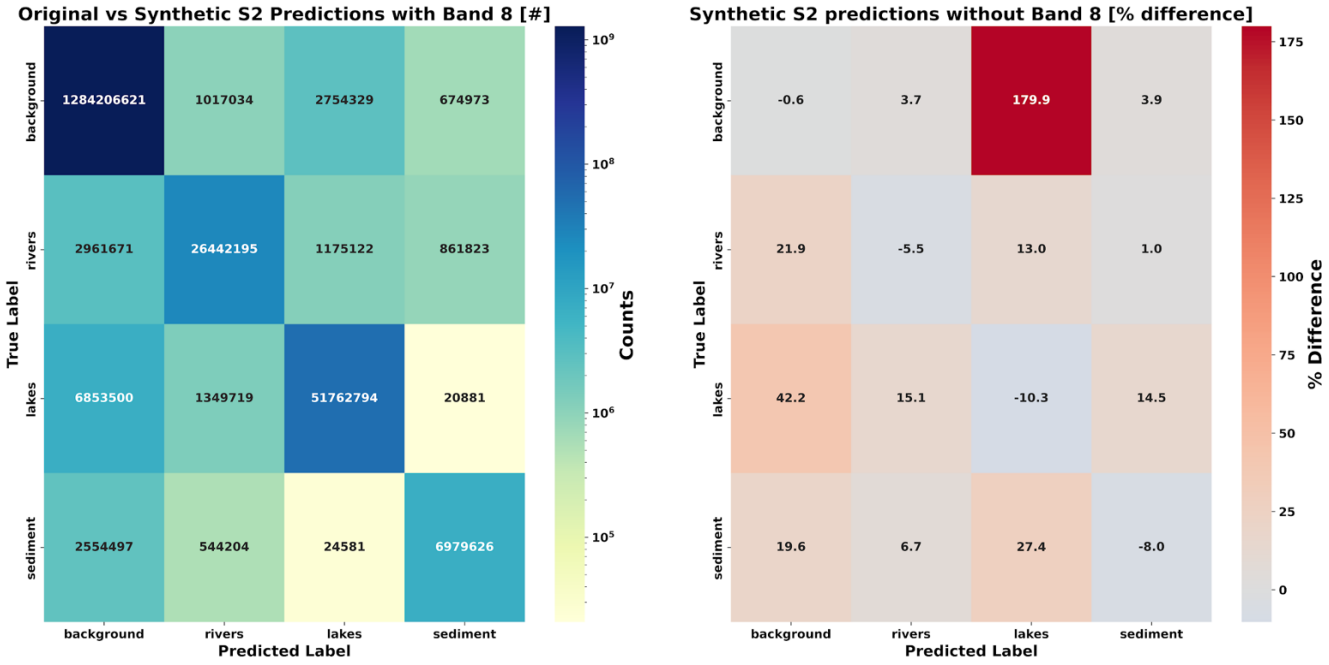


Figure 9. Confusion matrices. Left: comparison of class predictions in results obtained from native Sentinel-2 imagery (True Labels) vs. those obtained from synthetic Sentinel-2 imagery (Predicted Labels).

On the right of Figure 9, we see the impact of the removal of cloud-free data fusion from the process. Here, the impact is directly translated into a % difference with respect to the left part of Figure 9. We immediately see a 179% increase in background pixels predicted as lake pixels. This can be directly attributed to the creation of false water bodies in the synthetic images in the absence of data fusion (top two rows in Figure 7). We also see a 42% increase in lake pixels being predicted as background. This can be directly attributed to the creation of false patches of bare ground in the absence of data fusion (bottom 2 rows in Figure 7). We note a smaller increase of 21% for river patches being predicted as background. This can be attributed to wider channels where wind-ripple effects can initiate. Finally, we also see an increase of 19% of sediment bars predicted as background. This is more difficult to attribute, and the issue of sediment classification will be discussed below.

Table 2 presents the validation results on the large meandering and braided river subset of our validation data. This data has a total of 8,228,656, 106,666, and 7,279,917 pixels for rivers, lakes, and bars, respectively. We can see that the performance metrics for rivers and bars are similar to those in Table 1, indicating that the model performs equally well when faced with changing river environments. We see lower performance for the lakes class and a more pronounced degradation in the absence of B8 data. However, in this dataset composed of river scenes, we note the relatively small pixel count of the lake class and the absence of large lakes in this validation subset. Performance of the water class is slightly

reduced when compared to Table 1, which is consistent with the lower performance of the lake class. In the last row, we can see that when compared to the classification output of full archive cloud-free Sentinel-2 imagery in bands 8, 4, and 3, kappa scores are low. This indicates that the classification outputs of our optimally generated synthetic imagery have not mimicked the band 8 data, which is common to both.

Table 2. Validation kappa scores for meandering and braided rivers. Also shown is the comparison with full 3-band cloud-free Sentinel-2 imagery.

	Water	Rivers	Lakes	Gravel Bars
With cloud-free B8	0.89	0.88	0.80	0.76
Without cloud-free B8	0.86	0.86	0.62	0.72
Sentinel-2 cloud-free image bands 8, 4, and 3	0.46	0.42	0.53	0.46

Table 3 shows the results of a comparison of our merged water class to the ground truth benchmark datasets of Wieland et al. [9], S1S2-Water. We used the Sentinel-2 water masks in order to test our synthetic Sentinel-2 images against their native counterparts. We see scores ranging from 0.93 to 0.97, showing that our best synthetic imagery, produced with data fusion, delivers very strong classification performance when processed with the pipeline of Carbonneau et al. [2]. We also see a marked degradation of performance if the synthetic imagery is produced in the absence of the full archive cloud-free band 8 data fused to the synchronous Sentinel-1 imagery.

Table 3. F1, precision, recall scores, IoU, and kappa scores of water predictions compared to S1S2-Water, with and without image fusion.

	Precision	Recall	F1	IoU	Kappa
With cloud-free B8	0.97	0.95	0.96	0.93	0.95
Without cloud-free B8	0.80	0.79	0.76	0.70	0.73

4. Discussion

We find that the inclusion of multi-temporal data fusion in our style transfer approach enhanced the classification performance of the resulting synthetic imagery. The inclusion of long-term, full archive data from band 8 of Sentinel-2 helped the image translation model avoid some key errors that can be traced to the greater sensitivity of SAR data to soil moisture and wind-ripple effects. These effects, typical of SAR remote sensing, break a fundamental assumption in the analysis of water bodies from remotely sensed data: water appears as a dark feature in the landscape. By incorporating a long-term image in the near-infrared part of the spectrum, our style transfer model has learned to mitigate these SAR errors and provide a better translation to optical Sentinel-2 data. This effect can be clearly seen in Figure 7, where large areas can be mistranslated and falsely interpreted as water or ground in the absence of NIR data from Sentinel-2. This results in stronger classification performances in all cases. This work therefore shows that the fusion of long-term Sentinel-2 data in the near-infrared band 8, a band which is critical to water remote sensing, can disambiguate SAR data and lead to better performance. We also find that our model architecture performs well in cases where small-scale features can change over the 8-year acquisition period of the cloud-free imagery, thus leading to cloud-free mosaic channels that do not closely match contemporary conditions, seen for example in the middle row of Figures 5 and 6. The examination of meandering and braiding river classification performance in Table 2 shows a sustained

performance for the river and gravel bar classes. Crucially, the third line of Table 3 shows that when we classify a 3-band (8, 4, 3) cloud-free image constructed from a full 8 years of data for a given study site, we obtain very different results when compared to our synthetic imagery using VV and VH Sentinel-1 bands. We argue that this shows that our synthetic images are not learning to mimic small-scale features from the fused band 8 data and that, as intended, they are learning to produce fine-scale detail based on the Sentinel-1 data and only using the Sentinel-2 band 8 data to eliminate large-scale errors associated with moisture and wind-ripple effects. Specifically, this means that the B8 blurring module of the architecture (Figure 4) is performing the intended task of removing small-scale features from the band 8 data in order to force the model training to learn fine-scale detail from the Sentinel-1 inputs.

Errors in the final class product (Table 1) are acceptable. For our main class of interest, rivers, we lose 4.5% of river pixels to the background class, 2.7% are confused with gravel bars, and 3.7% and 2.7% with lakes. The confusion with the background and gravel bars is likely due to a residual presence of wind ripple on medium channels of a scale smaller than our blurring of the band 8 data (see Figure 4), which increases the brightness of these water patches, leading to a classification of these patches as a ground class. This loss is greater for lakes, potentially due to some remaining effects of wind ripple and soil moisture, but this class is not our priority and is only included in our classification because Carbonneau and Bizzi [3] found in early stages of work that the inclusion of a specific semantic class for lakes improved performance for rivers. The largest errors in Table 1 relate to the detection of gravel bar pixels. We see in Figure 9 (left) that 25% of gravel bars predicted from native Sentinel-2 images are confused with the background class if we use synthetic imagery. Having higher errors for this class is not new. Carbonneau and Bizzi [3] faced similar issues. From an ontological perspective, the gravel class is arguably the most weakly defined. Rivers and lakes share similar spectral characteristics. Specifically, the use of an infrared band in the input data allows both human observers and models to distinguish water as patches that are dark in infrared, red, and green, whilst vegetation is bright in the infrared and darker in reds and greens. Furthermore, rivers and lakes tend to have more distinct shapes. Fluvial sediment patches do not benefit from such clear distinctions. Fallow fields and bare ground patches that are not connected to a water body can have similar spectral characteristics. This makes the identification of gravel patches with purely spectral criteria highly error-prone. In terms of shape, gravel bars can have a variety of shapes. Whilst some shapes, such as point bars, are effectively learned by the model, visible bars that result from changes in water level can have very diverse shapes that become hard to capture with convolutional or even ViT (Vision Transformer) approaches. Additionally, the contrast of the water-connected edge of a gravel bar will vary with both the slope and the turbidity of the water. If a gravel bar has a low slope, then the shallow submerged portion of the bar will be visible through the water. This lowers the contrast of the dry gravels compared to the wet gravels and makes delineation more difficult. This is compounded by the levels of turbidity, which can make the water more or less clear and change the Secci depth, thus making the shallow submerged portion of a bar more or less visible. These factors combine to make the gravel class the most difficult class to predict.

We find that our F1 result of 0.96, along with an IoU of 0.93, compares favorably to other reports of water classification performance using Sentinel-1 data. Fakhri and Gknatsios [37] report a best F1 score of 0.84 in a flood detection study in New South Wales, Australia. Zhang et al. [38] report an IoU of 0.83 in a flood detection study of Hainan Island, China. Ghosh et al. [39] develop global-scale flood detection models using Sentinel-1 data

with a range of deep architectures. When validated against a flood event in Florence, they report a best IoU of 0.75. Finally, Zhang et al. [40] use a range of models and SAR datasets to detect flood inundation extents. They report a best IoU of 0.812 and a best F12 score of 0.86. However, we should note that these studies are all focused on flood mapping and often include urban areas. Comparisons should be treated with caution. Nevertheless, our results compare favorably to these findings, which suggests that the fusion of band 8 data from Sentinel-2 to VV and VH data from Sentinel-1 makes a significant contribution to semantic classification performance and differentiates this work from others.

Overall, we view the levels of error in our results as acceptable in the context of the specific application of this workflow. The objective of this work was never to replace cloud-free Sentinel-2 imagery. We recognize that a cloud-free Sentinel-2 image will always have superior quality when compared to a Sentinel-1 equivalent and that, for the purpose of semantic classification, Sentinel-2 imagery should be prioritized over Sentinel-1. However, in cases where cloud cover is present, the use of Sentinel-1 data to synthesize Sentinel-2 data is a valuable complement. Particularly for our area of interest, fluvial geomorphology, the ability to examine river response and changes at times of peak discharge, presumably under cloud cover, is a novel and important step. The errors reported here can inform interpretation. When comparing semantic class inventories from Sentinel-1 and -2, variations in river water area in the range of $\pm$10% should not be interpreted as significant. In the case of gravel bars, the size of the error, 25%, is such that any interpretations of bar area changes should be treated with caution and further validated by manual measurement.

In terms of computational load, we find that our chosen UNet architecture delivered fast inference with a low computational footprint. We deployed the algorithm on a modest workstation with an older XEON ES-260 CPU running at 2.1 Ghz and equipped with an NVIDIA GEFORCE 1080Ti GPU with 11 Gb of RAM. Our first step is to acquire the full global dataset for cloud-free Sentinel-2 band 8 images. Given that the cloud-free composites are created from the full archive of Sentinel-2 data, the processing requirement is high, and the download requires approximately 3 weeks. This set of single-band images in 8-bit format and with a spatial resolution of 20 m requires 70 Gb of storage at high compression. Fortunately, this is a process that only needs to be carried out once. Then, we tested inference speed on data from the Po basin in Northern Italy. We found that our system can infer synthetic imagery at a rate of approximately 100 km^2 per second. This would therefore lead to a total processing time of 17 days for the 148 million km^2 land surface area of the globe on our relatively modest older-generation GPU. However, in studies specifically focused on river corridors, we found that using existing datasets [41,42] to establish river corridors can cut the area to process by as much as 80%. This means that, in theory, global inference of synthetic imagery can be achieved in as little as ~3 days. However, in practice, we have found that the bottleneck of this pipeline is the Sentinel-1 VV and VH download speeds from Google Earth Engine. These can be variable, and depending on the current traffic on Google servers, individual users may have two or three parallel workers to execute jobs. Global-scale downloads of the needed Sentinel-1 data require an estimated 7–10 days. This is satisfactory because it aligns well with our Sentinel-2 classification workflow, which requires ~10 days to process global-scale data [2].

Another key limitation that readers should note pertains to additional uses of the synthetic imagery generated with our method. Our choice of an inverse weighted RMSE loss and of a perceptual loss component calculated from an existing trained model designed for the semantic classification of rivers and lakes will, by definition, create synthetic imagery that is tailored to our model and, at best, water classification. An obvious weakness of

perceptual loss functions is that they produce models trained to mimic the performance of the model used to calculate the perceived loss. This means that the synthetic imagery presented here is less well suited to applications not focused on water features. For example, preliminary assessments of the similarity of NDVI values calculated from native Sentinel-2 imagery and matching synthetic imagery show errors as large as 0.2, which are quite significant within the {-1,1} set of NDVI values. Readers interested in applying the methods shown here to non-water-facing problems are encouraged to re-train the style transfer model with a suitable loss function.

5. Conclusions

We developed a novel model capable of style transfers from Sentinel-1 to Sentinel-2 imagery. In addition to the commonly used VV and VH polarization of Sentinel-1, we fused a cloud-free mosaic image of Sentinel-2 band 8 (NIR). This allows us to produce synthetic Sentinel-2 imagery even if cloud cover is 100%. Our method is data-sparse and computationally efficient. It uses input imagery with a spatial resolution of 20 m with a relatively small attention UNet model of 18.1 trainable parameters to produce outputs with a spatial resolution of 10 m. Instead of producing the full 13 channels of Sentinel-2 imagery, our model only synthesizes three channels in NIR, red, and green (bands 8, 4, and 3). Finally, the style transfer model is trained with a customized loss function that gives more weight to dark pixels (e.g., water) and also uses a perceptual loss component that evaluates error based on a pre-trained model from Carbonneau [2]. This results in a model that is trained specifically to generate images that mimic the original Sentinel-2 data when processed with the pipeline of Carbonneau [2]. This combination of workflows is therefore capable of surveying the world's rivers in any cloud conditions. By lifting the requirement for cloud-free imagery, global-scale river surveys become much more capable of capturing high-flow events, which cause disasters and perform geomorphic processes.

Funding: This project did not receive any funding.

Data Availability Statement: The training and validation data used here require 1.1 Tb of storage and are not easily hosted with a permanent DOI. But the data can be made available upon request. A demonstration notebook with the GEE download code, the style transfer inference code, and the final model weights is available on GitHub [43]. The validation set (267 sites) rasters for semantic classes predicted from native Sentinel-2 imagery and from synthetic Sentinel-2 imagery were made available to the journal.

Acknowledgments: We thank three anonymous reviewers for their time and contributions to the manuscript. ChatGPT v4.5 in agent mode was used to produce initial reference lists relating to style transfer from SAR to optical imagery and water mapping using Sentinel-1. All references suggested by ChatGPT were individually consulted and checked for accuracy.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Billson, J.; Islam, M.D.S.; Sun, X.; Cheng, I. Water Body Extraction from Sentinel-2 Imagery with Deep Convolutional Networks and Pixelwise Category Transplantation. *Remote Sens.* **2023**, *15*, 1253. [CrossRef]
2. Carbonneau, P.E. Global Semantic Classification of Fluvial Landscapes with Attention-Based Deep Learning. *Remote Sens.* **2024**, *16*, 4747. [CrossRef]
3. Carbonneau, P.E.; Bizzi, S. Global Mapping of River Sediment Bars. *Earth Surf. Process. Landf.* **2024**, *49*, 15–23. [CrossRef]
4. Grill, G.; Lehner, B.; Thieme, M.; Geenen, B.; Tickner, D.; Antonelli, F.; Babu, S.; Borrelli, P.; Cheng, L.; Crochetiere, H.; et al. Mapping the World's Free-Flowing Rivers. *Nature* **2019**, *569*, 215–221. [CrossRef]
5. Linke, S.; Lehner, B.; Ouellet Dallaire, C.; Ariwi, J.; Grill, G.; Anand, M.; Beames, P.; Burchard-Levine, V.; Maxwell, S.; Moidu, H.; et al. Global Hydro-Environmental Sub-Basin and River Reach Characteristics at High Spatial Resolution. *Sci. Data* **2019**, *6*, 283. [CrossRef]

6. Moortgat, J.; Li, Z.; Durand, M.; Howat, I.; Yadav, B.; Dai, C. Deep Learning Models for River Classification at Sub-Meter Resolutions from Multispectral and Panchromatic Commercial Satellite Imagery. *Remote Sens. Environ.* **2022**, *282*, 113279. [CrossRef]
7. Nyberg, B.; Henstra, G.; Gawthorpe, R.L.; Ravnås, R.; Ahokas, J. Global Scale Analysis on the Extent of River Channel Belts. *Nat. Commun.* **2023**, *14*, 2163. [CrossRef]
8. Valman, S.J.; Boyd, D.S.; Carbonneau, P.E.; Johnson, M.F.; Dugdale, S.J. An AI Approach to Operationalise Global Daily PlanetScope Satellite Imagery for River Water Masking. *Remote Sens. Environ.* **2024**, *301*, 113932. [CrossRef]
9. Wieland, M.; Fichtner, F.; Martinis, S.; Groth, S.; Krullikowski, C.; Plank, S.; Motagh, M. S1S2-Water: A Global Dataset for Semantic Segmentation of Water Bodies from Sentinel- 1 and Sentinel-2 Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 1084–1099. [CrossRef]
10. Carbonneau, P.E.; Dugdale, S.J.; Breckon, T.P.; Dietrich, J.T.; Fonstad, M.A.; Miyamoto, H.; Woodget, A.S. Adopting Deep Learning Methods for Airborne RGB Fluvial Scene Classification. *Remote Sens. Environ.* **2020**, *251*, 112107. [CrossRef]
11. Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-Resolution Mapping of Global Surface Water and Its Long-Term Changes. *Nature* **2016**, *540*, 418–422. [CrossRef] [PubMed]
12. Carbonneau, P.; Bizzi, S. Seasonal Monitoring of River and Lake Water Surface Areas at Global Scale with Deep Learning. *Res. Sq.* **2022**. [CrossRef]
13. Bioresita, A.; Puissant, A.; Stumpf, A.; Malet, J.-P. A Method for Automated and Rapid Mapping of Water Bodies and Floods Using Sentinel-1 GRD SAR Data. *Remote Sens.* **2018**, *10*, 217. [CrossRef]
14. Huth, J.; Gessner, U.; Klein, I.; Yésou, H.; Lai, X.; Oppelt, N.; Kuenzer, C. Analyzing Water Dynamics Based on Sentinel-1 Time Series—A Study for Dongting Lake Wetlands in China. *Remote Sens.* **2020**, *12*, 1761. [CrossRef]
15. Jiang, C.; Zhang, H.; Wang, C.; Ge, J.; Wu, F. Water Surface Mapping from Sentinel-1 Imagery Based on Attention-UNet3+: A Case Study of Poyang Lake Region. *Remote Sens.* **2022**, *14*, 4708. [CrossRef]
16. Oakes, G.; Hardy, A.; Bunting, P.; Rosenqvist, A. RadWet-L: A Novel Approach for Mapping of Inundation Dynamics of Forested Wetlands Using ALOS-2 PALSAR-2 L-Band Radar Imagery. *Remote Sens.* **2024**, *16*, 2078. [CrossRef]
17. Shen, G.; Fu, W.; Guo, H.; Liao, J. Water Body Mapping Using Long Time Series Sentinel-1 SAR Data: A Case Study of Poyang Lake. *Water* **2022**, *14*, 1902. [CrossRef]
18. Twele, A.; Cao, W.; Plank, S.; Martinis, S. Sentinel-1-Based Flood Mapping: A Fully Automated Processing Chain. *Int. J. Remote Sens.* **2016**, *37*, 2990–3004. [CrossRef]
19. Wang, Z.; Xie, F.; Ling, F.; Du, Y. Monitoring Surface Water Inundation of Poyang Lake and Dongting Lake in China Using Sentinel-1 SAR Images. *Remote Sens.* **2022**, *14*, 3473. [CrossRef]
20. Xing, L.; Tang, X.; Wang, H.; Fan, W.; Wang, G. Monitoring Monthly Surface Water Dynamics of Dongting Lake Using Sentinel-1 Data at 10 m. *PeerJ* **2018**, *6*, e4992. [CrossRef] [PubMed]
21. Gulácsi, A.; Kovács, F. Sentinel-1–Imagery-Based High-Resolution Water Cover Detection on Wetlands, Aided by Google Earth Engine. *Remote Sens.* **2020**, *12*, 1614. [CrossRef]
22. Chamatidis, I.; Istrati, D.; Lagaros, N.D. Vision Transformer for Flood Detection Using Satellite Images from Sentinel-1 and Sentinel-2. *Water* **2024**, *16*, 1670. [CrossRef]
23. 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]
24. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv* **2018**, arXiv:1611.07004.
25. Tsai, Y.-H.; Hung, W.-C.; Schulter, S.; Sohn, K.; Yang, M.-H.; Chandraker, M. Learning to Adapt Structured Output Space for Semantic Segmentation. *arXiv* **2020**, arXiv:1802.10349. [CrossRef]
26. Simonetti, D.; Pimple, U.; Langner, A.; Marelli, A. Pan-Tropical Sentinel-2 Cloud-Free Annual Composite Datasets. *Data Brief* **2021**, *39*, 107488. [CrossRef]
27. Ao, D.; Dumitru, C.O.; Schwarz, G.; Datcu, M. Dialectical Generative Adversarial Networks for SAR Image Translation: From Sentinel-1 to TerraSAR-X. *Remote Sens.* **2018**, *10*, 1597. [CrossRef]
28. Çolak, E.; Sunar, F. A Comparison between CycleGAN Based Feature Translation and Optical-SAR Vegetation Indices. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2022**, *XLIII-B3-2022*, 583–590. [CrossRef]
29. Guo, Z.; Guo, H.; Liu, X.; Zhou, W.; Wang, Y.; Fan, Y. Sar2color: Learning Imaging Characteristics of SAR Images for SAR-to-Optical Transformation. *Remote Sens.* **2022**, *14*, 3740. [CrossRef]
30. 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]
31. Do, J.; Lee, J.; Kim, M. C-DiffSET: Leveraging Latent Diffusion for SAR-to-EO Image Translation with Confidence-Guided Reliable Object Generation. *arXiv* **2024**, arXiv:2411.10788.

32. Zhang, M.; Zhang, P.; Zhang, Y.; Yang, M.; Li, X.; Dong, X.; Yang, L. SAR-to-Optical Image Translation via an Interpretable Network. *Remote Sens.* **2024**, *16*, 242. [CrossRef]
33. Li, M.; Lin, J.; Ding, Y.; Liu, Z.; Zhu, J.-Y.; Han, S. GAN Compression: Efficient Architectures for Interactive Conditional GANs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 9331–9346. [CrossRef] [PubMed]
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762.
35. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and Flexible Image Augmentations. *Information* **2020**, *11*, 125. [CrossRef]
36. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. *arXiv* **2016**, arXiv:1603.08155. [CrossRef]
37. Fakhri, F.; Gkanatsios, I. Quantitative Evaluation of Flood Extent Detection Using Attention U-Net Case Studies from Eastern South Wales Australia in March 2021 and July 2022. *Sci. Rep.* **2025**, *15*, 12377. [CrossRef]
38. Zhang, Z.; Xiong, J.; Li, X.; Li, Y.; Liu, J. A SAR-Based Flood Mapping Approach: Application of SAR-SIFT Registration and Modified DeepLabV3 Segmentation in Flood Hazard Assessment. *Geocarto Int.* **2025**, *40*, 2512188. [CrossRef]
39. Ghosh, B.; Garg, S.; Motagh, M.; Martinis, S. Automatic Flood Detection from Sentinel-1 Data Using a Nested UNet Model and a NASA Benchmark Dataset. *PFG-J. Photogramm. Remote Sens. Geoinf. Sci.* **2024**, *92*, 1–18. [CrossRef]
40. Zhang, Z.; Giezendanner, J.; Mukherjee, R.; Tellman, B.; Melancon, A.; Purri, M.; Gurung, I.; Lall, U.; Barnard, K.; Molthan, A. Assessing Inundation Semantic Segmentation Models Trained on High- versus Low-Resolution Labels Using FloodPlanet, a Manually Labeled Multi-Sourced High-Resolution Flood Dataset. *J. Remote Sens.* **2025**, *5*, 0575. [CrossRef]
41. Yamazaki, D.; Ikeshima, D.; Sosa, J.; Bates, P.D.; Allen, G.H.; Pavelsky, T.M. MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resour. Res.* **2019**, *55*, 5053–5073. [CrossRef]
42. Lehner, B.; Grill, G. Global River Hydrography and Network Routing: Baseline Data and New Approaches to Study the World's Large River Systems. *Hydrol. Process.* **2013**, *27*, 2171–2186. [CrossRef]
43. Carbonneau, P.E. S1-to-S2-Style-Transfer: A Model and Demo Script to Predict Sentinel 2 Images Based on Sentinel 1 Inputs. Available online: https://github.com/PCarbonneauDurham/S1-to-S2-Style-transfer (accessed on 29 August 2025).

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.