International Journal of Semantic Computing
Vol. 14, No. 1 (2020) 55–69

DOI: 10.1142/S1793351X20400036

Improved Semantic Segmentation of Water Bodies and Land in SAR Images Using Generative Adversarial Networks

M. M. Manohara Pai* and Vaibhav Mehrotra†

Department of Information and Communication Technology
Manipal Institute of Technology
Manipal Academy of Higher Education
Manipal, India
**mmm.pai@manipal.edu*
†*vaibhavmehrotra2197@gmail.com*

Ujjwal Verma‡

Department of Electronics and Communication Engg
Manipal Institute of Technology
Manipal Academy of Higher Education
Manipal, India
ujjwal.verma@manipal.edu

Radhika M. Pai‡

Department of Information and Communication Technology
Manipal Institute of Technology
Manipal Academy of Higher Education
Manipal, India
radhika.pai@manipal.edu

The availability of computationally efficient and powerful Deep Learning frameworks and high-resolution satellite imagery has created new approach for developing complex applications in the field of remote sensing. The easy access to abundant image data repository made available by different satellites of space agencies such as Copernicus, Landsat, etc. has opened various avenues of research in monitoring the world's oceans, land, rivers, etc. The challenging research problem in this direction is the accurate identification and subsequent segmentation of surface water in images in the microwave spectrum. In the recent years, deep learning methods for semantic segmentation are the preferred choice given its high accuracy and ease of use. One major bottleneck in semantic segmentation pipelines is the manual annotation of data. This paper proposes Generative Adversarial Networks (GANs) on the training data (images and their corresponding labels) to create an enhanced dataset on which the networks can be trained, therefore, reducing human effort of manual labeling. Further, the research also proposes the use of deep-learning approaches such as U-Net and FCN-8 to perform an efficient segmentation of auto annotated, enhanced data of water body and land. The experimental results show that the U-Net model without GAN achieves superior performance on SAR images with pixel accuracy

‡Corresponding authors.

of 0.98 and F1 score of 0.9923. However, when augmented with GANs, the results saw a rise in these metrics with PA of 0.99 and F1 score of 0.9954.

Keywords: GANs; SAR images; semantic segmentation; U-net; deep learning.

1. Introduction

The surface water is an essential component in understanding and deriving insights about local ecological and hydrological processes. Surface water, in contrast to atmospheric or groundwater, largely refers to water on the surface of earth such as rivers, lakes and wetlands. Surface water is often subjected to external forces that result in expansions, contractions and changes in appearance, lending a dynamic component to the flow of the water. Extreme changes to surface water can have serious repercussions such as floods which is currently the most common natural disaster to affect the world. Thus, it is imperative to develop new approaches to detect and constantly monitor and predict future water levels.

The recent large-scale proliferation of remote sensing satellites such as the Sentinel-1, Landsat, and Radarsat resulted in regular monitoring of the earth at high-frequency periodic intervals. Further, these satellites are equipped with high-resolution microwave sensors capable of imagery in all terrain conditions as well as showing invariance to day and night cycles. One of the significant advantages is their ability to penetrate thick cloud cover.

Detection of surface water in Synthetic Aperture Radar (SAR) images until now has largely been addressed by elaborate image processing algorithms. Some commonly applied approaches include the watershed algorithm [1], thresholding [2] and morphological profiling followed by the traditional machine learning algorithms such as Support Vector Machines (SVMs). Although these algorithms have been shown to perform effectively for a specific polarization, the model does not generalize across polarizations. In addition, the presence of foreign objects such as bridges results in gaps in the output [1]. This and the significant time investment in terms of hand tuning paves the way for more robust approaches such as neural networks.

The advent of very deep neural networks in the past few years along with off-the-shelf libraries for learning algorithms has enabled easy building of end-to-end models to perform common computer vision tasks such as image recognition, object detection and image segmentation. Deep learning methods offer a significant advantage over traditional image processing pipeline, in that there is little domain knowledge or insight required. One such instance of image segmentation is semantic segmentation in which specific regions of an image are automatically categorized as one among several predetermined classes. Semantic segmentation has found significant success in areas such as self-driving cars [3] as well as diagnostics of medical images [4], and analysis of videos acquired by Unmanned Aerial Vehicles [5, 6]. Recently, SAR image was analyzed for segmentation of land and waterbodies [7]. However, this approach only considers U-net architecture for semantic segmentation

with limited dataset. In this paper, the dataset is augmented using GANs, eliminating time consuming manual annotations, thus improving the performance of the model.

The primary contribution in this paper is as follows:

(1) The usage of the U-Net [9] and FCN [10] architecture to perform semantic segmentation of surface water and land using high-resolution SAR data.
(2) Show the effectiveness of transfer learning, a modern deep learning paradigm which reuses knowledge learnt from similar tasks.
(3) Further improve the U-Net methodology by augmenting the dataset of manually annotated images through the use of Generative Adversarial Networks (GANs) to generate SAR images and corresponding masks.

1.1. *Water body representation in SAR images*

SAR images are typically represented as the grayscale images, wherein the associated intensity value of each pixel is denoted by the proportion of microwave which is back scattered. Land, which is usually rough, appears bright with high intensity. Water appears dark since most of the incident radar energy is scattered away. The significant contrast difference can thus be exploited for efficient segmentation. An inherent problem in SAR images is Speckle Noise, a form of multiplicative noise that corrupts SAR images by altering backscatter. Therefore, speckle noise can distort the river edges, thus, making it difficult to accurately determine the boundaries. Fortunately, robust algorithms such as the Median filter and the Lee filter exist to reduce speckle noise [8].

1.2. *Using generative adversarial networks to create more data*

In this paper, U-net architecture has been used. U-Net proves to be a well-suited model because U-Net works with very few training images and detect boundaries of an irregular and rough nature.

Using GANs additional SAR images with the corresponding labels can be generated, thus, increasing the size of the dataset and hence, enhancing the performance of U-Net architecture.

GANs comprise two models, a Generator G and a Discriminator D. The generator aims to generate new data from the latent space provided similar to the expected one. On the other hand, the discriminator is fed data from the generator as well as the real dataset, and its aim is to recognize if the data are 'fake' (from the generator) or 'real' from the dataset.

Through various iterations, these models act as adversaries. The generator learns how to generate data which are closer to the 'real' dataset, whereas the discriminator gets better at identifying 'fake' data. An ideal state is reached when the generator becomes so good that the discriminator cannot distinguish between data from the dataset and the generator. In such a state, the discriminator predicts the data to be

fake or real with a probability of 0.5 and then the data from the generator can be used as a substitute of the real data.

For the purpose of this paper, we use Deep-Convolutional Generative Adversarial Networks (DC-GAN) [11].

The rest of this paper is organized as follows. Section 2 reviews similar works, Sec. 3 details with the proposed U-Net and FCN-based SAR segmentation, and experimental results are presented and analyzed in Sec. 4. Finally, conclusions are drawn in Sec. 5.

2. Related Work

Primary work in the field of semantic segmentation of SAR images has been towards target recognition and road segmentation. Cui *et al.* [12] used region-based convolutional neural networks for target detection in large-scene images. The objective of their model was to detect targets such as tanks and armored cars, as provided in the MSTAR dataset. This methodology involved a fast sliding method to slice and resize the images that are input to the model. An average accuracy of 94.67% was recorded.

Yang *et al.* [13] used Conditional Random Fields on region adjacency graphs for the semantic labeling of SAR images. Gabor filters were employed to extract texture information, whereas to exploit back-scattering intensity information, gamma distribution and histogram cues were used. The highest accuracy of 86.5% was reached when all the above-mentioned techniques were used.

Henry *et al.* [14] used deep fully convolutional neural networks (FCNs) for road segmentation in SAR images. They found that by adding a tolerance rule towards spatially small mistakes, FCNs proved to be an effective model for road segmentation, overcoming the major difficulty of isolating thin objects in a speckled environment. However, the model had difficulty in generalizing over a variety of patterns and would fail in applications, wherein the contour of the water bodies is extremely irregular.

A Deep Learning approach based on a modified U-Net architecture has been shown to work by Zhengxin Zhang *et al.* [15], for the extraction of the road using the Massachusetts roads dataset. Relaxed precision and recall were used as the evaluation parameters along with a break-even point, a point where the relaxed precision and recall were equal. The deep residual U-Net or ResNet showed a break-even point at 0.9187, outperforming the conventional U-Net with a break-even point of 0.9053.

River channel segmentation has been explored in [1] using an image processing approach (Watershed segmentation). The primary drawback of using the watershed segmentation algorithm is irregular and jagged boundaries of rivers. Similarly, objects such as bridges and ships can cause the algorithm to get "stuck" within the local high contrast region of the river.

3. Proposed Methodology

In this section, the proposed framework for surface water and land segmentation is presented. In this study, Sentinel 1 SAR images from the European Open Access hub [16], which provides free access to the Sentinel family of products, were manually collected. Using Sentinels Application Platform (SNAP) [17], the Refined Lee filter [8] is applied to despeckle the SAR images. Subsequently, the open source python library *labelme* was used to manually annotate the image. Each pixel is labeled either as water body (river, lake, backwater, etc.) or land. Finally, data-augmentation techniques are applied to generate new training samples (Fig. 1).

3.1. *Speckle noise reduction*

One of the challenges of segmenting SAR images is the presence of the speckle noise. To reduce the effect of speckle noise, filtering techniques that preserve the boundaries of the river are applied to the SAR images. Ardhi Wicaksono Santoso *et al.* [8] performed a comparative study of filters based on properties such as Speckle Index (SI), Average Difference (AD), Equivalent Number of Looks (ENL), and determined the Lee filter to have the best metrics. The Refined Lee filter [18] is an improvement over the original Lee filter which dynamically adjusts the number of pixels used in the sliding window by employing the K-Nearest Neighbor algorithm.

3.2. *Semantic segmentation architecture*

We evaluate the performance of two different deep learning networks proposed for semantic segmentation. The first architecture, the so-called FCNs proposed by Long *et al.*, [10] extended image classification networks by replacing the fully connected layers with a decoder network. The decoder network applies deconvolution operations to create an image with the same height and width dimensions of the input. FCNs are highly flexible with respect to the encoder architecture and can be trained using a variety of pre-trained classification networks such as the VGG (FCN-VGG16), GoogLeNet (FCN-GoogLeNet) or the FCN-AlexNet. FCNs are implemented in three variants in accordance with the spatial precision of their output. We implement the FCN-8 architecture which successively upsamples a series of concatenated upsamples and max-pooled layers to produce a precise 8x segmentation map. The U-Net architecture proposed by Ronneberger [9] demonstrated its effectiveness for biomedical image segmentation. It consists of two paths, a contracting path and an expansive path, which correspond to the encoder and decoder architectures, respectively. The additional skip connections between the encoder and decoder halves improve localization and generate highly precise segmentation maps.

3.2.1. *Transfer learning*

Training very deep neural network architectures is usually not a feasible task given the requirements of computational power and limitations on the size of the dataset.

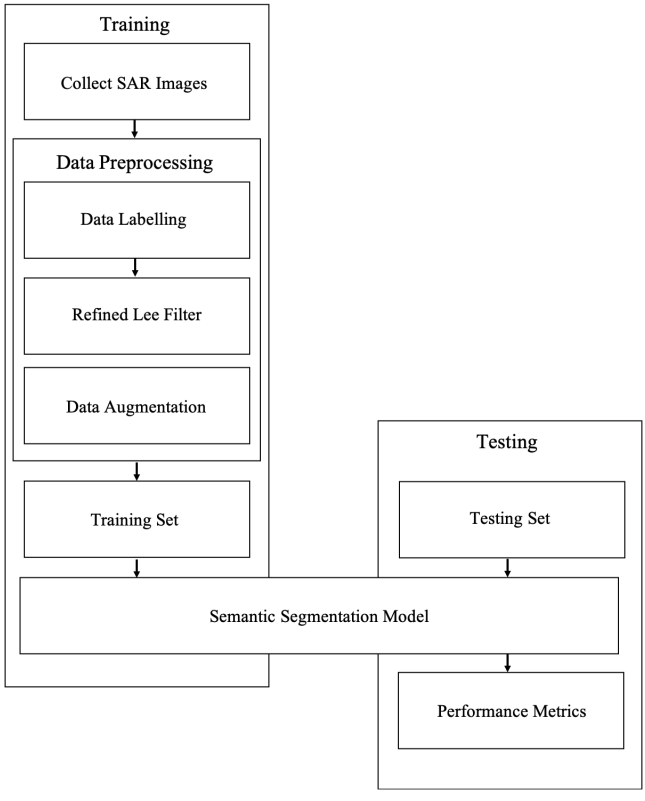


Fig. 1. Proposed methodology.

Transfer learning is based on the principle of reusing the weights of a pre-trained model for a similar task. Such an initialization has been proven to work better than the random weight initialization [19].

In this paper, it is proposed to compare the results obtained by learning the U-Net model from scratch and using the pre-trained weights as learnt by the U-Net model on the ISBI 2015 Cell Tracking dataset. In addition, another semantic segmentation architecture FCN is used to segment land and water body.

3.3. *Augmenting with GANs*

The raw SAR images along with their corresponding masks are fed into the GAN where the generator has been given the latent space and the discriminator is fed real images and generated images. Once this GAN reaches an acceptable level of performance, we will obtain new SAR image patches along with their corresponding masks. These images, combined with the initial dataset are fed into the U-Net for further analysis.

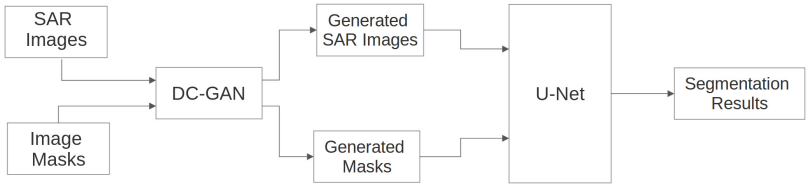


Fig. 2. Using DC-GANs to augment segmentation process of U-Net.

Deep Convolutional GANs (DC-GANs) are used to create more patches of SAR images. The ground truth masks and raw SAR images are fed into the GAN (Fig. 2). The generator maps the underlying distribution of the SAR data as well as the corresponding masks to generate more images. To feed the GAN in the first place, heavy augmentation is required which might lead to a skewed output in which case the generated images will look like the training dataset to a higher degree than what is required.

The DC-GAN is trained with batch size of 128, with batch normalization to avoid any large spikes within the batch. The Adam optimizer is used and a sparse softmax cross-entropy with logits loss function is used in the implementation with an initial learning rate of 0.0002.

4. Results and Discussion

4.1. *Experimental procedure*

Dataset: A subset of the publicly available SAR data from the European Copernicus Satellite mission is utilized to evaluate the performance of the proposed framework. This study utilizes 40 level 1 Ground Range Detected (GRD) images acquired over land using the Interferometric Wide (IW) swath mode with a 5×20 m resolution. Image data from coastal areas of India (Mangalore–Udupi region) were collected and rescaled appropriately to produce tiles of 512×512 pixels. Subsequently, areas of interest, which included images with river pixels, were selected. For training and testing, a dataset of 30 and 10 images, respectively, was created. The training set also includes only "river" and only "land" images. The training set contains 12% of river pixels and the test set 19% of river pixels. The class imbalance problem is handled by optimizing the Dice Loss function for the U-Net model [20]. Labeling the images was performed to create the ground truth of the respective SAR images. Some of the images in the training set along with the ground truth are displayed in Fig. 3.

The proposed algorithm was implemented in python 2.7 using the keras functional API on a Intel Xeon with a Tesla V100 GPU, 32 GB RAM on a system running CUDA 9.0.

Data Augmentation: Manual labeling of a large number of river boundaries in SAR images is a laborious and time-consuming task. Therefore, data augmentation is

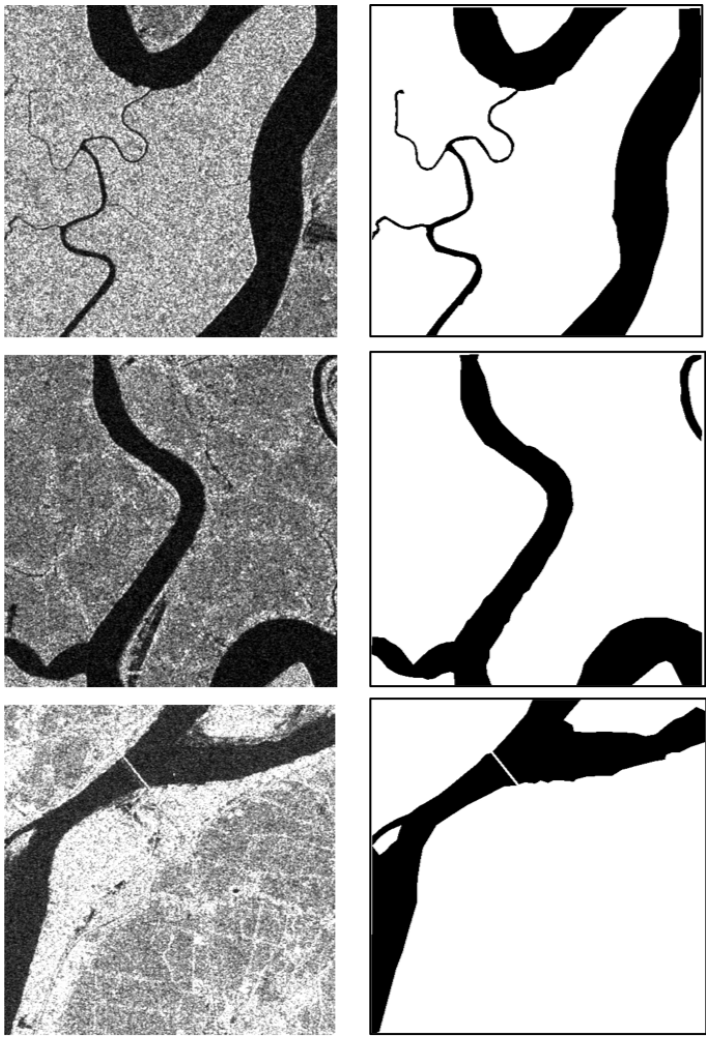


Fig. 3. Examples from the training set, column (a) corresponds to SAR images and column and (b) is the ground truth.

the preferred choice in such a scenario. Data augmentation is a technique that applies transformations such as rotation, translation, scaling, etc. to improve the usage of the annotated data and achieve invariance [21] with respect to width and height shifts along with rotation. Similar to tissue deformation, river water flow is also susceptible to extensive contouring and changes in width. In this study, the data is augmented by rotating, translating, zooming, horizontal flipping and shearing of the original data. The images are augmented in real time by randomly selecting the parameters to transform. The range or parameters allowed can be seen in Table 1.

Table 1. Augmentation parameters.

Rotation Range	0.2
Width Shift Range	0.05
Height Shift Range	0.05
Zoom Range	0.05
Horizontal Flip	True

After the generation of additional SAR images using the GANs, the entire dataset still goes through augmentation to provide the adequate features to the neural network for training.

Training: The weights of the network were randomly initialized with a normal distribution with mean zero and standard deviation $\sqrt{2/n}$. This initialization has been proven by [22] to allow for very deep architectures to converge. In all the layers, a zero padding is applied such that the dimension of the output is the same as the input (same padding). The Adam optimizer, with a learning rate of 10^{-4}, is used to learn the network parameters. The U-net model was trained for five epochs, each of 2500 steps with a batch size of 4. The U-Net after augmenting with GAN was still trained in the same way.

Loss Function: For the U-Net model, the soft-dice loss function [20] is a typically preferred loss function to reduce bias in predictions. It is especially useful when the training data suffer from a class imbalance problem. It is formulated as follows:

$$\text{Dice Loss} = \frac{2|A \cap B|}{|A| + |B|},$$

where numerator $A \cap B$ represents the intersection of the sets A and B and the denominator, the number of elements of A and B, respectively. In this study, the soft-dice loss function is used because of the class imbalance problem as discussed earlier.

4.2. *Evaluation metrics*

To evaluate the algorithms, several metrics, namely, Precision, Recall, Mean Intersection over Union (MIoU) as well as the Pixel Accuracy (PA) are used [10]. Precision is the fraction of the water-body pixels which are labeled correctly and Recall is the fraction of all the labeled water-body pixels that are correctly predicted. They are formulated as follows:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}},$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}.$$

Since neither Precision nor Recall is sufficient to completely describe the performance of a model, the F1 score, a weighted harmonic mean of the two metrics is also

employed. It is given as follows:

$$F1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

The models are also evaluated with respect to the MIoU as well as the PA. The MIoU measures the similarity between the labeled image and the ground truth. PA simply reports the number of pixels correctly classified by the model. The two metrics are as follows:

$$\text{MIoU} = \frac{(1/C)\sum_i n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}},$$

$$\text{PA} = \frac{\sum_i n_{ii}}{\sum_i t_i},$$

where C is the number of classes, n_{ji} is the number of pixels of class i mistakenly classified as belonging to class j. t_i represents the number of pixels belonging to class i.

4.3. *Generating SAR images using GANs*

The DC-GAN was able to capture the underlying distribution of the noise in the SAR data and recreate such images. When the pipeline was fed with the ground truths and raw SAR data simultaneously, we were able to create images and their masks. The noisy texture of the SAR images has been adequately captured by the GAN as seen by visual inspection (Fig. 4). The output of the DC-GAN is SAR images of 32×32 resolution.

4.4. *Results of neural network architectures for semantic segmentation*

We evaluate the three network architectures (U-Net, transfer U-Net and FCN) on 10 test images with respect to the metrics defined above. For transfer learning, the pretrained model was trained on the 2015 ISBI Cell Tracking challenge data. The quantitative results of the models, can be seen in Table 2. Qualitatively, the visual outputs of the segmentation models can be seen in Fig. 5.

As is expected, the U-Net model, which learns its weights from scratch, displays the highest performance with respect to the metrics defined. It achieves a high precision and recall and thus maintains a high F1 score as well.

We see that FCNs perform the worst especially in the areas where the river channels are very thin. In addition, FCNs seem unable to capture fine detail of the boundaries which result in a blocky appearance and discontinuities as seen in the last image and second last image, respectively.

We would also like to highlight the effectiveness of the U-Net architecture in detecting ships. Although we do not explicitly label ships in the training dataset, the

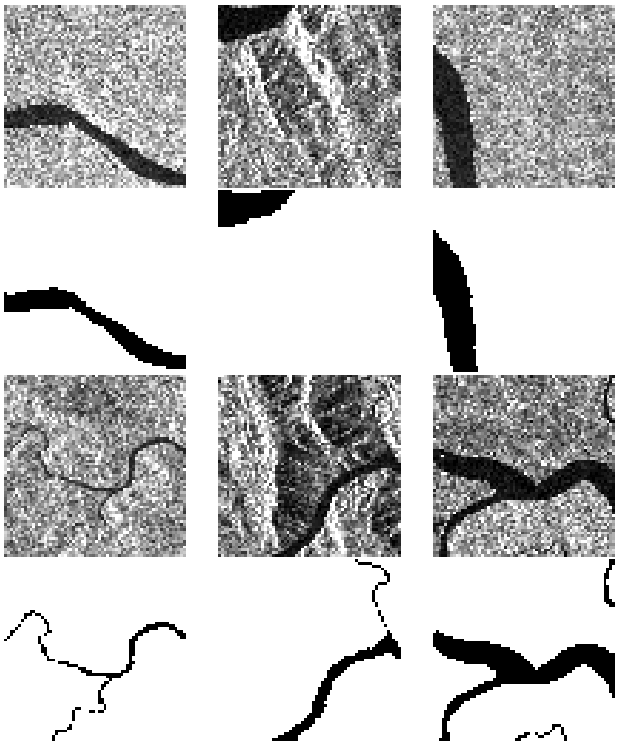


Fig. 4. SAR images and corresponding masks generated using GAN. The actual output size is 32×32 but it has been upscaled for better visualization.

Table 2. Performance comparison.

Architecture	Precision	Recall	F1	MIoU	PA
U-Net	0.9927	0.9919	0.9923	0.9551	0.9876
FCN	0.9877	0.9820	0.9848	0.9189	0.9757
Transfer U-Net	0.9943	0.9881	0.9912	0.9512	0.9859

network is effectively able to segment ships from the surrounding water. We note that FCNs were not able to capture this detail.

Results of Transfer Learning: We would like to highlight an interesting result, that is, the surprising effectiveness of the Transfer Learning approach. We extensively experimented with retraining selected decoder layers. Retraining only the last 1×1 convolution layer results in predictions that are significantly noisy. Similarly, retraining only the second last convolution layer produces a model with lesser noise and so forth. We discover that eventually retraining the last five layers architecture results in a model that performs comparatively to the U-Net model which learned its weights from scratch. Thus, we see that the spatial-and low-level features map

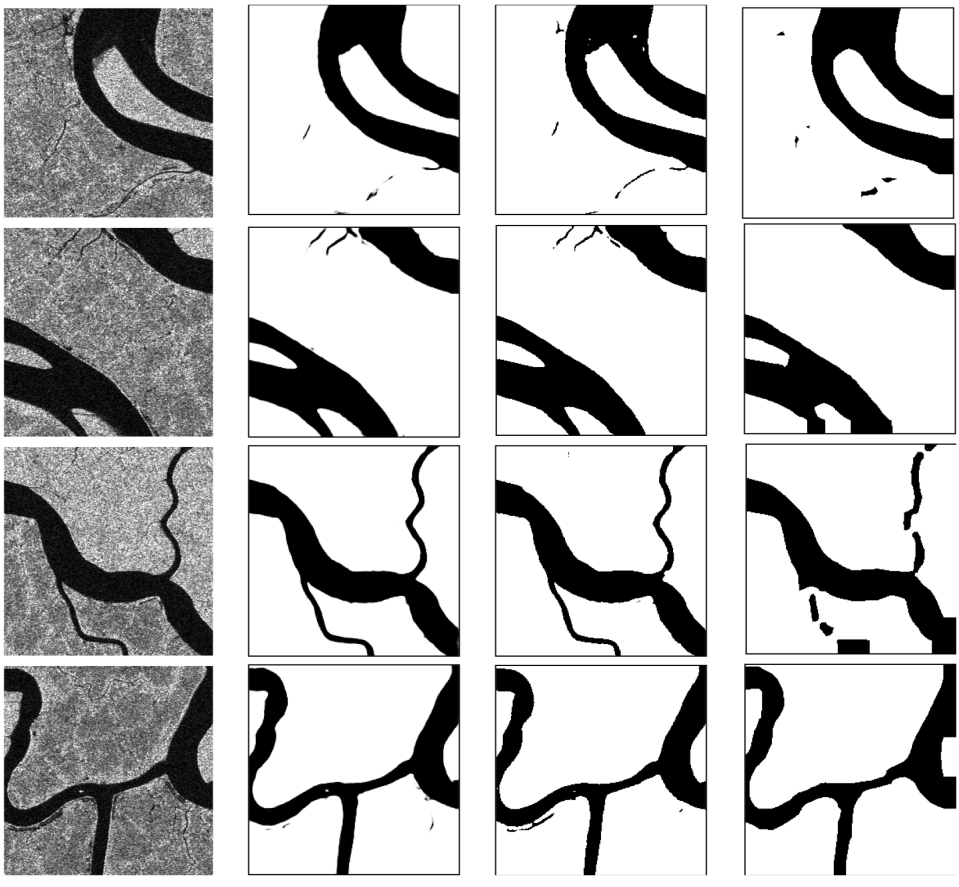


Fig. 5. Experimental results of proposed architecture. Column (a) is the SAR image captured by the Sentinel-1 satellite. Column (b) is the segmentation map produced by the U-Net model. Column (c) is the map produced by the transfer learning approach and column (d) is the map produced by the FCN model.

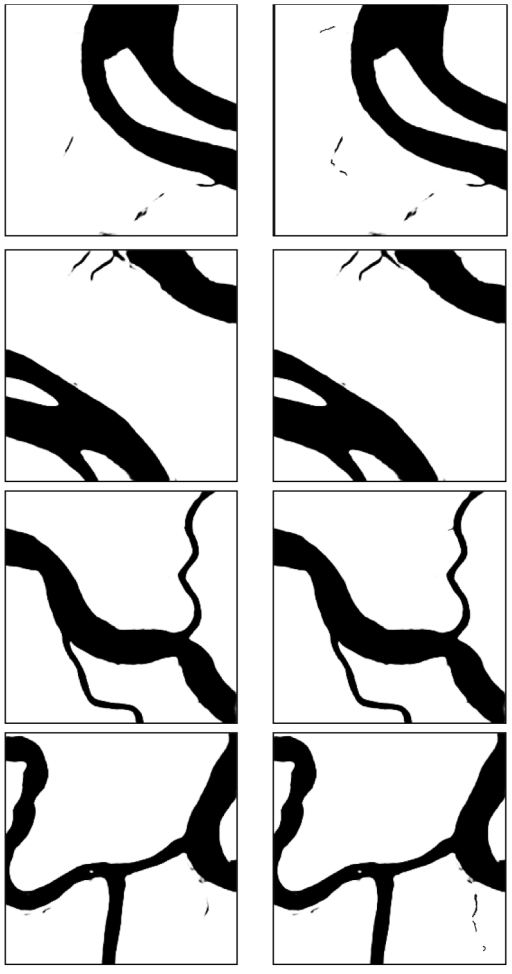
very well from the biomedical image segmentation problem over to surface water segmentation in SAR data.

4.4.1. *Results of augmenting U-Net with GAN*

The Vanilla U-Net architecture was trained with a combination of the original image-mask dataset and the generated image-mask dataset, therefore increasing the size of the training set even before augmentation was done. We evaluated the two methods on 10 test images with respect to the metrics defined above. We trained the U-Net model on 30 images that were manually annotated and the U-Net + GAN model is trained on 10 additional images that were generated by the DC-GAN. The

Table 3. Performance comparison after augmenting with GANs.

Architecture	Precision	Recall	F1	MIoU	PA
U-Net	0.9927	0.9919	0.9923	0.9551	0.9876
U-Net + GAN	0.9959	0.9949	0.9954	0.9641	0.9902



(a) (b)

Fig. 6. Augmenting the size of the dataset using GANs. Column (a) is the semantic segmentation result of U-Net and (b) is U-Net augmented with GANs.

quantitative results can be seen in Table 3. For visual inspection, qualitative results can be seen in Fig. 6.

Both the architectures perform very well on the SAR images and obtain a good F1 score of 0.99. Unlike the traditional image processing approaches such as Watershed

segmentation, U-Net models can identify the fine details. Moreover, the transfer U-net architecture is able to even segment rivers whose width is small. However, Vanilla U-Net fails to identify small width rivers.

5. Conclusion

In this paper, a robust methodology is proposed for an efficient and highly precise segmentation of surface river water and land. Two different implementations of U-Net architecture are studied on the SAR images, one in which U-net is trained from scratch (Vanilla U-Net) and other in which pre-trained weights are used (Transfer U-Net). In addition, an FCN is also trained on the same data and performances are compared. Experimental results show that both the U-Net architectures gave similar performance in terms of F1 score, PA and mean IoU, which was superior to the performance shown by the FCN, specially interms of mean IoU. However, transfer U-Net is able to identify very minute details in the image such as small rivers etc. This methodology is augmented using GANs to create addition training data. The results show improvement in the accuracy across various parameters. Therefore, the large manual overhead of labeling SAR images for ground truth is removed because the GAN creates images that can be used instead. This opens the door for created larger datasets.

One limitation however to this approach is the possibility of false positives, that is, the model may identify water in regions of relatively low intensity. For our future work, we would extend this model for multi-class classification and introduce information from panchromatic satellite imagery for verification.

References

[1] M. Ciecholewski, River channel segmentation in polarimetric sar images: Watershed transform combined with average contrast maximisation, *Expert Syst. Appl.* **82** (2017) 196–215.

[2] J. Li and S. Wang, An automatic method for mapping inland surface waterbodies with radarsat-2 imagery, *Int. J. Remote Sensing* **36**(5) (2015) 1367–1384.

[3] M. Siam, S. Elkerdawy, M. Jägersand and S. Yogamani, Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges, pp. 1–8, 2017. Available at https://doi.org/10.1109/ITSC.2017.8317714.

[4] H. R. Roth, C. Shen, H. Oda, M. Oda, Y. Hayashi, K. Misawa and K. Mori, Deep learning and its application to medical image segmentation, *Medical Imaging Technology* **36** (2018) 63–71.

[5] S. Girisha, M. Manohara Pai, U. Verma and R. M. Pai, Performance analysis of semantic segmentation algorithms for finely annotated new UAV aerial video dataset (Manipal UAVid), *IEEE Access* **7** (2019) 136239–136253.

[6] S. Girisha, M. Pai, U. Verma and R. Pai, Semantic segmentation of UAV aerial videos using convolutional neural networks, in *Second IEEE Int. Conf. Artificial Intelligence and Knowledge Engineering*, 2019, pp. 21–27.

[7] M. Pai, V. Mehrotra, S. Aiyar, U. Verma and R. Pai, Automatic segmentation of river and land in sar images: A deep learning approach, in *IEEE Second Int. Conf. Artificial Intelligence and Knowledge Engineering*, 2019, pp. 15–20.

[8] A. W. Santoso, D. Pebrianti, L. Bayuaji and J. M. Zain, Performance of various speckle reduction filters on synthetic aperture radar image, in *4th Int. Conf. Software Engineering and Computer Systems*, 2015, pp. 11–14.

[9] O. Ronneberger, P. Fischer and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention* (Springer International Publishing, 2015), pp. 234–241.

[10] J. Long, E. Shelhamer and T. Darrell, Fully convolutional networks for semantic segmentation, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[11] A. Radford, L. Metz and S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks.

[12] Z. Cui, S. Dang, Z. Cao, S. Wang and N. Liu, Sar target recognition in large scene images via region-based convolutional neural networks, *Remote Sens.* **10**(5) (2018). Available at http://www.mdpi.com/2072-4292/10/5/776.

[13] W. Yang, L. Chen, D. Dai and G.-S. Xia, Semantic labelling of sar images with conditional random fields on region adjacency graph, *Radar, Sonar and Navig. IET* **5** (2011) 835–841.

[14] C. Henry, S. M. Azimi and N. Merkle, Road segmentation in sar satellite images with deep fully convolutional neural networks, *IEEE Geosci. Remote Sens. Lett.* **15**(12) (2018) 1867–1871.

[15] Z. Zhang, Q. Liu and Y. Wang, Road extraction by deep residual u-net, *IEEE Geosci. Remote Sens. Lett.* **15**(5) (2018) 749–753.

[16] European copernicus open access hub. Available at https://scihub.copernicus.eu.

[17] Sentinel application platform. Available at http://step.esa.int/main/toolboxes/snap/snap-faq/.

[18] A. S. Yommy, R. Liu and S. Wu, Sar image despeckling using refined lee filter, *2015 7th Int. Conf. Intelligent Human-Machine Systems and Cybernetics*, 2015, pp. 260–265.

[19] J. Yosinski, J. Clune, Y. Bengio and H. Lipson, How transferable are features in deep neural networks? arXiv:1411.1792. Available at http://arxiv.org/abs/1411.1792.

[20] F. Milletari, N. Navab and S. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *Fourth Int. Conf. 3D Vision*, 2016, pp. 565–571.

[21] A. Dosovitskiy, J. T. Springenberg, M. Riedmiller and T. Brox, Discriminative unsupervised feature learning with convolutional neural networks, in *Advances in Neural Information Processing Systems 27*; Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (eds.) (Curran Associates, 2014), pp. 766–774. Available at http://papers.nips.cc/paper/5548-discriminative-unsupervised-feature-learning-with-convolutional-neural-networks.pdf.

[22] K. He, X. Zhang, S. Ren and J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in *IEEE Int. Conf. Computer Vision*, 2015, pp. 1026–1034.