Automatic Segmentation of River and Land in SAR Images: A Deep Learning Approach

Manohara Pai M. M
Department of Information and Communication Technology
Manipal Institute of Technology
Manipal Academy of Higher Education
Manipal, India
mmm.pai@manipal.edu

Vaibhav Mehrotra
Department of Information and Communication Technology
Manipal Institute of Technology
Manipal Academy of Higher Education
Manipal, India
vaibhavmehrotra2197@gmail.com

Shreyas Aiyar
Department of Information and Communication Technology
Manipal Institute of Technology
Manipal Academy of Higher Education
Manipal, India
shreyas.aiyar96@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

***Abstract*—The ubiquitousness of satellite imagery and powerful, computationally efficient Deep Learning frameworks have found profound use in the field of remote sensing. Augmented with easy access to abundant image data made available by different satellites such as LANDSAT and European Space Agency's Copernicus missions, deep learning has opened various avenues of research in monitoring the world's oceans, land, rivers, etc. One significant problem in this direction is the accurate identification and subsequent segmentation of surface-water in images in the microwave spectrum. Typically, standard image processing tools are used to segment the images which are time-inefficient. However, in recent years, deep learning methods for semantic segmentation is the preferred choice given its high accuracy and ease of use. This paper proposes the use of deep-learning approaches such as U-Net to perform an efficient segmentation of river and land. Experimental results show that our approach achieves vastly superior performance on SAR images with pixel accuracy of 0.98 and F1 score of 0.99.**

***Index Terms*—Semantic Segmentation, SAR image, U-Net, Deep Learning.**

I. Introduction

Observation of 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 subject 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 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 traditional machine learning algorithms such as Support Vector Machines (SVM's). 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.

978-1-7281-1488-0/19/$31.00 ©2019 IEEE
DOI 10.1109/AIKE.2019.00011

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]. In this paper, it is proposed to develop an efficient methodology to tackle the problem of surface water detection from satellite SAR images.

The primary contribution in this paper is as follows:

1) The usage of the U-Net 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.

A. *Representation of Water 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 [5].

B. *Motivation for U-Net Architecture*

In this paper, it is proposed to demonstrate the effectiveness of U-Net architecture [6] using a manually labelled dataset of SAR images. Each pixel is thus labelled either as river/water bodies or land. U-Net proves to be a well-suited model because:

1) U-Net works with very few training images due to the effectiveness of data augmentation approaches.
2) It has proven to detect boundaries of an irregular and rough nature with very high accuracy.

The rest of this paper is organised as follows. Section II reviews similar work, section III details the proposed U-Net-based SAR segmentation, and experimental results are presented and analysed in Section IV. Finally, conclusions are drawn in Section V.

II. Related Work

Primary work in the field of semantic segmentation of SAR images has been towards target recognition and road segmentation. Cui et al. [7] 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 armoured cars, as provided in the MSTAR dataset. The 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. [8] used Conditional Random Fields on region adjacency graphs for the semantic labelling of SAR images. Gabor filters were employed to extract texture information whereas to exploit backscattering 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. [9] used deep fully-convolutional neural networks for road segmentation in SAR images. They found that by adding a tolerance rule towards spatially small mistakes, fully convolutional neural networks (FCNNs) 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 generalising 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. [10] for the extraction of the road using the Massachusetts roads data set. 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 .9187, outperforming the conventional U-Net with a break-even point of .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.

III. 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 [11], which provides free access to the Sentinel family of products, was manually collected. Using Sentinel's Application Platform (SNAP) [12], the Refined Lee filter [5] is applied to despeckle the SAR images. Subsequently, the open source python library *labelme* was used to manually annotate the image. Finally, data-augmentation techniques are applied to generate new training samples.

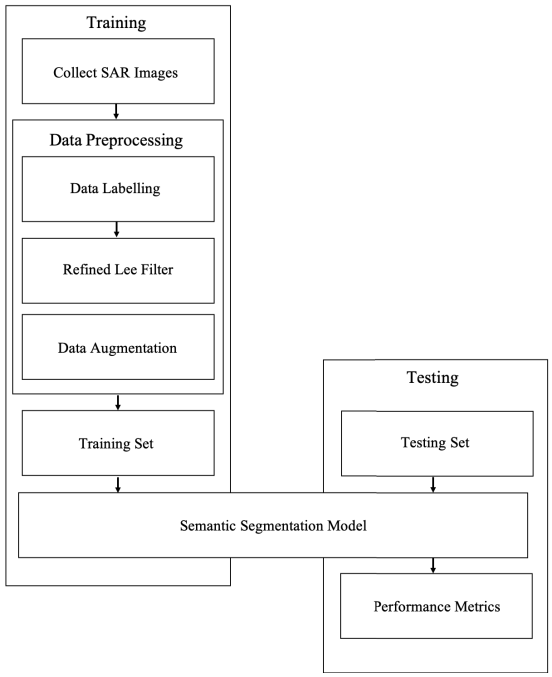

Fig. 1. Proposed Methodology

A. *Despeckling with Refined Lee Filter*

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. [5] performed a comparative study of filters based on properties such as Speckle Index (SI), Average Difference (AD), Equivalent Number of Looks (ENL), and have determined the Lee filter to have the best metrics. The Refined Lee filter [13] 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 Neighbour algorithm.

B. *Semantic Segmentation Architecture*

In this work, the performance of U-Net architecture [6] is studied for semantic segmentation of surface-water and land. The U-Net architecture is a popular architecture for semantic segmentation originally proposed for biomedical image segmentation. It consists of three paths, a contracting path and an expansive path which correspond to the encoder and decoder architectures respectively. The presence of a bridge acts as an intermediate layer between the two paths. The additional copy and crop operations between the encoder and decoder halves improve localisation and generate highly precise segmentation maps while retaining high-level semantic information.

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. Transfer learning is based on the principle of reusing the weights of a pre-trained model for a similar task. Such an initialisation has been proven to work better than random weight initialization [14]. 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.

IV. Results & Discussion

A. *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 5x20m resolution. Image data from coastal areas of India (Mangalore-Udupi region) was collected and rescaled appropriately to produce tiles of 512x512 pixels. Subsequently, areas of interest, which included images with river pixels, were selected. For training and testing, created a dataset of 30 and 10 images respectively. 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 optimising the Dice Loss function for the U-Net model [15]. Labelling 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 Figure 2.

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 labelling of a large number of river boundaries in SAR images is a laborious and time-consuming task. Therefore, data augmentation is 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 [16] 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. Finally, 50,000 images are obtained which are used to train the network.

Training: The weights of the network were randomly initialised with a normal distribution with mean zero and standard deviation $\sqrt{2/n}$. This initialisation has been proven by [17] 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 optimiser, with a learning rate of 10^{-4} is used to learn the network parameters. The U-net model was trained for 5 epochs, each of 2500 steps with a batch size of 4.

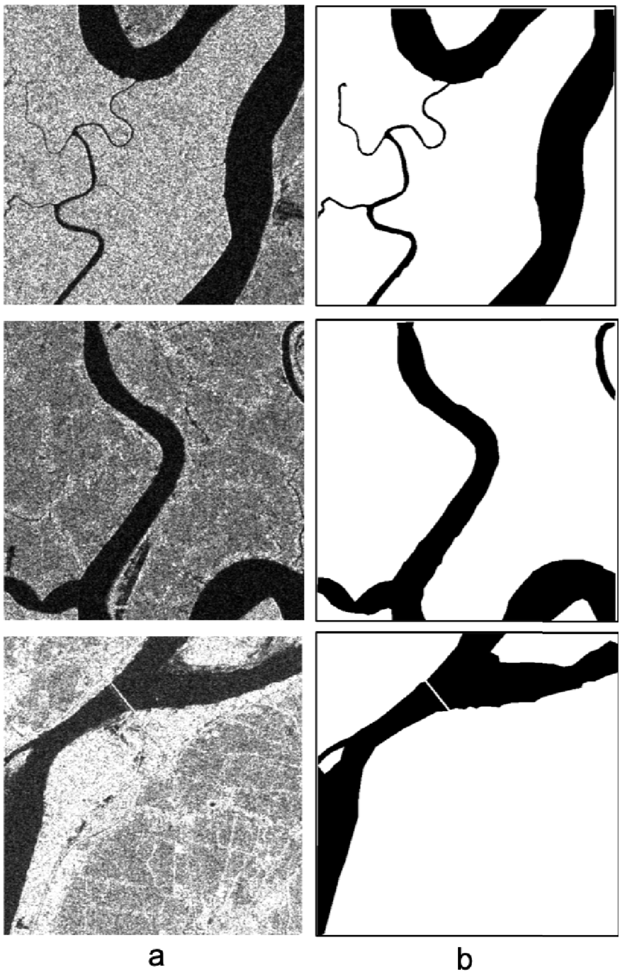

Fig. 2. Examples from the training set, column **(a)** corresponds to SAR images and column **(b)** is the ground truth.

Loss Function: For the U-Net model, the soft-dice loss function [15] is a typically preferred loss function to reduce bias in predictions. It is especially useful when the training data suffers from a class imbalance problem. It is formulated as follows:

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

The 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.

B. Evaluation Metrics

To evaluate the algorithms, several metrics namely Precision, Recall, Mean Intersection over Union (MIoU) as well as the Pixel Accuracy (PA) is used [18]. Precision is the fraction of the water-body pixels which are labelled correctly and Recall is the fraction of all the labeled water-body pixels that are correctly predicted. They are formulated as follows:

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

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

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:

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

The models are also evaluated with respect to the Mean Intersection over Union (MIoU) as well as the Pixel Accuracy (PA). The MIoU measures the similarity between the labelled image and the ground truth. Pixel accuracy simply reports the number of pixels correctly classified by the model. The two metrics are as follows:

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

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

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.

C. Results

In this work, two models are used: Vanilla U-Net and Transfer U-net. The Vanilla U-Net architecture is trained on our SAR images dataset, while Transfer U-Net is trained on the 2015 ISBI Cell Tracking challenge and fed the weights generated for our transfer learning task. The results of the transfer learning task represent the best results i.e the one obtained by retraining the decoder of the U-Net model. The network architectures are evaluated on 10 test images with respect to the metrics defined above. The quantitative results of the models is shown in Table I. Qualitatively, the visual outputs of the segmentation models can be seen in Figure 3.

TABLE I
PERFORMANCE COMPARISON

Architecture	Precision	Recall	F1	MIoU	PA
Vanilla U-Net	0.9927	0.9919	0.9923	0.9551	0.9876
Transfer U-Net	0.9943	0.9881	0.9912	0.9512	0.9859

Both the architectures perform very well on 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.

One more interesting result is surprising effectiveness of the Transfer Learning approach. An extensive experimentation with retraining selected decoder layers is done. Retraining only the last 1x1 Convolution layer results in predictions that are significantly noisy. Similarly by retraining only the second last convolution layer produces a model with lesser noise and so forth. It is discovered that increasing the number of trainable layers in the decoder architecture results in successively better performance. Performance peaks on retraining the entire decoder architecture which results in a model that performs comparatively to the U-Net model that learned its weights from scratch. It is seen that the spatial features map very well from the biomedical image segmentation problem over to surface water segmentation in SAR data.

V. Conclusion

In this paper, a robust methodology is proposed for an efficient and highly precise segmentation of surface river water and land. In addition, two different implementation of U-Net architecture is studied on SAR images, one in which U-net is trained from scracth (Vanilla U-Net) and other in which pretrained weights are used (Transfer U-Net). Experimental results show that the both architectures gave similar performance in terms of F1 score, pixel accuracy and mean IoU. However, transfer U-Net is able to identify very minute details in the image such as small rivers etc. 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 Systems with Applications*, vol. 82, pp. 196 – 215, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417417302622

[2] J. Li and S. Wang, "An automatic method for mapping inland surface waterbodies with radarsat-2 imagery," *International Journal of Remote Sensing*, vol. 36, no. 5, pp. 1367–1384, 2015. [Online]. Available: https://doi.org/10.1080/01431161.2015.1009653

[3] M. Siam, S. Elkerdawy, M. Jägersand, and S. Yogamani, "Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges," pp. 1–8, 2017. [Online]. Available: 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," *arXiv e-prints*, p. arXiv:1803.08691, Mar. 2018.

[5] A. W. Santoso, D. Pebrianti, L. Bayuaji, and J. M. Zain, "Performance of various speckle reduction filters on synthetic aperture radar image," in *2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)*, Aug 2015, pp. 11–14.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Springer International Publishing, 2015, pp. 234–241.

[7] 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 Sensing*, vol. 10, no. 5, 2018. [Online]. Available: http://www.mdpi.com/2072-4292/10/5/776

[8] 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 & Navigation, IET*, vol. 5, pp. 835 – 841, 11 2011.

[9] C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in sar satellite images with deep fully convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 12, pp. 1867–1871, Dec 2018.

[10] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, May 2018.

[11] "European copernicus open access hub." [Online]. Available: https://scihub.copernicus.eu

[12] "Sentinel application platform." [Online]. Available: http://step.esa.int/main/toolboxes/snap/snap-faq/

[13] A. S. Yommy, R. Liu, and S. Wu, "Sar image despeckling using refined lee filter," *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 2, pp. 260–265, 2015.

[14] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *CoRR*, vol. abs/1411.1792, 2014. [Online]. Available: http://arxiv.org/abs/1411.1792

[15] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, Oct 2016, pp. 565–571.

[16] 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, Inc., 2014, pp. 766–774. [Online]. Available: http://papers.nips.cc/paper/5548-discriminative-unsupervised-feature-learning-with-convolutional-neural-networks.pdf

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 1026–1034.

[18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

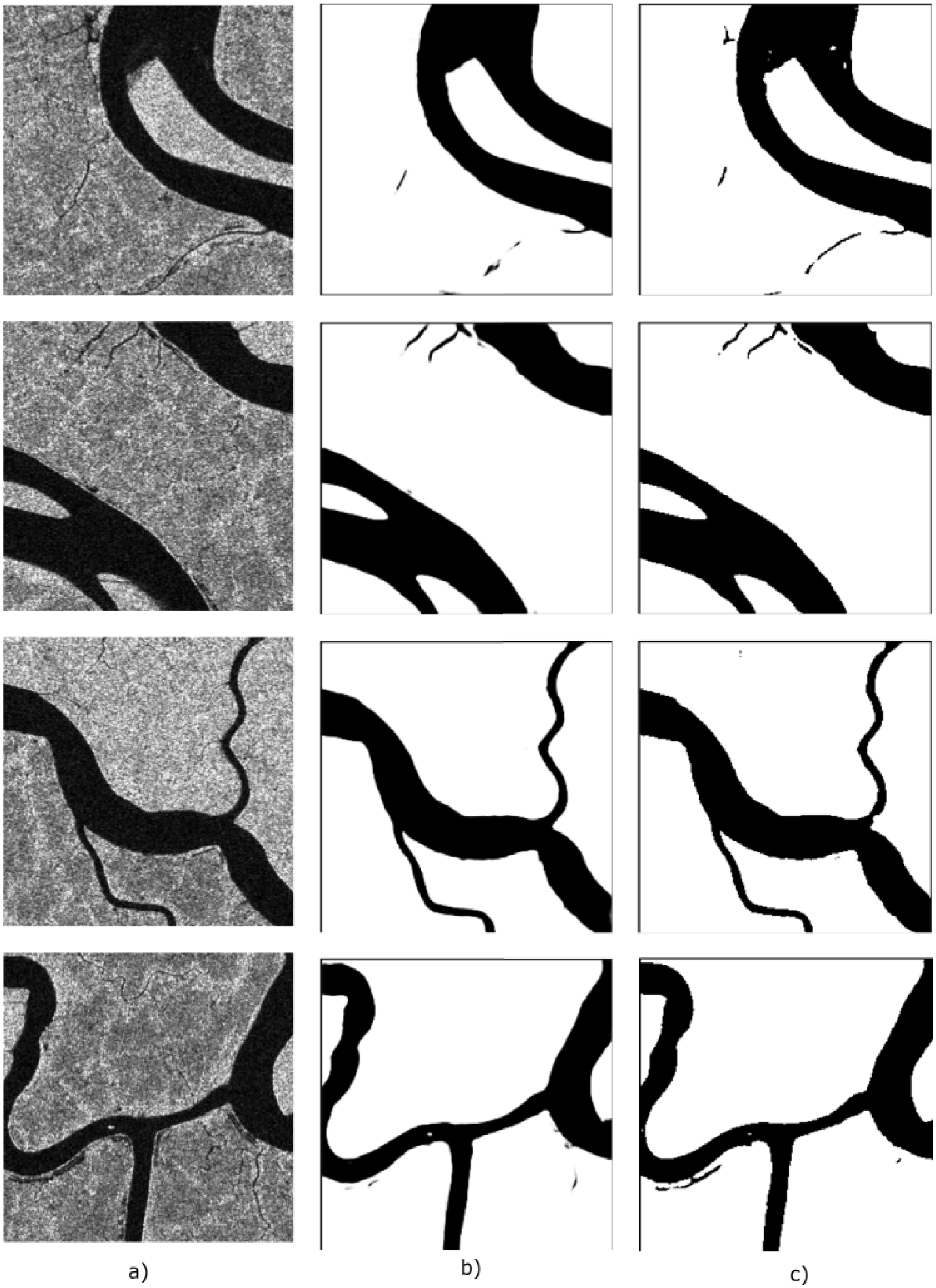

Fig. 3. Experimental Results of Proposed Architecture. Column **(a)** is the SAR image captured by the Sentinel-1 satellite. Column **(b)** is the segmentation map produced by the Vanilla U-Net model. Column **(c)** is the map produced by the Transfer U-Net model.