http://dx.doi.org/10.5755/j01.eie.26.6.25849 ELEKTRONIKA IR ELEKTROTECHNIKA, ISSN 1392-1215, VOL. 26, NO. 6, 2020

A New Classification Approach with Deep Mask R-CNN for Synthetic Aperture Radar Image Segmentation

Ridvan Yayla[1, *], Baha Sen[2]
[1]*Department of Computer Engineering, Bilecik Seyh Edebali University, 11230, Bilecik, Turkey*
[2]*Department of Computer Engineering, Ankara Yildirim Beyazit University, 06010, Ankara, Turkey*
ridvan.yayla@bilecik.edu.tr

Abstract—**In this paper, a hybrid classification approach which is combined with a more deep mask region-convolutional neural network and sparsity driven despeckling algorithm is proposed for synthetic aperture radar (SAR) image segmentation instead of the classical segmentation methods. In satellite technology, synthetic aperture radar images are strongly used for a lot of areas, such as evaluating air conditions, determining agricultural fields, climatic changes, and as a target in the military. Synthetic aperture radar images must be segmented to each meaningful point in the image for a quality segmentation process. In contrast, synthetic aperture radar images have a lot of noisy speckles and these speckles should be also reduced for a quality segmentation. Current studies show that deep learning techniques are widely used for segmentation methods. High accuracy and fast results can be obtained with deep learning techniques for image segmentation. Mask region-convolutional neural network can not only separate each meaningful field in the image, but it can also generate a high accuracy prediction for each meaningful field of synthetic aperture radar images. The study shows that smoothed SAR images can be classified as multiple regions with deep neural networks.**

Index Terms—**Image segmentation; Neural networks; Radar imaging; Synthetic aperture radar.**

I. INTRODUCTION

Synthetic aperture radar (SAR) images are widely used in satellite technology. In current satellite technology, the SAR images are used for detection of a target in the military, changing air condition maps, determining the agriculture terrains. Thanks to SAR images, the desired targets can be hit with a high accuracy in the military by using unmanned air vehicles and the people can be informed of air conditions in advance with the early warning system that is based on SAR images [1]. Moreover, the agriculture terrains are periodically observed for its crop performance thanks to SAR image processing [2]. Segmentation methods play an important role in the evaluation of meaningful parts of the SAR image. In recent years, the segmentation methods could have been applied for a high accuracy rate by the development of deep learning methods and high graphic processing unit (GPU) resources.

According to that SAR images have a high-quality resolution and complexity, separating to meaningful parts of SAR images is quite difficult [3]. This complexity of the SAR image is a big problem for a quality segmentation. Smoothed and uncomplicated SAR images help more quality and high accuracy segmentation [4]. Recent studies show that the researchers focus on the SAR image despeckling. A multi-scale convolutional neural network model is proposed for SAR image semantic segmentation in [5]. The model contains to noise removal, convolutional, feature concatenation, and classification stages. As a different study, a deep learning approach that is called “Image Despeckling Convolutional Neural Network” (ID-CNN) uses a set of convolutional layers for automatically removing speckle from the input noisy images [6]. In our study, the sparsity-driven despeckling (SDD) method is used for smoothing process. The SDD method smoothens SAR image speckle noises and edges [7].

Region-based segmentation methods are useful for extracting to meaningful parts of the SAR images [8]. Instead of the whole image data are tested, the most effective method is to analyse a sample of the part of the image [9]. Convolutional Neural Network (CNN) is the one of the most used deep learning architectures for object detection and image segmentation in image processing.

As an initial step, region-based segmentation methods extract to free form regions from an image. Secondly, it describes these regions, and finally it follows to segmentation using a recognition pipeline [10]. Region-based image segmentation is related to pixel similarity and homogeneity. Mask Region-based Convolutional Neural Network (Mask R-CNN) is one of the useful segmentation methods which has been inspired by the Faster R-CNN algorithm in recent years. It is presented that the Mask R-CNN method detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance [11]. While Mask R-CNN creates to bounding box with region proposal network (RPN) for predicted regions and performs to classification and bounding box regression by using region of interest (RoI)

Manuscript received 22 March, 2020; accepted 3 September, 2020.

branch, it also adds to predicted segmentation masks in predicted regions. Mask R-CNN can also separate to same objects or meaningful fields in an image with different masking by using to instance segmentation. In our approach, we studied on a deeper Mask R-CNN framework that is based on matterport implementation for the smoothed SAR images by using to trained input weights of CNN.

II. Related Works

A. Sparsity Driven Despeckling (SDD) Method

SAR images have a lot of speckle noises. The noisy speckles of SAR images constitute an obstacle for image segmentation. Due to the speckle noises, pointless pixel similarity is a big problem in the segmentation process. For this reason, the speckle noises should be reduced for a high-quality SAR image segmentation [12].

The SAR image speckles are reduced by using SDD minimization method. In our study, Moving and Stationary Target Acquisition and Recognition (Mstar) clutter dataset is handled. It contains to 100 SAR images with different angles of the regions. All image noises of the dataset are reduced by using SDD method and the smoothed SAR images are used for a high-quality segmentation in the Mstar dataset preparing process. The new Mstar dataset is called “Mstar SDD dataset”. A few samples of original and smoothed SAR images of the Mstar dataset are shown in Fig 1.

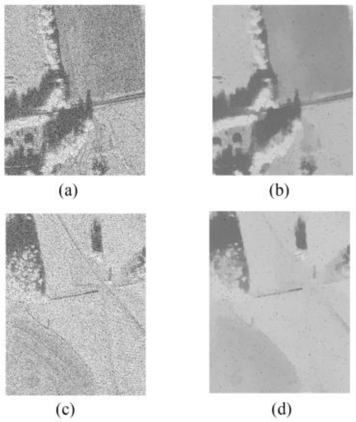


Fig. 1. Samples of Mstar database SAR images: (a) Mstar HB06271 sample; (b) Mstar HB06271 (smoothed sample) with SDD; (c) Mstar HB06158 sample; (d) Mstar HB06158 (smoothed sample) with SDD.

While various techniques, such as the Lee filter, the Lee refined filter, the Frost filter, and the Kuan filter for despeckling to the speckles, have been proposed, the new despeckling methods, such as Bayesian denoising method and Markov random field (MRF) method, have been proposed also [13]. Moreover, wavelet-based algorithms and methods are proposed with the increased complexity for reducing the speckle noises of SAR images.

Sparsity driven despeckling (SDD) method is developed for reducing to the edge and point noises of the SAR images with L_0 and L_1 norms by using a single parameter and less execution time [14]. In the SDD method, the SAR image despeckling optimization problem is defined in (1)

$$\hat{F} = \arg\min_{F} J(F). \tag{1}$$

The cost function J(F) is written in a matrix-vector form as follows [7]

$$\begin{aligned}\tilde{J}(\tilde{F}) &= (v_F - v_G)^T (v_F - v_G) + \\ &+ (v_F^T C_x^T W_x C_x v_F + v_F^T C_y^T W_y C_y v_F),\end{aligned} \tag{2}$$

where v_F and v_G are the vector presentations of $\tilde{F}$ and $\tilde{G},$ respectively. G is the observed speckled image. When C_x and C_y express to Toeplitz matrices, W_x and Wy are diagonal matrices [14]. The matrix-vector form in (2) enables a special iterative optimization method where a linear system is solved in each step via (3):

$$Av = v, \tag{3}$$

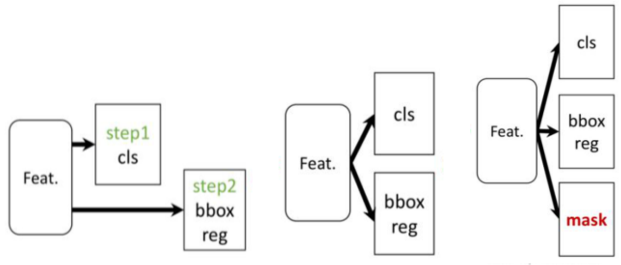
$$A^n = I + C_x^T W_x^{(n)} C_x + C_y^T W_y^{(n)} C_y. \tag{4}$$

A^n is the weight matrix of nth iteration computed using $W_x^{(n)}$ and $W_y^{(n)}$ based on structural vector $v_F^{(n)},$ such that I is an identity matrix and n is the iteration number [14]. Vector $v_F^{(n)}$ is used for the smooth process by using to A^n weight matrix in the SDD method. This smoothed output image is used as a matrix for the convolutional input layer in our study.

B. Mask R-CNN

Mask R-CNN is a deep neural network that is based on R-CNN, Fast R-CNN, and Faster R-CNN algorithms, for instance, segmentation [15], [16].

R-CNN generates to independent regions proposal by using the selective search algorithm. Each region proposal is given to CNN so that it generates to features as a feature extractor from each region (Fig. 2). After passing through CNN, R-CNN extracts a feature vector for each region proposal, and finally support vector machine (SVM) is applied for classifying to the desired region with extracted features from CNN [17]. Fast R-CNN uses a single deep CNN to extract features for the entire image once unlike R-CNN. The whole proposal regions are sent to CNN architecture and it runs for all proposal regions [18]. Every proposal region works on CNN architecture by using a technique that is called “Region of Interest (ROI) Pooling”. The last CNN is used in Faster R-CNN for Region Proposal Network (RPN) that depends on the calculated features of the image instead of using the selective search algorithm CNN for Region Proposal Network (RPN) that depends on the calculated features of the image instead of using the selective search algorithm [19].



(a) (b) (c)

Fig. 2. a) –R-CNN (slow); b) – Fast/-er R-CNN; c) – Mask R-CNN.

The most prominent difference between Fast R-CNN and Faster R-CNN is RPNs. According to the usage of last convolutional layers, RPNs reduce the computational requirements of the overall inference process and they decide where the meaningful fields are in the image. The RPN quickly and efficiently scans every location to assess whether further processing needs to be evaluated in a given region. The proposal anchor boxes are the bounding boxes that are predicted for each meaningful parts with different square sizes. In this way, Faster R-CNN decreases to calculation time and gives faster results.

III. PROPOSED METHOD

In the proposed solution, the SDD method is combined with the Mstar database for the initial hybrid study. When the Mstar SAR images are investigated, most of images have a big complexity. For instance, when the forests and their shadows are compared with the roads in SAR images, these regions have a big pixel similarities and the whole input images are smoothed with SDD method. In this way, the Mstar SDD dataset images that are smoothed and of reduced complexities are created one by one. This dataset is converted to numpy arrays for the CNN input layers by using to Mask R-CNN algorithm. The Mstar SDD database images will be given as an input matrix of the region-based convolutional input layer.

Secondly, the five class identifiers are defined for the intelligent classification process. The SAR image regions are defined as the most observed five classes that are forest, building, road, tree(s), and terrain. The undefined region of the SAR images is defined as background.

Thirdly, Mask R-CNN algorithm is used for detecting regions of the SAR images. Instead of the RoI Pooling method, the RoI align method is used in Mask R-CNN [11]. When the RoI pooling method uses feature map bounds of quantized integer coordinates that cause the more segmentation losses, the RoI align method uses feature map bounds of non-quantized floating coordinates that help to fewer segmentation losses. Regular RoI pooling changes the topology of the features and it causes a misalignment between feature map outputs and RoIs. Using regular RoI pooling would negatively impact the ability to predict pixel-accurate masks. Bilinear interpolation is used for estimating the exact values of the input features at four fixed locations in each RoI bin. In this way, the results only aggregate then for per bin. The loss function output of Mask R-CNN is expressed in (5)

$$L_{mask-rcnn} = L_{cls} + L_{bbox} + L_{mask}, \tag{5}$$

where L_{cls} is loss function, L_{bbox} is prediction box loss function, and L_{mask} is mask loss function. The loss function is observed in the segmentation process and it is aimed that the segmentation and prediction losses are reduced to masking with different backbone networks. While pre-trained Mask R-CNN weights are given to the initial convolutional inputs in the first running time, the trained weights that are observed at the end of the first running time are given to initial convolutional inputs in the second running time for executing deeper CNN. It is aimed that higher accuracy and less loss results are obtained for SAR image segmentation by using deeper region-based CNN. The basic steps of the proposed method are shown in Fig. 3.

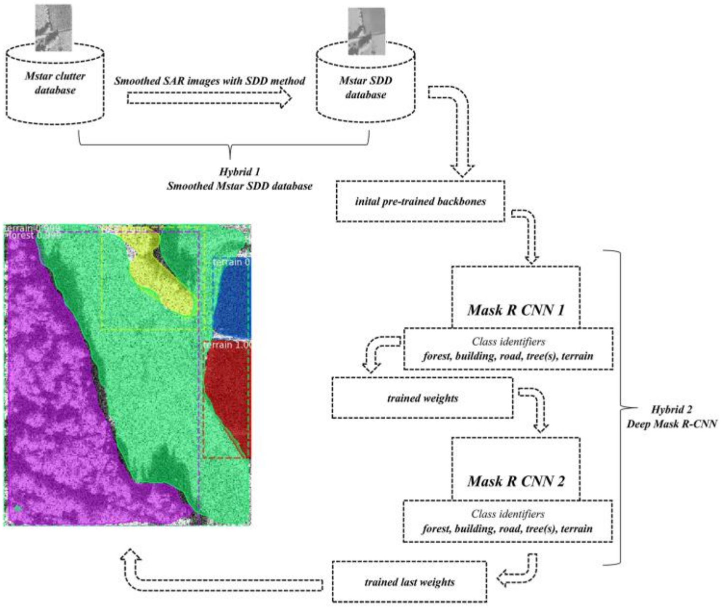


Fig. 3. The basic steps of the proposed method.

IV. MATERIALS

A. Dataset

In our study, the Mstar public clutter dataset is handled for SAR image segmentation. It contains 100 SAR images with different locations and angles [20]. The Mstar dataset is very useful for region-based image segmentation thanks to its low-resolution and size. 80 % of SAR images of Mstar clutter dataset are randomly divided for the training process, and 20 % of SAR images are randomly divided for the validation process.

SAR images are smoothed by using SDD algorithm in related work. Thanks to the SDD algorithm, the masking edges can be detected and drawn more clearly. After the smooth process, visual geometry group (VGG) image annotator is used for determining the regions and its polygonal edges [21]. It is a simple and useful annotation software that can generate a single JavaScript Object Notation (json) file. The json file holds to coordinates of the polygonal regions and information of these regions. These defined regions are sent to the convolutional neural networks as an input neuron.

When the Mstar clutter images are investigated, the *forest, road, tree* or *trees, terrain,* and *building regions* of the SAR images are mostly observed for an intelligent classification. *The five identifier classes* are created based on Mask R-CNN matterport implementation in our study [22].

B. Classification

Current studies show that the Mask R-CNN algorithm contains useful region classification predictions, masking, and a bounding box regression. However, the Mask R-CNN can predict a single object or region detection of an input image, it can also predict to multi-detection of different or same objects. Thanks to instance segmentation, the same regions of the input image are masked with different masking.

While the Mask R-CNN matterport implementation is executed, the pre-trained coco weights are used for the initial weights. The pre-trained coco weights help to start initial weights that are required for the input layer of the region-based convolutional neural network. Moreover, the segmentation processes are observed with different pre-trained models, such as vgg16, resnet50, resnet101, and inceptionv3.

C. Deep Mask R-CNN

In our study, a SAR classification algorithm that is based on Mask R-CNN implementation is created with pyhton. Initially, the coco model weights are used for the input layer of CNN.

SAR classification algorithm is basically in the following steps:

Initial step:
Step 1. Prepare to Mstar SDD database.
Step 2. Create to dataset directory system.
Step 3. Determine to desired regions of SAR images with VGG image annotator.
Training step:
Step 4. Determine to classification regions (*forest, building, road, tree(s), terrain*).
Step 5. (If it is initial step) set to default backbone networks weights for training process (initial weights - coco or other backbones).
(If the algorithm is executed one time)
set the last obtained keras weights to algorithm.
Step 6. Train to algorithm until total epoch.
Evaluation step:
Step 7. During the training, observe to losses.
Step 8. Generate to SAR classification weights for each iteration. If the algorithm loss, masked loss, and bounding box loss are not sufficiently reduced, go to step 5. If it is enough, go to step 9.
Step 9. Display to segmented and masked SAR image.

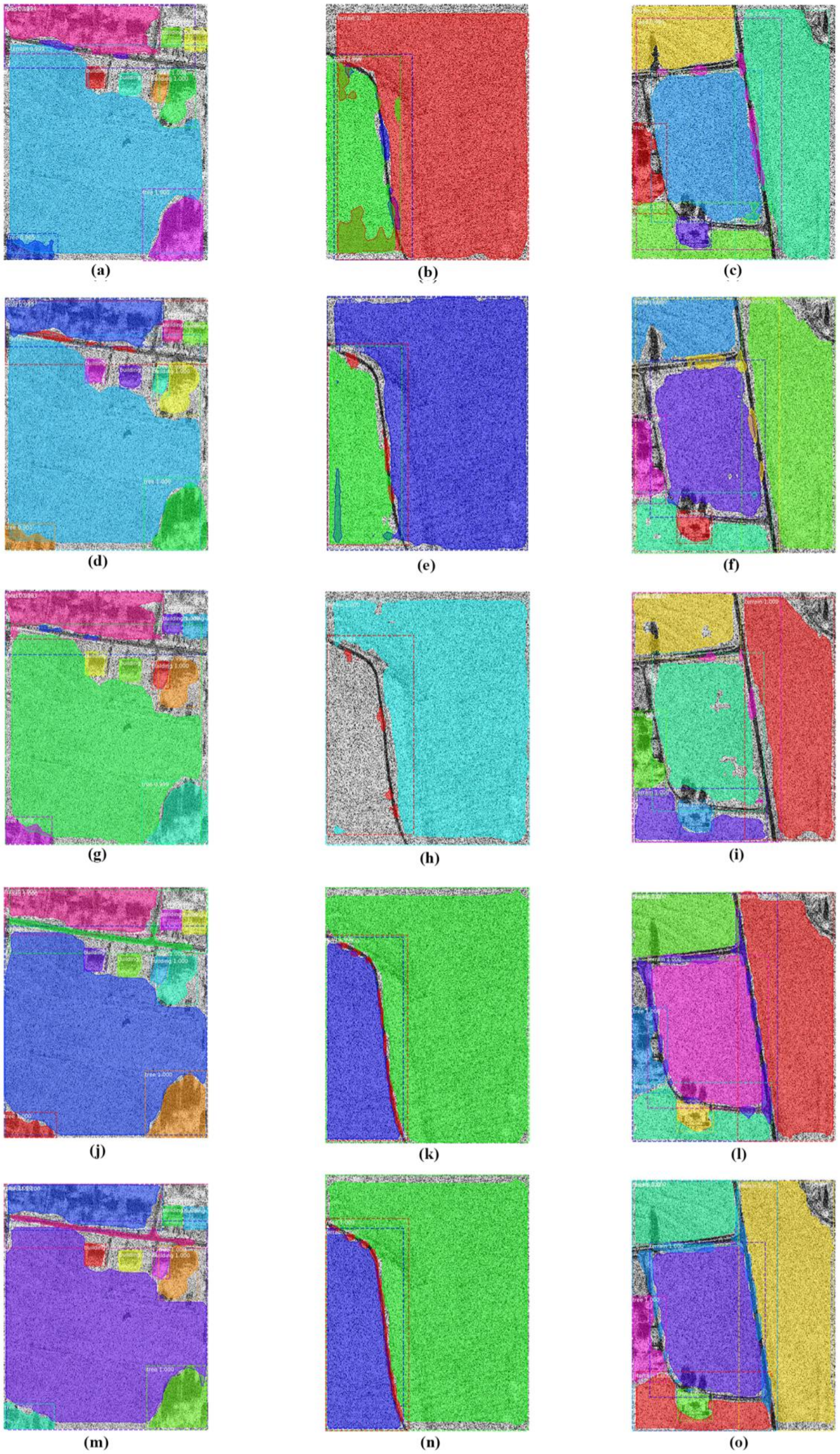
V. EXPERIMENTS

We studied on python virtual environment based on Mask R-CNN matterport implementation. In our experiment, CPU: Intel (R) Core (TM) 4 cores, memory: 16 GB, GPU: i7-2600 NVIDIA GeForce GTX1070 Ti 8 GB GDDR5 hardware configuration is used, and the experiment is built with python 3.6, tensorflow 1.13, keras 2.1 in virtual environment. All experiments are applied with 60 total epochs and 500 per epoch. The running time is approximately 4 hours and 30 minutes for each training process.

In addition, when vgg16, inception v3, resnet50, resnet101, and coco backbone networks, which is combined Deep Mask R-CNN, are compared to each other, Deep Mask R-CNN is trained with high accuracies and low loss rates in resnet50 and coco models. Moreover, the bounding box and predictions are also obtained with high accuracy. Deep Mask R-CNN performance results are shown in Table I, and the segmentation results for the five different backbone models are shown in Fig. 4. When the trained weights that are obtained from Mask R-CNN 1 are applied as initial weights for Mask R-CNN 2, the high accuracy and low loss results are obtained. The final accuracy (*accuracy 2*) and loss rates (*loss 2*) are also shown in Table I.

TABLE I. DEEP MASK R-CNN EXPERIMENTAL RESULTS.

DEEP MASK R CNN	MASK R-CNN 1					
	Backbone	**loss 1**	**mrcnn class loss 1**	**mrcnn bbox loss 1**	**mrcnn mask loss 1**	**Accuracy 1 (%)**
	Vgg16	0.5206	0.0900	0.0889	0.2516	47 %
	Inceptionv3	0.5033	0.0850	0.0862	0.2472	49 %
	Resnet101	0.4731	0.0794	0.0752	0.2423	52 %
	Resnet50	*0.1019*	*0.0270*	*0.0095*	*0.0499*	*89 %*
	Coco	*0.0994*	*0.0263*	*0.0091*	*0.0487*	*90 %*
	MASK R-CNN 2					
	Backbone	**loss 2**	**mrcnn class loss 2**	**mrcnn bbox loss 2**	**mrcnn mask loss 2**	**Accuracy 2 (%)**
	Vgg16	0.2347	0.0261	0.0321	0.1428	76 %
	Inception v3	0.2410	0.0278	0.0335	0.1446	75 %
	Resnet101	0.2131	0.0268	0.0296	0.1260	78 %
	Resnet50	*0.0651*	*0.0193*	*0.0046*	*0.0355*	*93 %*
	Coco	*0.0651*	*0.0183*	*0.0045*	*0.0367*	*93 %*



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m) (n) (o)

Fig. 4. Segmentation results with different backbones: (a)–(c) Vgg16, (d)–(f) inceptionv3, (g)–(i) Resnet101, (j)–(l) Resnet50, (m)–(o) Coco model.

According to experimental results, when the Mask R-CNN framework is applied deeper as the proposed method, it is observed that the class, masking, and bounding box losses are reduced by Deep Mask R-CNN.

VI. CONCLUSIONS

In this study, a new hybrid classification method for SAR image segmentation is proposed. According that the SAR images have a noise complexity, the smoothed SAR images provide convenience for a quality segmentation. While the SDD method provides to reduction of the speckle noises in the SAR image by using the mixed norm of L_0 and L_1 with a single parameter for fast computing, Deep Mask R-CNN provides a high-quality segmentation with multi-region predictions.

While most of the current studies focus on detecting a single region of the SAR image segmentation, multi-region detection of SAR image segmentation is trained. Multi-region segmentation can be applied for the complexity images with robust segmentation features of the Mask R-CNN [23].

The SDD method for the input layers of CNN is used in the proposed method and SAR images are classified with multi-region segmentation via Deep Mask R-CNN. The study shows that no matter which backbone network is trained for SAR image segmentation, the algorithm losses of the Deep Mask R-CNN are reduced and the accuracy of the algorithm increases.

VII. FUTURE STUDIES

SAR images are of vital importance for the satellite technology. It is especially used for target detection in military and air condition maps for early warning systems. Our study focuses on five regions (forest, building, tree, terrain, and road) of the 100 SAR images based on Mstar dataset as a sample study. These regions can also be extended for more multi-regions, such as lakes, oceans, seas, city centers, mountains, etc. Moreover, more successfully segmentation and masking can be obtained by training more different SAR images that are smoothed with SDD method. Multiple segmentation can also be applied in less running time and loss rates with Deep Mask R-CNN.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] B. Sen, M. Peker, A. Cavusoglu, and F. V. Celebi, “A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms”, *Journal of Medical Systems*, vol. 38, 2014. DOI: 10.1007/s10916-014-0018-0.

[2] U. Yuzgec, Y. Becerikli, and M. Turker, “Dynamic neural-network-based model-predictive control of an industrial baker’s yeast drying process”, *IEEE Transactions on Neural Networks*, vol. 19, pp. 1231–1242, 2008. DOI: 10.1109/TNN.2008.2000205.

[3] 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, pp. 1867–1871, 2018. DOI: 10.1109/LGRS.2018.2864342.

[4] R. Wen, Ch.-B. Chng, and Ch.-K. Chui, “Augmented reality guidance with multimodality imaging data and depth-perceived interaction for robot-assisted surgery”, *Robotics*, vol. 6, no. 2, pp. 1–18, 2017. DOI: 10.3390/robotics6020013.

[5] Y. Duan, X. Tao, Ch. Han, X. Qin, and J. Lu, “Multi-scale convolutional neural network for SAR image semantic segmentation”, in *Proc. of 2018 IEEE Global Communications (Globecom)*, Abu Dhabi, Dec. 2018, pp. 1–6. DOI: 10.1109/GLOCOM.2018.8647657.

[6] P. Wang, H. Zang, and V. M. Patel, “SAR image despeckling using a convolutional neural network”, *IEEE Signal Processing Letters*, vol. 24, pp. 1763–1767, 2017. DOI: 10.1109/LSP.2017.2758203.

[7] C. Ozcan, B. Sen, and F. Nar, “Early-exit optimization using mixed norm despeckling for SAR images”, *in Proc. of 23nd Signal Processing and Communications Applications Conference (SIU)*, Malatya, 2015, pp 779–782. DOI: 10.1109/SIU.2015.7129944.

[8] S. Arisoy and K. Kayabol, “Mixture-based superpixel segmentation and classification of SAR images”, *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 11, pp. 1721–1725, 2016. DOI: 10.1109/LGRS.2016.2605583.

[9] T. Zhou, S. Ruan, and S. Canu, “A review: Deep learning for medical image segmentation using multi-modality fusion”, *Array*, vols. 3–4, pp. 1–11, 2019. DOI: 10.1016/j.array.2019.100004.

[10] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, “A review of semantic segmentation using deep neural networks”, *International Journal of Multimedia Information Retrieval*, vol. 7, pp. 87–93, 2017. DOI: 10.1007/s13735-017-0141-z.

[11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN”, in *Proc. of 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 2017, pp. 2980–2988. DOI: 10.1109/ICCV.2017.322.

[12] C. Ozcan, B. Sen, and F. Nar, “Fast feature preserving despeckling” in *Proc. of 22nd Signal Processing and Communications Applications Conference (SIU)*, Trabzon, 2014, pp. 1007–1010. DOI: 10.1109/SIU.2014.6830402.

[13] H. Xie, L. E. Pierce, and F. T. Ulaby, “SAR speckle reduction using wavelet denoising and Markov random field modeling”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 2196–2212, 2002. DOI: 10.1109/TGRS.2002.802473.

[14] C. Ozcan, B. Sen, and F. Nar, “Sparsity-driven despeckling for SAR images”, *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 115–119, 2016. DOI: 10.1109/LGRS.2015.2499445.

[15] R. Yayla and B. Sen, “Research on region-based convolutional neural network for semantic segmentation”, in *Proc. of 8th International Conference on Advanced Technologies (ICAT'19)*, Sarajevo, 2019, pp. 244–249.

[16] K. Radhakrishnan, C. Chaithanya, and S. Priya, “Multiple components detection in motherboard using Mask R-CNN”, *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 5, pp. 1764–1767, 2019.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, in *Proc. of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 2014, pp. 580–587. DOI: 10.1109/CVPR.2014.81.

[18] R. Girshick, “Fast R-CNN”, in *Proc. of 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1440–1448. DOI: 10.1109/ICCV.2015.169.

[19] S. Ren, K. He, R. Girshick, and J. Sun “Faster R-CNN: Towards real-time object detection with region proposal networks”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 2257, 2017. DOI: 10.1109/TPAMI.2016.2577031.

[20] *Mstar Public Clutter Dataset*. [Online]. Available: https://www.sdms.afrl.af.mil/index.php?collection=mstar&page=clutter

[21] A. Dutta and A. Zisserman “The VIA annotation software for images, audio and video”, in *Proc. of 27th ACM International Conference on Multimedia*, Nice, 2019, pp. 2276–2279. DOI: 10.1145/3343031.3350535.

[22] W. Abdulla, “Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow”, Mask R-CNN matterport implementation, 2017. [Online]. Available: https://github.com/matterport/Mask_RCNN

[23] F. V. Celebi, M. Yucel, H. H. Goktas, and K. Danisman, “Intelligent modelling of alpha (α) parameter; Comparison of ANN and ANFIS cases”, *Optoelectronics and Advanced Materials - Rapid Communications*, vol. 7, pp. 470–474, 2013.