DNN-PolSAR: Urban Image Segmentation and Classification using Polarimetric SAR based on DNNs

Soumyadip Sarkar[a], Farhan Hai Khan[a], Sayanti Dutta[a], Raisa Chatterjee[b], Tripti Kumari[c], Deeptendu Santra[a], Nirmalya Misra[a], AritroPal Choudhury[a], Shobhit Kumar[a], Tamesh Halder[d], Dipjyoti Paul[e], Sajal Sarkar[f], Rintu Kumar Gayen[a], Arundhati Misra Ray[g], Debashish Chakravarty[d]

[a]Institute of Engineering & Management, Kolkata, India, 700091
[b]Department of CSE, Jadavpur University, India, 700032
[c]Department of CSE, Indian Institute of Information Technology (IIIT), Ranchi, Jharkhand, India -834010
[d]Department of Mining Engineering, Indian Institute of Technology, Kharagpur, India, 721302
[e]Office H304, Department of Computer Science, University of Crete, Heraklion, Crete 70013
[f]Information Security Dept., Power Grid Corporation of India Ltd, Gurgaon, India, 122001
[g]Space Application Center, ISRO, India, 380015

Abstract. Synthetic Aperture Radar (SAR) image segmentation and classification is a popular technique for learning and detection of objects such as buildings, trees, monuments, crops water-bodies, hills, etc. SAR technique is being used for urban development and city-planning, building control of municipal objects, searching best locations, detection of changes in the existing systems, etc. using polarimetry based on Deep Neural Networks. In this paper, we proposed a technique for Urban Image Segmentation and Classification using Polarimetric SAR based on Deep Neural Networks (DNN-PolSAR). In our proposed DNN-PolSAR technique, we use Mask-RCNN, LinkNet, FPN, and PSPNet as model architectures, whereas ResNet50, ResNet101, ResNet152, and VGG-19 are used as backbone networks. We, first, apply polarimetric decomposition on airborne Uninhabited Aerial Vehicle Synthetic Aperture (UAVSAR) images of urban areas and then the decomposed images are fed to DNNs for segmentation and classification. We then simulate DNN-PolSAR considering different hyper-parameters and compare the obtained scores of hyper-parameters against used model architectures and backbone networks. In comparison, it is found that DNN-PolSAR based on FPN model with ResNet152 performed the best for segmentation and classification. The mean Average Precision (mAP) score of the DNN-PolSAR based on FPN with a pixel accuracy of 90.9% is 0.823, which outperforms *other Deep Learning models*.

Keywords: Polarimetric SAR, FPN, PSPNet, Mask-RCNN, LinkNet, Image Segmentation.

1 Introduction

Segmentation and classification of an image is a process of splitting and categorizing the image into different parts based on the predefined category of objects. In this process, each pixel in an image is categorized based on the predefined labels of objects. Image segmentation has historically been used primarily for recognizing scenes in which similar objects can be placed more accurately. However, recently image segmentation is being used in different fields such as medical imaging, autonomous driving, etc. very successfully. Therefore, image segmentation can also be used for satellite image and Polarimetric Synthetic Aperture Radar (PolSAR) of urban cover areas for categorization and analysis.[1,2] PolSAR is a very popular technique in remote sensing and it is being used in wide ranges of applications namely segregation and classification in GIS, remote sensing, etc. It is also used for mapping of areas such as forest, vegetation, urbanized areas, etc. Data generated from PolSAR provides SAR resolutions, which help to understand images in forms of scattering components such as surface scattering, volume scattering, helix scattering, double-bounce scattering, and wire scattering. Based on these scattering components, PolSAR helps to understand classification of objects. For example, it is seen that PolSAR generates more prominent helix and double-bounce scattering components for images of urban areas.[3–5] Hence, here in this work, we consider double-bounce scattering and helix scattering components for classification of objects in images of urban areas. With growth of urbanization and increasing population in urban areas, tracking, studying and analysis of the urban cover areas have become very essential, particularly in terms of locating and classifying objects such as buildings, crops water-bodies and hills. So, accurate locating and classification of different objects using images of urban areas is important for designing quick and reliable solutions.[6] However, urban image segmentation and classification is a very challenging task even using SAR polarimetry. Hence, it is more challenging task to segment and classify images of urban cover areas. This is because urban cover relatively shows small part of total surface. Fortunately, a huge collections of satellite imagery datasets is available freely which can be used for image segmentation and classification of urban cover areas. Image segmentation and classification of urban cover areas using PolSAR is very difficult task due to urban structures, whose orientation is not in line of sight (LoS) of the radar. However, recognition of such areas is important for a number of reasons such as disaster relief, urban planning, and environmental monitoring. But, it is not possible to feed the scattering of images of urban covers areas taken using the Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) into a neural network. This is because it is required to employ a set of decompositions of images in order to retrieve various information using scattering such as surface scattering, double bounce scattering, volume scattering, helix scattering, wire scattering, etc. It is also required to identify different areas such as grassland, urban areas, hills, etc. with different scattering information and components from the images, so the data obtained from scattering become significant. A scattering component allows us to determine what kind of area is captured in particular images. A grassland, for example, may have high values for surface scattering, while an urban area may have high values for both double bounce and helix scattering. With the application of these decomposition techniques to the UAVSAR raw scattering matrix elements, different areas tend to exhibit different characteristics, which can be used to perform image segmentation and classification. Based on the above mentioned principle, in this paper, we present a technique for Polarimetric SAR (PolSAR) image segmentation and classification of Radar Satellite Imagery of urban areas using Deep Neural Networks (DNNs) such as SPNet, LinkNet, FPN, and Mask-RCNN based on different backbone networks such as such as EffcientNet, DenseNet, MobileNet, Inception, ResNet, and VGG19 (discussed in Subsection 3.1). In our proposed technique, we first apply polarimetric decomposition on airborne Uninhabited Aerial Vehicle Synthetic Aperture (UAVSAR) images of urban areas and then the decomposed images are fed to DNNs for segmentation and classification. We simulate our proposed technique and accordingly obtain simulation results using different DNNs. The major contribution of this work are as follows:

- We introduce new architectures consisting of models and backbone networks for urban classification and perform rigorous evaluation of all the machine learning classifiers in the field of Remote Sensing. We propose a technique to identify and detect buildings, grassland, hills from the PolSAR images.

- Presented and described the best and most effective Deep Learning methods for PolSAR image segmentation and classification.

- We obtain simulation results of our proposed technique with different backbone networks such as EffcientNet, DenseNet, MobileNet, Inception, ResNet, and VGG19.

- We carried out an extensive comparison of the current state-of-the-art models on the same datasets for segmentation and classifications of urban area covers.

The reminder of the paper is organized as follows. Section 2 presents Related Work in the area of satellite image segmentation and classification. In Section 3, we discuss the architectures and backbones of the models, which are used in four different experimentation presented in our paper. An overview of the datasets used for training and validation is discussed in Section 4. In Section 5, results of experimentation based on different databases and discussion on the results are presented. Finally, we concluded our paper in Section 6.

2 Related Works

Segmentation of PolSAR images of urban cover areas poses unique challenges such as partial visibility of surfaces, different scattering, etc. However, only good thing is that area structure of the these images are well defined. But, design and development of algorithms for analysis and classifications of such as images requires focused research for opening possibilities for the application of Remote Sensing in various fields. In view of this, A numbers of works have been proposed for image analysis and classification using semantic segmentation. In this section, we briefly present and discuss some of the advancements in classification approaches for PolSAR as well some of the examples where architectures from a different area of study has been successfully applied in remote sensing. A recent study conducted by De et al[7] to build a Deep Learning based novel technique for classification of urban areas. The information in the augmented dataset used in this work is transformed using a stacked auto-encoder, before feeding it to a neural network for classification. This technique achieved an accuracy of 91.3%, which was an enhancement in performance as compared to the techniques present at that time. In,[8] Cui et al proposed an architecture comprising Dense Attention Pyramid Network (DAPN), Region Proposal Network (RPN) and a detection network for multi-scale ship detection in SAR images. Here, DAPN was used to extract multi-scale fused features for generating and detecting to use in the subsequent iterations of the technique. The top-down densely connected networks are used to get concatenated feature maps of lower layers. The proposed method provided an accuracy of 89.8%, which was 11% higher than the previous models on the SAR ship detection data set (SSSD). DAPN was also 20% more than the faster R-CNN.[9] They also show that the top-down pyramid structure with attention is very effective in obtaining the feature maps which contained more spatial and semantic information. Recently, Mohanty et al presented applications of Mask-RCNN[10] on the segmentation and detection of building on Google Maps Satellite Imagery Data. Authors found the results to be impressive with a fina loss value of 0.15 for the instance image segmentation model. A research Wang et al. explored the problems in the classification of PolSAR images due to the presence of nonlinear data.[11] This study proposed a kernel sparse representation based classification approach. This kernel function technique solves the problems caused by the nonlinear features. This helps in attaining more accurate results in the task of classification. This study used an Airborne SAR dataset from San Francisco, United States of America. In,[12] Femin et. al. proposed an approach for detecting buildings using CNNs from satellite images. In this work, different building footprints from the images were identified using CNN method. The proposed work was also detected different shapes and colors. The detection accuracy by this approach for building was found to be 83%. On the other hand, Wang et al. introduced a deep feature extraction approach,[13] where multilevel polarimetric feature vector is extracted using a PAO_PTD_CNN. The authors extracted superpixels using simple linear iterative clustering (SLIC) from the feature vector for classification map. Finally, the result is obtained combining the superpixel map and the deep feature classification vector with Kappa Score of 0.86. The authors of[14] mentioned that the semantic segmentation can also be implemented for high-resolution PolSAR images using neural network architecture such as MP-ResNET which contains three concurrent semantic embedding branches and uses a multi-scale feature fusion design in decoder to use each encoding branch. The authors noticed that MP-ResNet improves the aggregation of context information compared to baseline Fully Convolution Network (FCN). The suggested method based on MP-ResNet surpasses numerous state-of-the-art methods in all accuracy with a mean F1 of 92.25% and IoU of 89.60% in classification using the Gaofen Dataset. Zhao et. al. showed in[15] that segmentation can also be achieved using edge information based on spectral graph partitioning. Here, the authors defined segmentation as a three-part process namely edge information extraction, edge-based similarity matrix analysis, and Normalised Cut. This method overcame the pepper-salt phenomenon along with much more complete and the boundaries of the segments. The method of[16] by Ouahabi et. al. aimed to improve the segmentation efficiency without compromising the accuracy using Fully Convolution dense Dilated Network model. Here, the authors found that the Low resolution and contrast, shadow interference as well as differences in size and position of the abnormal tissue are the challenges that hinder the process of obtaining the segmentation of ultrasound images. Yuanyuan et al. in his work[17] explores how different classification algorithms are affected by the choice of polarimetric parameters such as Alpha, HAAlpha_T11, Shannon entropy, VanZyl3_Vol, Neuman_delta_mod, Barnes2_T33, Barnes1_T33, and entropy.

3 Backbone Network and Model Architecture

3.1 Backbone Networks

A backbone network is mainly used to extract network feature for classification of objects. Here, in this paper we have used ResNet152,[18] ResNet101,[19] ResNet50, and VGG-19[20] backbone networks for features extraction.

3.2 Model Architectures

In this Subsection, we explain model architectures such as M-RCNN, PSPNet, FPN, and LinkNet used for classification in our work. The model architectures classify the extracted features using the base model from the deep neural backbone networks discussed in 3.1.

3.2.1 MR-CNN

The M-RCNN[21] was developed as an extension to the Faster-RCNN[9] which has been widely used so far for various object detection purposes. The F-RCNN/M-RCNN as output yields an object's label along with the object's bounding box. F-RCNN uses a feature extractor block that extracts the features from the image. These features are then used to train the bounding box regressor and the classifier. The Mask RCNN as the name suggests extends F-RCNN by training a binary mask in parallel with the bounding box regressor and object classifier. The first stage of the Mask-RCNN (like the F-RCNN) is the Region Proposal Network (RPN). Each bounding box is paired with an objectness score that denotes the probability score of the object. The second stage of the M-RCNN is called the head of the network. In F-RCNN this head is generally a stack of convolution layers and a dense layer for bounding box regression. M-RCNN in parallel to this bounding box learning algorithm uses a stack of convolution layers for Mask representation. This parallel task makes it theoretically faster and more accurate than other segmentation models.

3.2.2 Feature Pyramid Network (FPN)

A FPN is a fully convolutional feature extractor that takes a single-scale image of any size as input and produces correspondingly sized feature maps at several layers.[22] The model comprises two distinctive parts such as a conventional convolutional network (like VGG-19 or ResNet50) that acts as a feature extractor and a deconvolutional network with compatible feature sizes. However, there is a crucial difference between these two parts: the convolutional network goes from bottom to top whereas the flow in the deconvolutional network goes from top to down. The blocks in the convolutional network are connected in the deconvolutional network by linear multiplication. The output of blocks in the deconvolutional layer are connected to individual convolution layers which are not directly connected. These layers are transformed into a stack of layers. This dataset undergoes some upsampling and activation to give us an image map.

3.2.3 LinkNet

The LinkNet is a lightweight network architecture designed for performing segmentation tasks with a special focus on processing time.[23] Instead of a typical auto-encoder style segmentation model where the spatial semantics are first extracted using encoder blocks and then the decoder uses this spatial information for spatial categorization. This method has a certain downside in terms of both computation and accuracy. The pooling and strided convolution used in encoders may result in some loss of spatial information. So instead, the LinkNet algorithm uses skip connections from one encoder block to the corresponding block to prevent the loss of information at each stage. This idea of semantic information preservation is very similar to an U-Net except in this case the results of the encoder are added to the results of the corresponding decoder block instead of performing feature concatenation. For experimentation, we will be using the model proposed in the original LinkNet paper. The model uses four encoder blocks and four corresponding decoder blocks. There are two special blocks of fully convolutional neural networks at the beginning and end of the network to preserve the dimensions of the image.

3.2.4 Pyramid Scene Parsing Network (PSPNet)

The PSPNet of[24] is a model used for semantic segmentation. Its specialty is that it uses a pyramid parsing module. Different region-based context aggregation is used by this module to exploit global context information. The final predictions are made more reliable due to the presence of local and global clues together. Given an input image, the feature map can be extracted using a pre-trained CNN, using a dilated network strategy. The final size of the feature map is reduced to 1/8th of the input image. A pyramid pooling module is then applied on the top of the map, for gathering context information. A four-level pyramid is used where the pooling kernels cover the whole, half of, and small parts of the image. The results from the pooling kernels are then concatenated to form a global prior. In the next step this prior is concatenated to the original feature map. The obtained result is finally passed through a stack of convolutional layers to generate the final prediction.

Table 1 Decomposition Method and Polarimetric Parameter

Decomposition Method	Polarimetric Parameter		
Cloude[25]	Cloude_T11	Cloude_T22	Cloude_T33
H/A/Alpha[26]	Entropy H/A/A_T11	Anisotropy H/A/A_T22	Shannon Entropy H/A/A_T33
VanZyl3[26]	VanZyl3_Vol	VanZyl3_Odd	VanZyl3_Dbl
Neuman[27]	Neuman_delta_mod	Neuman_delta_pha	Neuman_tau
FreeMan2[28]	FreeMan2_Vol	FreeMan2_Ground	
FreeMan[29]	FreeMan_Vol	FreeMan_Odd	Freeman_Dbl
Huyen[30]	Huyen_T11	Huyen_T22	Huyen_T33
Bhattacharya[31]	Frey_Dbl	Frey_Hlx	Frey_Odd
Singh[32]	Singh_6SD1	Singh_G4U2_Vol	Singh_G4U2_Odd
Barnes1[33]	Barnes1_T11	Barnes1_T22	Barnes2_T33
Barnes2[33]	Barnes2_T11	Barnes2_T22	Barnes2_T33
Pauli[25]	Pauli_a	Pauli_b	Pauli_c
Holm1[34]	Holm1_T11	Holm1_T22	Holm1_T33
Holm2[34]	Holm2_T11	Holm2_T22	Holm2_T33
Arri3_NNED[35]	Arii_NNED_Vol	Arii_NNED_Odd	Arii_NNED_Dbl
An_Yang3[36]	An_Yang3_Vol	An_Yang3_Odd	An_Yang3_Dbl
An_Yang4[37]	An_Yang4_Vol	An_Yang4_Odd	An_Yang4_Dbl
Yamaguchi3[38]	Yamaguchi3_Vol	Yamaguchi3_Odd	Yamaguchi3_Dbl
Yamaguchi4[39]	Yamaguchi3_Vol	Yamaguchi3_Odd	Yamaguchi3_Dbl

4 Datasets

Datasets play an important and crucial role in any machine learning algorithms for segmentation and classification. In our proposed technique too datasets play major roles in segmentation of classification of images of urban cover areas. We have mentioned in Section 1 that a huge collections of satellite imagery datasets is available for image segmentation and classification of urban cover areas. Therefore, in order to train our proposed algorithm, we have used PolSAR images of Lancaster, Palmdale and Rosamond city from airborne UAVSAR. However, we have considered only building classes for semantic and instance segmentation from these datasets using Deep Learning over various polarimetric decompositions. It is also to be mentioned that as similar of,[17] we have used different polarization parameters such as Alpha, HAAlpha_T11, Shannon en-

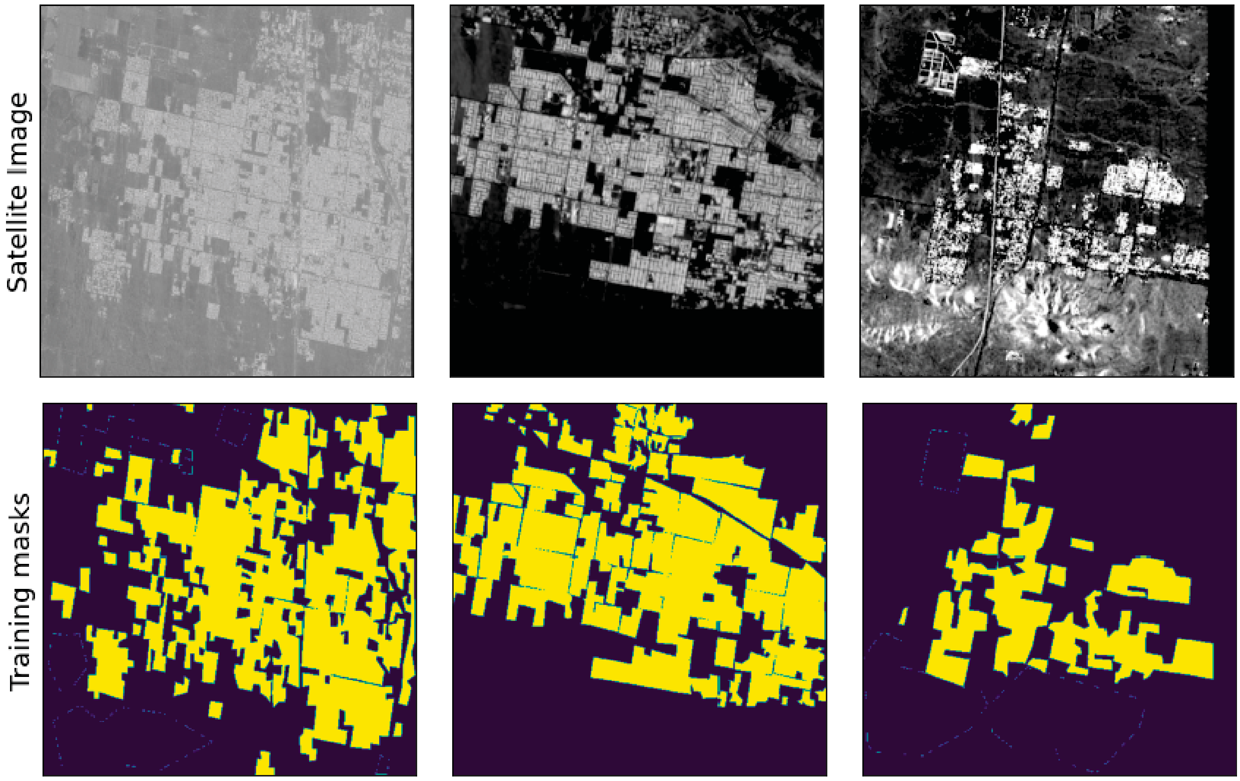

Fig 1 Sample Datasets and their Corresponding Masks for Urban Areas.

Table 2 Description of the Datasets

Location	Coordinates		Region/Country	Datasets	
	Latitude	Longitude		Train	Test
Lancaster	40.037° N	76.305° W	Pennsylvania, USA	71	33
Rosamond	34.8641° N	118.1634° W	Karen County, California, USA	60	64
Palmdale	34.3452° N	118.62° W	Los Angeles, California, USA	50	17

tropy, VanZyl3_Vol, Neuman_delta_mod, Barnes2_T33, Barnes1_T33, and entropy to improve the classification accuracy of our proposed technique. We use PolSARPro v6.0 Software Suite[40] for decomposition results in our proposed work. In Table 1, we shows all the decomposition methods and corresponding polarimetric parameters thoes were applied on the datasets with our proposed technique. It is required to be mentioned that we also performed image augmentation using random rotation and image flipping to generate more data before passing them through the model. We generated 3 transformed images from each image with image size of $1331 \times 1101 \times 3$ for enhancing our datasets. The enhanced datasets are used for training based on PolSAR images of Lancaster, Palmdale, and Rosamond cities of USA. The reason of usage of datasets from different cities for increasing segmentation accuracy by introducing variance in the datasets. Detail of used datasets are given in Table 2. It can be seen from Table 2 that there are 71, 60, and 50 training datasets for Lancaster, Palmdale, and Rosamond respectively. But, we used 33, 64, and 17 test datasets for Lancaster, Palmdale, and Rosamond respectively. Figure 1 shows sample datasets of satellite images (upper of parts of the Figure) as well as corresponding masks (the lower parts of the Figure) of the satellite images. From Figure 1, it can be seen that our proposed technique based on used datasets correctly segmented and classified urban areas from the images. In the lower parts of the Figure 1, the yellow color masks represent the presence of urban areas for the given satellite images. Details of results with our proposed technique and discussion on the results are given in the following Section 6.2.

5 Proposed Technique

Proposed technique to be discussed providing a block diagram of all important steps. The block diagram of out proposed scheme is shown in Figure 2.

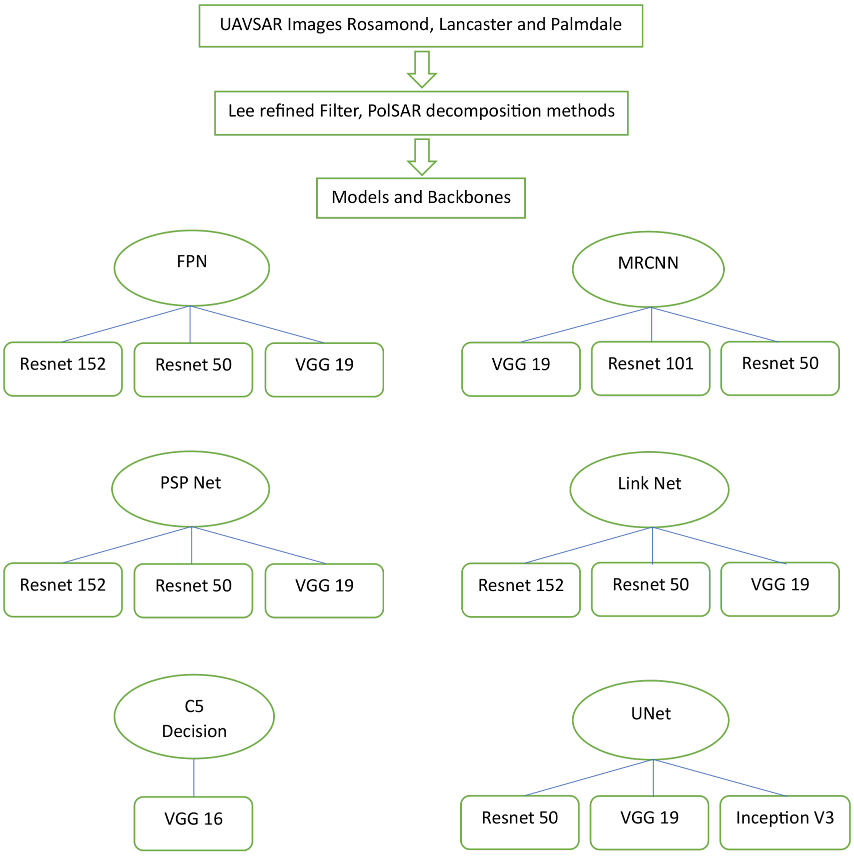

Fig 2 Block Diagram of Our Proposed Scheme

6 Simulation Studies

In this Section, we have presented simulation of our proposed technique and discussed on obtained results. It is to be mentioned that the main motivation of our work is to understand the learning capacity and rate of convergence against the above mentioned architectures with different backbone networks. In order to do that we have considered four different architectures as well as four different backbone networks to obtain unbiased results. We have used *Intersection Over Union (IoU) score*, *Pixel Accuracy*, *F1 Score*, *Cohen's Kappa Score*, *Area Under the Curve*, *Recall*, *Precision*, and *Mean Average Precision(mAP)* hyper-parameters as performance metrics to obtain simulation results. We have also discussed these metrics to draw performance comparison of different architectures and backbone networks.

Table 3 Pixel Accuracy, Cohen's Kappa Score, and IoU Score of all Considered Model Architectures and Backbone Networks

Model Architecture	Backbone Network	Pixel Accuracy	Cohen's Kappa Score	IoU Score
FPN	**ResNet152**	0.909	**0.806**	**0.799**
	ResNet50	0.901	0.786	0.774
	VGG-19	0.909	0.805	0.796
MR-CNN	**ResNet101**	0.897	0.781	0.780
	ResNet50	0.638	0.057	0.069
	VGG-19	0.811	0.621	0.667
PSPNet	**ResNet152**	0.885	0.755	0.753
	ResNet50	0.895	0.771	0.758
	VGG-19	0.893	0.772	0.763
LinkNet	**ResNet152**	**0.910**	0.805	0.791
	ResNet50	0.464	0.127	0.419
	VGG-19	0.715	0.461	0.573
Unet	**VGG-19**	**0.918**	0.54	0.48
	InceptionV3	0.815	0.03	0.12
	ResNet50	0.947	0.54	0.67
C5	**VGG-16**	0.717	0.39	0.44

6.1 Simulation Environment

We have simulated our proposed technique using (Python3 on a Kaggle, Colab notebook and R language in a computer with 64 GB RAM. Simulated the proposed technique is conducted using model architectures namely MR-CNN, FPN, LinkNet, PSPNet, C5 decision tree and Unet against the ResNet152, ResNet101, ResNet50, Inception V3, VGG16 and VGG-19 backbones networks. However, we provided results of best performing backbone networks for each considered model architectures.

6.2 Results and Discussion

In Table 3, we presented simulation results of Pixel Accuracy, Cohen's Kappa Score, and IoU Score hyper-parameters for all the considered model architectures and backbone networks. The Unet and C5 based our old work also compared here.[41] On the other hand Table 4 presents simulation results of Mean Average Precision, Area Under the Curve, Recall, Precision, and F1 Score hyper-parameters for all the considered model architectures and backbone networks. From Table 3, it can be seen that Cohen's Kappa Score and IoU Score for FPN with ResNet152 is highest compared to all other scores. Similarly, Table 4 shows that mAP, AuC and F1 Scores for FPN with ResNet152 and VCG19 are highest respectively. Therefore, it may be concluded that the FPN gives the best accuracy among all models proposed in this paper. The high F1 score and AuC scores for the top three models confirm that the FPN architecture performs best among all other architectures. It gives pixel accuracy above 90% for three backbones, namely ResNet152, VGG-19, and ResNet50. On the other hand, it can be seen from Table 3 that the Pixel Accuracy of 91%

Table 4 mAP, AuC, Recall, Precision, and F1 score all Considered Model Architectures and Backbone Networks

Model Architecture	Backbone Network	mAP	AuC	Recall	Precision	F1 Score
FPN	**ResNet152**	**0.823**	0.965	0.917	0.850	0.882
	ResNet50	0.808	0.963	0.879	0.861	0.870
	VGG-19	0.817	**0.968**	0.928	0.843	**0.884**
MR-CNN	**ResNet101**	0.809	0.950	0.897	0.839	0.867
	ResNet50	0.394	0.634	0.075	0.647	0.134
	VGG-19	0.672	0.937	0.973	0.671	0.794
PSPNet	**ResNet152**	0.768	0.960	0.934	0.795	0.859
	ResNet50	0.784	0.961	0.896	0.835	0.865
	VGG-19	0.788	0.957	0.907	0.824	0.864
LinkNet	**ResNet152**	0.822	0.966	0.895	**0.868**	0.881
	ResNet50	0.419	0.618	**1.000**	0.411	0.583
	VGG-19	0.579	0.890	0.968	0.570	0.718

for LinkNet with ResNet152 is highest. But Table 4 shows that the values of Recall and Precision are highest for LinkNet with ResNet50 and ResNet152 respectively. The pixel accuracy, cohen's kappa, IoU score, AuC, Recall Precision, and F1 Score of MR-CNN with ResNet101 is better compared to other backbone networks. Based on these values of parameters, it can be inferred that the MR-CNN model architecture is the largest model used here in terms of the number of trainable parameters and consequently this architecture takes more time to train the system considering all used images than other considered model architectures. Finally, the Pixel Accuracy with ResNet50 and Cohen's Kappa Score and IoU Score with VGG19 for PSPNet are better compared to other two backbone networks. The values of AuC with ResNet50, Recall with ResNet152 Precision and F1 Score with ResNet50 are best. It can specifically be inferred about PSPNet that the PSPNet architecture performs well with VGG-19, ResNet50 and ResNet152 respectively as its backbone networks. The UNet and C5 decision trees results can be found at Halder1 which have lower performance than FPN. The prediction masks with our proposed technique are given in Figure 3, 4 and 5. It can be seen from Figures that the top three performing model architectures are FPN with ResNet152, M-RCNN with ResNet101, and PSPNet with VGG19. This is because as we have seen that values of considered parameters for FPN are highest with ResNet152, for M-RCNN are better with ResNet101 and for PSPNet are also best with VGG19. The pixel accuracy, cohen's kappa, IoU score, AuC, Recall Precision, and F1 Score of MR-CNN with ResNet101 is better compared to other backbone networks. Finally, we present the convergence analysis of our proposed technique. We shows the convergence of the model architectures against each considered backbone networks. The CLAHE, Gaussian Blurr and different types of augmentation like translation, Rotation, Flipping have been done. Prediction by FPN With Resnet50, ResNet152, VGG192 as backbone FPN with ResNet152 is highest compared to all other scores. Similarly, Table 4 shows that mAP,AuC and F1 Scores for FPN with ResNet152 and VCG19 are highest respectively. Therefore, it may be concluded that the FPN gives the best accuracy among all models proposed in this paper.The high F1 score and AuC scores for the top three models confirm that the FPN architecture per-forms best among all other architectures. It gives pixel accuracy above 90for three backbones, namely ResNet152, VGG-19, and ResNet50. The Pixel Accuracy of 91% for LinkNet with ResNet152 is

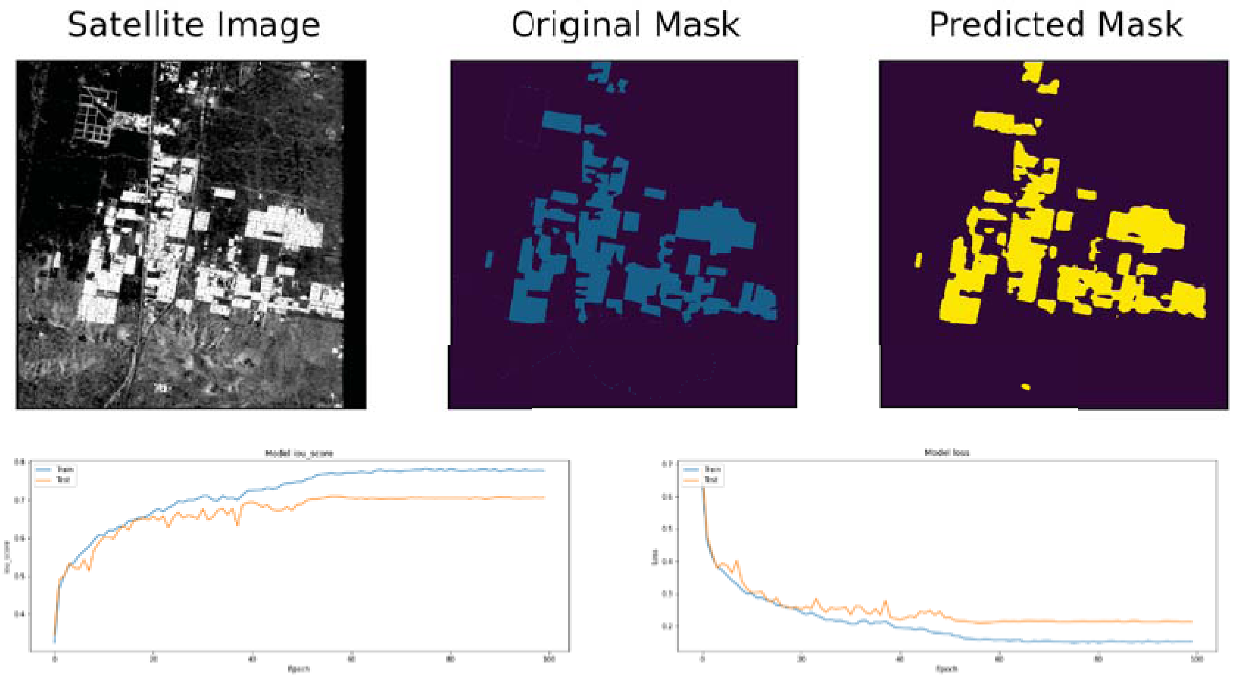

Fig 3 Results of prediction by FPN with ResNet-152 backbone.

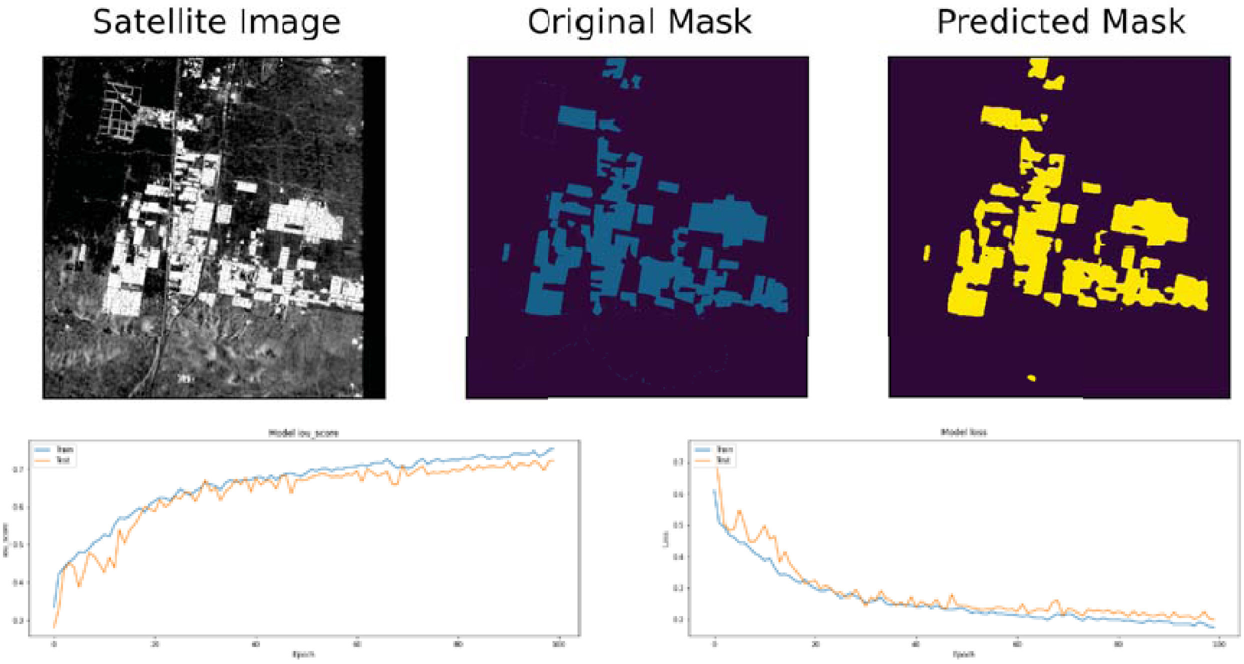

Fig 4 Results of prediction by PSPNet with VGG-19 backbone.

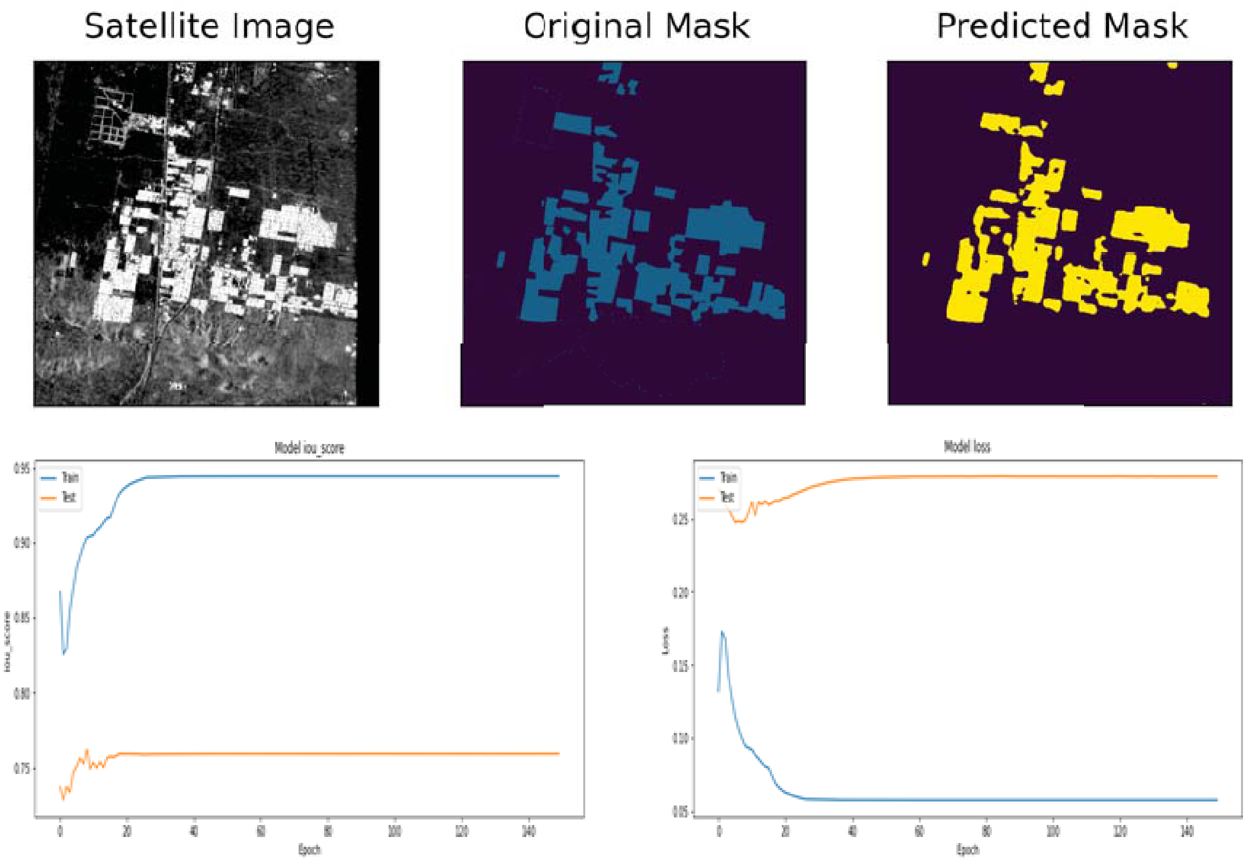

Fig 5 Results of prediction by MRCNN with ResNext-101 backbone.

highest.The values of Recall and Precision are highest for LinkNet with ResNet50 and ResNet152 respectively. The Pixel Accuracy with ResNet50 and Cohen's Kappa Score and IoU Score with VGG19 for PSP-Net are better compared to other two backbone networks. The values of AuC with ResNet50, Recall with ResNet152 Precision and F1 Score with ResNet50 are best. It can specifically be inferred about PSPNet that the PSPNet architecture performs well with VGG-19, ResNet50 and ResNet152 respectively as its backbone networks. We also present a graph plotting the convergence time for all model architectures against each backbone networks in Figure Linknet has lowest precision so the prediction mask is omitted.

7 Conclusion

This paper presents an image segmentation and classification technique of urban cover areas using Polarimetric SAR (PolSAR) which works based on Deep Neural Networks (DNNs) such as PSPNet, LinkNet, FPN, and Mask-RCNN. Here, we first applied polarimetric decomposition on airborne Uninhabited Aerial Vehicle Synthetic Aperture (UAVSAR) images of urban areas and then the decomposed images are fed into DNNs for segmentation and classification. Four different experimentations are carried out using four different databases and models such as PSPNet, LinkNet, FPN, and Mask-RCNN and then results obtained from the experimentations are compared with different backbone networks such as ResNet152, ResNet101, ResNet50, and VGG19. In comparison, it is seen that the FPN model with the ResNet152 as a backbone network obtained the best results on considered performance metrics such as mean Average Precision Score (mAP)

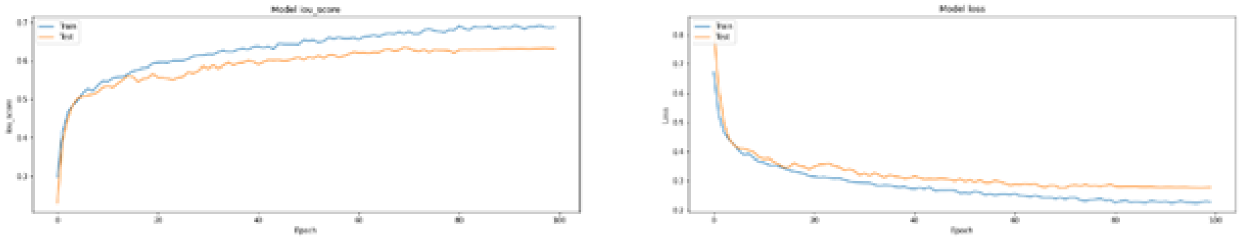

Fig 6 FPN with resnet50 backbone

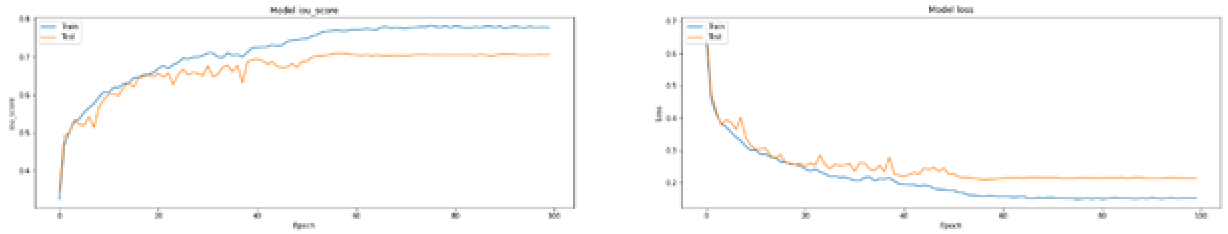

Fig 7 FPN with resnet152 backbone

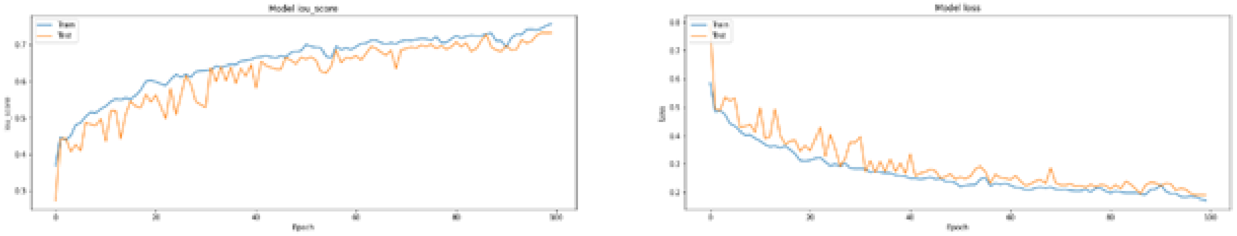

Fig 8 FPN with VGG192 backbone

and pixel accuracy. Specifically, it achieves the pixel accuracy of 90.9% and the mAP score of 0.823 and outperforms other Deep Learning models. In the future, the authors would like to explore for integrating the proposed technique for change detection and classification of multi-class objects in the domain of image processing.

References

1 Z. Niu, G. Hua, X. Gao, *et al.*, "Context aware topic model for scene recognition," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2743–2750, IEEE (2012).

2 S. A. Taghanaki, K. Abhishek, J. P. Cohen, *et al.*, "Deep semantic segmentation of natural and medical images: a review," *Artificial Intelligence Review* **54**(1), 137–178 (2021).

3 T. Zhou, Z. Li, and J. Pan, "Multi-feature classification of multi-sensor satellite imagery based on dual-polarimetric sentinel-1a, landsat-8 oli, and hyperion images for urban land-cover classification," *Sensors* **18**(2), 373 (2018).

4 S.-W. Chen and C.-S. Tao, "Polsar image classification using polarimetric-feature-driven deep convolutional neural network," *IEEE Geoscience and Remote Sensing Letters* **15**(4), 627–631 (2018).

5 Y. Zhang, J. Zhang, X. Zhang, *et al.*, "Land cover classification from polarimetric sar data based on image segmentation and decision trees," *Canadian Journal of Remote Sensing* **41**(1), 40–50 (2015).

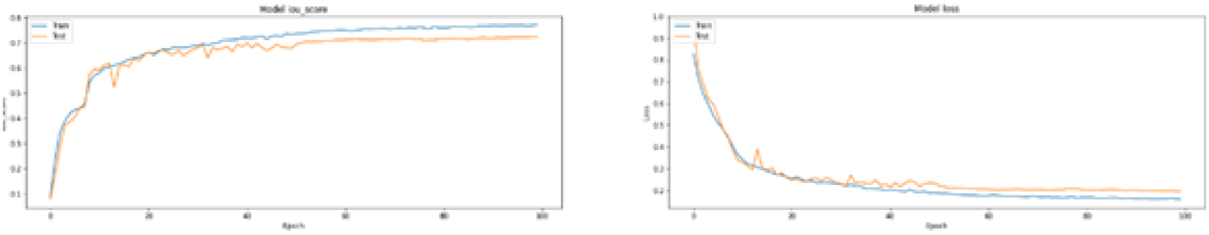

Fig 9 Results of prediction by linknet with inception3 as backbone.

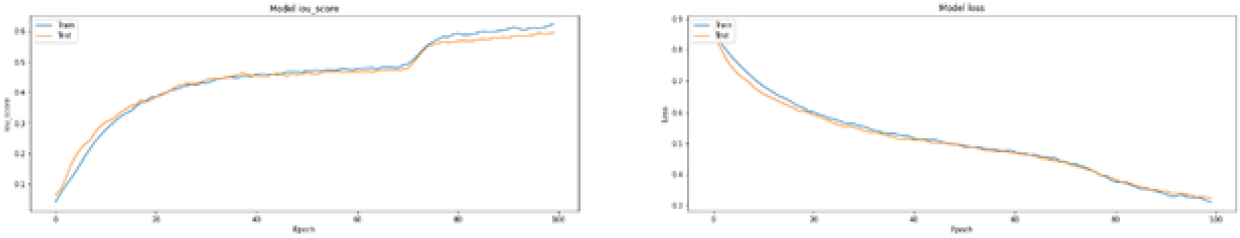

Fig 10 Results of prediction by linknet with mobilenet as backbone.

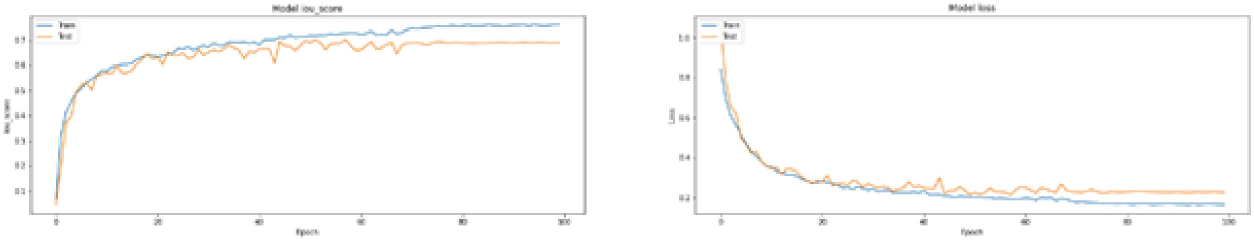

Fig 11 Results of prediction by linknet with Resnet152 as backbone.

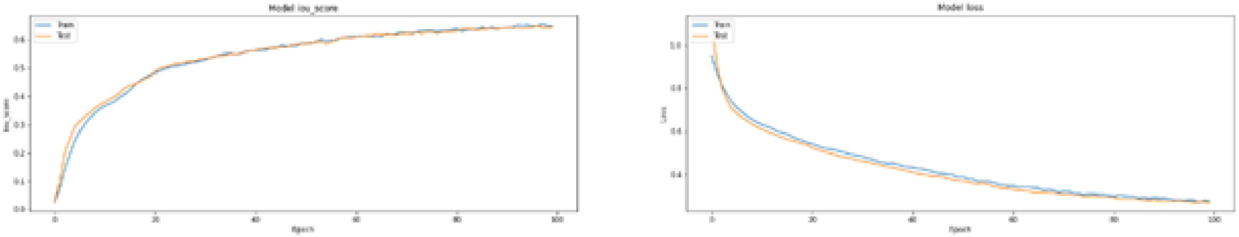

Fig 12 Results of prediction by link-net with efficient-net as backbone.

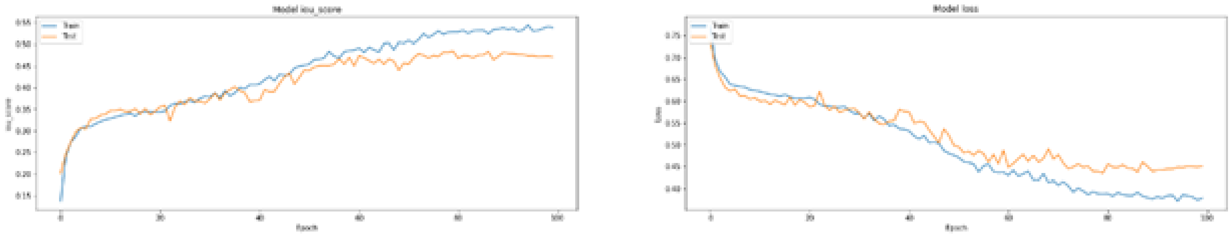

Fig 13 results of prediction by linknet with VGG19 as backbone.

6 S. De, L. Bruzzone, A. Bhattacharya, *et al.*, "A novel technique based on deep learning and a synthetic target database for classification of urban areas in polsar data," *IEEE Journal of*

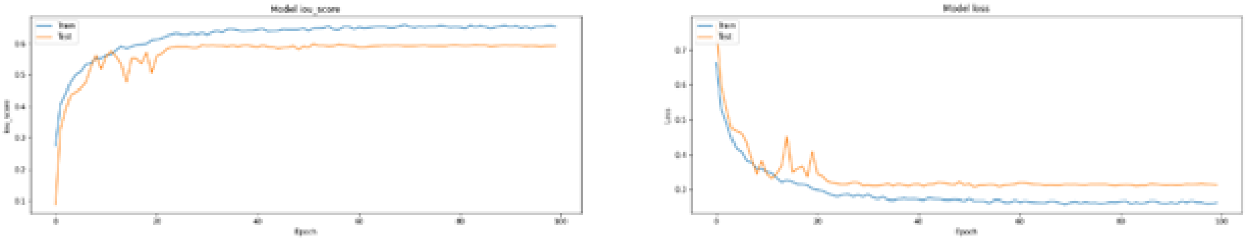

Fig 14 Results of prediction by Pspnet with inception2 as backbone.

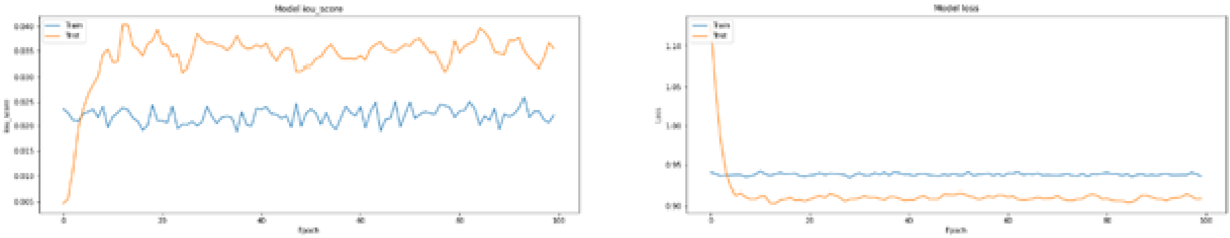

Fig 15 Results of prediction by Pspnet with mobilenet as backbone.

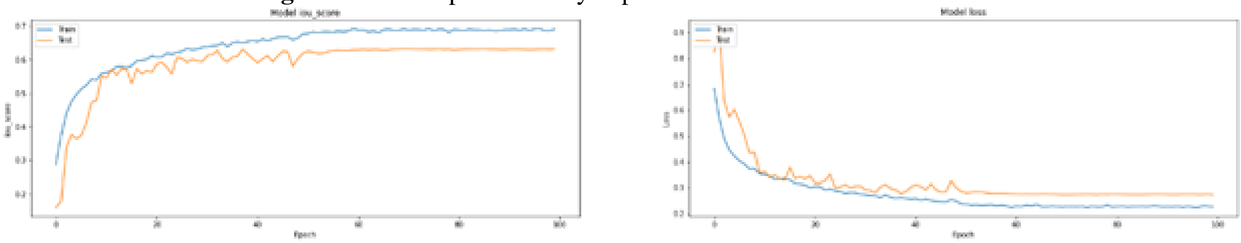

Fig 16 Results of prediction by pspnet with Resnet50 as backbone.

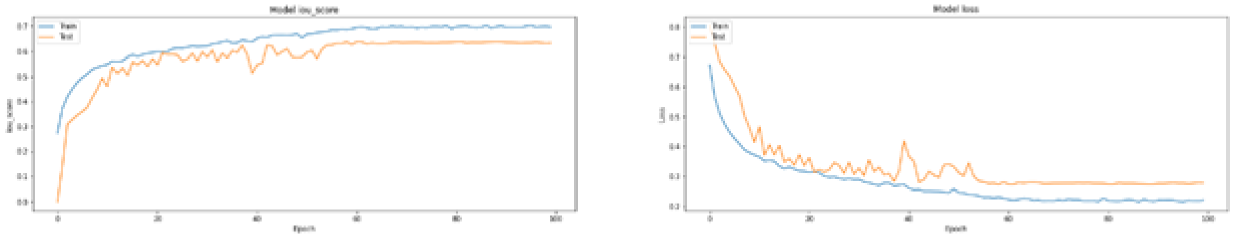

Fig 17 Results of prediction by pspnet with resnet152 as backbone.

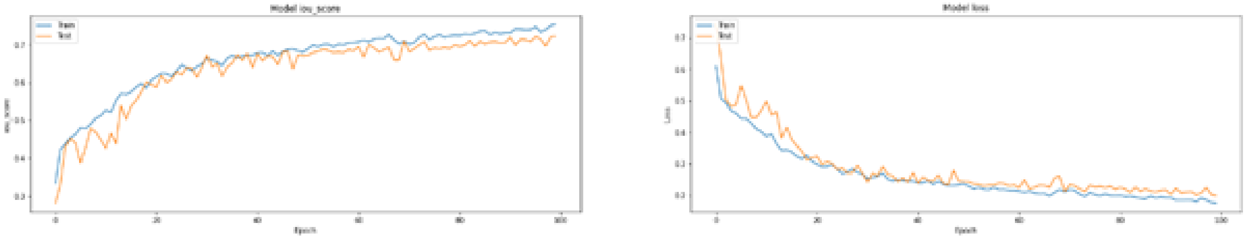

Fig 18 results of prediction by PSPNet with VGG19 as backbone.

Selected Topics in Applied Earth Observations and Remote Sensing **11**(1), 154–170 (2017).

7 S. De, L. Bruzzone, A. Bhattacharya, *et al.*, "A novel technique based on deep learning and a synthetic target database for classification of urban areas in polsar data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **11**(1), 154–170 (2018).

8 Z. Cui, Q. Li, Z. Cao, *et al.*, "Dense attention pyramid networks for multi-scale ship detection

in sar images," *IEEE Transactions on Geoscience and Remote Sensing* **57**(11), 8983–8997 (2019).

9 S. Ren, K. He, R. Girshick, *et al.*, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems* **28**, 91–99 (2015).

10 S. P. Mohanty, J. Czakon, K. A. Kaczmarek, *et al.*, "Deep learning for understanding satellite imagery: An experimental survey," *Frontiers in Artificial Intelligence* **3**, 85 (2020).

11 X. Wang, L. Zhang, B. Zou, *et al.*, "Polarimetric sar image classification based on kernel sparse representation," in *Compressive Sensing VII: From Diverse Modalities to Big Data Analytics*, **10658**, 106580L, International Society for Optics and Photonics (2018).

12 A. Femin and K. Biju, "Accurate detection of buildings from satellite images using cnn," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 1–5, IEEE (2020).

13 X. Wang, Z. Cao, Z. Cui, *et al.*, "Polsar image classification based on deep polarimetric feature and contextual information," *Journal of Applied Remote Sensing* **13**, 1 (2019).

14 L. Ding, K. Zheng, D. Lin, *et al.*, "Mp-resnet: Multipath residual network for the semantic segmentation of high-resolution polsar images," *IEEE Geoscience and Remote Sensing Letters* , 1–5 (2021).

15 L. Zhao and E. Chen, "Segmentation and classification of polsar data using spectral graph partitioning," in *MIPPR 2013: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications*, **8921**, 89210E, International Society for Optics and Photonics (2013).

16 A. Ouahabi and A. Taleb-Ahmed, "Deep learning for real-time semantic segmentation: Application in ultrasound imaging," *Pattern Recognition Letters* **144**, 27–34 (2021).

17 Y. Chen, X. He, J. Wang, *et al.*, "The influence of polarimetric parameters and an object-based approach on land cover classification in coastal wetlands," *Remote Sensing* **6**(12), 12575–12592 (2014).

18 K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).

19 S. Xie, R. Girshick, P. Dollár, *et al.*, "Aggregated residual transformations for deep neural networks," *arXiv preprint arXiv:1611.05431* (2016).

20 K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).

21 P. Burlina, "Mrcnn: A stateful fast r-cnn," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 3518–3523 (2016).

22 T.-Y. Lin, P. Dollár, R. Girshick, *et al.*, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125 (2017).

23 A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, 1–4, IEEE (2017).

24 H. Zhao, J. Shi, X. Qi, *et al.*, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2881–2890 (2017).

25 S. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Transactions on Geoscience and Remote Sensing* **34**(2), 498–518 (1996).

26 E. Pottier and J.-S. Lee, "Application of the h/a/alpha polarimetric decomposition theorem for unsupervised classification of fully polarimetric sar data based on the wishart distribution," in *SAR workshop: CEOS Committee on Earth Observation Satellites*, **450**, 335 (2000).

27 M. Neumann, L. Ferro-Famil, and E. Pottier, "A general model-based polarimetric decomposition scheme for vegetated areas," in *Proceedings of the 4th International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (ESRIN), Frascati, Italy*, 26–30, Citeseer (2009).

28 A. Freeman, "Fitting a two-component scattering model to polarimetric sar data from forests," *IEEE Transactions on Geoscience and Remote Sensing* **45**(8), 2583–2592 (2007).

29 A. Freeman and S. Durden, "A three-component scattering model for polarimetric sar data," *IEEE Transactions on Geoscience and Remote Sensing* **36**(3), 963–973 (1998).

30 J. R. Huynen, "Stokes matrix parameters and their interpretation in terms of physical target properties," in *Polarimetry: Radar, infrared, visible, ultraviolet, and X-ray*, **1317**, 195–207, International Society for Optics and Photonics (1990).

31 A. Bhattacharya, A. Muhuri, S. De, *et al.*, "Modifying the yamaguchi four-component decomposition scattering powers using a stochastic distance," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **8**(7), 3497–3506 (2015).

32 G. Singh and Y. Yamaguchi, "Model-based six-component scattering matrix power decomposition," *IEEE Transactions on Geoscience and Remote Sensing* **56**(10), 5687–5704 (2018).

33 R. Barnes, "Roll-invariant decompositions for the polarization covariance matrix," in *Proceedings of the Polarimetry Technology Workshop, Redstone Arsenal, AL, USA*, **1618** (1988).

34 W. A. Holm and R. M. Barnes, "On radar polarization mixed target state decomposition techniques," in *Proceedings of the 1988 IEEE National Radar Conference*, 249–254, IEEE (1988).

35 M. Arii, J. J. van Zyl, and Y. Kim, "Adaptive model-based decomposition of polarimetric sar covariance matrices," *IEEE Transactions on Geoscience and Remote Sensing* **49**(3), 1104–1113 (2011).

36 W. An, Y. Cui, and J. Yang, "Three-component model-based decomposition for polarimetric sar data," *IEEE Transactions on Geoscience and Remote Sensing* **48**(6), 2732–2739 (2010).

37 W. An, C. Xie, X. Yuan, *et al.*, "Four-component decomposition of polarimetric sar images with deorientation," *IEEE Geoscience and Remote Sensing Letters* **8**(6), 1090–1094 (2011).

38 Y. Yamaguchi, G. Singh, C. Yi, *et al.*, "Comparison of model-based four-component scattering power decompositions," in *Conference Proceedings of 2013 Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, 92–95, IEEE (2013).

39 Y. Yamaguchi, T. Moriyama, M. Ishido, *et al.*, "Four-component scattering model for polarimetric sar image decomposition," *IEEE Transactions on Geoscience and Remote Sensing* **43**(8), 1699–1706 (2005).

40 E. Pottier, F. Sarti, M. Fitrzyk, *et al.*, "Polsarpro-biomass edition: The new esa polarimetric sar data processing and educational toolbox for the future esa & third party fully polarimetric sar missions," in *ESA Living Planet Symposium 2019*, (2019).

41 T. Kumari, F. H. Khan, T. Halder, *et al.*, "Semantic segmentation of urban areas in polarimetric sar imaging using deep neural networks and decision trees," in *2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS)*, 479–482 (2021).