ACN-UNET Model for Semantic Segmentation

Atul Agarwal
Vellore Institute of Technology, Vellore
atul.agarwal2019@vitstudent.ac.in

Hemaksh Chaturvedi
Vellore Institute of Technology, Vellore
hemakshamit.chaturvedi2019@vitstudent.ac.in

Manav Nanwani
Vellore Institute of Technology, Vellore
manavanand.nanwani2019@vitstudent.ac.in

Akila Victor
Vellore Institute of Technology, Vellore
akilavictor@vit.ac.in

***Abstract*—The amount of cargo being transported using waterways is continually growing. As the number of ships rises, incidents at sea such as ecologically disastrous ship accidents, illegal fishing, piracy, drug smuggling, and illegal cargo movement become more prevalent. As a result, several agencies, ranging from environmental protection organizations to insurance companies and national government authorities, have been bound to pay closer attention to the open seas. The requirement for Ship Detection arises as a result of this cause. We utilize satellite photos to keep track of traffic on the sea since they provide more coverage.**

Seeing the wide scale need and requirement of automatic detection of ships in oceans and see, in this paper, we have proposed a novel approach. The proposed methodology is a modification of the UNET model called ACN-UNET, consisting of four layers at each step, giving an output of a segmented image in the satellite image.

Keywords—Ship Detection; Image Processing; Semantic Segmentation; Airbus Ship Detection Dataset; Deep Learning; ACN-UNET Algorithm; Convoluted Neural Networks

I. Introduction

Detection of ships has been a key-point in understanding and regulating the traffic at sea. Because of this many different algorithms have been tried and tested to get the desired result, the authors Fei Wu, Zhiqiang Zhou , Bo Wang, and Jinlei Ma have suggested a densely connected multiscale neural network (DCMSNN) based on Faster-RCNN [1] which annotates the ships present in the optical satellite images giving them an accuracy of 96.7%.

Convolutional neural networks (CNNs) have been the dominant technique for object identification, classification, and segmentation as computer technology and deep learning have improved [2]–[4]. The performance of features extracted using neural networks is superior to those extracted by hand .Many deep learning-based detection techniques have quickly emerged in recent years [5-6].

In this paper we have approached the problem by making a modification to the UNET model. Originally the UNET model consists of two layers of Convolutional Layers for encoding or decoding at each level[14]. The encoder helps us in understanding the 'what' part of the problem in the image whereas the decoder helps us understand the 'where' part of the problem in the image. In our proposed model we have included part of the decoder operation in the encoder side as you can see in Figure 5, which helps the model to figure out the 'what' and the 'where' simultaneously.

The dataset that we have used to implement our proposed architecture is the 'Airbus Ship Detection', present on kaggle[15] which holds more than 2,00,000 images, each having a size of 768 x 768 pixels.

II. LITERATURE REVIEW

We will discuss and implement a modification to the UNET model to detect ships in satellite images. The use of these images to detect ships has been studied under image segmentation as well as object detection problem statements. In a research [1], to achieve both multiscale and multi-scene SAR ship detection, they have suggested the use of a densely connected multiscale neural network (DCMSNN) based on Faster-RCNN. This is further divided into two subnets, first being the RPN (region proposal subnetwork) and the second is the detection subnetwork. This is particularly used to classify the fused map into regression or classification based on respective anchors. In [7], a lightweight network based on the SSD structure was developed specifically for SAR images. They further offered two feature representation optimization modules based on LSSD to improve detection accuracy. A bi-directional feature fusion module for semantic aggregation and feature reuse was also proposed. A high resolution SAR Images Dataset was used in [8] for ship detection and instance segmentation. Their model, HRSID, was developed for CNN-based detectors, which eliminated the weakness of the previous SAR ship dataset. A large-scale SAR imagery was used to test the migratory ability of the model. This basically shows the superiority of HRSID over SSDD.

A combination of Dense Feature Pyramid Network for feature fusion and Rotation Region Detection Network for prediction are used as a framework in [9]. This model also features reuse and ensures the effectiveness of detecting multi-scale objects. In [10] they have used PolSAR Images for ship detection using deep convolutional neural networks. Their model consists of four main processes: preprocessing, DCNN based sea-coast-ship classifier, modified Faster T-CNN ship detector, and then target fusion. After testing on 4 different dataset, they concluded that their model could generate proposals of different size from multi-level feature

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

maps as compared to the conventional R-CNN. In [11], a neural network DCNet trained on 6000 images collected from Doning port, China, the authors demonstrated the prediction of coordinates and classes of bow, cabin and stern on the ships. They were also able to detect the ships in different weather conditions.

The same dataset as ours was used by [12]. The authors implemented 2 different models for image segmentation and classification. The first one included Keras Unet and the second model was developed using FastAI Unet. Their model was better equipped to deal with low resolution images. The weakness of their models was that it required larger processing time and also their model was not trained to its full capacity. Other methods included the use of You Only Look Once(YOLO) Algorithm [13] which worked by combining the target area prediction and target category prediction as a regression problem. The next one [15] also used the same dataset from Airbus Ship Detection but their first process was to build a classifier to separate images consisting of ships from the images not consisting of ships. The second step involved the use of Unet with ResNet 18 encoder for image segmentation and object detection.

III. DATA PREPROCESSING

A. *Satellite Image Processing*

The original dataset contains close to 200000 images, each of 768 x 768 pixels size. However there is a class imbalance here as only around 42000 images contain ships in them. Hence, the first step in our methodology is to choose only those images which contain ships in them and use these images to train our model. We manually segregated these images from the original dataset.To optimize the performance of our model, we have applied a smoothing filter and changed the colour model of the images to HSV (Hue Saturation Value). To achieve this we have implemented the functions made available by the open source python library Scikit-Learn. The functions used are 'Gaussian Filter' and 'RGB to HSV Colour Transformation'. The first operation applied is the RGB to HSV Transform. The ships contained in our input images are expected to have different or distinct colour ranges and luminosities when compared to their background. We hope to exploit these differences using HSV image transforms to improve the efficiency of our segmentation process. The Gaussian Filter is applied next to the transformed images to smoothen the images and reduce the noise in the images. The gaussian filter is also needed to produce a less pixelated image.

B. *Mask Images Preprocessing*

The masks are provided in Running Length Encoding as a string data type for every individual ship contained in an image. To decode our masks correctly, our first step was to implement a function that would read the encoded pixels for each ship contained in the image from the dataset and to merge them correctly. To decode running RLE pixels the following method is used:

$$1(3)0(2)1(2) = 1110011$$

Here it is known that every pixel is followed by the number of times it must appear in the given image for the current sequence. After decoding every RLE string for a ship, the obtained image is converted to binary. That is every pixel value of the image is converted to either 0 or 1, based on whether the pixel contains a ship (1) or no (0). After converting every RLE string to a binary mask all the individual ship masks for a given image are merged. This is achieved by taking a 768 x 768 null matrix and then converting a given pixel in the null matrix to 1 if that a pixel in that same position in the image matrix is 1 for any of the other individual ship masks.

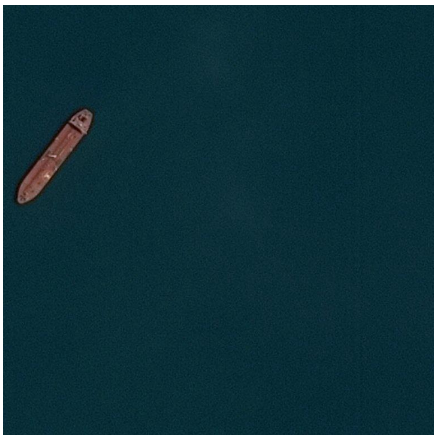

Figure 1. Original Input Image before Preprocessing

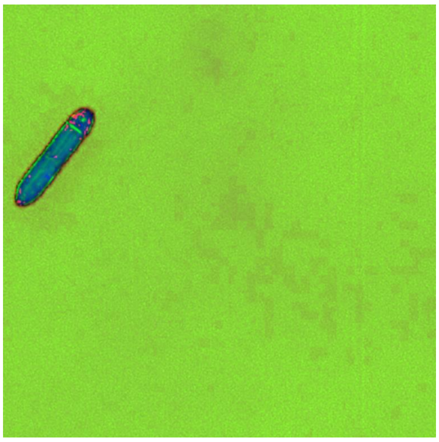

Figure 2. Preprocessed Input Image

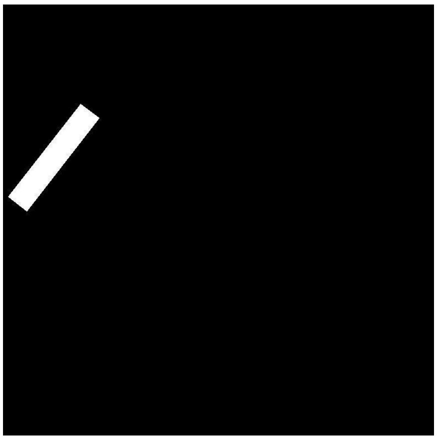

Figure 3. Preprocessed Input Mask

IV. UNET ARCHITECTURE

In the original UNET architecture, an input image first passes through the encoder branch and then the decoder branch to achieve high accuracy semantic segmentation for the given problem. The input image passes through two 2 dimensional Convolution Layers of kernel size 3 x 3 with RELU Activation Function Layers.. The in channels of the images are 3 x 3 and the out channels are 64. This step is followed by a 2 x 2 Max Pooling layer to highlight the important features of the images. This process is repeated 3 more times where the in channels for a convolutional layer is equal to the out channels of the previous block and the out channels of the current block is equal to (current input channels x 2). Finally a final pair of Convolutional Layers are applied to the image, with the number of out channels equal to 1024. After this step the model has figured out the 'WHAT' of the image. That is, the model has identified the ship in the image and now needs to find the 'WHERE' of the ships. To achieve this Up Convolutional Layers are used with kernel size 2 x 2. This step is followed by applying two Convolutional Operations on the image again. This time however, the number of channels are halved after every block of convolutions applied to the image. But before applying the Convolutional Layers block, Skip Connections are used to get the output of the corresponding step from the 'Encoder' branch of the UNET model and join this output with the current input to double the size of the current image. The 'Up-Convolutional' blocks too are applied four times on the image. Finally a 1 x 1 Convolutional Operation is applied to get the final output image from the model. This image is the segmentation map of the input image.

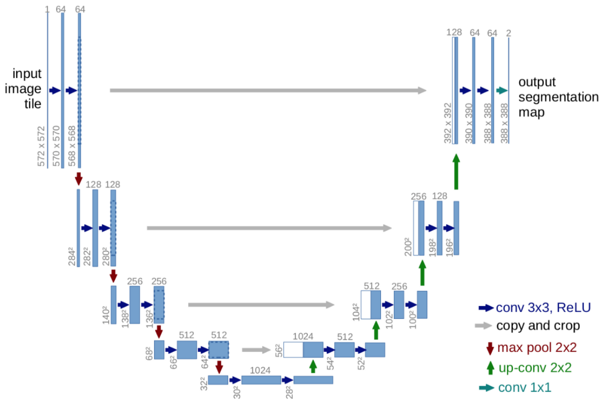

Figure 4. Original UNET Architecture

V. PROPOSED ACN-UNET ARCHITECTURE

In our proposed ACN-UNET Architecture, we intend to use four 2 dimensional Convolutional Layers in a block for the 'Encoder' branch of the UNET model. so, if our input image has 3 channels then in our proposed model for the first block the Convolutional Layer is used to increase the channels of the image to 64. Next another Convolutional Layer is applied to increase the channels by 10 to 74. This is followed by the third Convolutional Layer which will half the channels back to 64. A Dropout layer is added after this layer which will randomly convert a given image pixel to 0 based on a probability which is set to 0.6. The final Convolutional Layer is applied to the image next which keeps the number of channels the same i.e 64. This process is repeated 3 more times where the number of features/channels is increased to 128, 256, 512 and finally 1024. We take this approach as we believe that this would improve the models accuracy in better finding the position of the ships in our given input images. The 'Skip Connections' and the 'Decoder' Branch of the proposed model follow the same approach as the original UNET Architecture.

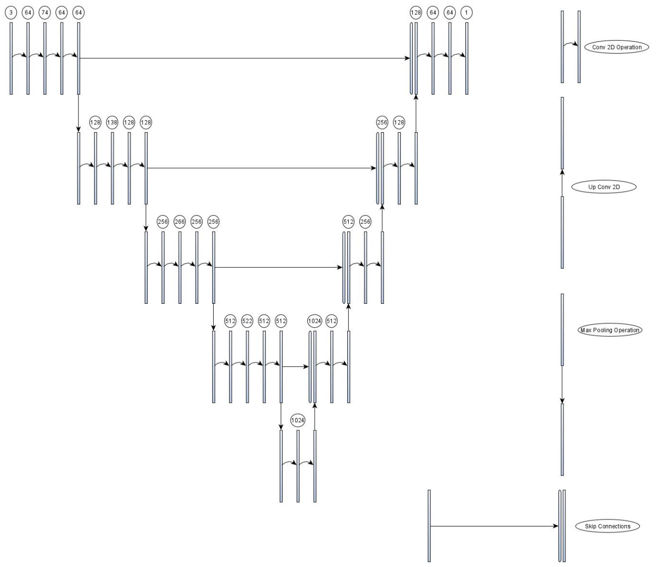

Figure 5. Proposed ACN-UNET Architecture

VI. RESULT

Our proposed model leads to a direct improvement over the original architecture in the predictions made for semantic segmentation under the given constraints. The model was run on Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz Systems. with 8GB RAM

TABLE I. RESULTS

Metric	Algorithms	
	Original UNET Architecture	***Proposed ACN-UNET Architecture***
Pixel Accuracy	91.13 %	95.66 %
Dice Score	0.27	0.33

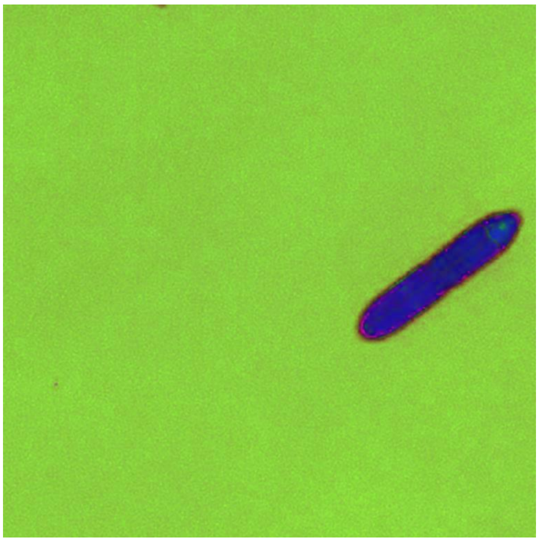

Figure 6. Input Image to Model

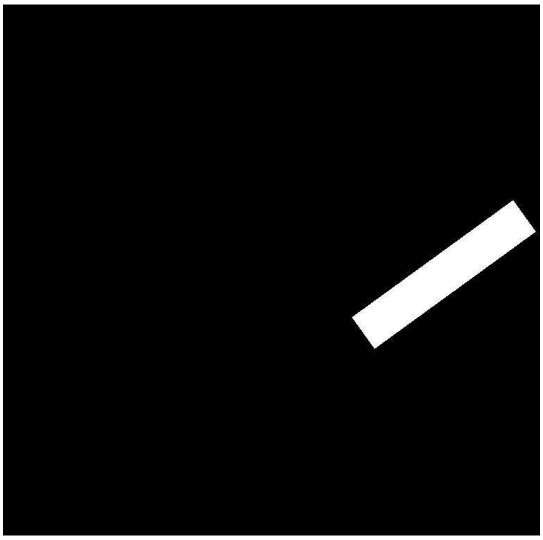

Figure 7. Expected Output

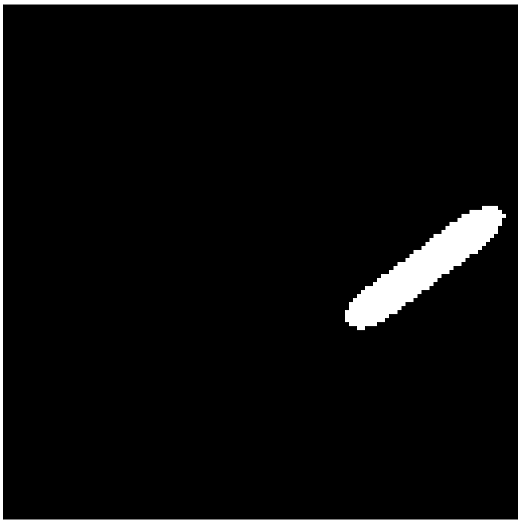

Figure 8. Predicted Image

Acknowledgment

The guidance and assistance provided by Dr. Akila Victor was greatly appreciated. We would also like to extend our gratitude to our university for their support throughout the course of this project.

References

[1]J. Jiao et al., "A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection," in IEEE Access, vol. 6, pp. 20881-20892, 2018, doi: 10.1109/ACCESS.2018.2825376.

[2]Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[3]Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[4]Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). PMLR.

[5]Nie, X., Duan, M., Ding, H., Hu, B., & Wong, E. K. (2020). Attention mask R-CNN for ship detection and segmentation from remote sensing images. IEEE Access, 8, 9325-9334.

[6]Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125).

[7]X. Zhang et al., "A Lightweight Feature Optimizing Network for Ship Detection in SAR Image," in IEEE Access, vol. 7, pp. 141662-141678, 2019, doi: 10.1109/ACCESS.2019.2943241

[8]S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su and J. Shi, "HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation," in IEEE Access, vol. 8, pp. 120234-120254, 2020, doi: 10.1109/ACCESS.2020.3005861.

[9] Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., & Guo, Z. (2018). Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. Remote Sensing, 10(1), 132.

[10] Fan, W., Zhou, F., Bai, X., Tao, M., & Tian, T. (2019). Ship detection using deep convolutional neural networks for PolSAR images. Remote Sensing, 11(23), 2862.

[11] Zhao, H., Zhang, W., Sun, H., & Xue, B. (2019). Embedded deep learning for ship detection and recognition. Future Internet, 11(2), 53.

[12] Karki, S., & Kulkarni, S. (2021, February). Ship Detection and Segmentation using Unet. In 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (pp. 1-7). IEEE.

[13] Li, X., & Cai, K. (2020, August). Method research on ship detection in remote sensing image based on Yolo algorithm. In 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS) (pp. 104-108). IEEE.

[14]Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[15] https://www.kaggle.com/c/airbus-ship-detection

[16] Hordiiuk, D., Oliinyk, I., Hnatushenko, V., & Maksymov, K. (2019, April). Semantic segmentation for ships detection from satellite imagery. In 2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO) (pp. 454-457). IEEE.