2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-7281-5791-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAECT49130.2021.9392463

Ship Detection and Segmentation using Unet

Sagar Karki, Siddhivinayak Kulkarni
School of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, India
sagarkarki136@gmail.com, siddhivinayak.kulkarni@mitwpu.edu.in

***Abstract*---Maritime surveillance is very crucial for every nation even if it is surrounded by land. The surveillance can not only inform about the threats and illegal activities but also manage the traffic and regulate a lot of movement on water. With more and more satellites with Synthetic-Aperture Radar (SAR) and other capabilities to capture high-resolution images being launched, automation of ship detection from these images is now a necessity. The experiments are performed on satellite image datasets where various encoders and loss functions for the Unet model are evaluated based on segmentation results over the test dataset. Our research and analysis of these encoders and loss functions for the Unet model lead us to certain conclusions as to which situation would benefit from which of the loss function and encoder and their combination. The research shows that the modes are robust and can perform well even with difficult backgrounds like clouds, waves etc. Our best model archived 0.823 on F2 score at different IOU thresholds on Airbus ship detection challenge test dataset.**

Keywords—Unet, Ship segmentation, Fastai, EfficientNet, GoogleNet

I. INTRODUCTION

The increase in launch of satellites and heavy data collected from these satellites aid in better maritime monitoring. Due to the high volume of image data, there arises a need to automate the process of analysing these images to have quick and efficient responses to suspicious activities on water. Maritime monitoring can not only help us in detecting illegal movements on the water surface but also warn us of potential threats and disasters. This can also help standard vehicles and hence improve the survival rate in case of disaster in the middle of the water body. These activities need quick and efficient responses which can be addressed by ship detection models proposed in this paper. The comparisons and the result displayed are from the best possible behaviour of the models using the training methods presented in the Models Architecture and training details section of this paper. The results of the experiments warrant further investigation of these models and training methods for maritime surveillance. This paper is divided into 5 sections. Section 1 explains the Literature review. Section 2 describes the Methodology followed by Section 3 which explains the Experiments. Section 4 gives detailed Analysis and Discussion of the results of the experiments. The final section provides the Conclusion and future scope of the research.

II. LITERATURE REVIEW

The use of satellite images to detect ships has been studied as both segmentation as well as object detection problem statement. In a research [1], the inshore ship detention method proposes the use of iterative bounding box regression. This method generates small and precise ROIs. The method proposed is specific to inshore detection. In [2] SAR images were used for ship detection. The results of the proposed method of using modified Faster R-CNN performed well for large ships in images but did not perform well on the small ships. In [3] Faster RCNN trained on 800x 800 resolution images that had 1282 images with ships and 2303 images without ships (with land areas), the authors demonstrate that the use of land and sea segmentation leads to missing out on some docked ships and their approach helped avoid those cases.

Other methods that include the use of You Only Look Once(YOLO)[4] and an incarnation of the Faster R-CNN framework with Feature Pyramid Network (FPN) as a feature extractor [5]. In [6] Eigen-ellipse discrimination is designed to exclude false alarm objects in near shore and harbour area. The authors use length-width ratio and geometric shape to eliminate non-ship targets. The test set has 205 images and out of them the method gives false alarm for 30 images and about 13 ships were missed[7]. This method is simple and can only be used as a baseline and is not suitable for deployment in the zones where the least false alarm rate is necessary. In [8] visual attention model with a local binary pattern is proposed where large images are cut into smaller ones and analysed. The classification is done using the traditional model of SVM. CNN is giving better results than SVM for feature extraction and classification tasks that involve calculating pixel to pixel-wise analysis to make a decision. The method is good at showing suspicious candidates but is not a deployment-ready approach to detect target ships accurately. Most traditional approaches include gray value statistics [9], texture analysis, and shape examination [10]. Experiments have also been performed using a blend of bottom-up visual attention mechanism and neighbourhood similarity method [11]. The multi-stage approach in [12] demonstrated the use of the traditional local binary pattern identification method on higher resolution of images to get the promising ship candidates. An early method suggested in [13] proposes the use of texture features that help in differentiating between sea and ships. The method relied on the confidence mapping of extracted ship candidates. The proposed method is insensitive to different waves, illumination changes, ships with different sizes, and bright/dark intensities. The drawback of analysis was a lesser number of test cases. In [14] the use of Tensorflow APIs is proposed which does not need heavy computation. The images that the model was trained on had only one ship per image, therefore the model has a high chance of performing well on images with only one big ship compared to multiple ships. The results of the model club multiple ships in an image and identify them as one ship .In experiments by Qiancong Fan et al. [26] the performance of Unet model is analysis and compared with mask RCNN .The model chives a F1 score of 0.912 on Compact polarimetric synthetic aperture radar (CP SAR) images and verify the advantages of Compact polarimetric over single and linear-dual polarization.

There has been heavy research concerning the segmentation of ships from image and detection of ships from image. The models that have been used mostly have Resnet as encoder for Unet models and regular Mask RCNN models. Very few loss functions have been explored to train segmentation models on satellite image data. This paper explores different encoders and loss functions to train segmentation models.

III. METHODOLOGY

In Figure 1 the blocks inside the blue dotted line represent the training phase and in the red dotted line represents the testing phase. “Image and segmentation mask” block represents the start of the input pipeline where the data is entered which is used for training. The “Pre-processing Input data” block represents the augmentation part of the input pipeline. The “Unet Model” box represents the Unet model . The “Test Image” is the input pipeline used for testing. The ”Segmentation Masks” block represents the segmentation masks generated by the Unet model. The “Classifier” block represents classifiers that help in removing false positives by classifying the images into image with ship and image without ship. The “Final decision CSV with run length encoded mask data” block represents the finally output csv that has the segmentation masks in run length encoding for images with shape

Experiments are performed using different training methods using different tools mainly Keras and Fastai to develop and train these models. Due to computation power limitations, the experiments in this paper have been performed with 224x224 resolution of the image. The dataset has 193K images and every image has a resolution of 768x768. The images are scaled down to perform our experiments

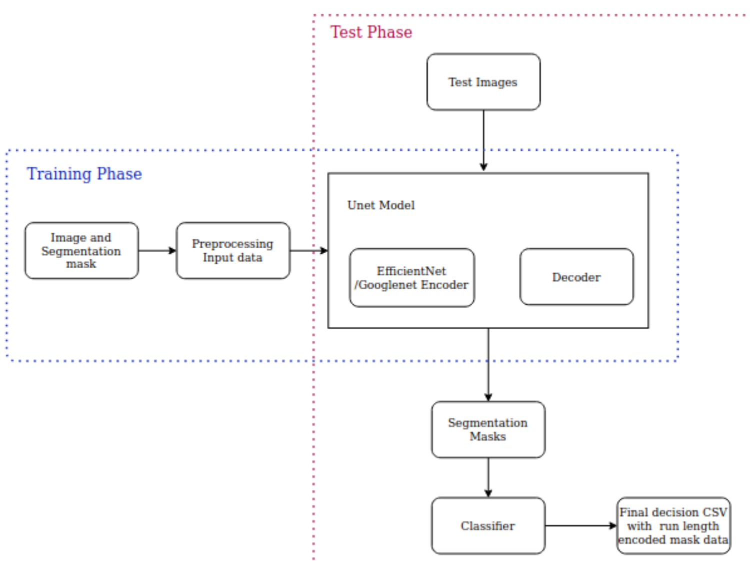

Fig. 1. Methodology Workflow

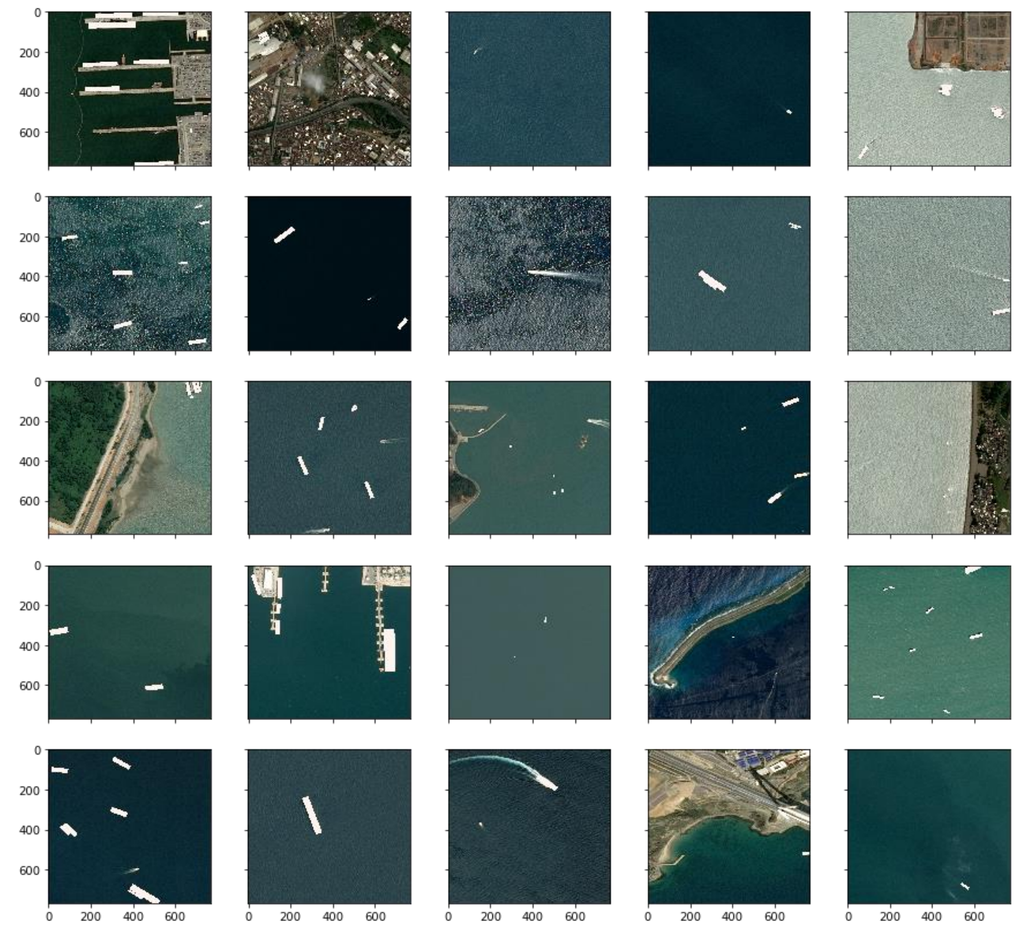

Fig. 2. Images with ships and masks overlaid on them

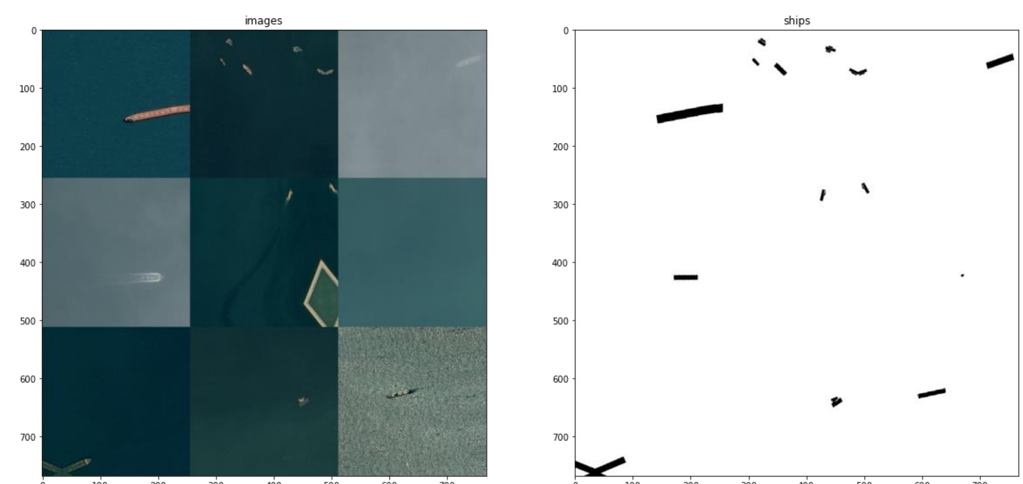

Fig. 3. Batch of images with ships and corresponding batch of mask

A. Dataset

The data is taken from Kaggle [15] which has 193K images out of which only 40K [Figure 2] images having 80K ships in total. The images that have ships may have one or more ships in it. The corresponding run-length encoded data is provided in a separate CSV file to generate masks from it for the image. The data has 15 k images for testing and Kaggle has the ground truth values. The models that perform well on the test data demonstrate their robustness and a rough estimate of its performance when it will be deployed. All the images are satellite images with 768* 768 resolution. All the training is done after scaling down the images to 224*224 images due to computational power limitations shown in Figure 2 and custom cropped images in Figure 3

IV. Experiments

Keras [16] [25] and Fastai [17] are used for building and training our models (Figure 1). Using Keras Unet [18] with EfficientNet B0 and B3 [19] as encoders are trained. Fastai is used to build an Unet model with googlenet[19,20] encoder and self-attention mechanism. All the models are trained using Adam optimizer [21].

Experiments are performed on Keras models by using Keras's ReduceLRonPlateau callback function for learning rate scheduling and later with cyclic learning rate [22]. For the Fastai model, we used a one cycle policy [23]. Keras models are trained with early stopping callback being monitored on validation loss. All the encoders used in the experiments are pre-trained models on Imagnet.

Keras models are initially trained using Binary Cross-Entropy (BCE) plus dice loss for 10 epochs using only images with ships in them to reduce training time. To improve the performance of the models and reduce false positives, models were further trained for 20 epochs on the same data using tversky loss [24] and assigning a heavy penalty to the false positives in the loss function.

The models were further trained on a dataset that had both ships and no ship images for 10 epochs. The models were then trained on different combinations of ratio of images with ships and no ships. In order to improve the performance the models were further trained for another 15 -17 epochs with different ratios of images with ships and no ships in the training batch.

Fastai model was trained by first freezing the encoder and only training the decoder using one cycle policy with initial Lr =1e-4 for 3 epochs only using images with ships (using 32K for training and 8K for validation). The whole model was further trained for 4 epochs. Experiments were further performed by using 90% of images with ships for training in a similar fashion as before. A classifier is trained to classify the image whether it has a ship or not using resnet34 to reduce false positives. This was done to avoid false positives. The Fastai segmentation model was trained using Focal loss plus dice loss function. The classifier is trained on binary cross-entropy loss. All the experiments are performed using Tesla K80 GPU

A. Analysis and Discussion

Quantitative evaluation is done based on the F2 Score at different Intersections over Union (IoU) thresholds. The metric sweeps over a range of IoU thresholds, at each point calculating an F2 Score. The threshold values range from 0.5 to 0.95 with a step size of 0.05: (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95). In other words, at a threshold of 0.5, a predicted object is considered a "hit" if its intersection over union with a ground truth object is greater than 0.5.

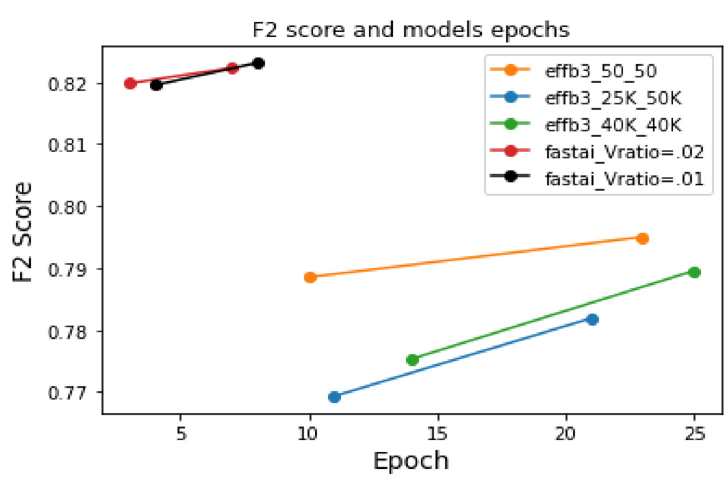

Fig. 4. Comparison of performance of different encoders in Unet and training method

In Figure 4 effb3_50_50 represents Unet with EfficientNet B3 encoder model trained with batches having an equal number of ships and no ship images. effb3_25K_50K represents Unet with EfficientNet B3 encoder model trained on a training dataset having 25K images with ship and 50K images without ships. effb3_40K_40K represents Unet with EfficientNet B3 encoder model trained on a training dataset having 40K images with ship and 40K images without ships. fastai_Vratio=0.2 represents Fastai Unet model with Googlenet encoder trained on images with ships and the validation set has 20% of the ship images .fastai_Vratio=0.2 represents Fastai Unet model with Googlenet encoder trained on images with ships and the validation set has 10% of the ship images .The experiments involve 3 different models on trained with different training methods and the results given by the Fastai model which is Unet with Googlenet as encoder trained on 224*224 resolution of ships only images achieved the best result when compared quantitatively using the F2 -score with other models trained according to the details given in Experiments section of this paper. The EfficientNet B0 encoder did not perform as good as the Unet with EfficientNet B3 as the encoder. All the results are shown after they have been filtered using a classification model that classifies images on whether or not it has a ship to reduce false positives. The Fastai models achieves superior results. The loss function has played a major role in improving the models F2 score but the evaluation metric does not always show the bigger picture. The evaluation metric used is sensitive to false positive. This method, although dependent on the IoU, is relative to the performance of the model at certain threshold values. If the model has more IoU at threshold 0.5 and less at 0.75 and another model has the reverse of this situation then even under such circumstances the F2 score will remain the same even if the one of them performs qualitatively better at a higher threshold. It can be inferred that even if the training set of the Fastai model was increased the model's performance improved slightly. This shows that the model can be trained with less data to achieve near similar performance when there is an increase of training data by 10%. Fine-tuning on higher resolution the models may further enhance the performance of the model.

B. *Qualitative Analysis*

Experiments and results suggest that even though the evaluation metric may not always showcase the information regarding qualitative evaluation, The Fastai Unet model with Goognet encoder model (Figure 5) performs better even at lower numbers of epochs and less training data. The ship segmentation masks generated by Fastai Unet model(Figure 5) are qualitatively better than the ones generated by Unet with EfficientNet encoder (Figure 6). It can be observed that the Unet with EfficientNet encoder is good for detecting big ships but it does not give good results when ships are close to each other and on small ships. In Figure 5 the ship mask are over laid on the image and ships are highlighted by bright white as compared to rest of deep blue background

Fig. 5. Results of Unet with Googlenet encoder

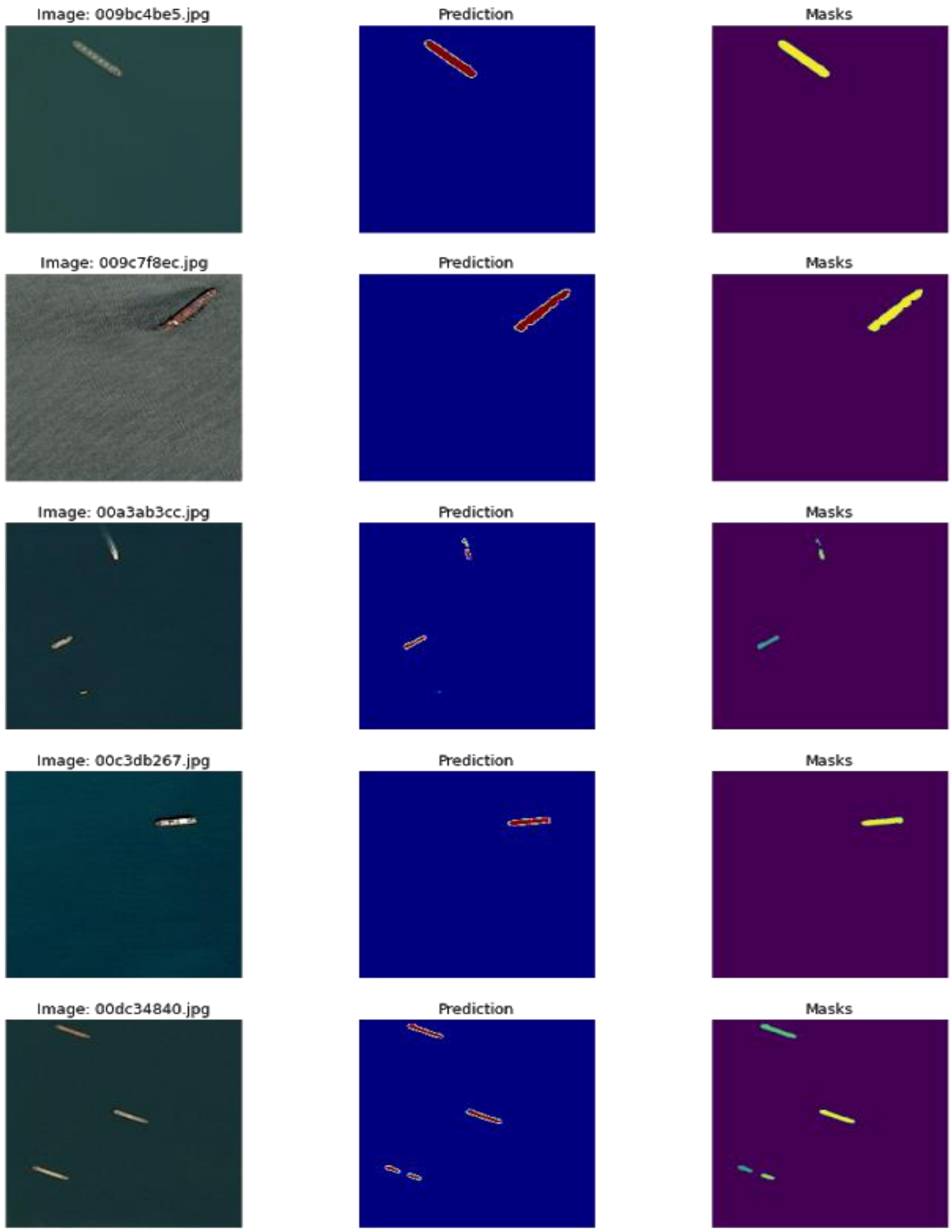

Fig. 6. Results of Unet with EfficientNet B3 encoder

V. Conclusion and Future Scope

This paper investigates training methods with different models using the Unet model. The EfficientNet encoder trained Tversky loss on further fine-tuning on higher resolution can help in tackling false alarms. The smaller and close ships are better segmented by the Googlenet encoder model. Both these models clubbed with a good classifier of whether there is a ship in the image or not further boosts their F2 score. The research in this paper provides a solution that can ease the work of analysing the satellite images and with further fine-tuning, the model can be deployed for maritime surveillance purposes. The future scope of the research includes clubbing these models along with the ship identification method , movement tracking and prediction methods to further enhance maritime surveillance.

References

[1] F. Wu, Z. Zhou, B. Wang and J. Ma, "Inshore Ship Detection Based on Convolutional Neural Network in Optical Satellite Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 11, pp. 4005-4015, Nov. 2018, doi: 10.1109/JSTARS.2018.2873190.

[2] D. Štepec, T. Martinčič and D. Skočaj, "Automated System for Ship Detection from Medium Resolution Satellite Optical Imagery," OCEANS 2019 MTS/IEEE SEATTLE, Seattle, WA, USA, 2019, pp. 1-10, doi: 10.23919/OCEANS40490.2019.8962707.

[3] Y. He, X. Sun, L. Gao and B. Zhang, "Ship Detection Without Sea-Land Segmentation for Large-Scale High-Resolution Optical Satellite Images," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018, pp. 717-720, doi: 10.1109/IGARSS.2018.8519391.

[4] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. 2015.

[5] Wang Y, Wang C, Zhang H, Dong Y, Wei S. Automatic Ship Detection Based on RetinaNet Using Multi-Resolution Gaofen-3 Imagery. Remote Sensing. 2019. p. 531. doi:10.3390/rs11050531

[6] W. Ao and F. Xu, "Robust Ship Detection in SAR Images from Complex Background," 2018 IEEE International Conference on Computational Electromagnetics (ICCEM), Chengdu, 2018, pp. 1-2, doi: 10.1109/COMPEM.2018.8496647.

[7] Zhao Ying-hai,Wu Xiu-qing,Wen Ling-yun,Xu Shou-shi, “Ship Target Detection Scheme for Optical Remote Sensing Images”, 2008

[8] Zhina Song, Haigang Sui and Yujie Wang, "Automatic ship detection for optical satellite images based on visual attention model and LBP," 2014 IEEE Workshop on Electronics, Computer and Applications, Ottawa, ON, 2014, pp. 722-725, doi: 10.1109/IWECA.2014.6845723.

[9] Zhao Ying-hai,Wu Xiu-qing,Wen Ling-yun,Xu Shou-shi, “Ship Target Detection Scheme for Optical Remote Sensing Images”, 2008

[10] C. Zhu, H. Zhou, R. Wang and J. Guo, "A Novel Hierarchical Method of Ship Detection from Spaceborne Optical Image Based on Shape and Texture Features," in IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 9, pp. 3446-3456, Sept. 2010, doi: 10.1109/TGRS.2010.2046330.

[11] F. Bi, B. Zhu, L. Gao and M. Bian, "A Visual Search Inspired Computational Model for Ship Detection in Optical Satellite Images," in IEEE Geoscience and Remote Sensing Letters, vol. 9, no. 4, pp. 749-753, July 2012, doi: 10.1109/LGRS.2011.2180695.

[12] F. Yang, Q. Xu, F. Gao and L. Hu, "Ship detection from optical satellite images based on visual search mechanism," 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, 2015, pp. 3679-3682, doi: 10.1109/IGARSS.2015.7326621.

[13] G. Huang, Y. Wang, Y. Zhang and Y. Tian, "Ship Detection Using Texture Statistics from Optical Satellite Images," 2011 International Conference on Digital Image Computing: Techniques and Applications, Noosa, QLD, 2011, pp. 507-512, doi: 10.1109/DICTA.2011.91.

[14] M. Kartal and O. Duman, "Ship Detection from Optical Satellite Images with Deep Learning," 2019 9th International Conference on Recent Advances in Space Technologies (RAST), Istanbul, Turkey, 2019, pp. 479-484, doi: 10.1109/RAST.2019.8767844.

[15] Airbus Ship Detection Challenge. Kaggle

[16] Keras Team. Keras: the Python deep learning API.

[17] Howard J, Gugger S. fastai: A Layered API for Deep Learning. 2020. doi:10.3390/info11020108

[18] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

[19] Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019.

[20] C. Szegedy *et al.*, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[21] 21. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2014.

[22] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Santa Rosa, CA, 2017, pp. 464-472, doi: 10.1109/WACV.2017.58.

[23] Smith LN. A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay. 2018.

[24] Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. 2017.

[25] Pavel Yakubovskiy. Segmentation Models .GitHub repository. GitHub. Yakubovskiy:2019

[26] Fan, Q.; Chen, F.; Cheng, M.; Lou, S.; Xiao, R.; Zhang, B.; Wang, C.; Li, J. Ship Detection Using a Fully Convolutional Network with Compact Polarimetric SAR Images. *Remote Sens.* 2019, *11*, 2171