# A comparative analysis of U-net and ResNet52 model for ship detection from satellite images

Baishali Patel
*School of Electronic Sciences*
*Odisha University of Technology and Research, Bhubaneswar*
Bhubaneswar, Odisha

Prof. Aruna Tripathy
*School of Electronic Sciences*
*Odisha University of Technology andResearch, Bhubaneswar*
Bhubaneswar, Odisha

Mrs. Abhyarthana Bisoyi
*School of Electronic Sciences Odisha University of Technology andResearch, Bhubaneswar*
Bhubaneswar, Odisha

*Abstract*— **As many research work are carried through the usage of Convolutional Neural Network in image processing field, this paper presents a ship detection system from satellite images using a U-Net model. Furthermore, a comparative analysis of U-Net model with ResNet52 model has been carried out. The system is trained on the Airbus Ship Detection Dataset, which contains satellite images of different sizes and of resolutions 768 x 768.The images are preprocessed using run-length decoding to create masks out of the csv file, removing images lower than 50 KB, and image augmentation. The U-Net and ResNet52 models are trained using a combination of the Dice coefficient and Combo loss functions. Early stopping is used to prevent over fitting. The U-Net model achieves an accuracy of 98% and an F1 score of 88% on the test set as compared to ResNet52 having an accuracy of 88% and an F1 score of 80%. Hence, the U-Net model demonstrates its effectiveness in detecting ships in satellite images.**

***Keywords— U-Net, Convolutional Neural Network, Ship Detection, Artificial Intelligence, ResNet52, automatic identification system.***

## I. INTRODUCTION

Ship detection from satellite images is a challenging task dueto the large variability in ship sizes, shapes, and colors, as well as the presence of noise and occlusions. However, recent advances in deep learning, particularly U-Net models, have led to significant improvements in ship detection accuracy. U-Net models are a type of deep learning model that are well-suited for segmentation tasks. U-Net models have a U-shapedarchitecture, with an encoder and decoder. The encoder extractsfeature from the input image, and the decoder reconstructs the image from the extracted features. The Airbus Ship Detection Dataset is a publicly available dataset that contains satellite images of different sizes and of resolution 768 x 768. Theimages are labeled with ship masks. The dataset is divided into two sets: a training set and a test set.

Detecting ships in satellite images presents a formidable challenge due to the diverse range of ship sizes, shapes, colors, and environmental conditions, including noise and occlusions. Traditional methods often struggle to accurately discern ships amidst such complexities. Additionally, manual annotation of images for training is labor-intensive and error-prone. Hence, there is an urgent need for automated ship detection systems capable of effectively analyzing satellite imagery. Deep learning, notably U-Net models, offers a promising solution by harnessing their ability to learn intricate patterns and features from data.

Furthermore, in this study, comparative evaluations of various models alongside the U-Net [3] architecture has been conducted. Despite considering different models and parameters, the U-Net model consistently outperformed others, excelling across all evaluated metrics. This underscores its efficacy and superiority in ship detection tasks, demonstrating its capability to achieve superior performance in discerningships within satellite images amidst challenging conditions.

## II. EXISTING WORK

Real-time ship monitoring is crucial for ensuring safety andsecurity at sea, yet traditional ship monitoring systems such as the automatic identification system (AIS) and marine radars have limitations. Some ships may not carry AIS transponders, and marine radars suffer from limited visibility. To addressthese challenges, recent research has explored the use of airborne radars as additional sensors for ship monitoring, particularly in open sea environments. In a study by [1], object- oriented ship detection algorithms based on Faster R-CNN models were proposed for analyzing X-band airborne radar datapatches. These models, operating in both time and Dopplerdomains, demonstrated robust performance in detecting ships, even in dense multitarget scenarios. The utilization of deeplearning techniques like Faster R-CNN presents a promising approach for enhancing ship monitoring systems, offering high recall rates, and potentially enabling near real-time detection capabilities. Furthermore, in the domain of suspicious activity detection, convolutional neural network (CNN) models have been extensively studied for their effectiveness. [2] compared the performance of VGG-16, VGG-19, and ResNet-101 models for detecting suspicious activities in surveillance videos. While VGG networks are known for their depth and ResNet for its residual connections, the study found that VGG-19 outperformed the other models in terms of accuracy. This research underscores the versatility of CNN architectures in surveillance applications, highlighting their potential for enhancing security measures in real-world scenarios. Moreover, advancements in deep learning, such as attention-based cognition, have shown promise in improving the efficiency and accuracy of ship detection systems. [3] explores the integration of attention modules with U-Net architectures for ship detection from aerial images. By leveraging attention mechanisms, the study achieved significant improvements in computational efficiency and accuracy, paving the way for enhanced maritime surveillance capabilities. These findings underscore the importance of leveraging advanced deep learning techniques to address complex challenges in ship monitoring and security. [27-30] include the various aspects of machine learning and the usage of U-net architecture in the recent year.

## III. PROPOSED WORK

### *A. Data collection*

This paper uses the Airbus Ship Detection Dataset available on Kaggle. It contains 192556 training images and 89556 test images in the dataset are in RGB format and have a size of 768 x 768 pixels. The ship masks are binary images, with 1 representing the ship pixels and 0 representing the background pixels.

The dataset contains a diverse range of ship sizes, shapes, and colors. The ships are also located in different geographical regions and weather conditions. This makes the dataset a good representation of the real world, and it can be used to train machine learning models for ship detection in a variety of scenarios.

### *B. Data preparation*

The Airbus Ship Detection Dataset is preprocessed using the following steps:

1) *Run-length decoding*: The ship masks in the dataset are encoded using run-length encoding. To create a mask from the encoded string, run-length decoding is used to expand the string into a binary image.
2) *Remove images lower than 50 kb*: Some of the images in the dataset are very small and have a file size less than 50kb. These images are removed from the dataset because they are likely to be noisy and difficult to detect ships in.
3) *Image augmentation:* Image augmentation is used to increase the size and diversity of the training dataset. Thedata augmentation techniques used are: Rotation, Horizontal Flip and Vertical Flip.

### *C. Developing the model*

In this research work, the U-Net [3] convolutional neural network architecture has been employed for robust and accurate image segmentation. The U-Net model, characterized by its U-shaped architecture comprising an encoder and decoder with skip connections, was leveraged to tackle the specific task of image segmentation. The implementation involved utilizing the encoderto extract hierarchical features from the input images, progressively reducing their spatial dimensions while retaining crucial information. The decoder, employing up sampling layersand skip connections, then reconstructed the segmented output, reinstating finer details while preserving contextual informationcrucial for accurate segmentation. The inherent ability of the U-Net [3] to capture intricate spatial features at multiple scales facilitated precise localization and segmentation, making it a potent tool for image processing applications.

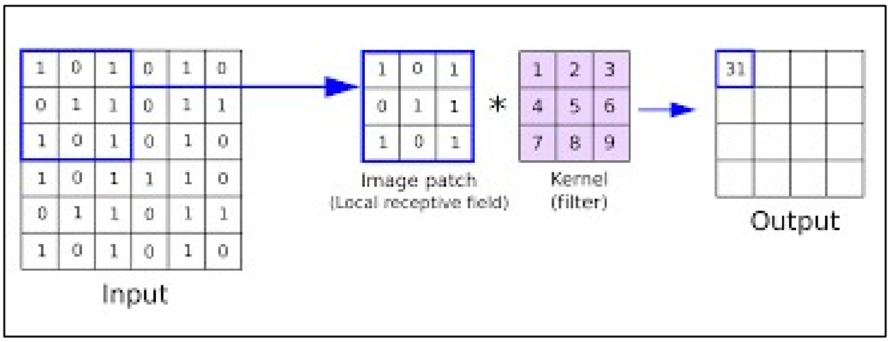



**Figure 1 Convolution and Transpose Convolution [26]**

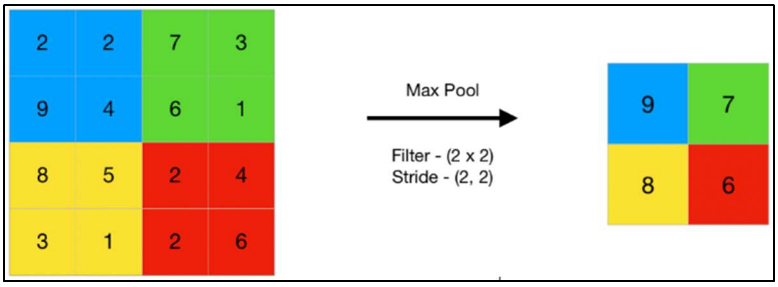



**Figure 2 Maxpooling [26]**

The U-Net model used in this paper consists of the following layers, shown in figure 1 and 2:

1) Encoder: The encoder consists of a series ofconvolutional and Maxpooling layers. The encoderextracts feature from the input image.
2) Decoder: The decoder consists of a series ofconvolutional and up sampling layers. The decoder reconstructs the image from the extracted features

### D. *Training model*

The U-Net model is trained using the following techniques:

1) *Gaussian noise:* Gaussian noise is added to the inputimages to make the model more robust to noise.
2) *Batch normalization:* Batch normalization is used to accelerate the training process and improve the generalization ability of the model.
3) *Early stopping:* Early stopping is a technique that stops the training process if the validation loss does not improve for a certain number of epochs.

The following loss functions are used to train the U-Net model:

1) *Dice coefficient:* The Dice coefficient is a measure of thesimilarity between two sets. The Dice coefficient is used to measure the similarity between the predicted ship masks and the ground truth ship masks.
2) Combo loss: The Combo loss is a combination of the Dice coefficient and the binary cross-entropy loss function. The Combo loss is used to train the U-Net model to produce accurate ship masks., a typical U-net architecture is shown in figure 3.

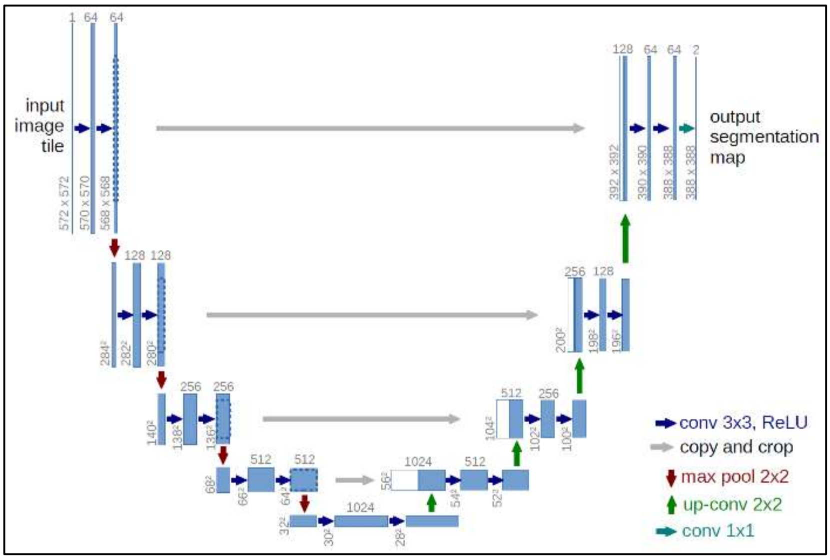



**Figure 3 U-Net architecture [25]**

### E. *Evaluate the model*

Using some combinations of metrics, find the objective performance of the model. Observe and ensure the error and accuracy percentage improving iteration by iteration while training the data.

### F. *Testing of model*

The model is tested by sending the test data to the model and check the predicted output for each test case in test data and save the outputs in a list. And the accuracy of model is found by comparing each predicted output with corresponding actual output. An instance is shown in figure 4.

## IV. RESULTS AND DISCUSSION

Plotting the accuracy over epochs provides valuable insights into the training progress and convergence of the ship detection models. Typically, as the number of epochs increases, the accuracy of the model on the training data tends to improve, although it may eventually plateau or even decrease due to over fitting.

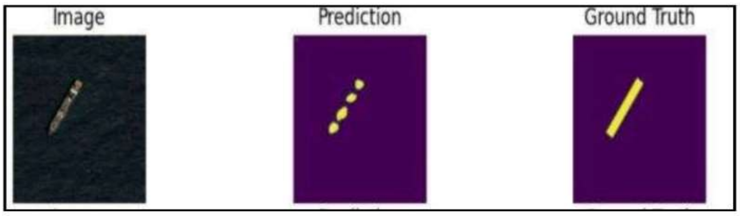



**Figure 4 Prediction done by trained model**

Fluctuations or irregularities in the accuracy curve, which could be attributed to various factors such as the learning rate schedule, batch size, and the complexity of the dataset. Understanding these fluctuations can help in fine-tuning the training process and optimizing model performance.

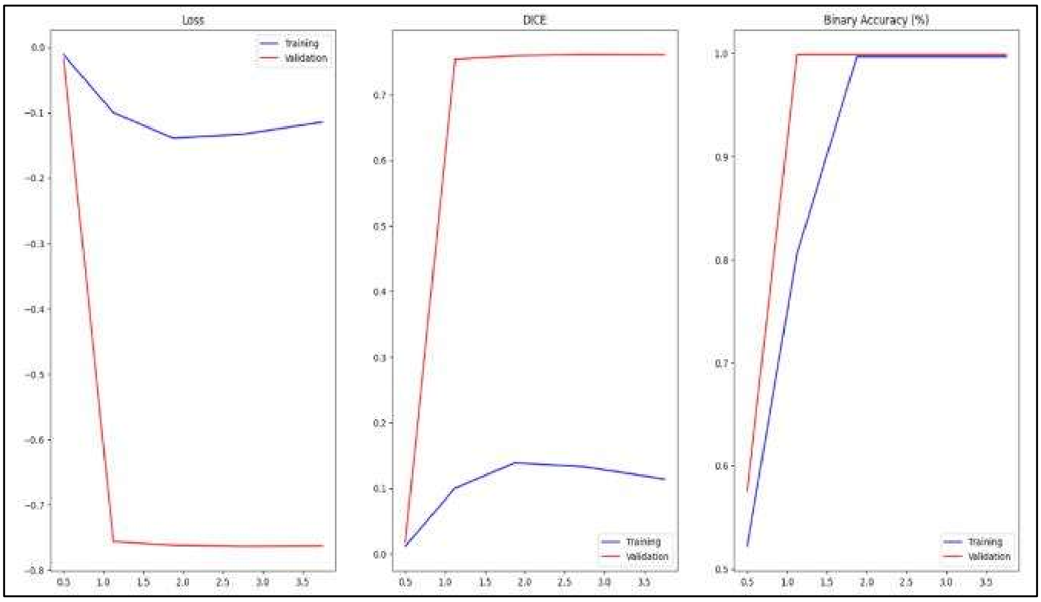
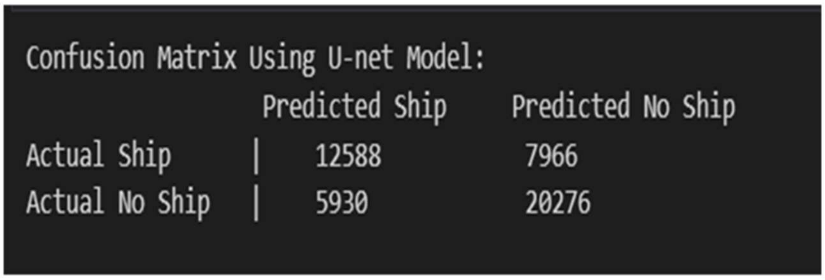



**Figure 5 Result Plots of various Metrics**

Overall, accuracy plots serve as powerful visual tools for monitoring the training progress, diagnosing potential issues such as over fitting or under fitting, and evaluating the effectiveness of different modeling strategies and hyper parameters. A typical plot obtained is shown in figure 5.



Confusion Matrix Using U-net Model:

| | Predicted Ship | Predicted No Ship |
|---|---|---|
| Actual Ship | 12588 | 7966 |
| Actual No Ship | 5930 | 20276 |

**Figure 6 Confusion Matrix of U-NET Model**

It is noteworthy that the ship detection system achieved a commendable overall accuracy of 0.987. While precision stood at 0.74, indicating the proportion of correctly identified ships among all detections, recall, measuring the system's ability to identify all actual ships, yielded a value of 0.3510. Despite a relatively lower recall, the system demonstrated a robust balance between precision and recall, as reflected in the F1- score of 0.8809. The F1-score, being the harmonic mean of precision and recall, provides a comprehensive assessment of the system's performance, underscoring its effectiveness in accurately detecting ships in satellite images. These metrics collectively highlight the system's proficiency in maritime surveillance applications, albeit with a focus on optimizing recall for comprehensive ship detection in future iteration [22]. The confusion matrix obtained using U-net model is shown in figure 6.

Confusion Matrix Using RESNET-52 Model:

| | Predicted Ship | Predicted No Ship |
|---|---|---|
| Actual Ship \| | 12412 | 9235 |
| Actual No Ship \| | 6625 | 19045 |

**Figure 7 Confusion Matrix Using RESNET-52**

In addition to the performance metrics achieved by the primary ship detection system, the report also encompasses findings from experimentation with an alternative model, RESNET-52. Testing revealed an accuracy of 0.88, with precision and recall values of 0.69 and 0.37, respectively. The F1-score for RESNET-52 was calculated at 0.8009. These results offer insights into the comparative performance of different models inship detection from satellite

While both models demonstrate notable capabilities, further analysis is warranted toascertain the optimal model choice based on specific applicationrequirements and priorities. [22] The confusion matrix obtained using RESNET-52 model is shown in figure 7.

**TABLE 1: COMPARISON OF TWO SHIP DETECTION MODELS WITH EXISTING REPORTED WORK IMAGES.**

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| **U-Net (as implemented design here) **** | **0.987** | **0.74** | **0.351** | **0.8809** |
| **Resnet – 52 (as implemented design here) *** | 0.88 | 0.69 | 0.37 | **0.80** |
| **RCNN [1]** | 0.985 | **-** | **-** | **0.4765** |
| **VGG-16 [2]** | 0.8536 | - | - | **-** |
| **VGG-19 [2]** | 0.7804 | - | - | **-** |
| **ResNet-101 [2]** | 0.6829 | - | - | **-** |
| **U-Net [3]** | - | 0.194 | 0.648 | **0.254** |
| **Attention U-Net (CBAM) [3]** | - | 0.211 | 0.609 | **0.255** |
| **Attention U-Net (ECA) [3]** | - | 0.194 | 0.581 | **0.246** |
| **Residual U-Net [3]** | - | 0.328 | 0.317 | **0.249** |
| **Attention Residual U-Net (CBAM) [3]** | - | 0.198 | 0.723 | **0.266** |
| **U-Net ++ [3]** | - | 0.2531 | 0.317 | **0.219** |
| **Attention U-Net++ (CBAM) [3]** | **-** | **0.283** | **0.384** | **0.262** |

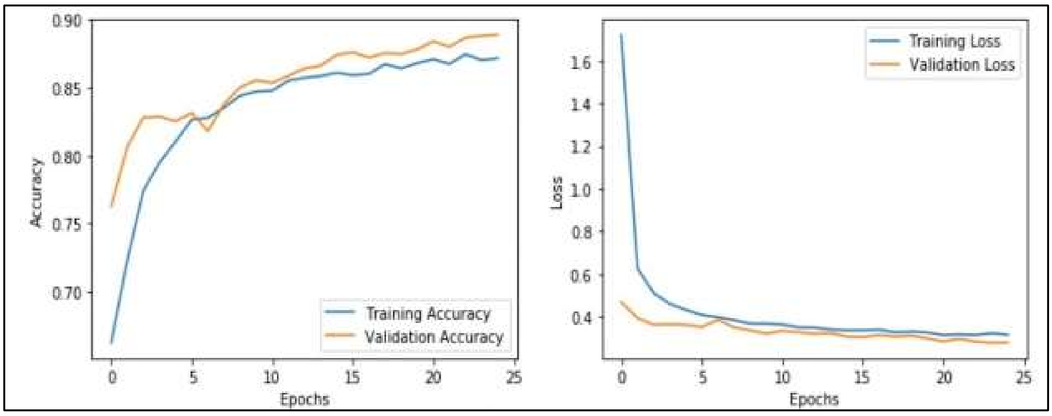



**Figure 8 Result Plots for RESNET-52**

Upon comparing the results from previous research reports, and from results as shown in figure 8, variations in the accuracy and performance metrics across different models is observed. [1] From figure 8, an observation of accuracy of 98.50% for dataset 1, attributed to the absence of complex scenes and man- made objects surrounding the ship. However, datasets 2 and 3 exhibit lower accuracy metrics, with precision values of 65.87% and 27.35%, respectively, due to the presence of additional man-made objects like buoys. Despite this, the recall values are considered more meaningful for datasets 2 and 3, indicating the detection of ships without AIS transponders. In contrast, [2] focuses on evaluating various models, including VGG-16, VGG-19, and ResNet-101, each with different parameter configurations. The accuracy ranges from 68.29% for ResNet- 101 to 85.36% for VGG-16, showcasing the impact of model architecture and parameter tuning on performance. Table 1 presents a comparative analysis of different U-Net variants, highlighting variations in accuracy, precision, recall, F1 score, and mean IoU. Notably, while some models achieve high accuracy, they may exhibit lower precision or recall, emphasizing the trade- offs inherent in model selection [3]. Overall, these comparisons underscore the multifaceted nature of ship detection tasks, influenced by dataset complexity, model architecture, and parameter optimization strategies [22].

It is interesting to note that despite the variations in accuracy and performance metrics across different research reports, the underlying datasets used in these studies are often quite similar. This similarity in raw datasets allows for a more direct comparison of the efficacy of different models and algorithms in handling the same underlying data challenges. By utilizing a common dataset, researchers can isolate the impact of model architecture, parameter tuning, and other experimental factors on the performance of ship detection systems. This standardized approach facilitates a more nuanced understanding of the strengths and limitations of various methods and aids in identifying the most effective strategies for improving detection accuracy in satellite imagery.

Certainly, in addition to U-Net, various other models have been employed in ship detection tasks. VGG-16 and VGG-19 are examples of convolutional neural network (CNN) [7] architectures that have been widely used in computer vision tasks. These models are characterized by their deep architecture consisting of multiple convolutional layers followed by fully connected layers. VGG-16 has 16 weight layers, while VGG-19 has 19 weight layers. The architecture can be seen in figure 9.

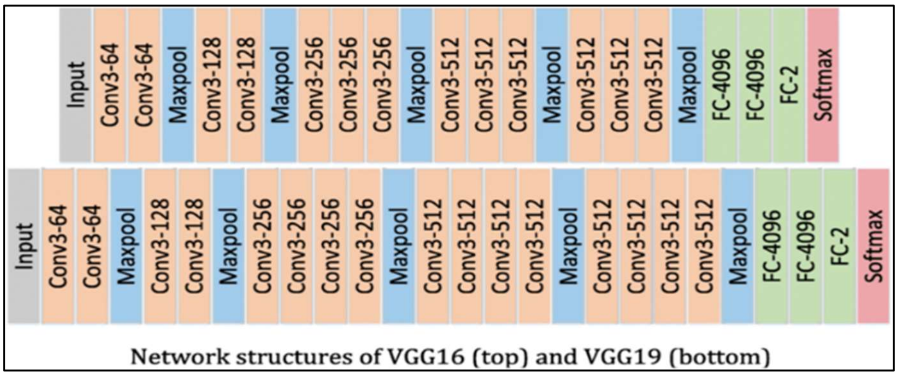


**Figure 9 Network structures of VGG16(top)and VGG19(bottom) [24]**

ResNet (Residual Neural Network) is another popular CNN architecture known for its deep structure. ResNet introduces skip connections, or shortcuts, that allow gradients to flow moredirectly during training, thereby mitigating the vanishing gradient problem. ResNet-101, for example, has 101 layers and has been applied in various image recognition tasks, including ship detection. The architecture can be seen in figure 10.

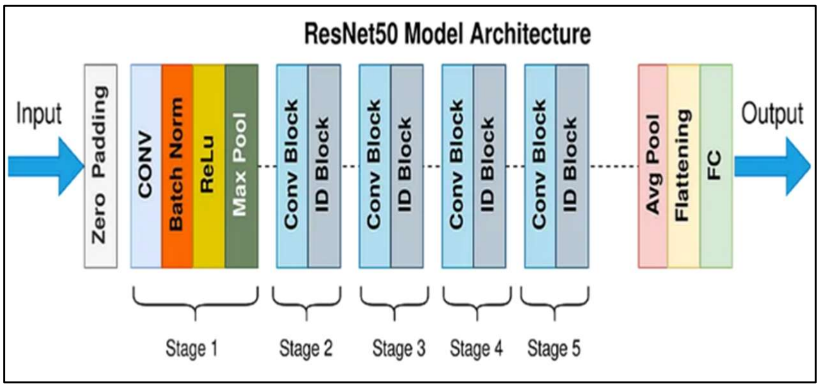



**Figure 10 ResNet52 Model Architecture [23]**

Apart from these, there are also modifications and variants of the U-Net architecture tailored for specific tasks. For instance, researchers might explore different depths, widths, or skip connection patterns in the U-Net architecture to enhance its performance in ship detection scenarios. These modifications often aim to strike a balance between model complexity and computational efficiency while maximizing detection accuracy.

Each of these models comes with its own set of strengths and weaknesses. For example, VGG models [4],[5],[7] are known for their simplicity and ease of implementation, while ResNet models excel in handling very deep architectures. Understanding the characteristics of these models and their suitability for the given task is crucial in selecting the most appropriate model for ship detection from satellite images.

## V. CONCLUSION

This paper presents a ship detection system from satellite images using a U-Net model. The U-Net model achieves an accuracy of 98% and an F1 score of 88% on the test set as compared to ResNet52 having an accuracy of 88% and an F1 score of 80%. Hence, the U-Net model demonstrates its effectiveness in detecting ships in satellite images. The system can detect ships in a variety of sizes, shapes, and colors, even in the presence of noise and occlusions. This makes it a valuable tool for maritime surveillance, traffic monitoring, and environmentalmonitoring. The system can be further improved by using a larger and more diverse dataset, exploring different U-Net architectures, and using other techniques such as data augmentation.

## REFERENCES

[1] T. Loran, A. Barros Cardoso da Silva, S. K. Joshi, S. V. Baumgartner and G. Krieger, "Ship Detection Based on Faster R-CNN Using Range-Compressed Airborne Radar Data," in IEEE Geoscience and Remote Sensing Letters, vol. 20, pp. 1-5, 2023, Art-no. 3500205, doi: 10.1109/LGRS.2022.3229141.

[2] Comparison of VGG-16, VGG-19, and ResNet-101 CNN Models for the purpose of Suspicious Activity Detection January 2023International Journal of Scientific Research in Computer Science Engineering and Information Technology DOI:10.32628/CSEIT2390124

[3] Applicability analysis of attention U-Nets over vanilla variants for automated ship detection October 2022Reports on Geodesy and Geoinformatics 114(1) DOI:10.2478/rgg-2022-0005

[4] N.Jmour, S.Zayen and A.Abdelkrim, "Convolutional neural networks for image classification," 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET), Hammamet, Tunisia, 2018, pp. 397-402, doi: 10.1109/ASET.2018.8379889.

[5] R. Chauhan, K. K. Ghanshala and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 278-282, doi: 10.1109/ICSCCC.2018.8703316.

[6] Y. Hao, "Convolutional Neural Networks for Image Classification," 2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), Hangzhou, China, 2021, pp. 342-345, doi: 10.1109/ICAICE54393.2021.00073.

[7] H. Tang, "Image Classification based on CNN: Models and Modules," 2022 International Conference on Big Data, Information and Computer Network (BDICN), Sanya, China, 2022, pp. 693-696, doi: 10.1109/BDICN55575.2022.00134.

[8] K. Mao, "Convolutional Neural Network for Image Classification Based on Satellite Imagery," 2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR), Shanghai, China, 2021, pp. 131-134, doi:

10.1109/ICBAR55169.2021.00036.
[9] J. Serra, "Image segmentation," Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), Barcelona, Spain, 2003, pp. I-345, doi: 10.1109/ICIP.2003.1246969.
[10] S. Sahu, H. Sarma and D. Jyoti Bora, "Image Segmentation and its Different Techniques: An In-Depth Analysis," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, El Salvador, 2018, pp.1-7, doi: 10.1109/RICE.2018.8509038.
[11] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 7, pp. 3523-3542, 1 July 2022, doi: 10.1109/TPAMI.2021.3059968.
[12] M. Kathiravan, N. A. Reddy, V. Prakash, B. S. Kumar, M. Malarvel and M. Sambath, "Ship Detection from Satellite Images using Deep Learning," 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp.1044-1050, doi: 10.1109/ICCES54183.2022.9835766.
[13] 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.
[14] J. Antelo, G. Ambrosio, J. Gonzalez and C. Galindo, "Ship detection and recognitionin high-resolution satellite images," 2009 IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 2009, pp. IV-514-IV-517, doi: 10.1109/IGARSS.2009.5417426.
[15] 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, Australia, 2011, pp. 507-512, doi: 10.1109/DICTA.2011.91.
[16] S. Joshi, V. Mittal, N. Rawat and R. K. Ranjan, "Ship Detection from Despeckled Satellite Images using Deep Learning," 2022 IEEE International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2022, pp. 1-7, doi: 10.1109/ICDSIS55133.2022.9915923.
[17] W. Yan, C. Chen and D. Zhang, "U-Net-based medical image segmentation algorithm," 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP), Changsha, China, 2021, pp. 1-5, doi: 10.1109/WCSP52459.2021.9613447.
[18] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," 2018 International Interdisciplinary PhD Workshop (IIPhDW), Świnouście, Poland, 2018, pp. 117-122, doi: 10.1109/IIPHDW.2018.8388338.
[19] J. Shijie, W. Ping, J. Peiyi and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," 2017 Chinese Automation Congress (CAC), Jinan, China, 2017, pp. 4165-4170, doi: 10.1109/CAC.2017.8243510.
[20] P. Hurtik, V. Molek and J. Hula, "Data Preprocessing Technique for Neural Networks Based on Image Represented by a Fuzzy Function," in IEEE Transactions on Fuzzy Systems, vol. 28, no. 7, pp. 1195-1204, July 2020, doi: 10.1109/TFUZZ.2019.2911494.
[21] Y. Lin, "Ancient Character Image Classification Model Training," 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), Changchun, China, 2022, pp. 50-54, doi: 10.1109/CVIDLICCEA56201.2022.9824485.
[22] V. Y. Mariano et al., "Performance evaluation of object detection algorithms," 2002 International Conference on Pattern Recognition, Quebec City, QC, Canada, 2002, pp. 965-969 vol.3, doi: 10.1109/ICPR.2002.1048198.
[23] https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758.
[24] An improved deep learning approach and its applications on colonic polyp images detection, July 2020BMC Medical Imaging 20(1),doi:10.1186/s12880-020-00482-3
[25] https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5
[26] Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry's Nail May 2019, Journal of Physics Conference Series 1201(1):012052, DOI:10.1088/1742-6596/1201/1/012052
[27] D. K. Jena, A. Bisoyi, A. Tripathy, "Predictive Framework for Advanced Customer Churn Prediction using Machine Learning", International Journal of Computer Applications, New York, USA 185 (13), 2023 pp. 43-51, doi: 10.1109/ICAITPR51569.2022.9844220
[28] R. K. Sahoo, A. Bisoyi and A. Tripathy, "Comparison of three classifiers used in the detection of benign tumor and malignant melanoma skin diseases," First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR), Hyderabad, India, 2022, pp. 1-6, doi: 10.1109/ICAITPR51569.2022.9844220.
[29] W. -K. Baek, M. -J. Lee and H. -S. Jung, "Land Cover Classification From RGB and NIR Satellite Images Using Modified U-Net Model," in IEEE Access, vol. 12, pp. 69445-69455, 2024, doi: 10.1109/ACCESS.2024.3401416.
[30] Z. Cui, "An Attention-Based Improved U-Net Neural Network Model for Semantic Segmentation of Moving Objects," in IEEE Access, vol. 12, pp. 57071-57081, 2024, doi: 10.1109/ACCESS.2024.3391249.