Reports on Geodesy and Geoinformatics, 2022, Vol. 114, pp. 9–14

DOI: 10.2478/rgg-2022-0005
Received: 22 May 2022 / Accepted: 28 September 2022
Published online: 25 October 2022

ORIGINAL ARTICLE

Applicability analysis of attention U-Nets over vanilla variants for automated ship detection

Pranshav Gajjar [1*], Manav Garg [1], Vatsal Shah [1], Pooja Shah [1] and Anup Das [2]

[1]Department of Computer Science and Engineering, Institute of Technology, Nirma Univeristy
[2]Space Application Centre, Indian Space Research Organisation

*19bce060@nirmauni.ac.in

Abstract

Accurate and efficient detection of ships from aerial images is an intriguing and difficult task of extreme societal importance due to their implication and association with maritime infractions, and other suspicious actions. Having an automated system with the required capabilities indicates a substantial reduction in the related man-hours of characterization and the overall underlying processes. With the advent of various image processing techniques and advancements in the field of machine learning and deep learning, specialized methodologies can be created for the said task. An intuition for the enhancement of existing methodologies would be a study on attention-based cognition and the development of improved neural architectures with the available attention modules. This paper offers a novel study and empirical analysis of the utility of various attention modules with U-Net and other subsidiary architectures as a backbone for the task of computationally efficient and accurate ship detection. The best performing models are depicted and explained thoroughly, while considering their temporal performance.

Key words: deep learning, attention, ship detection, U-Net

1 Introduction

Nowadays, the use of maritime transportation is growing at a breakneck speed. However, as transportation has increased, many new marine enterprises have emerged, increasing the enormous number of ships, marine pollution, and associated maritime offenses (Khan and Yunze, 2018). An automated system capable of monitoring sea traffic accurately and efficiently implicates immoderate societal impacts. The obtained system would lead to the replacement of excessive and extreme manpower which is required for constant monitoring and a successful deployment acts as the primary driver for the subsequently illustrated research. There is a need to reduce maritime infractions, related suspicious activities, and related pollution effects, hence the overall categorization accuracy and temporal trade-offs are important considered metrics and parameters (Khan and Yunze, 2018). With the availability of a variety of data gathering tools and technologies, aerial photographs using satellites, drones, and other related technologies can be acquired feasibly (Holloway and Mengersen, 2018). A sufficient quantity of the available image-based data also leads to possible experiments of predictive analysis with deep learning (Gajjar et al., 2022). The methodologies which revolve around deep neural architecture are perceived as a highly relevant science that has shown promising results in a multitude of related paradigms like oil spill detection (Bianchi et al., 2020; Mehta et al., 2021), moving and stationary target acquisition (Coman and Thaens, 2018), and image denoising (Yu and Sapiro, 2011; Zhang et al., 2016). The subsidiary techniques of deep learning are designed to imitate human cognitive abilities, and improvements in the architectural aspects and other additions might provide further improvements for a thorough predictive analysis.

There has been a considerable amount of work and implementations in the recent literature that helped this novel study. The relevant papers and articles associated with this study are further described in a condensed form. The use of satellite imagery to detect ships has been investigated as a problem statement for both segmentation and object detection. In the article by Zhu et al. (2010), a space-based optical picture was investigated for ship detection. To reduce the negative impacts of clouds, ocean waves, and tiny islands, the paper proposed basic shape analysis, picture segmentation, and a semi-supervised hierarchical classification algorithm. The study, as described in Yang et al. (2015), showed the application of the classic local binary pattern recognition method on higher resolution photos to find probable ship prospects using a multi-stage approach. An early technique proposed in the paper by Huang et al. (2011) involves the use of textural cues to distinguish between sea and ship. The confidence mapping of extracted ship candidates was used in the procedure. Different waves, light variations, ship sizes, and bright/dark intensities do not affect the suggested technique. The disadvantage of the analysis was that there were fewer test cases. Yang et al. (2015) conducted a ship detection investigation using the optical satellite image's visual search technique. A shape and neighborhood similarity analysis was conducted which eliminated false alarms such as clouds, islands, and waves using a global contrast model.

Li et al. (2017b) investigated the automated detection and identification of ships on the coast using big satellite photos. The scale-invariant feature transform (SIFT) record was leveraged for eliminating the geographical coordinate error between the port template and the test picture. The regions of interest (ROI) are then removed and horizontally aligned. The scroll window approach is then used to apply the learned multi models over the ROIs, resulting in ship candidates. Finally, the fusion approach is used to merge candidates. The study, as presented in Morillas et al. (2015), indicates that the use of breaking a high-dimensional Synthetic Aperture Radar (SAR) data sample into multiple fixed-size images was present and further offered a predictive analysis by leveraging the Support Vector Machine (SVM) methodology. In the paper by Mehta et al. (2022), the author proposed the use of V-Nets and augmentative approaches for a related field of oil spill detection and remote sensing. Another set of architectures that has shown utility in statements concerning associated domains and object detection is the Feature Pyramid Network, as evidenced by Shamsolmoali et al. (2021) and Zareapoor et al. (2021). The Faster-R-CNN methodology was proposed in the paper by Li et al. (2017a) to detect ships from SAR images. This methodology could adequately detect large ships but failed to detect smaller ships.

Analogous to psychology, the term attention can be understood as obtaining specific concentrations or areas of interest which the subject focuses on, concerning deep learning, technologies and additions associated with attention also perform similarly (He and Wang, 2022). The selective cognitive process does implicate superlative results and is a relevant and important area of research in applied machine learning. As considerably sparse research exists on the intersection of attention-assisted deep learning and the task of accurate ship detection, this paper focuses on various attention-based implementations and uses the famous and reliable U-Net architecture (Ronneberger et al., 2015) as a backbone for this study. This direction of neural architectures is highly inspired by the paper by Karki and Kulkarni (2021), in which the utility of U-Nets for this domain of accurate ship detection was proposed. By scouring the literature and available articles, various models were obtained that were based on the U-Net and established architectures that coupled various attention strategies with the neural network and its subsidiary approaches. A total of seven neural architectures were tested while considering their real-time utility and associated temporal analysis, which would enable us to obtain a usable computational trade-off and gauge the utility of attention modules and the tested approaches. There are four sections to this study. The Methodology is described in Section 2. The outcomes of the trials are analyzed and discussed in Section 3, which is followed by the conclusion and the future scope of this novel empirical study.

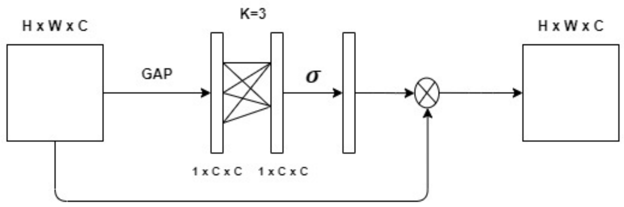

Figure 1. Pictorial representation for Efficient Channel Attention (ECA) and underlying procedures (Niu et al., 2021).

2 Methodology

With the motivation of testing and validating the domain-specific utility of attention mechanisms in deep learning and ship detection, a variety of models were tested. By checking the available literature and by understanding the possible combinations of attention modules and different U-Net variants, a total of seven different architectures are assessed. The architectures include the U-Net, Convolutional Block Attention Module (CBAM) based U-Net (Trebing et al., 2021), Efficient Channel Attention (ECA) based U-Net (Shan et al., 2021), Residual U-Net (Khan et al., 2022), CBAM based Residual U-Net (Wang et al., 2022), U-Net++ (Zhou et al., 2020) and CBAM based U-Net++ (Zhao et al., 2021). This section contains extensive information on the two attention mechanisms and the three U-Net types.

2.1 Attention Modules

The use of the attention layers is an enhanced cognition experience where the deep architecture is assisted in memorizing large sequences of data. The paper mainly experiments with two attention mechanisms, ECA (Wang et al., 2020), and CBAM (Woo et al., 2018).

2.2 Efficient Channel Attention

The methodology is built on the foundation of enforcing cross-channel interaction while requiring no dimensionality reduction during the computation of channel attention. ECA is a very lightweight and efficient channel attention module that does not require any dimensionality reduction during the computation of channel attention (Wang et al., 2020). The parametric overhead added by ECA is small and directly comparable to the kernel size used in the 1-D convolution layer, which is intrinsically present in ECA and can be calculated as follows: Several key core principles of ECA, such as Cross Channel Interaction (CCI) and avoiding Dimensionality Reduction (DR), will be discussed in detail in the following subsections (Wang et al., 2020). The avoidance of dimensionality reduction and the capacity to perform efficiently and adequately within the intrinsic spatial dimensions were the most significant advancements made by the aforementioned module. The functionality of ECA can be better understood by the Figure 1.

2.3 Convolutional Block Attention Module

Channel Attention Module (CAM) and Spatial Attention Module (SAM) are two consecutive sub-modules of the CBAM that are used in that specific order: CAM and the SAM. The authors of the original paper point out that CBAM is applied at every convolutional block in deep networks, this is proposed to acquire subsequent 'Refined Feature Maps' from the input intermediate feature maps (Woo et al., 2018). The functionality of CBAM can be better understood by the Figure 2.

Both the CAM and SAM play a crucial part in robust attention, and they are founded on the idea that channel attention states which

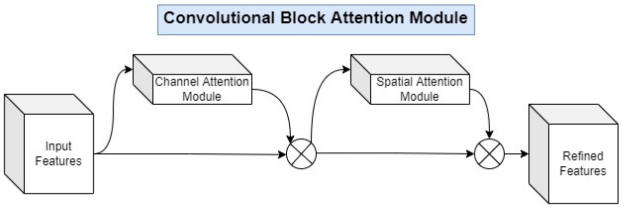

Figure 2. Overview of the Convolutional Block Attention Module (CBAM) and the underlying modules (Woo et al., 2018).

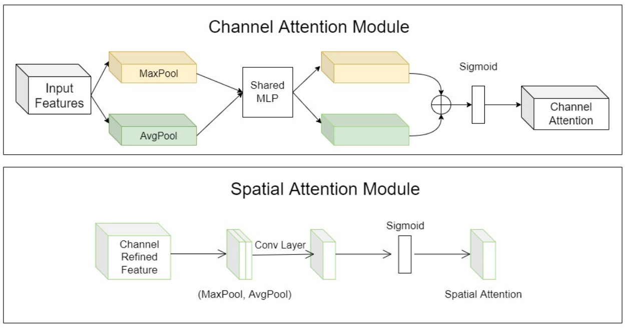

Figure 3. The Channel Attention Module (CAM) and the Spatial Attention Module (SAM) (Woo et al., 2018).

feature map is relevant for learning and improves, or further refines, the feature map in question. Meanwhile, spatial attention sends information about what is important to learn inside the feature map. Combining the two significantly improves the robustness of the Feature Maps and, as a result, justifies the large gain in model performance. In the pipeline being used, the Channel Attention Module is an integral component (Woo et al., 2018). The CAM essentially provides a weight for each channel and, as a result, enhances those specific channels that are most contributing to learning and, as a result, extends the prevalent model performance. The inherent working of CBAM and its underlying methodologies is further mentioned below.

The term 'spatial' refers to the domain space included inside a single feature map. Spontaneous spatial attention reflects the attention mechanism/attention mask on the feature map or a single cross-sectional slice of the tensor, as shown in the Figure 3. To function properly, the SAM must perform a three-fold sequential action. The first portion of it is referred to as the Channel Pool, and it consists of the Input Tensor being decomposed into two channels, with each of the two channels representing Max Pooling and Average Pooling across the channels, respectively (Woo et al., 2018).

3 U-Net

The paper leverages experiments with three main U-Net subtypes, the standard U-Net, the U-Net++, and the Residual U-Net. The last two were heavily inspired by the original architecture, which is currently perceived as an industry standard and a reliable model choice for a plethora of tasks. The U-Net architecture was developed primarily for the semantic segmentation of biomedical images (Ronneberger et al., 2015). The contraction path, also known as the encoder, and the extension path, also known as the decoder, form the architecture. The encoder consists of two 2×2 convolution operations. Each convolution operation is an unfilled convolution using a rectified linear unit (ReLU) after each convolution, followed by a 2×2 maximum pooling operation using stride 2 for downsampling. Performing downsampling doubles the feature channels. While in the decoder the feature map upsampling is followed by a 2×2 convolution ('up convolution') that halves the number of feature channels, a concatenation with a proportional cut feature map from the encoder, and two 3×3 convolutions, each of which is followed by ReLU. Cropping is used to prevent edge pixels from being lost each time you fold. Using a 1×1 convolution layer, map the feature vectors of each of the 64 components to the required number of classes in the final layer. The architecture has a total of 23 convolutional layers (Ronneberger et al., 2015). A graphical description of the architecture is mentioned in the Figure 4.

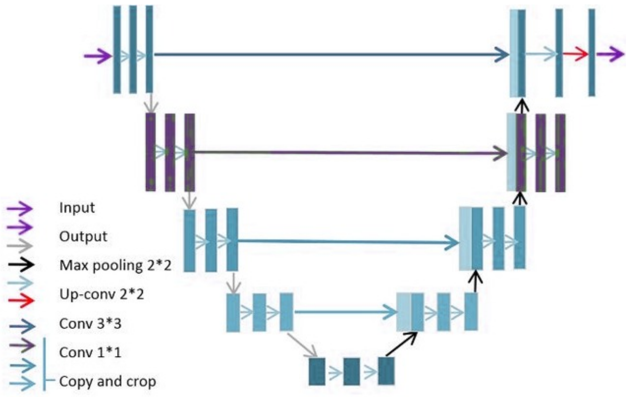

Figure 4. The standard U-Net architecture as mentioned in the article (Ding et al., 2019).

3.1 Residual U-Net

This is a fully collapsed neural network that enhances the standard U-Net design by allowing deep residual learning. The decoder and encoder algorithms are the same as those used by U-Net (Zhang et al., 2018). U-Net uses two 3×3 convolution functions, followed by a ReLU as the activation function. However, the residual UNet replaces these layers with pre-activated residual blocks. The encoder sends the input image through the encoder block. The encoder consists of three encoder blocks connected by the remaining pre-activated blocks and the output of each encoder block acts as a skip connection to the corresponding decoder block. To half the spatial dimensions, it employs the same methods as U-Net (Zhang et al., 2018). The bridge consists of the remaining pre-activated blocks with a stride value of 2. The decoder is made up of three decoder blocks, each of which doubles the spatial dimensions of the feature map and reduces the number of feature channels. The Residual U-Net decoder follows the same pattern as the U-Net decoder for upsampling and concatenation of feature maps with skip connections from the encoder block, and 1×1 convolution is used in the final layer to acquire the appropriate number of classes (Zhang et al., 2018).

3.2 U-Net++

The main motivation for introducing U-Net++ is that U-Net transfers the feature map directly from the encoder to the decoder network (Zhou et al., 2020). This concatenates semantically different features. U-Net++ aims to solve the described problem by including Dense block and convolutional layers between the encoder and the decoder network. Redesigned skip pathways are depicted in purple, thick skip connections are depicted in yellow, and deep supervision is depicted in blue in Figure 5. In U-Net++, the output of the convolution layer before the same high-density block is fused with the upsampled output of the convolution layer of the lower density block (Zhou et al., 2020). As a result, these skip routes assist in closing the semantic gap between encoder and decoder subpaths. Because of the dense convolution block alongside every skip path-

Table 1. Obtained empirical data for the proposed experimented methodologies

Model	Accuracy	Precision	Recall	F1 Score	Mean IoU
U-Net	0.972	0.195	0.648	0.254	0.143
Attention U-Net (CBAM)	0.978	0.211	0.609	0.255	0.159
Attention U-Net (ECA)	0.979	0.194	0.581	0.246	0.167
Residual U-Net	0.991	0.328	0.317	0.249	0.230
Attention Residual U-Net (CBAM)	0.974	0.198	0.723	0.266	0.1599
U-Net ++	0.992	0.2531	0.317	0.219	0.248
Attention U-Net++ (CBAM)	0.998	0.283	0.384	0.262	0.201

Table 2. The related temporal data

Model	Training Time [s]	Testing Time [s]
U-Net	219.58	5.890
Attention U-Net (CBAM)	222.01	5.873
Attention U-Net (ECA)	222.58	5.570
Residual U-Net	312.52	11.658
Attention Residual U-Net (CBAM)	372.62	16.140
U-Net ++	292.25	11.982
Attention U-Net++	490.716	18.968

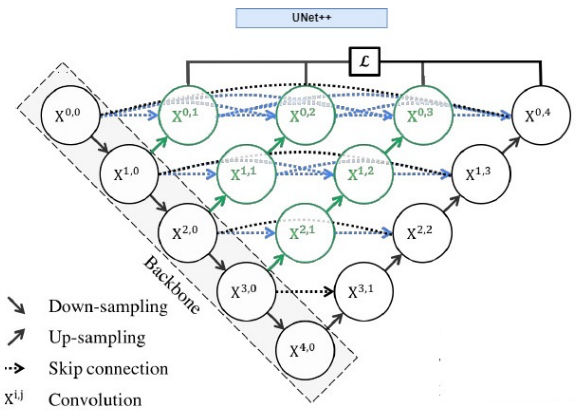

Figure 5. The U-Net++ architecture as depicted in the original paper (Zhou et al., 2020).

way, dense skip connections are furnished to growth gradient waft and make certain that each preceding characteristic map is accrued and arrives at the existing node. This produces feature maps with full resolution at many semantic levels. Deep supervision helps the model adjust the complexity between speed and performance (Zhou et al., 2020). The above-mentioned eccentricities provide a sufficiently balanced viewpoint for the inclusion of architecture in this study. The structure can be better understood by the Figure 5.

4 Results

The empirical analysis and the obtained results for the tested methodologies are further described in an extensive fashion. The paper leverages the Airbus Ship detection dataset, as used in the research Karki and Kulkarni (2021), by undersampling 12788 images and obtaining a relatively balanced class distribution for the leveraged subset. The images were rescaled to an image size of 256×256, to support computational constraints and an efficient training experience. Each tested neural architecture and methodology is tested on an identical train-test split which was designed using a stratified strategy to ensure a balanced training regimen and to have an unbiased validation of each tested neural architecture. The totality of the hyperparameters like the learning rate, batch sizes, and the associated constraints are obtained through sufficient experimentation. The train test distribution follows a ratio of 4:1 and is generated in an offline fashion. Each network is assessed by leveraging several metrics, like the percent accuracy, Precision, Recall, F1 score, and Mean IoU as available from the computational module presented in Trappenberg (2019), and is trained for an identical number of epochs.

From the Tables 1 and 2, it can be inferred that the utility of Attention mechanisms in standard neural strategies is justified for relatively simpler architectures like a vanilla U-Net, as an increase in 0.6 and 0.7 percent accuracy is observed for CBAM and ECA enhancements respectively. Similar trends are observed for the other tested metrics for both the attention strategies. When a vanilla U-Net was considered, despite higher training periods, 2.43 seconds and 2.68 seconds for CBAM and ECA respectively, the attention modules have shown a decrease in the overall testing efficiency. However, the modules failed when Residual U-Nets and U-Net++ were considered. The associated temporal characteristics for both the training phase and the real-time predictive phase were also trivially inadequate for CBAM-based attention enhancements. These results also explain that because of a smaller dataset in consideration and the relatively small sample size, the simpler methodologies could be trained to a sufficient extent. So, if implementations and utilities, where a sparse dataset size is present or dataset constraints are observed, leveraging a vanilla U-Net might not be the optimal architecture strategy and attention enhancements have substantial perks. But considering the recent literature, and the experiments in the scope of this paper, for a relatively larger sample space, computational capabilities, and data entities, experimentation concerning attention does implicate a plausible use case. To better apprehend the outcomes of these strategies more suitably, sample predictions of these networks are mentioned in the Figure 6.

5 Conclusion

This paper aimed to assess the applicability of various attention modules and strategies available in the recent literature for accurate and efficient segmentation and hence detection of ships from aerial images. The research leveraged a subset of the famous publicly available Airbus Ship detection dataset for obtaining a novel empirical study and heralds the behavior of various neural strategies in a relatively computationally constrained environment. This implicated the relative utility of these methodologies and helps us obtain a computational centric trade-off in various domains which resemble and corresponds to remote sensing and aerial object detection and have a sparse data sample size. The thorough experimentations

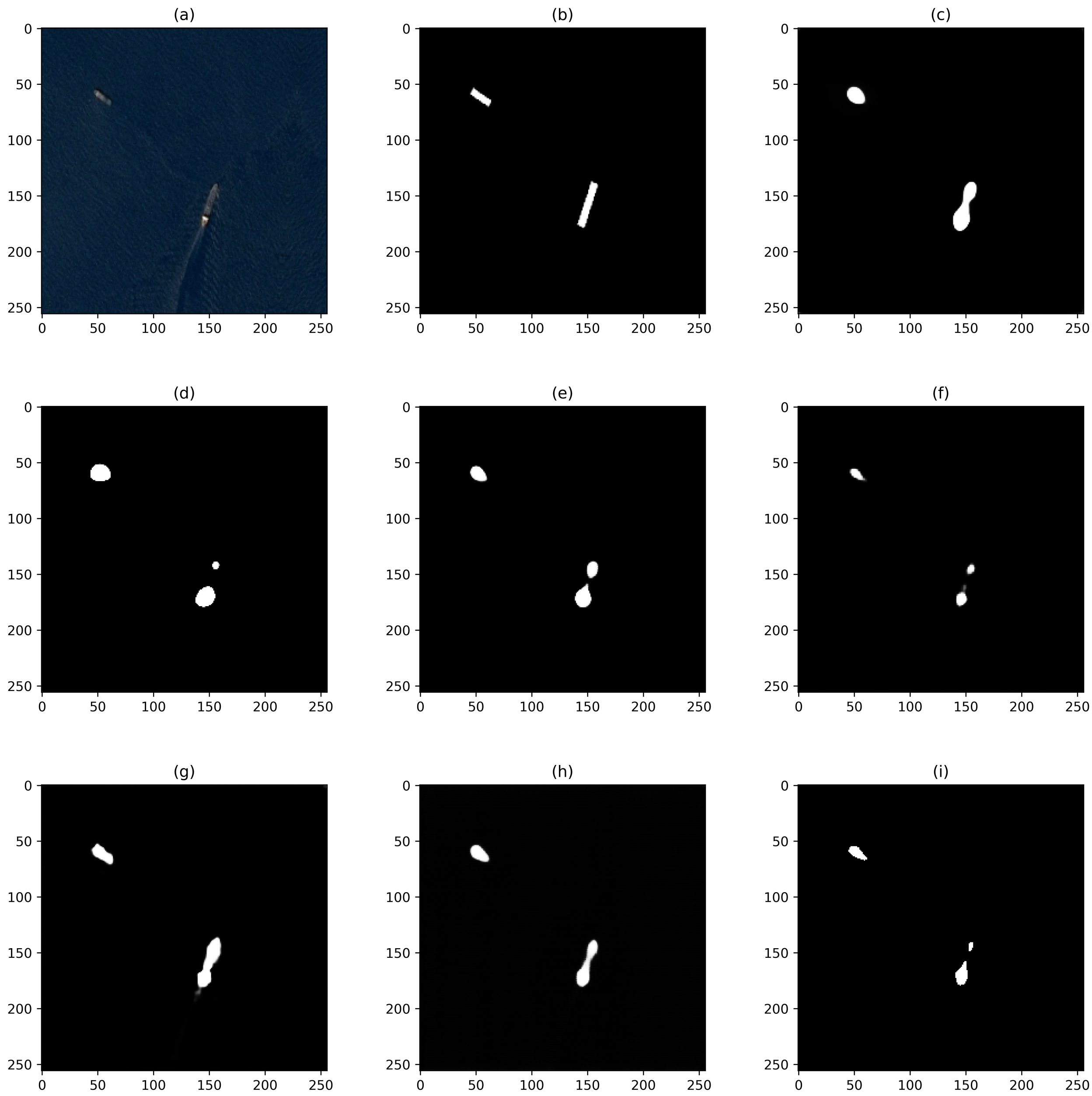

Figure 6. Here (a) represents the aerial view, (b) represents the ground truth, (c) depicts the results obtained from a standard U-Net, (d) represents the results obtained from a CBAM enhanced U-Net, (e) depicts the ECA enhanced U-Net, (f) indicates Residual U-Net, (g) represents CBAM enhanced Residual U-Net, (h) depicts a standard U-Net++, (i) represents the CBAM enhanced U-Net++ architecture.

can be leveraged to conclude the utility of attention to enhance a comparably simple architecture as non-trivial enhancements were observed, however in the case of relatively complex architectures like U-Net++ and Residual U-Nets, the technique has an inverse effect. For future work, the authors aim to leverage novel attention mechanisms for a larger spatial extent and aim to mitigate the computational predicaments and computing needs associated with attention-based cognition.

References

Bianchi, F. M., Espeseth, M. M., and Borch, N. (2020). Large-scale detection and categorization of oil spills from SAR images with deep learning.

Coman, C. and Thaens, R. (2018). A deep learning SAR target classification experiment on MSTAR dataset. In *2018 19th International Radar Symposium (IRS)*. IEEE, doi:10.23919/irs.2018.8448048.

Ding, Y., Chen, F., Zhao, Y., Wu, Z., Zhang, C., and Wu, D. (2019). A stacked multi-connection simple reducing net for brain tumor segmentation. *IEEE Access*, 7:104011–104024, doi:10.1109/access.2019.2926448.

Gajjar, P., Shah, P., Vegada, A., and Savalia, J. (2022). Triplet loss for chromosome classification. *Journal of Innovative Image Processing*, 4(1):1–15, doi:10.36548/jiip.2022.1.001.

He, Y. and Wang, S. (2022). SE-BLTCNN: A channel attention adapted deep learning model based on PSSM for membrane protein classification. *Computational biology and chemistry*, 98:107680, 35421797, doi:10.1016/j.compbiolchem.2022.107680.

Holloway, J. and Mengersen, K. (2018). Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sensing*, 10(9):1365, doi:10.3390/rs10091365.

Huang, G., Wang, Y., Zhang, Y., and Tian, Y. (2011). Ship detection using texture statistics from optical satellite images. In *2011 International Conference on Digital Image Computing: Techniques and Applications*. IEEE, doi:10.1109/dicta.2011.91.

Karki, S. and Kulkarni, S. (2021). Ship detection and segmentation using Unet. In *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*. IEEE, doi:10.1109/icaect49130.2021.9392463.

Khan, H. M. and Yunze, C. (2018). Ship detection in SAR image using yolov2. In *2018 37th Chinese Control Conference (CCC)*. IEEE, doi:10.23919/chicc.2018.8482863.

Khan, R. A., Luo, Y., and Wu, F.-X. (2022). RMS-UNet: Residual multi-scale UNet for liver and lesion segmentation. *Artificial Intelligence in Medicine*, 124:102231, doi:10.1016/j.artmed.2021.102231.

Li, J., Qu, C., and Shao, J. (2017a). Ship detection in SAR images based on an improved faster R-CNN. In *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*. IEEE, doi:10.1109/bigsardata.2017.8124934.

Li, W., Fu, K., Sun, H., Sun, X., Guo, Z., Yan, M., and Zheng, X. (2017b). Integrated localization and recognition for inshore ships in large scene remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 14(6):936–940, doi:10.1109/lgrs.2017.2688357.

Mehta, N., Shah, P., and Gajjar, P. (2021). Oil spill detection over ocean surface using deep learning: a comparative study. *Marine Systems & Ocean Technology*, 16(3-4):213–220, doi:10.1007/s40868-021-00109-4.

Mehta, N., Shah, P., Gajjar, P., and Ukani, V. (2022). Ocean surface pollution detection: Applicability analysis of V-Net with data augmentation for oil spill and other related ocean surface feature monitoring. In *Communication and Intelligent Systems*, pages 11–25. Springer.

Morillas, J. R. A., Garcia, I. C., and Zolzer, U. (2015). Ship detection based on SVM using color and texture features. In *2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*. IEEE, doi:10.1109/iccp.2015.7312682.

Niu, J.-Y., Xie, Z.-H., Li, Y., Cheng, S.-J., and Fan, J.-W. (2021). Scale fusion light CNN for hyperspectral face recognition with knowledge distillation and attention mechanism. *Applied Intelligence*, 52(6):6181–6195, doi:10.1007/s10489-021-02721-8.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, pages 234–241. Springer International Publishing, doi:10.1007/978-3-319-24574-4_28.

Shamsolmoali, P., Chanussot, J., Zareapoor, M., Zhou, H., and Yang, J. (2021). Multipatch feature pyramid network for weakly supervised object detection in optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13.

Shan, W., Wang, Y., and Lu, W. (2021). ECA-UNet: Denoise seismic data by learning from traditional method. In *First International Meeting for Applied Geoscience & Energy Expanded Abstracts*. Society of Exploration Geophysicists, doi:10.1190/segam2021-3583394.1.

Trappenberg, T. P. (2019). Machine learning with sklearn. In *Fundamentals of Machine Learning*, pages 38–65. Oxford University Press, doi:10.1093/oso/9780198828044.003.0003.

Trebing, K., Stanczyk, T., and Mehrkanoon, S. (2021). SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognition Letters*, 145:178–186, doi:10.1016/j.patrec.2021.01.036.

Wang, J., Yu, Z., Luan, Z., Ren, J., Zhao, Y., and Yu, G. (2022). RDAU-Net: Based on a residual convolutional neural network with DFP and CBAM for brain tumor segmentation. *Frontiers in Oncology*, 12, doi:10.3389/fonc.2022.805263.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, doi:10.1109/cvpr42600.2020.01155.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). CBAM: convolutional block attention module. In *Computer Vision – ECCV 2018*, pages 3–19. Springer International Publishing, doi:10.1007/978-3-030-01234-2_1.

Yang, F., Xu, Q., Gao, F., and Hu, L. (2015). Ship detection from optical satellite images based on visual search mechanism. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, doi:10.1109/igarss.2015.7326621.

Yu, G. and Sapiro, G. (2011). DCT image denoising: a simple and effective image denoising algorithm. *Image Processing On Line*, 1:292–296, doi:10.5201/ipol.2011.ys-dct.

Zareapoor, M., Chanussot, J., Zhou, H., Yang, J., et al. (2021). Rotation equivariant feature image pyramid network for object detection in optical remote sensing imagery.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2016). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. doi:10.1109/TIP.2017.2662206.

Zhang, Z., Liu, Q., and Wang, Y. (2018). Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, doi:10.1109/lgrs.2018.2802944.

Zhao, Z., Chen, K., and Yamane, S. (2021). CBAM-Unet++:easier to find the target with the attention module "CBAM". In *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*. IEEE, doi:10.1109/gcce53005.2021.9622008.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2020). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, doi:10.1109/tmi.2019.2959609.

Zhu, C., Zhou, H., Wang, R., and Guo, J. (2010). A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Transactions on Geoscience and Remote Sensing*, 48(9):3446–3456, doi:10.1109/tgrs.2010.2046330.