INVESTIGATING LOSS FUNCTIONS FOR SEGMENTING AND DETECTING SHIPS ON SAR IMAGERY

M. L. R. Lagahit [1, 2] * and M. Matsuoka [1, 2]

[1] Department of Architecture and Building Engineering, Tokyo Institute of Technology, Japan
[2] Tokyo Tech Academy for Super Smart Society, Tokyo Institute of Technology, Japan
* lagahit.m.aa@m.titech.ac.jp

KEY WORDS: Mobile Mapping, Point Cloud-Derived Imagery, Deep Learning, Semantic Segmentation, Loss Function

ABSTRACT:

In accordance with the United Nations (UN) Sustainable Development Goal (SDG) 16: Peace, Justice, and Strong Institutions, this study explores ship monitoring through the use of Synthetic Aperture Radar (SAR) for its potential applications to economic and security purposes. One method to extract ships through SAR-derived imagery is to employ the use of convolutional neural networks (CNN). However, the extraction of small features continues to be a challenging task for CNNs. To improve the performance in such cases, one way is to employ the use of an appropriate loss function, which helps guide the CNN model during training. In this paper, Focal Combo (FC) loss, a recent loss function designed for extreme class imbalance, will be investigated to analyze its effects when applied to ship extraction. In doing so, this paper also presents a thorough comparison of existing loss functions in their capability to segment and detect ships on SAR imagery. Making use of the U-Net model, our results demonstrate that by using FC loss we can observe an increase in segmentation of about 9% in terms of f3-score and a decrease in missed detections by about 17 ships (after post-processing) when compared to cross-entropy loss. Unfortunately, it has also shown a significant drop in precision of about 35% resulting in an additional 270 ships being incorrectly detected in the background. In future work, varying CNN models shall be tested to see if the pattern persists and several trials shall be conducted to assess consistency.

1. INTRODUCTION

1.1 Background

In 2015, the United Nations (UN) introduced the 2030 Agenda for Sustainable Development. Within this agenda lies 17 sustainable development goals (SDG), which "recognize that ending poverty and other deprivations must go hand-in-hand with strategies that improve health and education, reduce inequality, and spur economic growth – all while tackling climate change and working to preserve our oceans and forests" (United Nations, n.d.). One of which is goal 16: peace, justice, and strong institutions, which aims for building "effective, accountable and inclusive institutions at all levels", among others. This work anchors on this SDG, by providing a possible means to improve maritime surveillance increasing safety, and ensuring accountability in cases of issues in traffic or environment-related accidents.

In this paper, we tackle the task of ship detection on Synthetic Aperture Radar (SAR) imagery. SAR images have long been used for monitoring ships for economic and security purposes (Li et al., 2017). A method to automatically detect ships from SAR imagery is through the use of deep learning techniques like convolutional neural networks (CNN). However, CNNs are known to struggle when it comes to extracting poorly represented features in comparison to the background, much like the ships as compared to the open ocean. One way to work around this issue is by utilizing a loss function designed specifically for such extreme class imbalance, like Focal Combo (FC) loss.

1.2 Objective

The objective of this paper is to evaluate the effect of FC loss on the performance of a CNN model in detecting ships from SAR imagery. In doing so, we also present an evaluation of other loss functions as well as the exploration of the F_3-Score as an alternative assessment metric.

2. METHODOLOGY

2.1 Dataset and CNN Model

The SAR imagery dataset, SSDD, that was used for this paper was obtained from the works of (Li et al, 2017). The images were taken by RadarSat-2, TerraSAR-X, and Sentinel-1 satellites and contain both ships on/near the docks and ships in the open sea. For more details, please refer to their paper.

As for the convolutional neural network (CNN) model, we made use of U-Net (Ronneberger et al, 2015). Its structure is designed to have a sequence of symmetrically connected convolutions. Although originally designed for biomedical segmentation, it has long been applied to other datasets that range from point cloud-derived images to satellite imagery.

2.2 Loss Functions

As mentioned earlier, the loss function is one of the components of the model training procedure that takes the differences between initial predictions and labeled ground truth to adjust the weights inside the network. In this paper, we evaluate Focal Combo (FC) loss, as shown in Equation 1 (Lagahit et al., 2023).

$$\textit{Focal Combo Loss} =$$
$$\alpha\left(-w(1-P_t)^{\gamma}\log(P_t)\right)+(1-\alpha)\left(1-\left(\frac{2TP}{2TP+FP+FN}\right)^{\frac{1}{\beta}}\right), \quad (1)$$

FC loss was originally designed to improve road marking extraction from sparse point cloud-derived images, which was under the case of extreme class imbalance where the background largely outnumbers that of the target. Similarly, SSDD falls under the same condition wherein the pixels representing the background ocean or dock far outnumber those of the ships. In order to highlight our observations with FC loss, we also evaluate

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-321-2024

the dataset by employing other various loss functions designed to improve model performance when detecting small or poorly represented features such as weighted, focal, and combination loss functions that were also comprehensively evaluated in the paper of FC loss.

2.3 Evaluation Metrics

To evaluate the segmentation performance of our model we use precision and recall which are common segmentation metrics. Moreover, we also introduce the use of F_β-Score, as seen in Equation 2, where β=3 to give more priority to recall (Dalianis, 2017). Recall represents the number of actual positive cases that are correctly predicted as positive. In general, it would be preferred to have the model misclassify the background ocean pixels as ships rather than not being able to detect them.

$$F_\beta\text{-}Score = (1+\beta^2)\frac{Precision \times Recall}{\beta^2 \times Precision + Recall}, \tag{2}$$

In evaluating detection, we simply impose the assumption that each cluster in the images corresponds to a single ship and compare the number to that of the label or ground truth. This is based on the assumption that the model performs well enough to be able to detect the ships even if segmented only partially. This evaluation is a simple extension of further analysis of the resulting predictions.

3. RESULTS AND DISCUSSION

3.1 Ship Segmentation

Before proceeding to evaluate the segmentation results of a CNN trained in various loss functions, we will first compare the results of segmentation through CNN and traditional means – through watershed transformation. Watershed segmentation works by considering an image as a topographic surface, grouping regions of minima (the catch basins) that are separated by walls of maxima (the watershed lines) (Beucher, 2010). Table 1 shows the results of comparing the aforementioned techniques. It can be observed that segmentation through CNN largely outperforms that of the traditional in all aspects. A reason for this can be observed in Figure 1, where the lack of features to represent bounding lines that can segregate our target ships. This demonstrates the flexibility of CNN to segment features and supports the motivation to explore components related to its improvement.

Table 2 shows the evaluation results of U-Net trained on various loss functions. In our tests, focal combo loss yielded the third-lowest precision and third-highest recall, which means that the model tends to misclassify the pixels surrounding our target, giving importance to segmenting target pixels over maintaining correct boundaries. To get a better grasp of the segmentation results, we make use of the F_β-Score to combine precision and recall into one single evaluation metric. As stated in the methodology, in terms of F3-Score which gives more importance to recall, focal combo loss yielded the highest results.

Method	Precision	Recall	F_3-Score
Traditional	39.25	22.02	23.03
CNN	88.45	77.10	78.10

Table 1. Traditional vs CNN segmentation results (%).

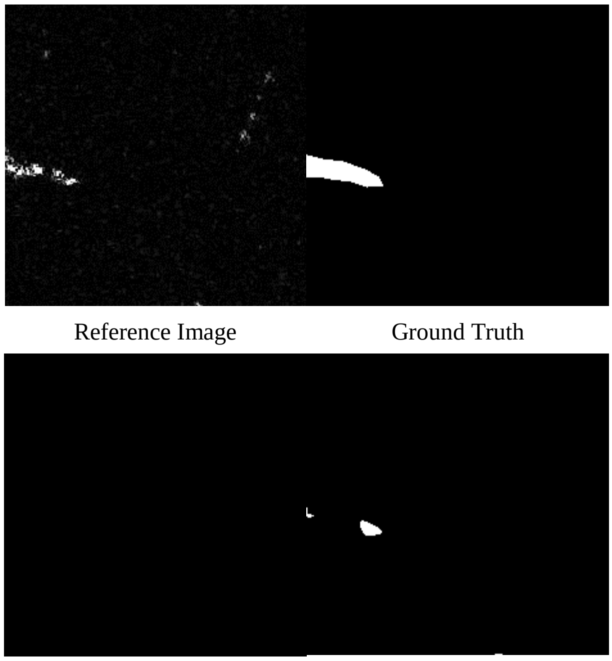


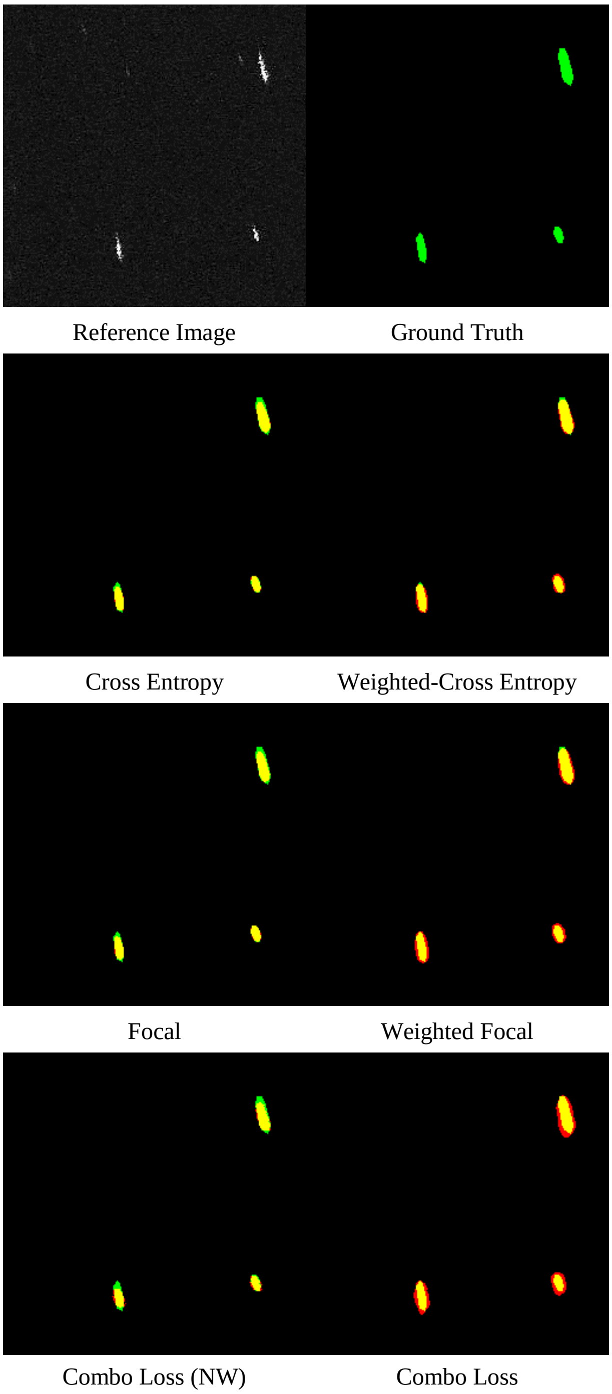
Figure 1. Sample segmentation ship segmentation using traditional and CNN methods.

Additionally, we have also observed that the addition of weights is a factor in causing the CNN to gain lower precision. We have also seen that focal versions of the weighted loss functions best their non-focal counterparts.

Loss Function	Precision	Recall	F_3-Score
Cross-Entropy	88.45	77.10	78.10
Weighted Cross Entropy	67.12	86.56	84.12
Focal Loss (γ=1)	86.42	78.31	79.05
Weighted Focal Loss (γ=1)	44.37	95.85	84.63
(Non-Weighted) Combo Loss (α=50)	86.55	64.16	65.86
Combo Loss (α=25)	48.25	93.93	85.81
(Non-Weighted) Focal Combo Loss (α=75, γ=1, β=2)	89.06	56.83	58.97
Focal Combo Loss (α=25, γ=1, β=2)	53.55	93.49	87.00

Table 2. U-Net performance on different loss functions. (%)

https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-321-2024 | © Author(s) 2024. CC BY 4.0 License.



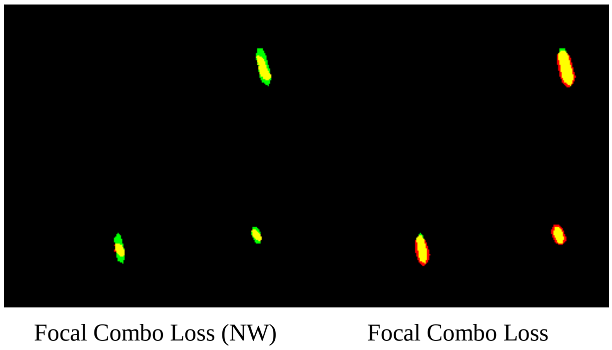


Figure 2. Sample segmentation results of using U-Net trained on various loss functions.

Taking a look at Figure 2, where green is the ground truth, red is the model's prediction, and yellow is the intersection between the two, as reflected by the evaluation metrics we can observe that weighted versions of the loss functions tend to overreach and misclassify surrounding pixels while non-weighted versions tend to only segment partially. We can see that most of the misclassifications, which cause lower precision, remain only on the edges of the ships, and very rarely can misclassifications be found on other areas of the image.

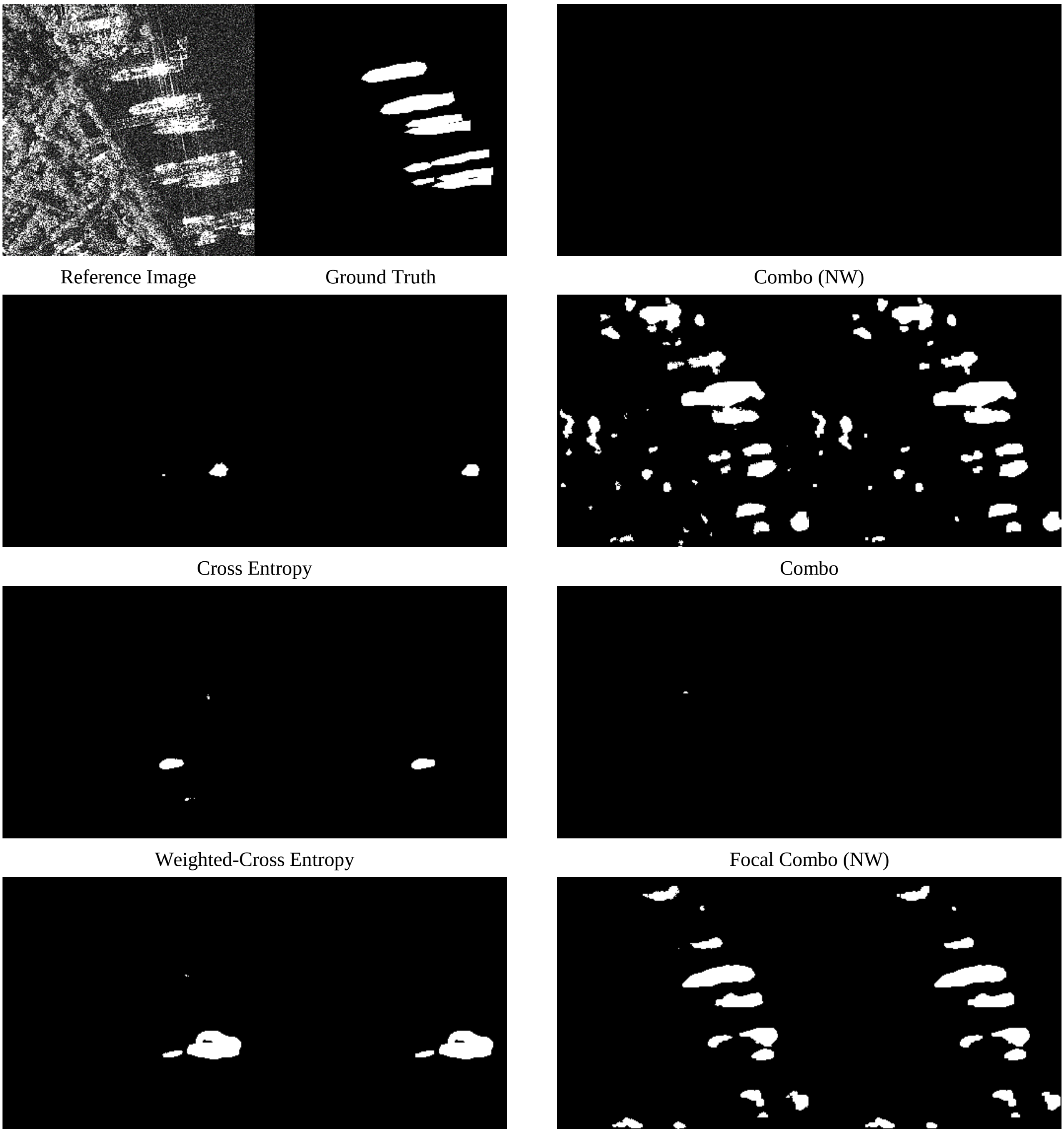
3.2 Ship Detection

As an extension, we also explore the influence of FC loss on ship detection. As we saw in Figure 3-2, even at partial segmentations all loss functions were able to guide U-Net in correctly detecting ships and determining their numbers. In this subsection, we impose the assumption that each segmented cluster represents a ship to simplify evaluation and thus can be used to represent the total number of ships in an image. In Table 3, we tally up those numbers by binning the misclassifications into background and missed errors. Background errors represent the number of clusters detected above those of the ground truth and missed errors represent the number of errors detected below. We can see that FC loss performs third highest in terms of least missed error but performs second lowest when it comes to least background error, which was as expected since FC attained low precision.

Loss Function	Total	Background	Missed
Reference	430	---	---
Cross-Entropy	**470**	**68**	28
Weighted Cross Entropy	545	126	**11**
Focal Loss (γ=1)	483	76	23
Weighted Focal Loss (γ=1)	561	152	21
(Non-Weighted) Combo Loss (α=50)	479	94	45
Combo Loss (α=25)	1242	825	13
(Non-Weighted) Focal Combo Loss (α=75, γ=1, β=2)	566	166	30
Focal Combo Loss (α=25, γ=1, β=2)	749	338	19

Table 3. U-Net detection performance on different loss functions (%)

https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-321-2024



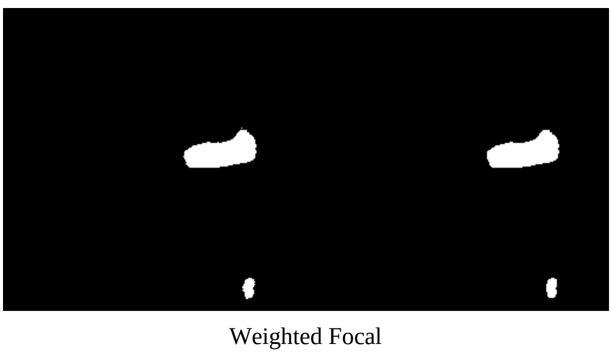


Figure 3. Except for the topmost row, sample segmentation results on various loss functions (left) before and (right) after morphological opening.

However, since our initial assumption fails when misclassifications form small clusters, we make use of the morphological opening operation (kernel = 3) to remove such noise and improve our assessment. We can see from Figure 3, that by applying such a filter such small clusters were successfully removed.

Loss Function	Total	Background	Missed
Cross-Entropy	457↓	60↓	33↑
Weighted Cross Entropy	501↓	97↓	26↑
Focal Loss (γ=1)	472↓	72↓	30↑
Weighted Focal Loss (γ=1)	524↓	117↓	23↑
(Non-Weighted) Combo Loss (α=50)	439↓	59↓	50↑
Combo Loss (α=25)	906↓	491↓	15↑
(Non-Weighted) Focal Combo Loss (α=75, γ=1, β=2)	482↓	92↓	40↑
Focal Combo Loss (α=25, γ=1, β=2)	658↓	126↓	11↓

Table 4. U-Net detection performance on different loss functions after morphological opening (%)

Loss Function	Precision	Recall	F_3-Score
Cross-Entropy	0.13↑	0.13↓	0.11↓
Weighted Cross Entropy	0.13↑	0.07↓	0.03↓
Focal Loss (γ=1)	0.11↑	0.11↓	0.09↓
Weighted Focal Loss (γ=1)	0.12↑	0.03↓	---
(Non-Weighted) Combo Loss (α=50)	0.40↑	0.13↓	0.02↑
Combo Loss (α=25)	1.14↑	1.02↓	0.90↓
(Non-Weighted) Focal Combo Loss (α=75, γ=1, β=2)	1.01↑	1.52↓	1.43↓
Focal Combo Loss (α=25, γ=1, β=2)	0.20↑	0.09↓	0.02↓

Table 5. Change in segmentation performance of U-Net on different loss functions after morphological opening (%)

After the removal of such noise, we can see from Table 4 that FC loss became the best-performing loss function in terms of least missed errors. More importantly, we can see that among other loss functions only FC loss gained an improvement in missed errors, which means that contrary to other loss functions noise is the main source of precision deterioration for FC loss. This is promising since such noise can be removed through simple methods such as the morphological opening that we have employed. To further support this claim, we also evaluated the post-processed segmentations in recall, precision, and f-score. In Table 5, we can see that that most of the changes remain at less than 1% which means that the clusters removed are only noise.

4. CONCLUSION

In this paper, we have successfully investigated the influence of focal combo (FC) loss on a CNN model in the tasks of ship segmentation and detection on SAR imagery. Moreover, at about 87% of F_3-Score, FC outperforms all other loss functions in ship segmentation. It has also been observed that, unlike other loss functions, most of what deteriorates the precision results of FC loss are small clusters that can be easily removed through traditional filtering like morphological opening. After post-processing, we can see that FC loss performs best in the least missed detections, with 11 out of 430 missed ships. In future work, further ablation studies should be provided including tests on varying structured CNN models to see pattern persistence and several more seeded trials to assess consistency.

REFERENCES

Beucher, S. N.d. Image Segmentation and Mathematical Morphology: The Watershed Transformation. people.cmm.minesparis.psl.eu/users/beucher/wtshed.html. (October 2023).

Dalianis, H. 2018. Evaluation Metrics and Evaluation. In: Clinical Text Mining. Springer, Cham.

Lagahit, M.L.R., Matsuoka, M. 2023. Focal Combo Loss for Improved Road Marking Extraction of Sparse Mobile LiDAR Scanning Point Cloud Derived Images using Convolutional Neural Networks. Remote Sensing, 15, 597.

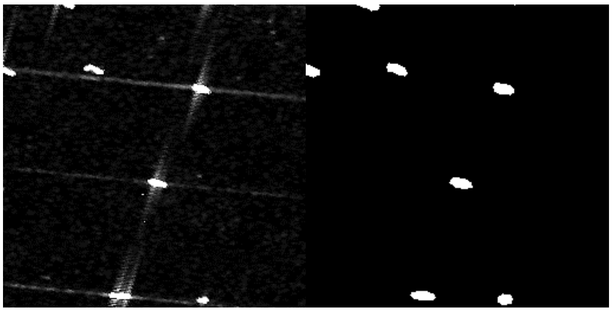
Li, J., Qu, C., and Shao, J. 2017. Ship detection in SAR images based on an improved faster R-CNN. 2017. 2017SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), 1-6.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Computing Research Repository. Retrieved October 2023, from the arXiv database.

United Nations. N.d. The 17 goals. sdgs.un.org/goals. (October 2023).

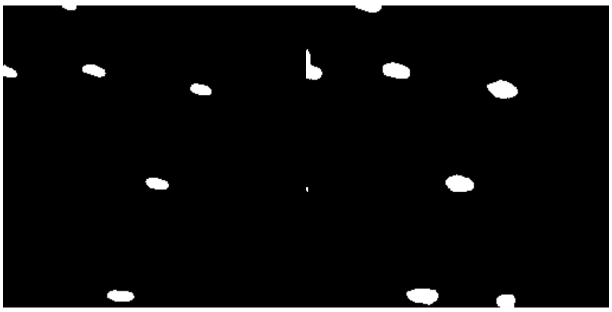
APPENDIX

Figure A-1 contains supplemental sample segmentation results that can also provide visual insights into detection. The resulting images below show clearly-spaced and well-defined features to see the background and missed errors. As an example, we can see that FC loss was able to detect 7 ships as compared to only 6 ships by cross-entropy, focal, and non-weighted combo loss.

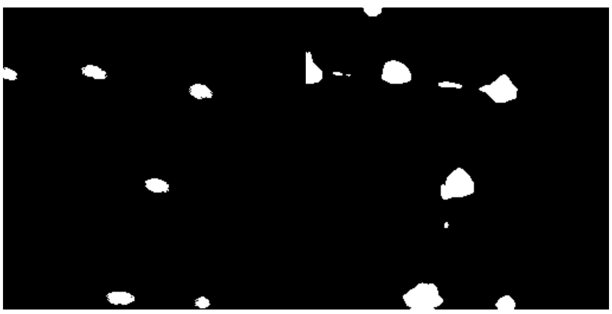


Reference Image Ground Truth

Cross Entropy Weighted-Cross Entropy



Focal Weighted Focal



Combo Loss (NW) Combo Loss

Focal Combo Loss (NW) Focal Combo Loss

Figure 4. Another sample segmentation results of U-Net on various loss functions.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-321-2024 | © Author(s) 2024. CC BY 4.0 License.