ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct244

A Comparative Study of Training Modifications for Small Object Detection in Satellite-Based Search and Rescue Missions

Gauri Todur[1]

[1]Santa Clara High School Santa Clara, USA

Publication Date: 2025/10/17

Abstract: Small object detection in high resolution satellite imagery for search and rescue (SAR) operations remains challenging, with targets sometimes 3-4 pixels in width, compared to full images of 1000-pixel resolution. Using the SaRNet dataset containing 2,552 satellite images from a real missing person search, we evaluated three modifications to a baseline Faster R-CNN Feature Pyramid Network architecture to improve the recall performance metric on small object detection. We tested (A) Focal Loss integration to address class imbalance since targets represent <0.16% of image area, (B) multi-scale training and testing at higher image resolutions (10-20% up-scaled) and (C) decreased anchor sizes. Results were mixed. Focal Loss was the only successful modification, improving small object recall by 4.4 percentage points (10.4% relative improvement) while also increasing recall on large objects. Surprisingly, both anchor optimization and multi-scale training degraded performance despite theoretical justification. Optimized anchor sizes decreased recall across all object sizes and caused the worst AR-d20 per- formance drop (-12.64 points), revealing that geometric anchor coverage doesn't guarantee detection improvement in transfer learning contexts. Multi-scale training decreased medium-sized object recall by 9.5 percentage points, contradicting recent super- resolution research. This work provides the first systematic evaluation of modifications of the baseline model for the SaRNet dataset towards improved small object detection. For operational SAR systems where lives depend on detection performance, our results recommend Focal Loss integration while cautioning against modifications that disrupt pre-trained model configura- tions.

Keywords: *Search and Rescue (SAR), Small Object Detection, Satellite Imagery, Faster R-CNN, Remote Sensing, SaRNet Dataset, Disaster Response, Focal Loss, Anchor Optimization.*

How to Cite: Gauri Todur (2025) A Comparative Study of Training Modifications for Small Object Detection in Satellite-Based Search and Rescue Missions. *International Journal of Innovative Science and Research Technology*, 10(10), 911-917. https://doi.org/10.38124/ijisrt/25oct244

I. INTRODUCTION

Search and rescue (SAR) operations in remote terrain re- quire rapid coverage of vast inaccessible areas where traditional ground searches are impractical or dangerous to manually search through. High-resolution satellites present a solution to this need due to their ability to survey areas within hours. However, a critical challenge lies in detecting small targets of interest, like missing persons, that occupy only a few pixels in satellite images.

This study uses the Search and Rescue dataset (SaRNet) from [1], which contains 2,552 satellite images of 1000×1000 pixels each, collected during a real search operation for a missing paraglider pilot. The dataset includes 4,206 bounding box annotations marking potential targets (paragliding wings, parachutes, etc.) identified by volunteers. The images were split into a training, validation and test set for deep learning application. Figure 1 presents the distribution of ground truth bounding box areas from the test set, demonstrating the predominance of small objects in SAR scenarios. The smallest bounding box in this test set was 12 square pixels in area.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct244

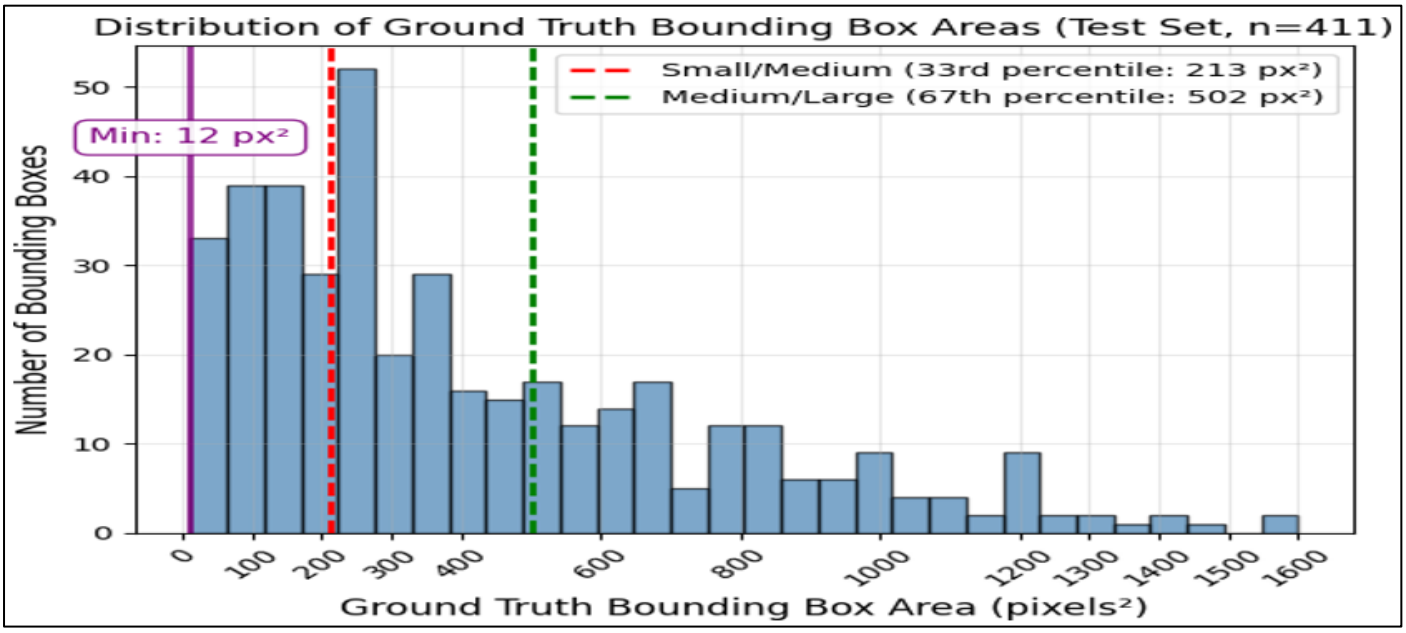

Fig 1 The Size Distribution of the Bounding Boxes in the SaRNet Test Dataset. Test Set Sizes were used for Size Categorization because the Models' Recall Scores, which are Calculated by Size Groups, are Based on Performance on the Test Set Images.

For this study, test set objects were categorized into three groups based on tertiles of the test set bounding box pixel area distribution as decided: "small" (< 213 pixels²), "medium" (213-501 pixels²), and "large" (> 501 pixels²). This categorization ensures equal representation, with 137 test bounding box objects in each size category. Figure 2 presents cropped satellite images of a small, medium, and large bounding box object from the SaRNet test set. Evidently, the smallest target appears nearly impossible to distinguish, while the medium and large ones are only marginally more visible. The cluttered and diverse landscapes surrounding the targets create further challenges regarding visibility.

The authors of [1] utilize the Faster Region-based Convolutional Neural Network (Faster R-CNN) with a ResNet-50 Feature Pyramid Network (FPN) backbone, implemented in Detectron2 [6], an open-source object detection model library, to assess the usability of their created dataset. The Faster- RCNN architecture, introduced by [2], is a two-stage object detection framework that has demonstrated robust performance across diverse detection tasks. The architecture consists of a Region Proposal Network (RPN) that generates object propos- als, followed by a classification and regression head that lo- calize and filter proposals into final detections. The ResNet-50 backbone leverages FPNs [7] to construct multi-scale feature representations. This hierarchical feature extraction enables detection of objects at different scales by utilizing feature maps from multiple network layers. The Faster-RCNN model was pre-trained on the MS-COCO dataset. In this research, the model titled "faster_rcnn_R_50_FPN_3x" in the Detectron2 model zoo with parameters defined by authors of [1] fine tuned on the labeled SaRNet dataset was considered the baseline model, since they found it to yield the highest performance on their dataset. Their custom performance metric, "AR-d20", is the Average Recall-Density to 20, which measures the average recall across detection density thresholds from 0 to 20 detections per square kilometer, providing an operationally relevant evaluation related to the human resources needed for verifying candidate detections in real SAR missions. The baseline model's AR-d20 was calculated to be 41.82.

Fig 2 Examples of SaRNet Test set Images Cropped Around a Labeled Small, Medium, and Large Sized Bounding Box Target Object (Based on this Study's Size Categorization).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct244

A method more commonly used to assess the performance of machine learning models is recall, defined as the number of correctly identified targets divided by the total number of targets to be found. This research calculated the recall of the baseline model for test set targets, grouped by object size. Recall is prioritized by [1] because in SAR operations, finding all possible targets for visual inspection is prioritized over being selective about predictions for the sake of accuracy. The data in Table 1 show a clear decrease in recall with decreasing object size, using the baseline model. Small object recall is only 42.3% which is 35.1 percentage points lower than the recall for large objects. In a real SAR mission, this means roughly 42.3% of small objects will be recovered by the baseline model. This performance gap motivated our investigation into training modifications specifically targeted at improving small object detection.

Table 1 Baseline Model Recall Scores

Baseline Model Recall (%) on Test Set Bounding Boxes		
Small	**Medium**	**Large**
42.3	71.5	77.4

This paper presents an evaluation of three modifications to the baseline Faster R-CNN architecture: integration of Focal Loss, increased image-scale during training and testing, and reduction of anchor sizes to better match small targets. The contributions of this research are the following:

- This work provides the first controlled comparison of targeted modifications specifically designed to increase recall of small object detection in satellite imagery in the SaRNet dataset.
- This study demonstrates the first successful integration of Focal Loss with the SaRNet dataset, achieving a 10.4% relative improvement in small object recall while increasing recall on larger objects.
- We provide evidence that geometric anchor-ground truth overlap do not directly translate to detection performance, revealing complicated interactions between pre-trained features and anchor scales in transfer learning scenarios.
- Our results reveal unexpected performance degradations when scaling up images for training and testing, revealing that multi-scale training benefits might not generalize across all remote sensing applications.

By evaluating the three model modifications through both recall and the SAR-specific AR-d20 measure, we show that detection improvements don't always trans- late to operational effectiveness in time-critical search scenarios.

II. METHODS

The three controlled modifications made to the baseline model are described in the following subsections. The Results section of this paper describes the training and test results of each of these modifications.

➢ *Baseline + Focal Loss*

Small object detection in satellite imagery suffers from extreme class imbalance, where the vast majority of image regions represent background landscape pixels, and only a small fraction contain target objects. This imbalance is partic- ularly severe for the objects in our chosen dataset occupying 3-40 pixel-wide regions in 1000×1000 images, representing less than 0.16% of the total image area.

[4] introduced Focal Loss specifically to address this chal- lenge in dense object detection scenarios. Traditional cross- entropy loss used by Faster R-CNN assigns equal importance to all training examples, allowing the overwhelming number of negative background samples to dominate the loss and gradient computations. This prevents the model from focusing on the rare but critical small object instances. Focal Loss addresses class imbalance by down-weighting the contribution of easily classified examples while maintaining full loss for difficult ex- amples, preventing the overwhelming number of background pixels from dominating the training signal. We replaced the baseline model's cross-entropy classification loss with Focal Loss, implementing the α -balanced variant proposed by [4], adopting their recommended hyperparameters: $\gamma = 2.0$ and $\alpha = 0.25$. The focusing parameter $\gamma = 2.0$ reduces the loss contribution from well-classified background regions by up to two orders of magnitude, while $\alpha = 0.25$ compensates for the severe foreground-background class imbalance inherent in satellite imagery.

➢ *Baseline + Multi-Scale Training*

Multi-scale training is a widely adopted technique in object detection that involves training models on images resized to different scales within each epoch. Specifically, images are randomly resized to one of several predefined scales, forcing the model to learn representations that generalize across scale variations. This approach exposes the network to objects at different resolutions, improving scale invariance and detection robustness across object sizes.

The baseline model employs multi-scale training with image scales ranging from 640 to 800 pixels, which downsize the SaRNet 1000×1000 pixel input images. This downsizing is particularly detrimental for small object detection, as it reduces already tiny 3-40 pixel objects, making them nearly impossible to detect reliably.

Research in satellite imagery object detection has demon- strated that small objects benefit significantly from training and testing at higher resolutions [5]. Inspired by these efforts, we modified the input scaling strategy to focus on maintaining or increasing object resolution. We ensured that our 1000×1000 pixel images are resized to maintain the original resolution or scaled up to 1100 or 1200 pixels during training. During inference, we consistently test at

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct244

1200 pixels, providing our small objects with at most 44% more pixels compared to the original resolution. This approach significantly improves the detectability of small objects like the search targets.

➢ *Baseline + Small Anchor Sizes*

In Faster R-CNN, anchor boxes are predefined rectangular regions of fixed sizes and aspect ratios that are systematically placed across the FPN feature maps. The RPN uses these anchors as reference templates, predicting for each anchor whether it contains an object and generates refined bound- ing box coordinates through regression offsets. The default Detectron2 anchor configuration uses anchor sizes (widths) of [32, 64, 128, 256, 512] pixels, designed for images containing larger objects. Anchor boxes must overlap sufficiently with ground truth objects to provide positive training examples [2]. When anchor boxes are too large relative to target objects, the Intersection over Union (IoU) scores become inadequately low, preventing the network from learning to detect these objects. For our dataset containing objects ranging in width of 3 to 40 pixels, the default 32-pixel minimum anchor size creates a fundamental mismatch. Small objects may fail to achieve the required IoU threshold with default anchors, rendering them invisible during training.

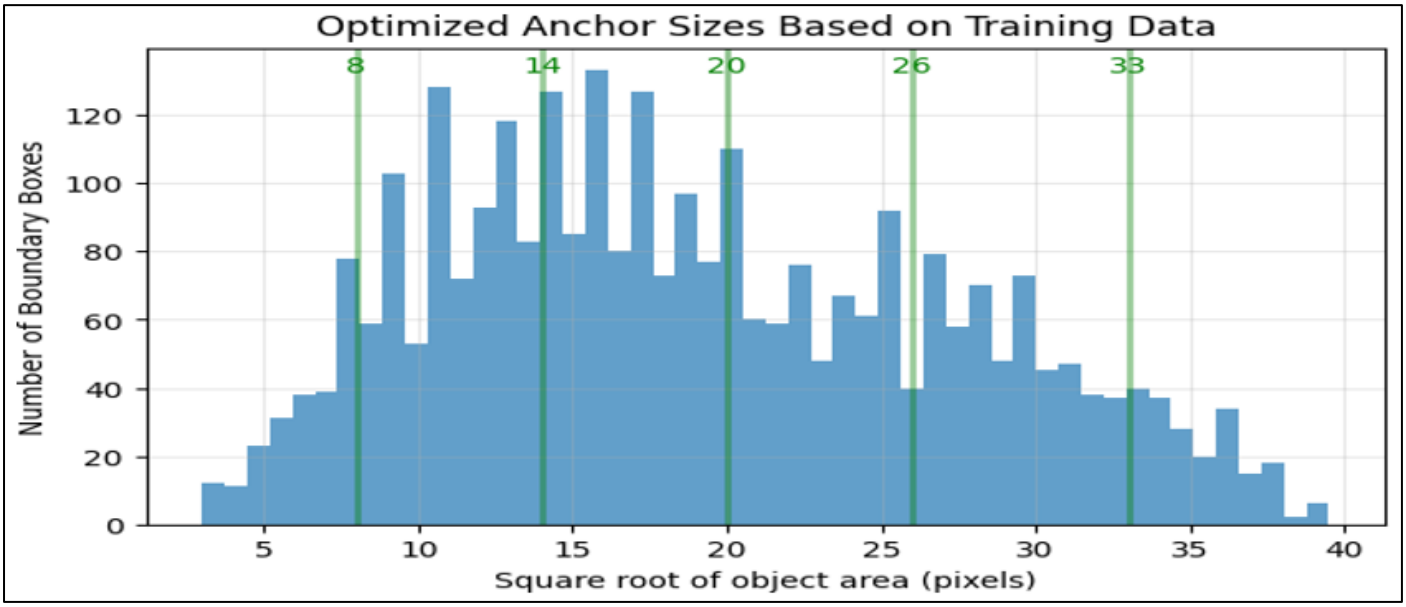

Fig 3 Distribution of Training Set Object Sizes (Square Root of Bounding Box area) with K-Means Optimized Anchor Sizes Shown as Vertical Lines.

Recent advances in small object detection emphasize the importance of data-driven anchor optimization, rather than manual selection. Zhao and Song (2024) found that using clustering analysis to identify optimal anchor sizes increased Faster R-CNN precision on aerial maritime search and rescue footage [3]. Following their practices for anchor optimization, we employed K-means clustering to determine optimal anchor sizes directly from our ground truth bounding box distribution. K-means clustering for anchor optimization operates by treating each ground truth bounding box as a data point rep- resented by its dimensions (width and height). The algorithm groups these bounding boxes into K clusters (in our case, K=5) by minimizing the within-cluster sum of squared distances. Each cluster's centroid represents an optimal anchor size that best represents the objects in that cluster. This unsupervised approach automatically finds the natural size groupings present in the dataset, removing the need for manual anchor size selection and ensuring that the chosen anchors closely match the actual object size distribution.

The algorithm was configured to find five anchor sizes that maximize coverage of our small object dataset. As shown in Figure 3., the K-means algorithm successfully identified five anchor positions that correspond to peaks and density concen- trations in the object size distribution of the training samples. The optimal anchor sizes were: [8, 14, 20, 26, 33] pixel widths. This configuration achieved 66.9% total coverage of all ground truth objects, with only 16 objects (0.5%) remaining uncovered. The per-anchor coverage analysis revealed varying effectiveness across the anchor spectrum: 41.4% coverage for 8-pixel anchors, 78.4% for 14-pixel anchors, 86.0% for 20- pixel anchors, 72.7% for 26-pixel anchors, and 55.8% for 33-pixel anchors. Overall, this configuration provides optimal coverage for our 3-20 pixel object range, with the smallest 8-pixel anchors enabling detection of the tiniest objects while the 33-pixel anchors accommodate the larger end of the size spectrum.

➢ *Training Procedure*

All three model modification experiments maintained iden- tical controlled variables with the baseline model to ensure fair comparison and isolate the impact of each modification. The baseline architecture consisted of a Faster R-CNN with ResNet-50-FPN backbone, pre-trained on MS-COCO and fine- tuned on the SaRNet training and validation

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct244

set images to find the labeled targets. Training was conducted using Stochastic Gradient Descent (SGD) with a base learning rate of 0.0001, batch size of 4 images per iteration, and 2 data loading workers. All models were trained for exactly 5,000 iterations without learning rate decay. The RPN used IoU thresholds of [0.2, 0.4] for proposal generation, while the Region of Interest (ROI) heads maintained a batch size of 128 regions per image. The standard loss configuration employed cross- entropy loss for classification and smooth L1 loss for bounding box regression, with identical loss weighting schemes (with the exception of the modified model using Focal Loss). All models used the same random seed initialization and were trained on identical hardware (single Tesla T4 GPU via Google Colab) to ensure computational consistency. The SaRNet dataset parti- tion remained constant with 70%/20%/10% splits for training, validation, and testing respectively. The SaRNet-specific AR- d20 metric was calculated for each model modification using the test set.

III. RESULTS

Table 2 Modified Models' Recall Score Comparison

	Recall (%) on Test Set Bounding Boxes					
Modified Model	**Small**	**ΔB**	**Medium**	**ΔB**	**Large**	**ΔB**
Baseline + Focal Loss	46.7	+4.4	69.3	-2.2	78.8	+1.4
Baseline + Multi-scale	39.4	-2.9	62.0	-9.5	77.4	+0
Baseline + Small Anchors	40.9	-1.4	66.4	-5.1	75.9	-1.5

Figure 4 presents the training set and validation set loss over the iterations the models were trained for. While the models could have trained for less iterations without significantly compromising performance as seen by their immediate drop in loss within the first few hundred iterations and steady plateau until the end, this research aimed to keep parameters like maximum iterations controlled to the baseline model to isolate the effects of the three primary model modifications. Table 2 presents the recall performance of each model modification across the three object size categories on the test set. Table 3 presents the AR-d20 metric results for each model. In both

Table 3 Modified Models' AR-d20 Score Comparison

Modified Model	**AR-d20**	**ΔB**
Baseline + Focal Loss	37.23	-4.59
Baseline + Multi-scale	36.06	-5.76
Baseline + Small Anchors	29.18	-12.64

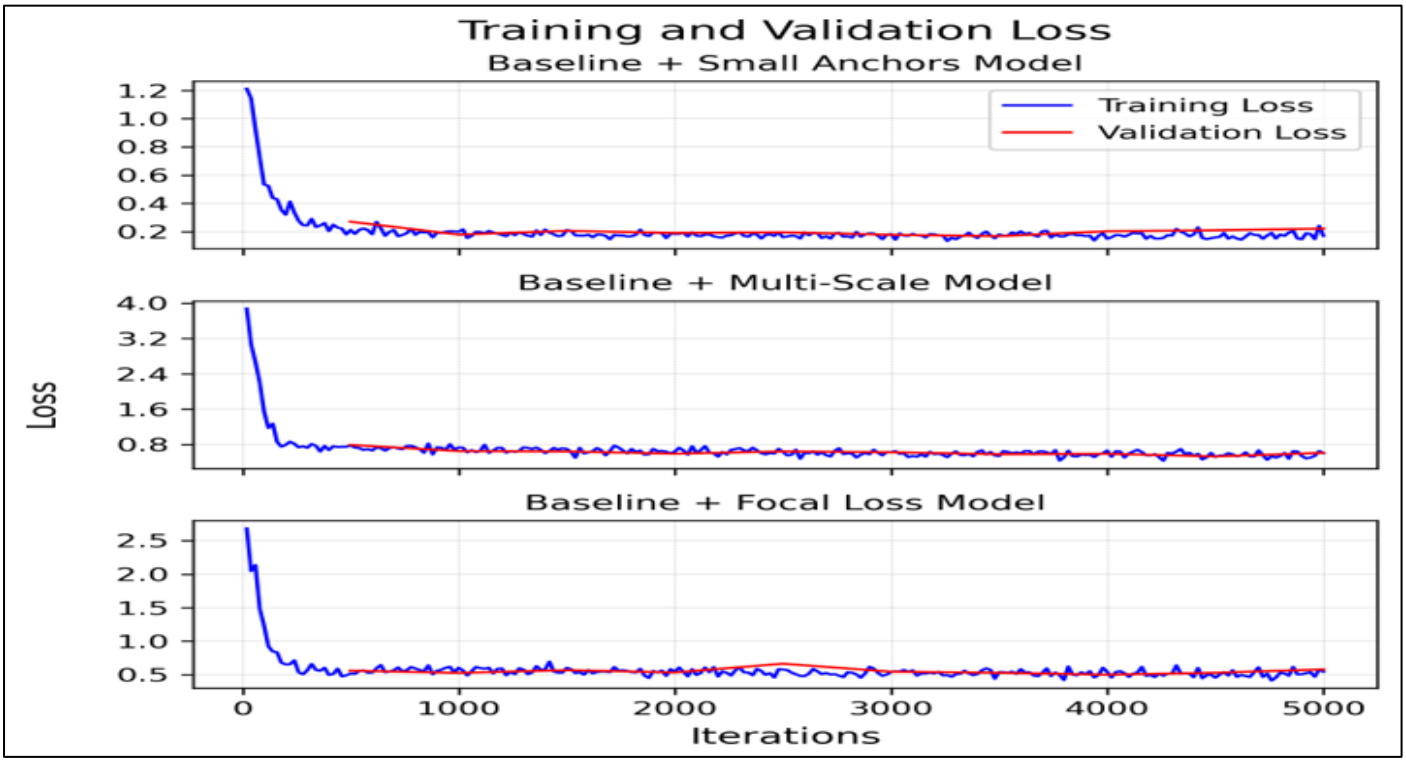

Fig 4 Training and Validation Loss Curves for Three Model Modifications Over 5,000 Iterations. All Models Exhibit Rapid Convergence within the First 500 Iterations Followed by Stable Plateaus. Notable Differences in Final Loss Values: Small Anchors (~0.1), Focal Loss (~0.3), and Multi-Scale (~0.5).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct244

- Note: y-Axis Scales Differ Between Subplots to Show Convergence Behavior.

Tables 2 and 3, ΔB refers to the change from the baseline model's performance for that particular metric.

The Baseline + Focal Loss integration emerged as the most effective single modification, demonstrating the only improvement in small object detection performance. Small object recall increased by 4.4 percentage points from the baseline model to 46.7% (relative increase of 10.4%). While medium-sized object performance decreased slightly (-2.2 percentage points), large object detection improved by 1.4 percentage points to 78.8%, achieving the highest performance in this category across all modifications. This pattern suggests that Focal Loss successfully addressed the class imbalance problem inherent in small object detection without severely compromising performance on larger, easier-to-find objects. In regards to its loss convergence, the close alignment between training and validation curves throughout the 5,000 iterations indicates learning without overfitting. Rather than driving loss toward zero through easy background predictions, the Focal Loss mechanism maintained meaningful loss contributions from challenging examples, explaining the moderate final loss values coupled with improved detection performance. The AR- d20 metric decreased modestly from 41.82 to 37.23 (-4.59), representing the smallest decrease among all modifications and suggesting that Focal Loss provides a decent balance between standard detection metrics and operational search requirements.

The Baseline + Multi-scale model performed below the baseline model metrics across all object categories. Small objects suffered a 2.9 percentage point decrease to 39.4% recall, while medium objects experienced the most severe impact with a 9.5 percentage point reduction to 62.0% recall. Large object performance remained unchanged at 77.4%. These results indicate that the chosen scale range (1000, 1100, 1200 pixels for training and 1200 pixels for testing) may have introduced domain shift effects that outweighed the predicted benefits of increased pixel resolution for small objects. The convergence analysis supports this interpretation, showing that Multi-scale training achieved the highest final loss (∼ 0.5) among all modifications. The elevated loss plateau suggests the model struggled to reconcile conflicting gradients from objects appearing at different sizes across training scales. Slight divergence between training and validation curves in later iterations indicates potential overfitting to the scale-augmented training distribution, which failed to generalize effectively to the test conditions. The AR-d20 performance decreased substantially from 41.82 to 36.06 (-5.76).

The Baseline + Small (decreased) Anchor size model showed modest negative impacts across all categories. Small object recall decreased by 1.4 percentage points to 40.9%, medium objects declined by 5.1 percentage points to 66.4%, and large objects dropped by 1.5 percentage points to 75.9%. The consistent degradation across all object sizes suggests that the K-means derived anchor sizes may have disrupted the feature map-anchor alignment optimized in the pre-trained model. This demonstrates the complexity of anchor opti- mization in transfer learning scenarios. Paradoxically, this modification achieved the lowest final training loss (∼0.1) while producing the worst overall performance. This can be attributed to the abundance of "easy" negative samples created by placing numerous small anchors across large image regions with sparse objects. The tight convergence between training and validation losses indicates the model learned stable but suboptimal predictions, focusing on confidently predicting background rather than improving object detection capability. The AR-d20 metric saw the most severe degradation among all modifications (-12.64 to 29.18), suggesting that optimized anchor coverage alone is not enough to maintain operational detection performance in transfer learning contexts.

IV. CONCLUSION

This systematic evaluation of small object detection strate- gies for satellite SAR operations reveals critical insights about the gap between theoretical optimization and practical perfor- mance. While Focal Loss was the only successful modification improving small object recall by 10.4% relative to baseline, both anchor optimization and multi-scale training unexpect- edly worsened performance. These counterintuitive results highlight fundamental challenges in adapting general computer vision techniques to specialized domains: geometric anchor coverage does not guarantee detection improvement when pre- trained features expect different anchor-feature relationships, and higher resolution training can introduce domain shift that negates the benefits of increased pixel detail.

For practical deployment in search and rescue operations where detection performance directly impacts human lives, our findings provide clear guidance: implement Focal Loss to handle extreme class imbalance while maintaining the default anchor configuration and training scales.

Future research should explore adaptive approaches that can reconcile pre-trained model expectations with domain- specific requirements, potentially through learnable anchor mechanisms or domain adaptation techniques that preserve the benefits of transfer learning while accommodating the challenges of small object detection in satellite imagery. Future work could also investigate combined approaches, particularly Focal Loss integration with other architectural changes.

This comprehensive evaluation provides the SAR com- munity with evidence-based recommendations for optimizing satellite imagery analysis systems, potentially reducing search times and improving outcomes in life-critical operations.

ACKNOWLEDGMENTS

This research was inspired by the author's participation in the MIT Beaver Works Summer Institute (BWSI)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct244

studying Remote Sensing for Disaster Response. The author would like to thank the BWSI instructors and guest speakers for the insights into the field of remote sensing for search and rescue applications, technical computer science lessons, and encouragement. This research was designed and conducted by the author. The code used in this research can be found here: https://github.com/GT1235/SearchAndRescueModels

REFERENCES

[1]. Thoreau, Michael & Wilson, Frazer. (2021). SaRNet: A Dataset for Deep Learning Assisted Search and Rescue with Satellite Imagery. 204-208. 10.1109/ISPA52656.2021.9552103.

[2]. S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" in IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 39, no. 06, pp. 1137-1149, June 2017, doi: 10.1109/TPAMI.2016.2577031.

[3]. Zhao, B., Song, R. Enhancing two-stage object detection models via data-driven anchor box optimization in UAV-based maritime SAR. *Sci Rep* 14, 4765 (2024). https://doi.org/10.1038/s41598-024-55570-z.

[4]. T. -Y. Lin, P. Goyal, R. Girshick, K. He and P. Dolla´r, "Focal Loss for Dense Object Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318-327, 1 Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[5]. J. Shermeyer and A. Van Etten, "The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp. 1432-1441, doi: 10.1109/CVPRW.2019.00184.

[6]. Yuxin Wu and Alexander Kirillov and Francisco Massa and Wan-Yen Lo and Ross Girshick, Detectron2, https://github.com/facebookresearch/detectron2, 2019.

[7]. Lin, Tsung-Yi et al. "Feature Pyramid Networks for Object Detection." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016): 936-944.