A. Dataset

The *SearchWing* dataset comprises 14,947 UAV-captured images (7,252 with objects) of maritime vessels taken at altitudes up to 750m. As shown in Figure 9 and Figure 10, most objects appear in images from >500m altitude, with over 50% being small or tiny objects. Images were captured in moderate sea conditions (Beaufort 3-4) with significant whitecaps (Figure 8), presenting greater detection challenges than *SeaDronesSee*. The size distribution of tiny objects differs notably from *SeaDronesSee*, with most under 15^2px (Figure 11). Data collection used dual gimbal-free *Raspberry Pi Camera v2* setup. The train-validation-test split maintains consistent distributions ($\pm 3\%$) of object sizes, altitudes, and conditions, with deliberate inclusion of unique vessels in validation/test sets. Split details are provided in Table 4.

Split-Name	# Objects	# Images	% With Objects	% Without Objects
Train	4175	3634	90%	10%
Validate	886	1619	50%	50%
Test	939	1619	50%	50%

Table 4. Dataset Split SearchWing

B. Qualitative Analysis

In Figure 12a and Figure 12b, we give a detailed overview of *FalconEye* prediction errors on the SearchWing validation set.

C. Ablation Study

We conduct a study similar to an ablation study to explore the impact of different modules in our system. We systematically assess how our proposed methods influence detection quality and inference speed, thereby guiding the final model construction based on these insights.

C.1. Choice of Backbone

We evaluate *MobileNetV2* and *MobileOne* backbones, both pre-trained on *ImageNet*. In this experiment no attention module is included. For *MobileNet*, we truncate after *block_6* rather than *block_6_expand_relu* as in [35] due to improved training stability. Results in Table 5 use *center-only slices* with 15% background samples for training, with evaluation on the full validation set. We report Precision@0.9Recall (P@0.9R) to assess false positives at high recall, and F1 score at its optimal threshold.

MobileOne significantly outperforms *MobileNetV2*, particularly in precision - a critical metric given our UAV bandwidth constraints. While we expected improvement from *MobileOne*, the magnitude exceeds typical architectural gains. One hypothesis is the increased model capacity (575k vs 80k parameters), though this warrants further study as it contradicts findings in [56].

As shown in Table 5, *MobileOne*'s reparameterization maintains prediction quality while achieving 6x GPU speedup through reduced parameters and memory access. Despite higher parameter count, *MobileOne*'s optimized design delivers faster inference than *MobileNet*.

C.2. Impact of Attention

We evaluate attention mechanisms' impact on model performance and speed, comparing our model without any attention module against variants with 8-Head MHSA, 4-Head MHSA, 4-Head MHSA with reduced channels (128→32), and SIM-AM[3]. As shown in Table 6, while attention modules improve AP (+0.01 to +0.05), MHSA variants significantly increase inference time (15x on GPU) and memory usage, limiting batch size to 8 tiles. SIM-AM emerges as a suitable solution, achieving the highest AP (0.89) with minimal latency increase (1.2x), making it the only viable attention alternative for our deployment scenario. The ONNX implementation of SIM-AM further reduces inference time to 82.5ms per tile, compared to 313ms in PyTorch.

[3] https://github.com/ZjjConan/SimAM

Figure 7. Dataset samples *SearchWing*. All images contain objects, some of which are tiny and hard to identify, showcasing the difficulty of the dataset.

Backbone	SearchWing			SeaDronesSee			#Params	MB	Latency GPU ms	Latency PI ms
	AP	P@0.9R	F1 best	AP	P@0.9R	F1 best				
MobileNetV2	0.65	0.2	0.64	0.79	0.45	0.71	**80k**	**0.32**	2.6	133*28
MobileOne (no-rep)	0.85	0.47	0.82	0.89	0.7	0.85	575k	2.3	8.1	388*28
MobileOne (rep)	**0.85**	**0.47**	**0.82**	**0.89**	**0.7**	**0.85**	148k	0.6	**1.4**	**66*28**

Table 5. Choice of Backbone. ”no-rep” is without reparameterization of the backbone.

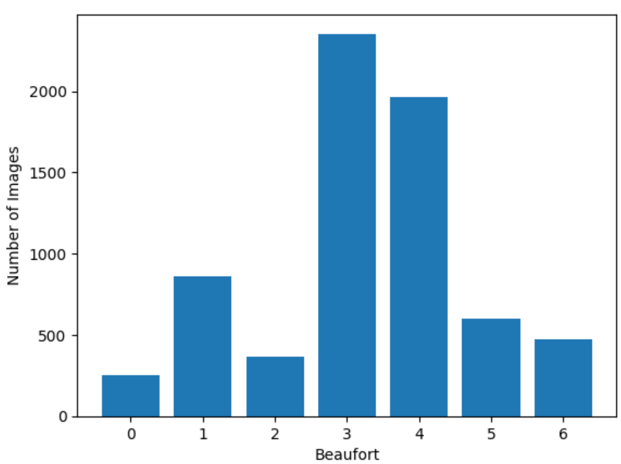


Figure 8. Beaufort Distribution SearchWing

Attention Module	AP	#Params	MB	Latency GPU ms*	Latency PI ms
None	0.85	148k	0.6	**1.4**	**274*28**
MHSA-8	0.89	214k	0.86	15.8	2420*28
MHSA-4	0.86	214k	0.86	15.7	1637*28
MHSA-4 reduced dim	0.85	**146k**	**0.59**	15.3	1371*28
SIM-AM	**0.89**	148k	0.6	1.7	313*28

Table 6. Impact of attention modules. * We utilize the max possible batch size - for ”none”, ”SIM-AM” = all tiles (28), for MHSA8 and MHSA4 = 8 tiles

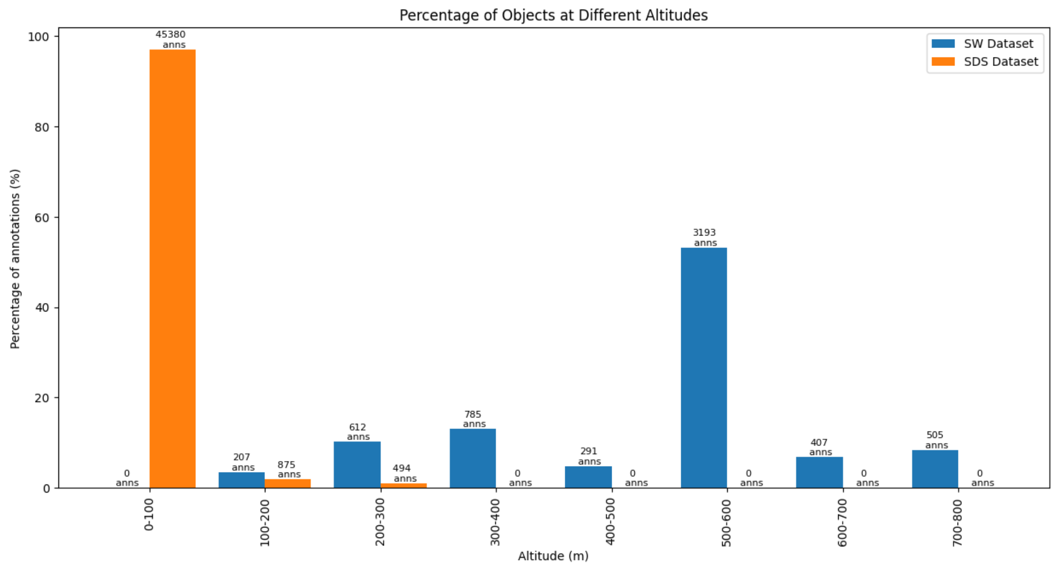


Figure 9. Altitude SearchWing(called SW dataset) vs. SeaDronesSee comparison. In percentages. Bars are annotated with the absolute sample count.

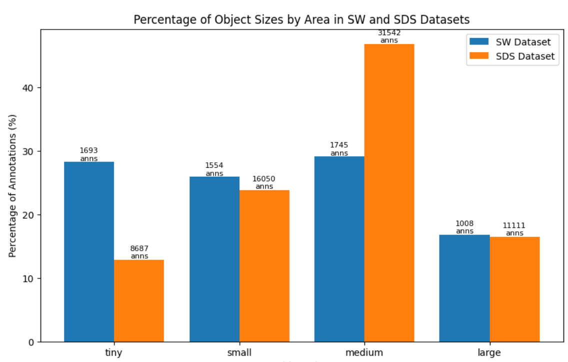
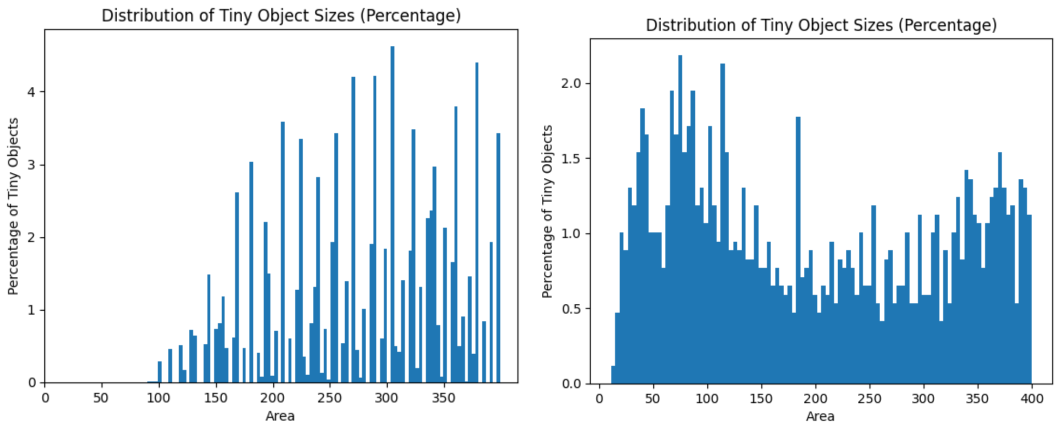


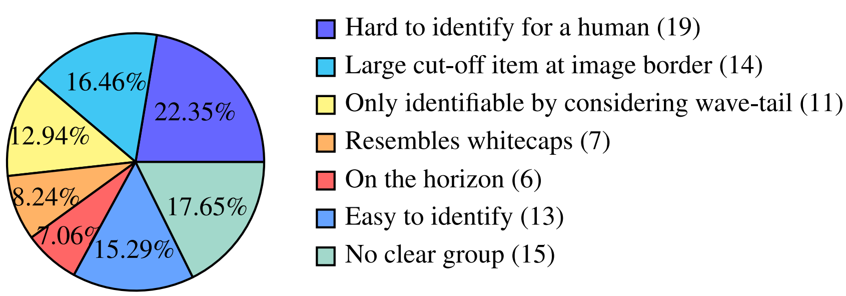
Figure 10. Object sizes SearchWing(called SW dataset) vs. SeaDronesSee comparison. In percentages. Bars are annotated with the absolute sample count.



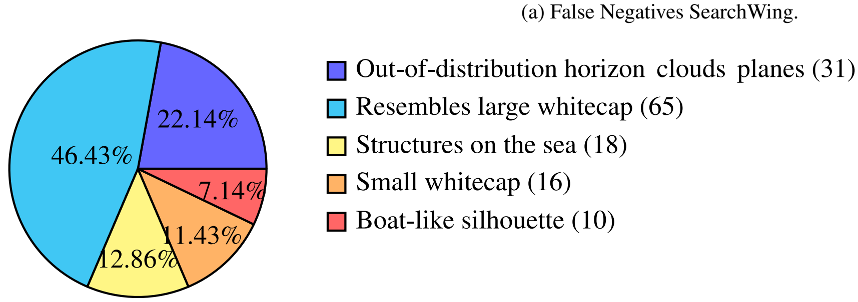
(a) Tiny Object Distribution SeaDronesSee.

(b) Tiny Object Distribution SearchWing.

Figure 11. Tiny Object Distribution Comparison: *SearchWing* contains significantly smaller objects.



(a) False Negatives SearchWing.



(b) False Positives SearchWing.

Figure 12. False Negatives and Positives Analysis.