Appendix

Training: networks were trained using data from Years 2015-2018, tested on Year 2019. The training set was split into 80%/20% for training and validation, repsectively. Both models used learning rate of $1 * 10^{-3}$ with the Adam optimisation algorithm [1], cross-entropy loss and early stopping [2] with patience of 7. We kept data augmentation for patches the same as in the previous work [3]. No data augmentation was used to train the event classifier.

Description	Act.	Output shape	Parameters
Input image	-	1 x 48 x 48	-
Conv 3 x 3	-	32 x 48 x 48	320
BatchNorm	ReLU	32 x 48 x 48	64
Conv 3 x 3	-	32 x 24 x 24	9 248
BatchNorm	ReLU	32 x 24 x 24	64
Conv 3 x 3	-	64 x 24 x 24	18 496
BatchNorm	ReLU	64 x 24 x 24	128
Conv 3 x 3	-	64 x 12 x 12	36 928
BatchNorm	ReLU	64 x 12 x 12	128
Conv 3 x 3	-	128 x 12 x 12	73 856
BatchNorm	ReLU	128 x 12 x 12	256
Conv 3 x 3	-	128 x 6 x 6	147 584
BatchNorm	ReLU	128 x 6 x 6	256
Conv 3 x 3	-	256 x 6 x 6	295 168
BatchNorm	ReLU	256 x 6 x 6	512
Conv 3 x 3	-	256 x 4 x 4	590 080
BatchNorm	ReLU	256 x 4 x 4	512
FC	-	1 x 512 x 1	2 097 153
BatchNorm	ReLU	1 x 512 x 1	2
Dropout (0.5)	-	1 x 512 x 1	-
FC	Softmax	1 x 6 x 1	3 073
Total			3 273 828

Table 1: Frame classifier architecture.

Description	Act.	Output shape	Parameters
Input confidences	-	1 x 6 x n	-
Conv 6 x 4	LReLU	128 x 6 x n	3 200
Conv 6 x 6	LReLU	128 x 6 x n	589 952
Adaptive Avg Pool	-	128 x 6 x 1	-
FC	Softmax	1 x 6 x 1	4 614
Total			597 766

Table 2: Event classifier architecture.

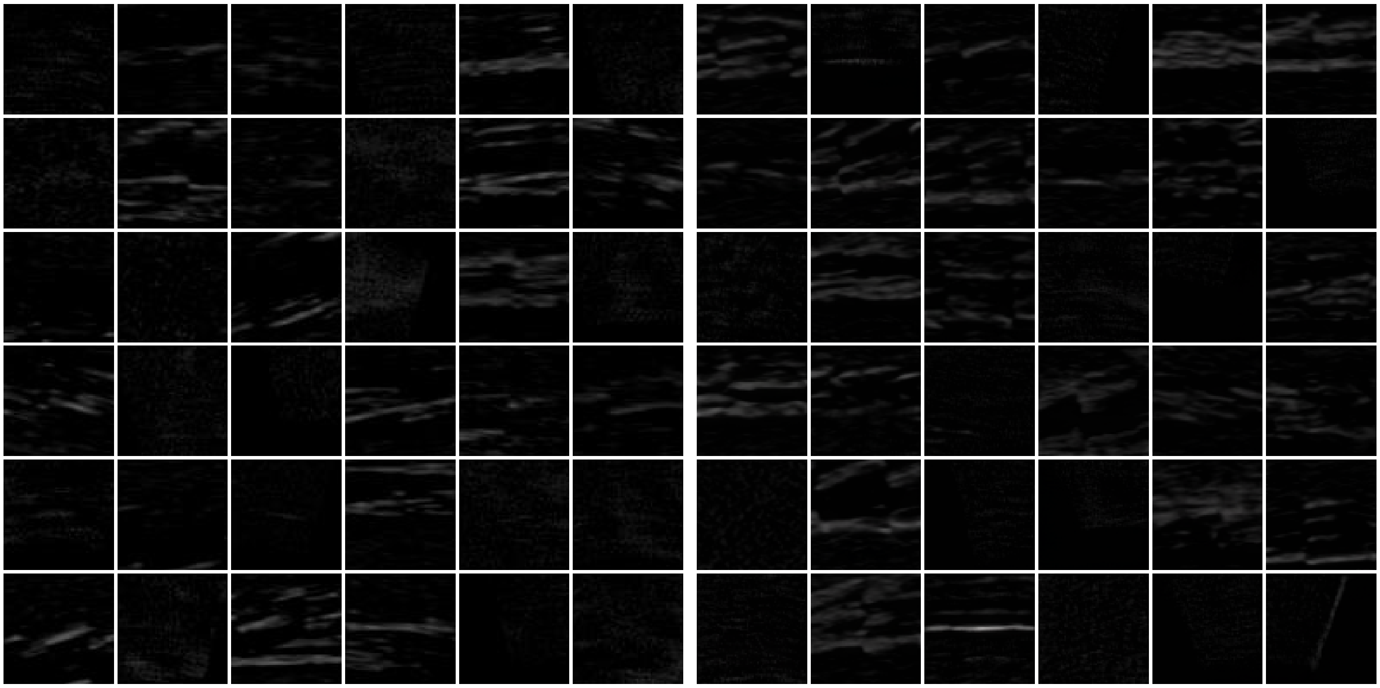

Figure 1: Examples of real (left) and generated (right) Background patches.

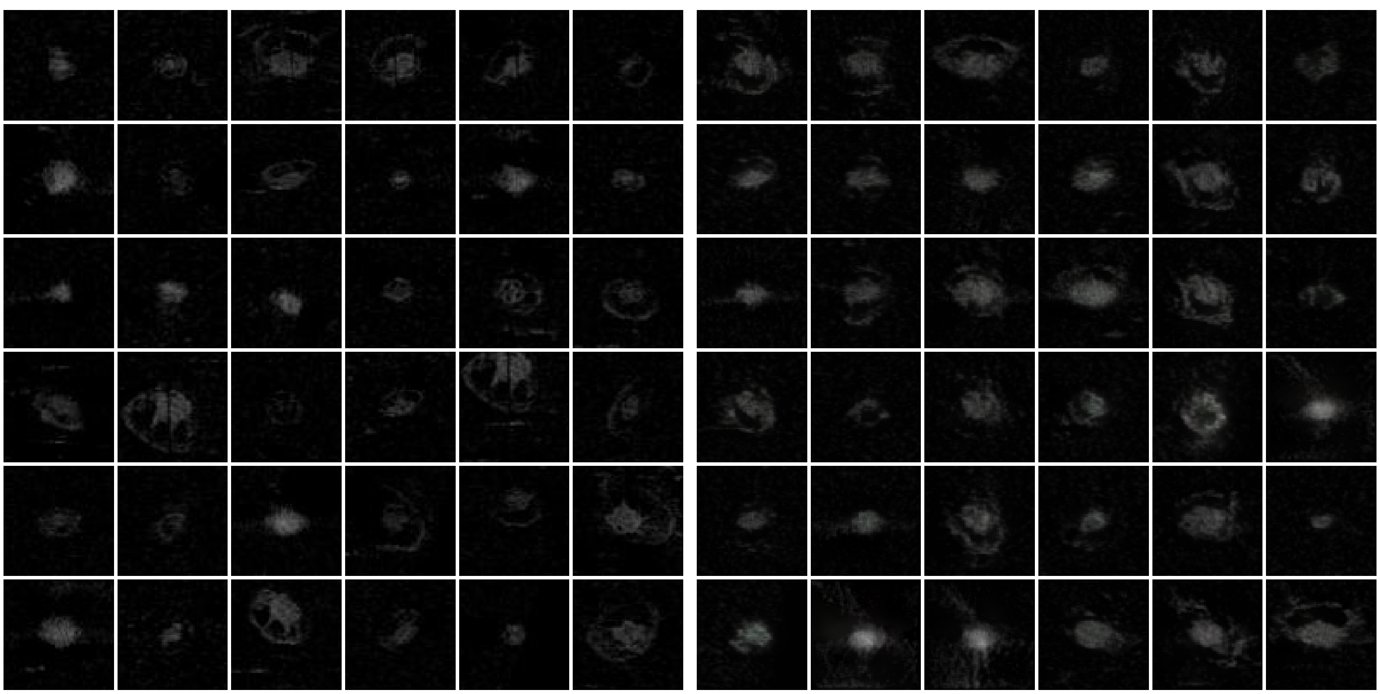

Figure 2: Examples of real (left) and generated (right) Jellyfish patches.

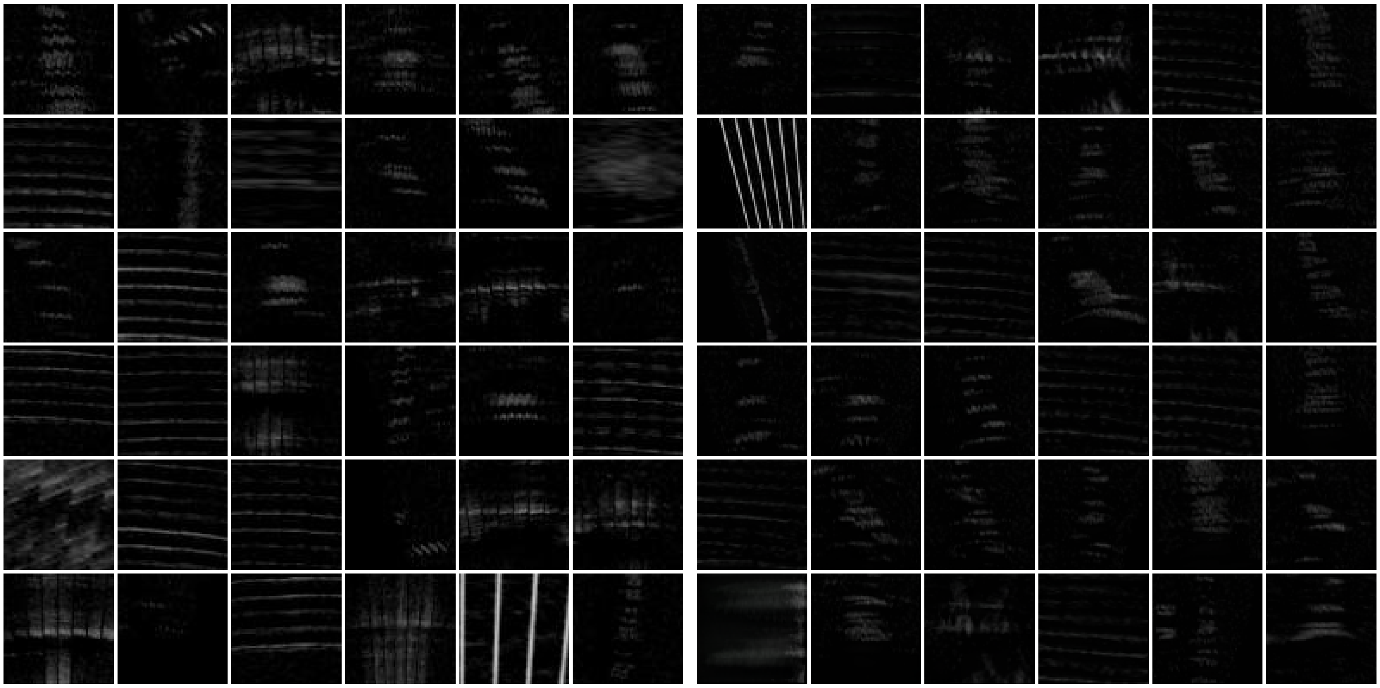

Figure 3: Examples of real (left) and generated (right) Artefacts patches.

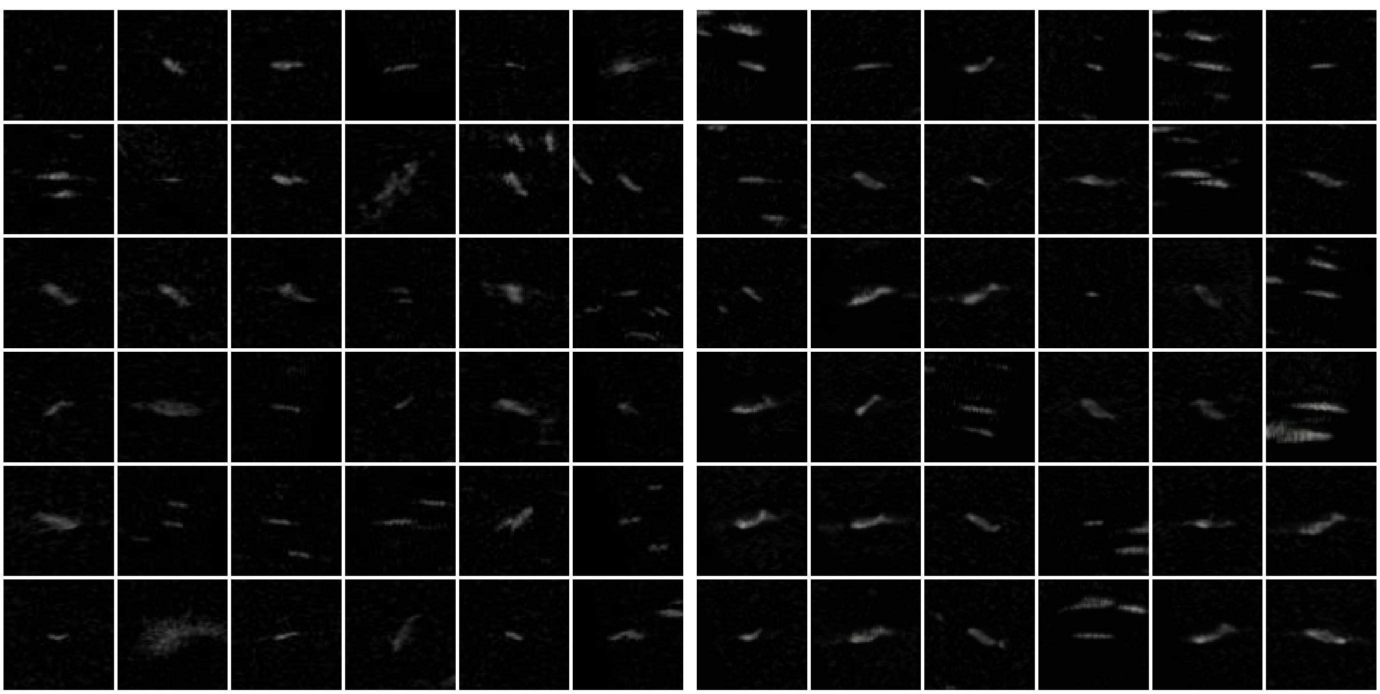

Figure 4: Examples of real (left) and generated (right) Fish patches.

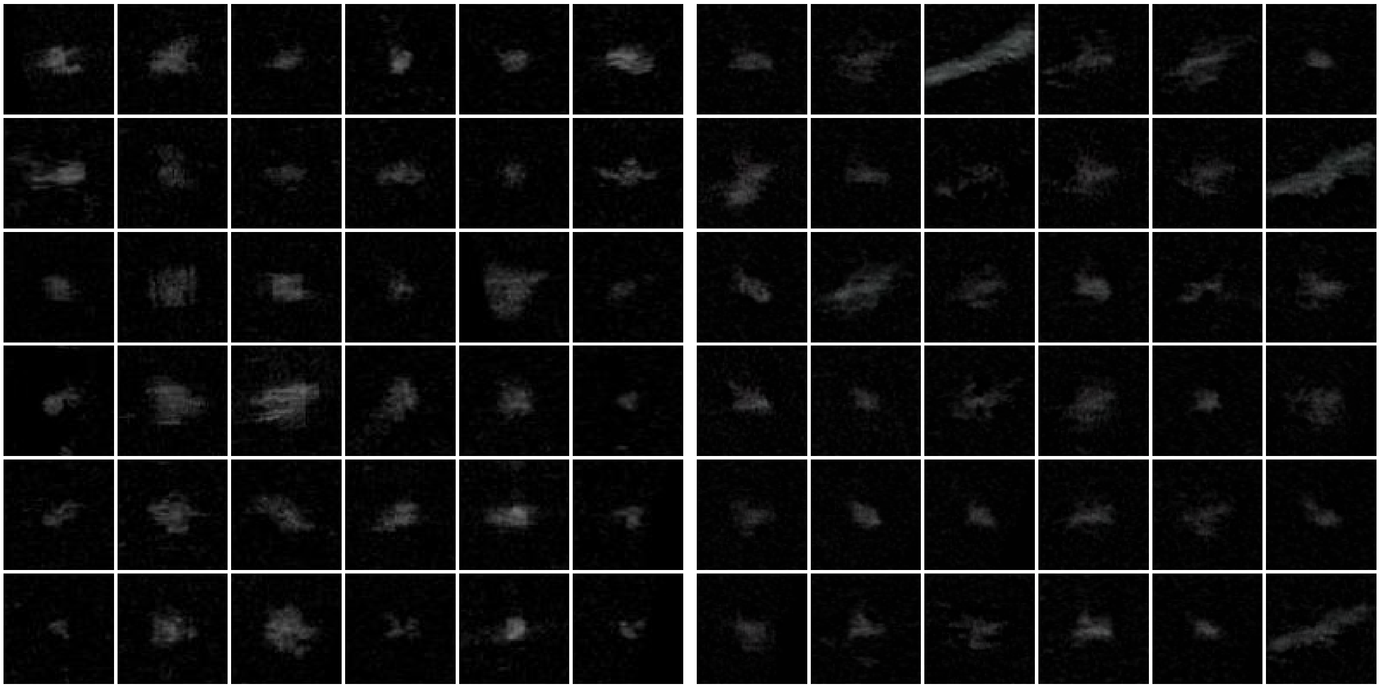

Figure 5: Examples of real (left) and generated (right) Seaweed patches.

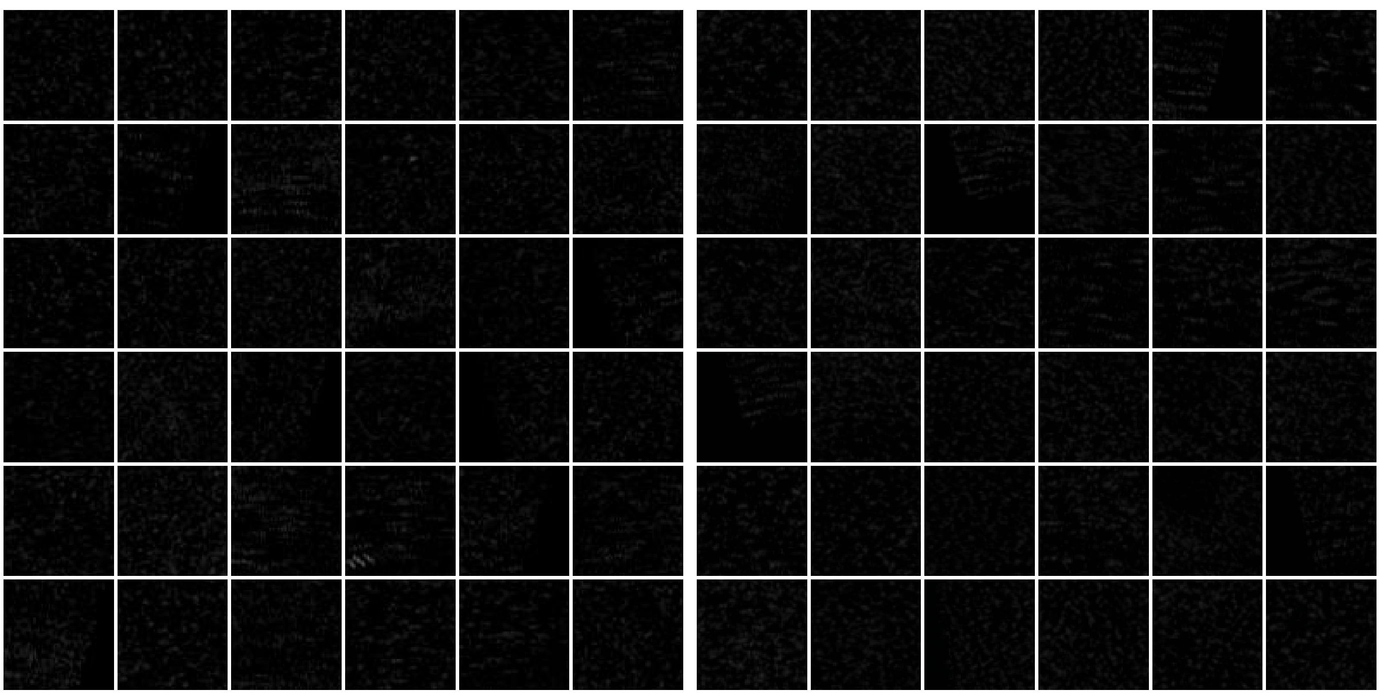

Figure 6: Examples of real (left) and generated (right) Sediment patches.

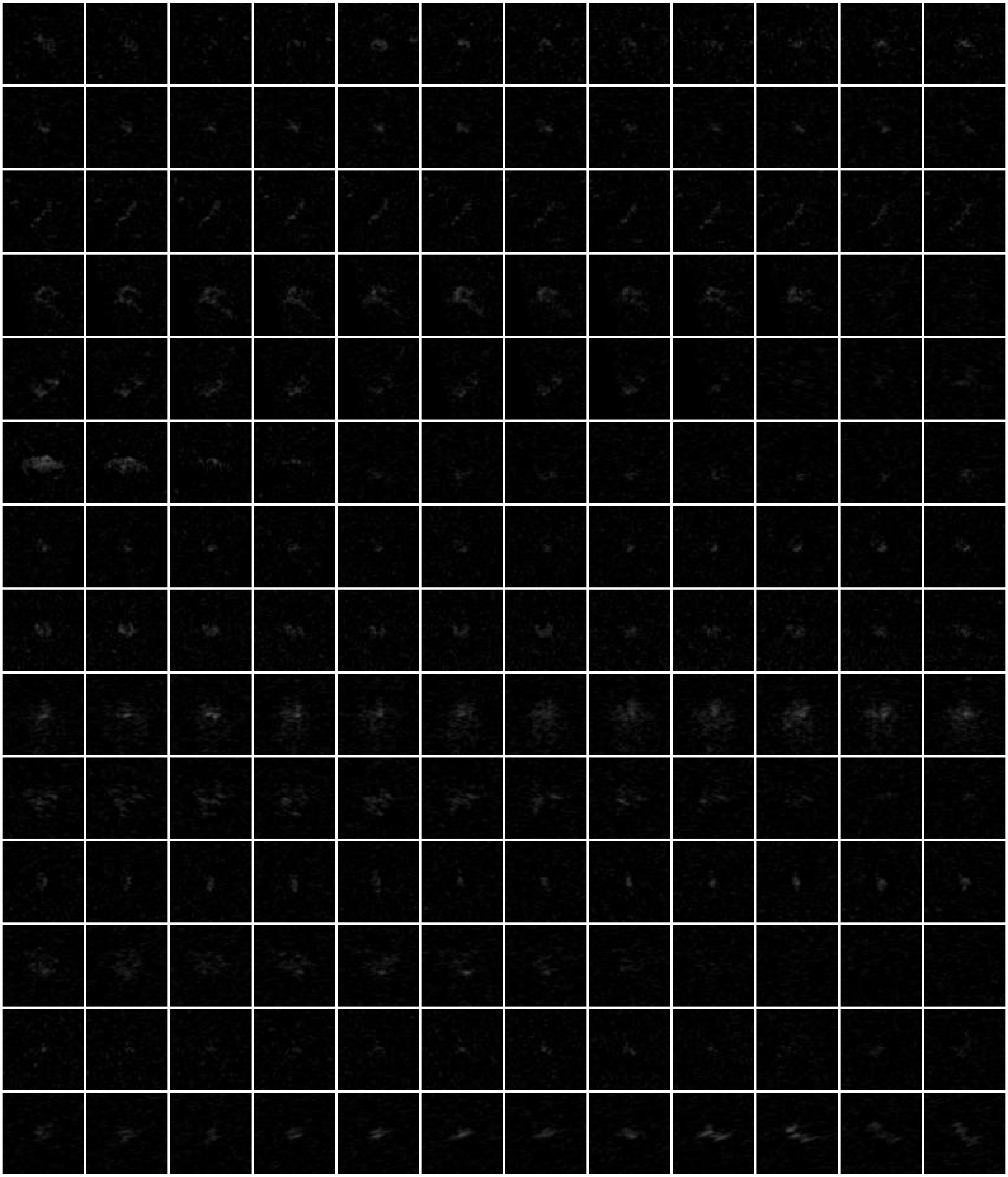

Figure 7: Excerpts from sequences of misclassified jellyfish. Each row represents a different object.

References

[1] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[2] L. Prechelt, "Early stopping-but when?," in *Neural Networks: Tricks of the trade*, pp. 55–69, Springer, 1998.

[3] G. French, M. Mackiewicz, M. Fisher, M. Challiss, P. Knight, B. Robinson, and A. Bloomfield, "Jellymonitor: automated detection of jellyfish in sonar images using neural networks," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*, pp. 406–412, Aug 2018.