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	-	$32 \ge 48 \ge 48$	320
BatchNorm	ReLU	$32 \ge 48 \ge 48$	64
Conv 3 x 3	-	$32 \ge 24 \ge 24$	$9\ 248$
BatchNorm	ReLU	$32 \ge 24 \ge 24$	64
Conv 3 x 3	-	$64 \ge 24 \ge 24$	18 496
BatchNorm	ReLU	$64 \ge 24 \ge 24$	128
Conv 3 x 3	-	$64 \ge 12 \ge 12$	36 928
BatchNorm	ReLU	$64 \ge 12 \ge 12$	128
Conv 3 x 3	-	$128 \ge 12 \ge 12$	73 856
BatchNorm	ReLU	$128 \ge 12 \ge 12$	256
Conv 3 x 3	-	$128 \ge 6 \ge 6$	147 584
BatchNorm	ReLU	$128 \ge 6 \ge 6$	256
Conv 3 x 3	-	$256 \ge 6 \ge 6$	$295\ 168$
BatchNorm	ReLU	$256 \ge 6 \ge 6$	512
Conv 3 x 3	-	$256 \ge 4 \ge 4$	590080
BatchNorm	ReLU	$256 \ge 4 \ge 4$	512
FC	-	$1 \ge 512 \ge 1$	$2 \ 097 \ 153$
BatchNorm	ReLU	$1 \ge 512 \ge 1$	2
Dropout (0.5)	-	$1 \ge 512 \ge 1$	-
FC	Softmax	$1 \ge 6 \ge 1$	$3 \ 073$
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	$128 \ge 6 \ge n$	$3\ 200$
Conv 6 x 6	LReLU	$128\ge 6\ge n$	$589 \ 952$
Adaptive Avg Pool	-	$128 \ge 6 \ge 1$	-
FC	Softmax	$1 \ge 6 \ge 1$	4 614
Total			597 766

Table 2: Event classifier architecture.



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

16	- 60			Ante	Ð		(書)	35	49	(à	No.
	1000	1			æ	A.			18	10	1
		40	4		٢	-	1983			1 miles	d>
1	(IR	Ŕ	4		(4		ē.		AN .	0	
÷	(đ.		ġ	<u>(</u> ه	a.	ġ.			ø		ø
-	a)	1	Ŵ	9		19			-	Ser.	

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

	and a second	新藤和 1111-			-				- States -		
			an.	AND			10				and and a second
						and a series			Calling	Streen .	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
			62 % NY	Topper-		and addition guilter	anna STAD Anna Anna MAT-	ann aitean aitean aitean aitean aitean			and a second
			2						情報作		North Contraction of the second
Entropy of the second s	i den a				R. (1995 - 1920			And .		1 21 24 21 2	1000

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

	*	+			<i>w</i>	4 1		-			-
-		*	Ja.	4.4				~	-		1
1		a.	ft.	1	14	. '	~			~	11
*	950	-		~	4	-		an a	~	16	No.
*	11	11		4	1	~		-	***		~
1			Ŧ	1	4	-	- 400-	1		~	-

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.

*@	40			Ð				194			1
		4	*		N		4				
		- 1		1							
	4	R_{m}^{2}	and the second s	R.	R	R.	12%) -	R	£r.		
42	4	1	÷.		4						
(*****	ά¢κ.	S. Pres									¢.
									4	4	24
-55	43	-94	7:3	34		4		67	1		14-1
		de la			1886 -						
		S.		All and a second							
		1		-62+		10 A			•		
						4	k				
5	P.	*	4	4			4	24	***	-	14.1

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.