University of St Andrews

Challenges in Automating Ecological Surveys with AI

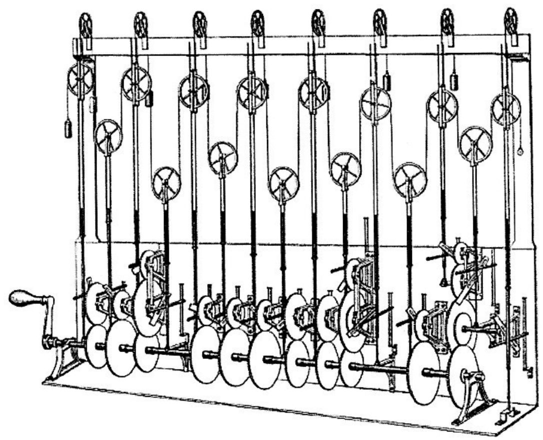
Kasim Terzic
School of Computer Science
University of St Andrews

About me

- Lecturer in Computer Science
  - I co-developed the CS5014 Machine Learning module
- Existing research projects in biology and conservation
  - animal re-ID from photos (with Biology)
  - salmon scale rings (with Marine Scotland)
  - seal pup counting (with SMRU and CREEM)
  - monitoring bird colonies (with SOI and Psychology)
- All work is part of larger collaborations

Machine learning in a nutshell

- Transform input to output using a sequence of operations
  - learning is a search for best parameters for these operations

Input →  → Output

- Classification: predicting categories
- Regression: predicting numbers

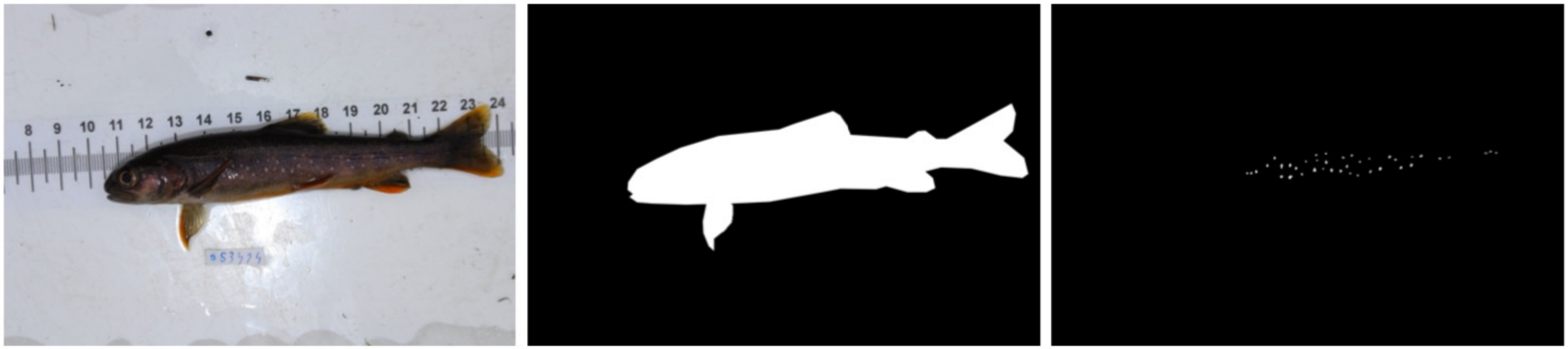
The photo shows
Sir Paul McCartney!

The temperature is
17.7 ºC, give or take

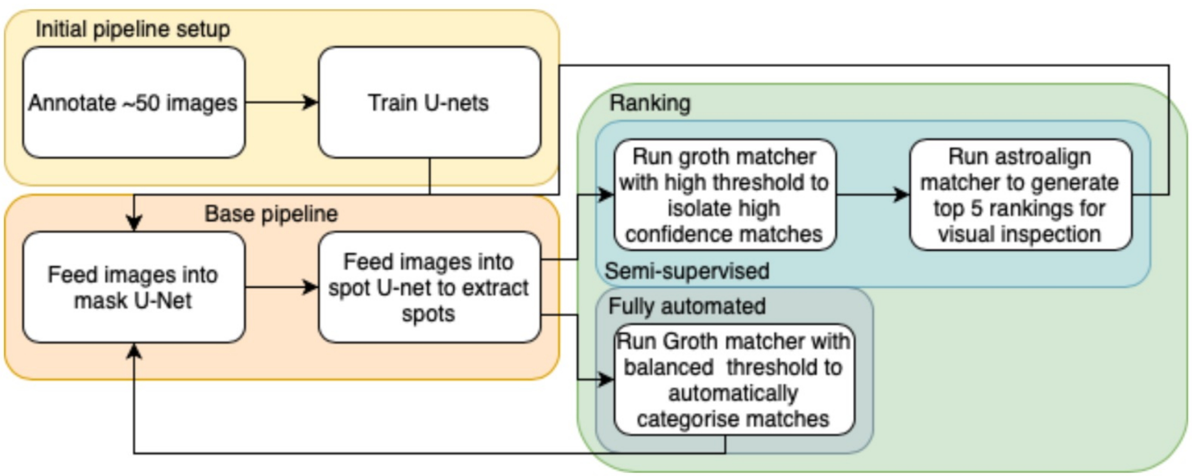
Vision-based Re-ID

- Re-ID of small benthic Arctic charr
  - spring-fed systems of caves in Iceland
  - sampled in June and August
- Fish are caught and measured
  - PIT tags, fin clips, photographs
- There is evidence of tag loss
  - visual confirmation by expert needed
  - ...based on characteristic spot patterns
  - this is very inefficient!

Spot-based re-identification



(a) Base image (b) Mask of fish location (c) Mask of spot location



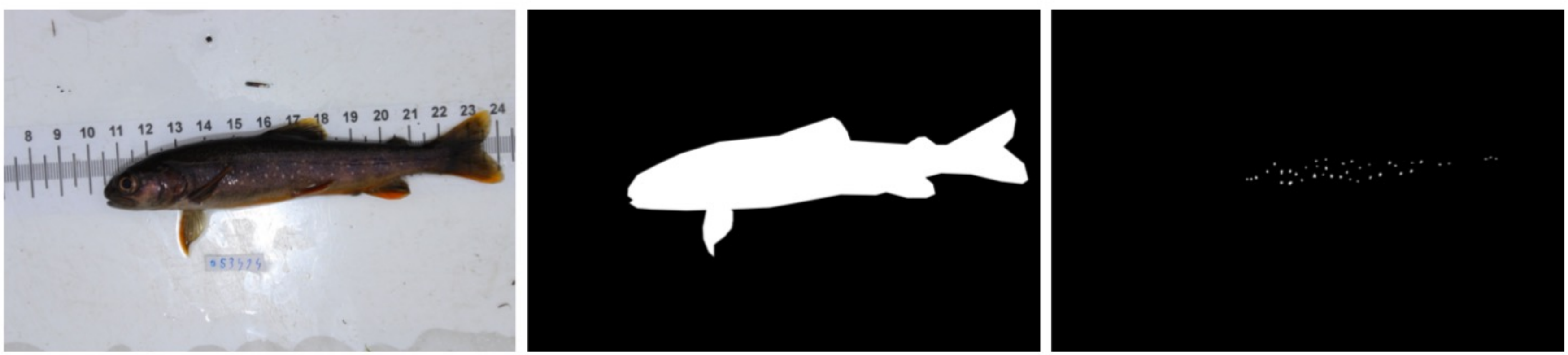
James Thorburn (SOI), Michael Morrissey, Lizy Mittel (Biology), with Jalaj Khandelwal, Ignacy Debicki

Why not end-to-end deep learning?

- Deep learning is about inferring intermediate features from data
  - train a classifier on large existing dataset
  - replace the final few layers (transfer learning)
  - optimise on new problem
- We tried this three times
  - multiple students and supervisors, multiple deep network architectures
  - multiple datasets (flapper skate, arctic charr)
  - train a classifier, replace the final few layers (transfer learning)
- No success
  - relatively small amount of data
  - features are encoded in patterns, harder to learn

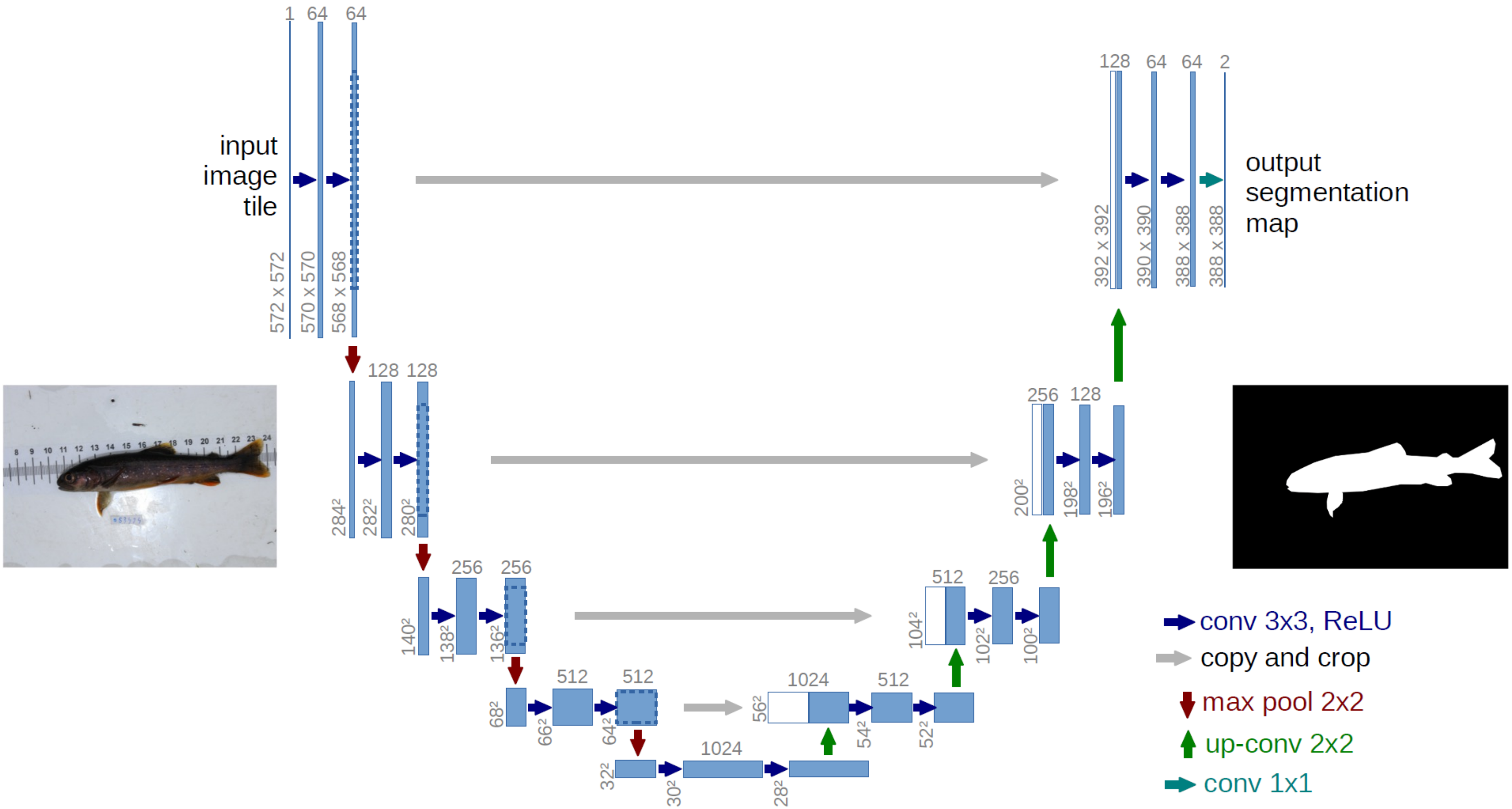
Fish and spot annotation

- Detecting spots has two steps
  - find the fish in the image (improves robustness against noise)
  - find the spots on the fish
- We hand-annotated 52 images
  - for training and for evaluation of algorithms



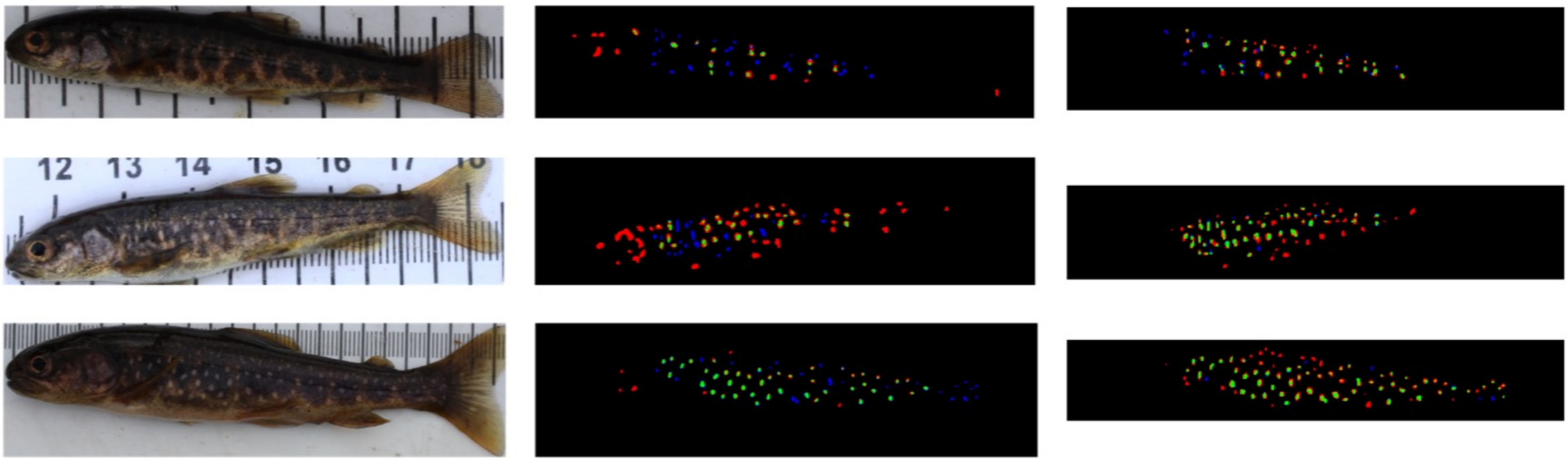
(a) Base image (b) Mask of fish location (c) Mask of spot location

Deep learning algorithm (U-Net)



Spot detection results

Fish specimen | Baseline spot segmentation | U-net spot segmentation

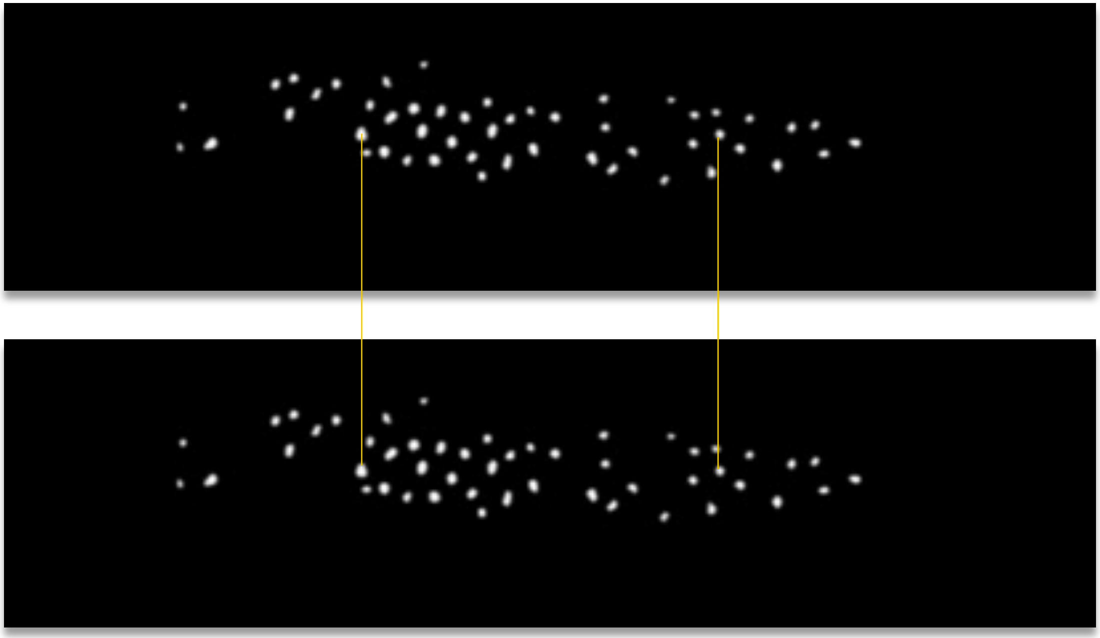


Legend: Correct, False positive, False negative

Baseline F1 score: 0.902 (fish), 0.418 (spots)
U-Net F1 score: 0.972 (fish), 0.566 (spots)

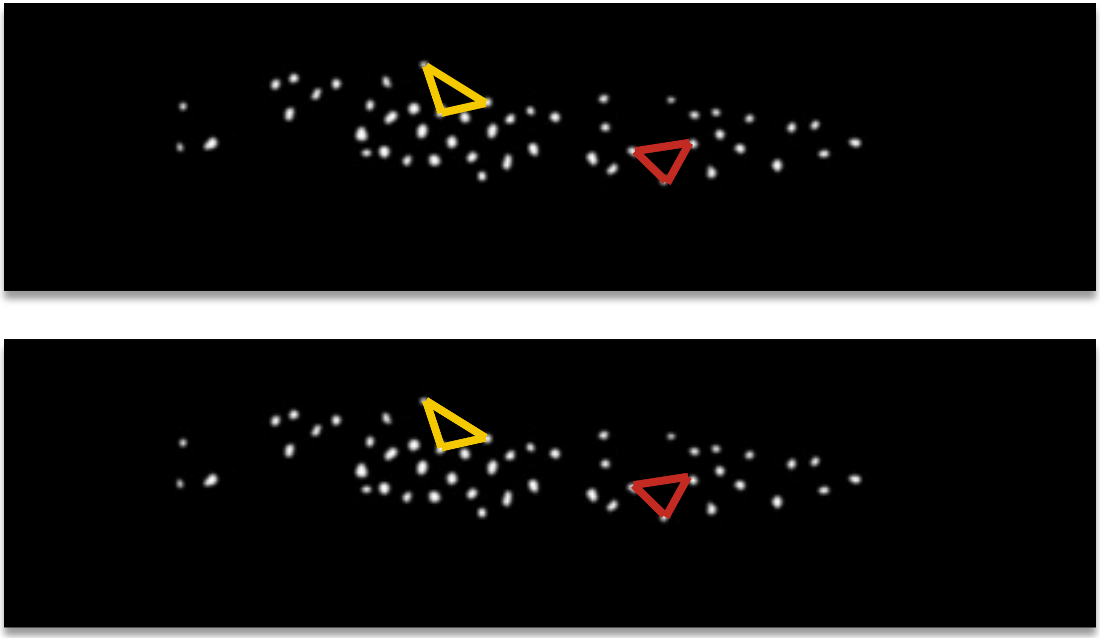
Difficulty of matching individual spots

- Individual spots are very hard to match to each other this way
  - locally, they all look similar to each other
  - and there is a lot of variation in terms of precise segmentation

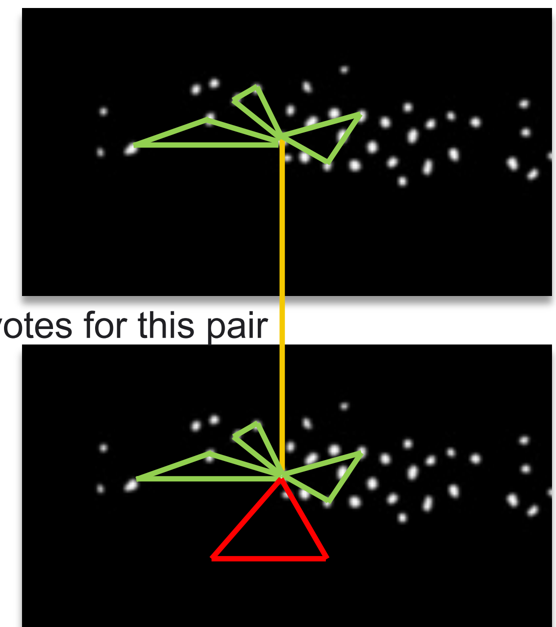


Local features based on fish spots

- Basic idea: Form triangles from spot centres
  - these can be characterised and compared in terms of similarity
  - angle ratios, handedness, orientation, etc.



Modified Groth matcher example

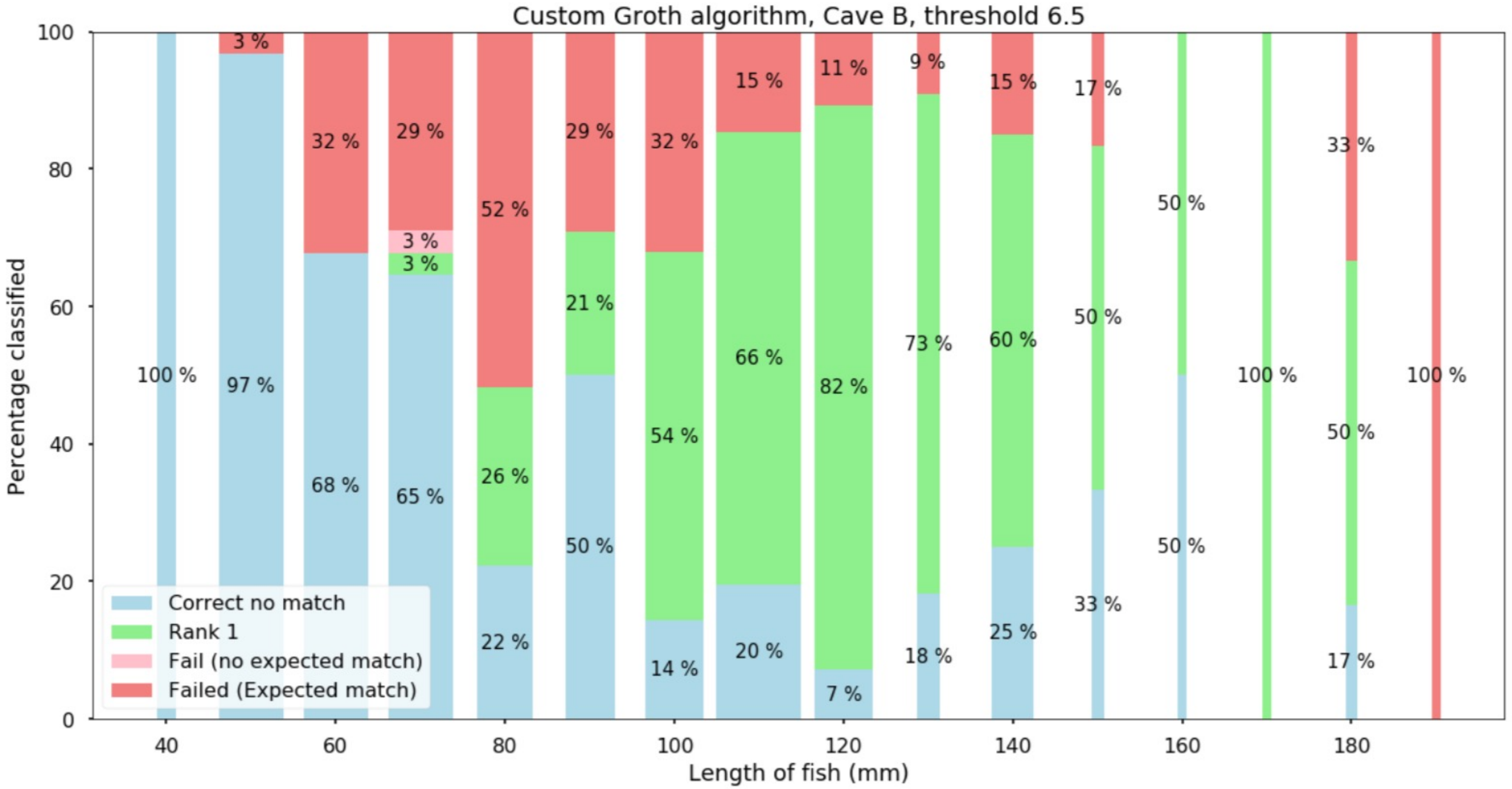


- Each point belongs to several triangles
  - some are correctly matched (green)
  - some are not (red)

- A matched pair of triangles contains three pairs of matched vertices
  - each triangle pair votes for each vertex pair

- After voting, we sort pairs of vertices
  - apply a threshold on number of votes
  - count the number of successfully matched points for every pair of images

Re-identification results (Groth)



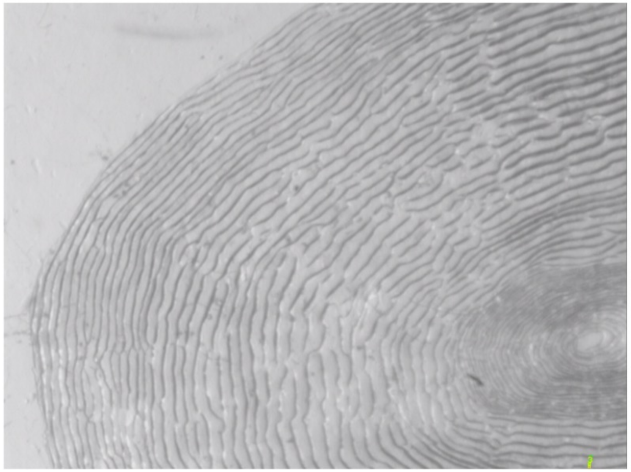
Consequences for survival estimation

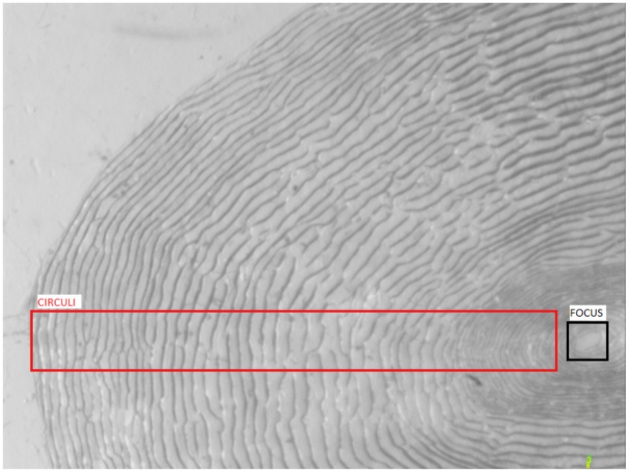
- We identified mismatches in original data
  - 20 in Cave A (10 due to PIT tag loss = 6.5%)
  - 22 in Cave B (12 due to PIT tag loss = 7.5%)
  - the rest were missing tag data or data entry mistakes
- We recalculated survival rate ϕ
  - increase of 4 percentage points
- June 2012 is a special case
  - no tagging was performed
  - ...so we examine it separately

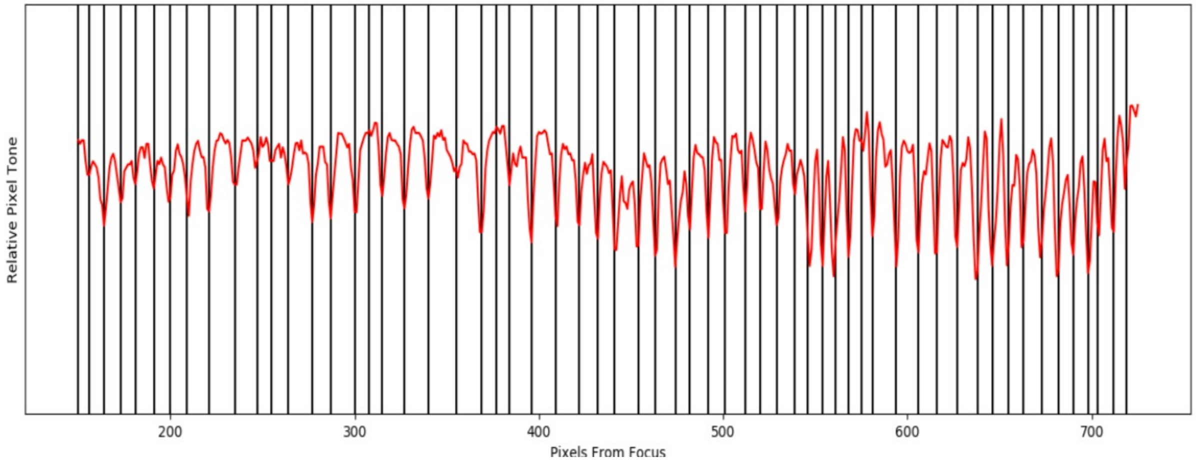
(a) original capture histories – August 2012–2019	
parameter	estimate
ϕ	0.74 (0.71 – 0.77)
p	0.53 (0.48 – 0.57)
(b) corrected capture histories – August 2012–2019	
parameter	estimate
ϕ	0.78 (0.75 – 0.81)
p	0.53 (0.49 – 0.58)
(c) corrected capture histories including June 2012	
parameter	estimate
ϕ	0.77 (0.74 – 0.80)
p	0.49 (0.44 – 0.53)

Re-identification of individuals from images using spot constellations: a case study in Arctic charr (*Salvelinus alpinus*)
IT Dębicki, EA Mittell, BK Kristjánsson, CA Leblanc, MB Morrissey, Royal Society Open Science 8 (7), 201768, 2021

Analysis of salmon scale rings





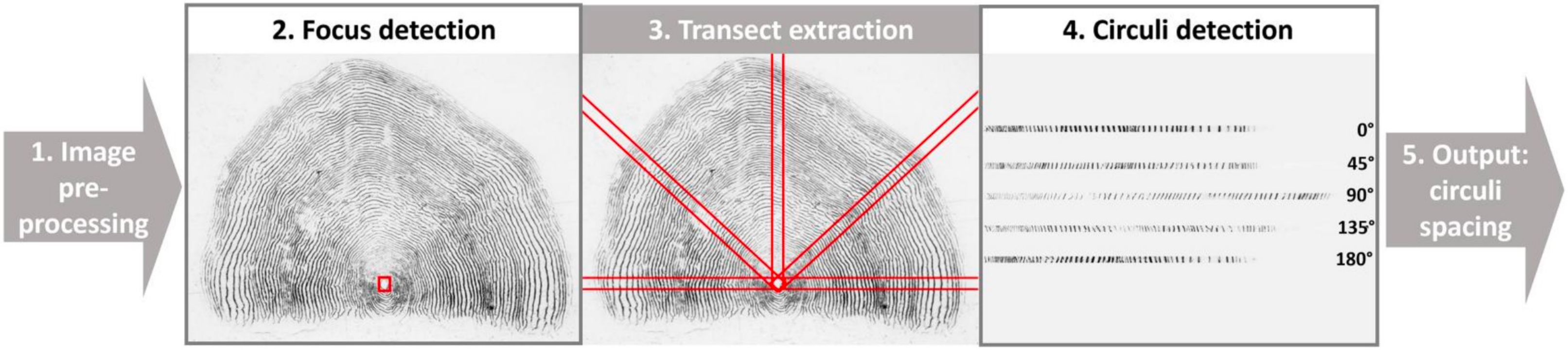


Chris Todd (Biology), Nora Hansen, James Ounsley (MSS), with Joe Guthrie, Bruno Caneco

Scale reading tool

- Developed by Bruno Caneco and James Ounsley
  - automates the scale reading process
  - finds the focus, creates transects, detects circuli, extracts spacings

- Reduces the work from minutes to seconds per image
  - deployed in Scottish Marine Directorate



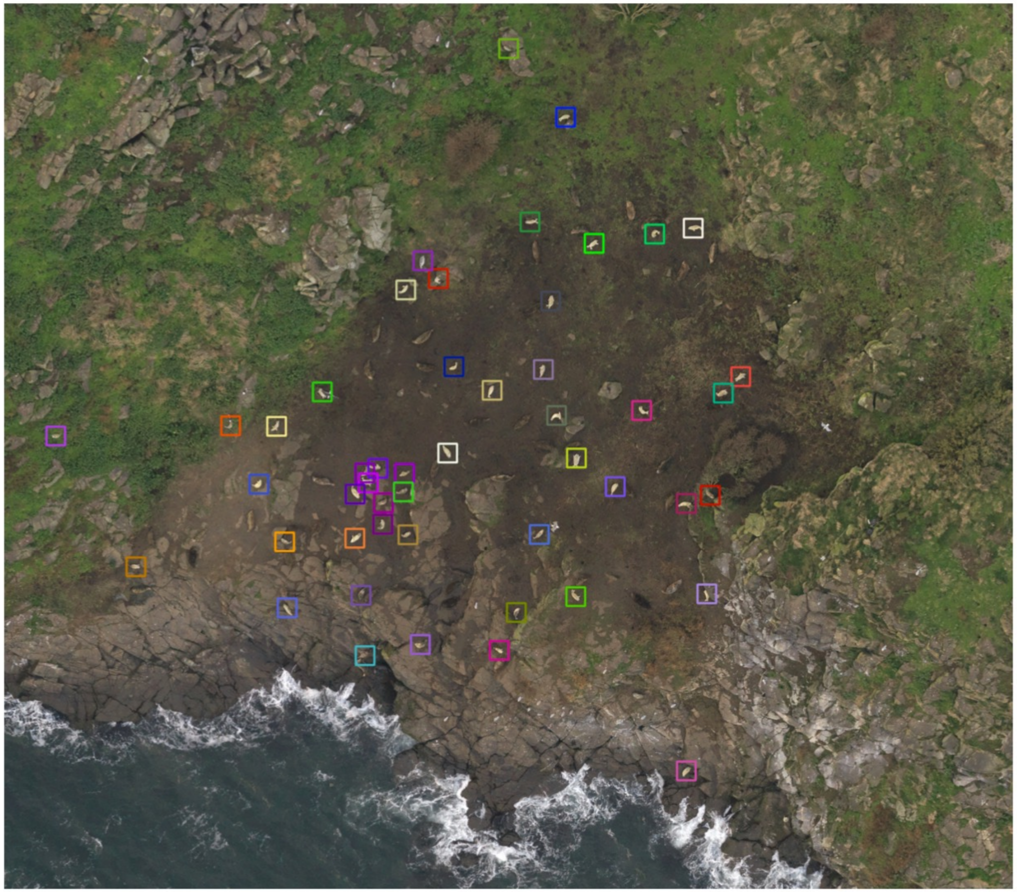
Performance aproaches human experts

Table 3. Total number of correct detections (true positives; TP), incorrect detections (false positives; FP), missed detections (false negatives; FN), and the average precision, recall and F1 of the circuli detector and of a second labeller when compared to labeller one (ground truth) across 87 transects.

	Circuli detector	**Human labeller**
TP	5944 (66; 47–94)	6060 (66; 49–100)
FP	617 (5; 1–16)	546 (4; 0–16)
FN	375 (2; 0–13)	259 (1; 0–12)
Precision	0.90 (0.93; 0.78–0.98)	0.95 (0.95; 0.77–1)
Recall	0.94 (0.97; 0.84–1)	0.96 (0.98; 0.89–1)
F1	0.92 (0.94; 0.82–0.98)	0.94 (0.96; 0.84–0.99)

Median values of TP, FP, FN, and F1 per transect are provided in brackets, followed by the 5th and 95th percentile.

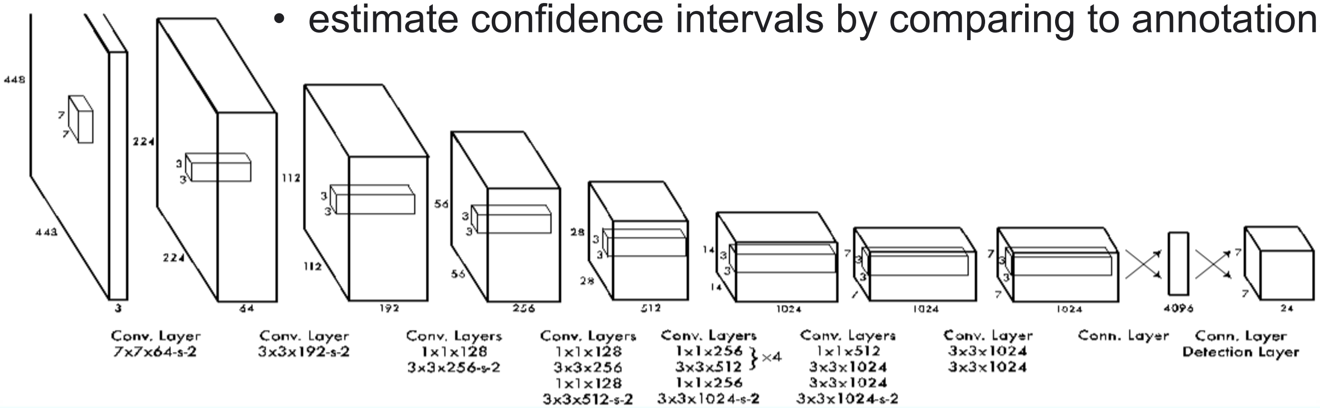
Automated seal pup counting



Chris Morris, Callan Duck (SMRU), Len Thomas (CREEM), with Samuel Pavlik, Dodo Denes, Sophie Bickerton

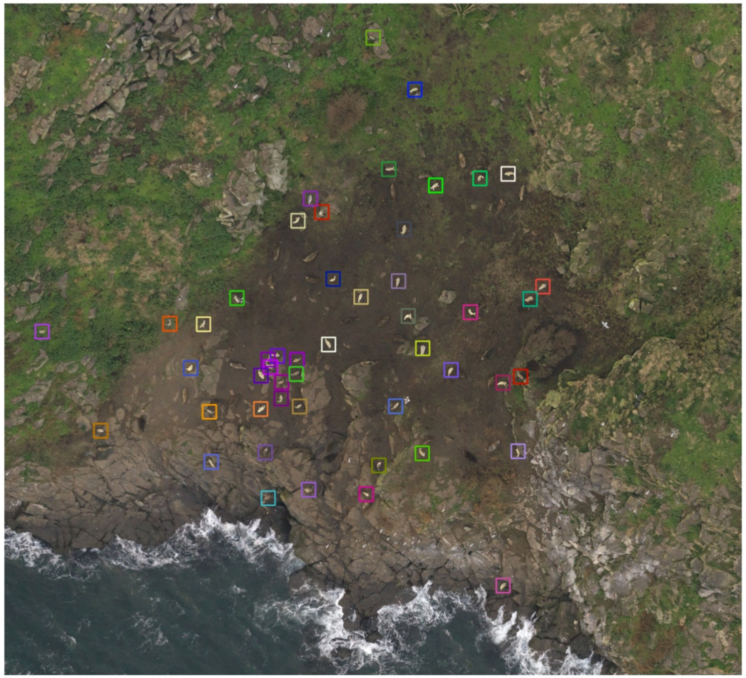
Current pipeline

- Cut large images into smaller tiles programmatically
  - convert annotations to correspond to coordinates in smaller tiles
- Train a deep CNN single-shot detector (YOLO v3)
  - on about 200GB of images
- Use precision and recall to estimate seal number
  - estimate confidence intervals by comparing to annotations



Predictions on new data

- Model is trained on past data
  - using annotated GIS-based gold standard from SMRU
- ...and applied on new images
  - bounding boxes with confidences
- Automatically convert to GIS format
  - for visualisation and inspection
- Counts obtained from detections



Current results

- Trade-off between recall/precision
  - More reliable when prioritising recall
- Estimate prediction intervals for both
  - non-parametric bootstrap model
- Use PIs to estimate true number
  - using threshold with tightest intervals
- Currently too wide
  - 1000 detections corresponds to 17-35 seals

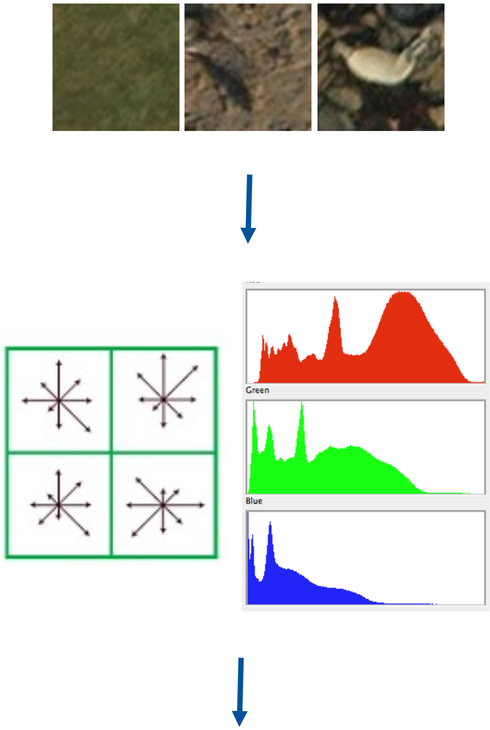
Threshold	Recall	Precision	AP
0.01	0.9971	0.0132	0.9036
0.1	0.9679	0.2479	0.9000
0.2	0.9336	0.5168	0.8866
0.3	0.8942	0.7091	0.8622
0.4	0.8484	0.8269	0.8268
0.5	0.7972	0.8962	0.7824
0.6	0.7347	0.9403	0.7249
0.7	0.6546	0.9668	0.6484
0.8	0.5447	0.9839	0.5411
0.9	0.3429	0.9930	0.3414

	Mean	Mean 95%	SD	SD 95%
Recall	0.969	(0.962, 0.978)	0.031	(0.024, 0.037)
Precision	0.026	(0.016, 0.034)	0.035	(0.020, 0.051)

Chris Morris, Callan Duck (SMRU), Len Thomas (CREEM), with Samuel Pavlik, Dodo Denes, Sophie Bickerton

Refining the detections

- Improve precision in a second stage
  - starting with detections from YOLO
- Based on classic computer vision methods
  - Histogram of Oriented Gradients (HOG)
  - Colour histogram
- Ran as part of CS5014 Machine Learning
- Over 95% accuracy on test data
  - over 20,000 examples



Classifier (SVM)

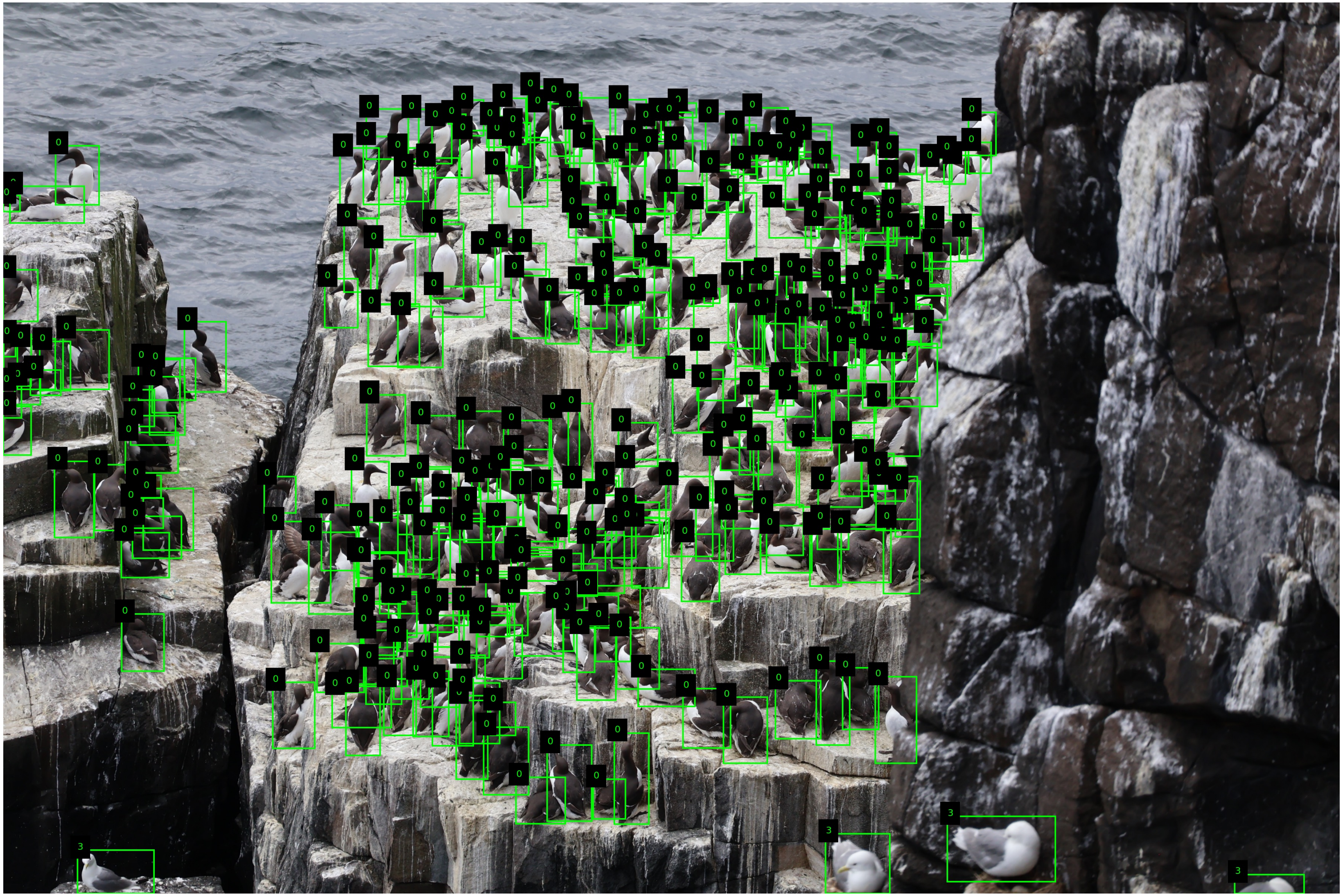
Monitoring cliff-nesting seabird colonies

- PhD student started in May 2023
  - joint with Schools of Biology and Psychology
- Monitoring seabird colonies
  - St Abb's Head and Isle of May
- Initial results on detection and counting
  - YOLOv4, RetinaNet
- Our goal is to understand behaviours
  - challenges in terms of tracking across video
  - need to re-ID individuals/nests across trips

Charis Hanna, Mark James (SOI), Karen Spencer (Psy)

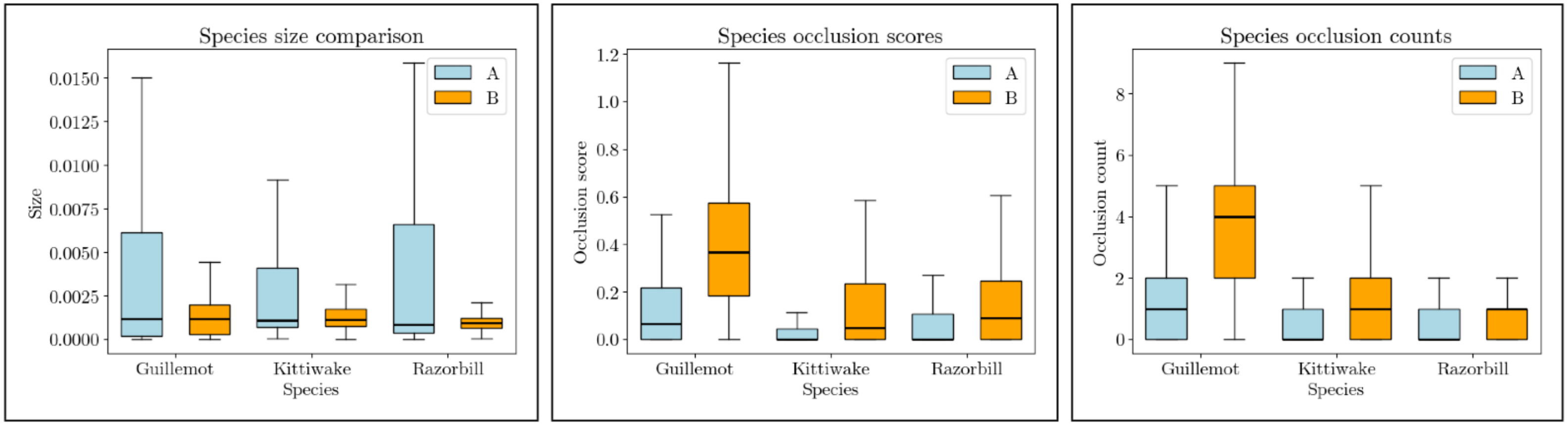
Data collection

- Existing annotated datasets
  - generally individual birds, easily identified and annotated
  - CUB, NABirds, iNaturalist
- Data collection is very hard
  - numerous trips to difficult-to-reach places
  - manual photography
  - annotation is hard and labour-intensive
- Our new dataset
  - ~1000 images from Isle of May (more to come)
  - ~200 annotated images (~100k individual birds)



Challenges for standard algorithms

- Some very challenging features
  - crowding and occlusion
  - large class imbalances
  - large range of sizes



(a) (b) (c)

State of the art in object detection

- Off-the-shelf object detectors do not do well on cliff data
  - reasonable performance, but not good enough for monitoring
  - transfer learning is only partially helpful

- Tested on two datasets, with similar results

		YOLO			RetinaNet			Faster R-CNN		
Species	No.	D	C	N	D	C	N	D	C	N
Shag	54	.402	.500	.333	.66	.750	.842	.526	.477	.416
Fulmar	40	.667	**1.00**	1.00	.76	**.833**	.738	.867	**.867**	1.00
Kittiwake	329	.841	*.751*	**.916**	.89	*.976*	**.936**	.749	*.921*	**.973**
Razorbill	171	.762	.953	.941	.91	.997	.967	.971	1.00	1.00
Puffin	181	.920	*.911*	.960	.92	*.962*	.934	.970	*.856*	.912
Gannet	453	.694	.664	**.781**	.82	.799	**.766**	.848	.898	**.793**
Guillemot	651	*.723*	*.742*	*.788*	.75	*.839*	*.778*	.726	*.755*	*.734*

Some conclusions so far

- Availability of data is key for deep learning algorithms
  - this usually means appropriately annotated data
  - algorithms with lots of good annotated data did well

- Black box solutions showed partial success
  - good at specialist tasks (detection, segmentation)
  - not very good at learning intermediate features from weak annotation

- Combination of domain and ML knowledge did best
  - arctic charr re-ID and seal detection
  - expert annotations and guidance to produce custom algorithms

Some challenges

- Need for annotations
  - opportunities for weakly supervised and self-supervised algorithms
- Need for more robust evaluation
  - confidence intervals, repeatable and reliable evaluation
  - this, in turn, requires well-annotated data
- Need for interdisciplinary collaborations
  - domain expertise and involvement is essential
- Need for standardised tools that can be adopted widely


University of
St Andrews
www.st-andrews.ac.uk