

Employing Neural Network Models to Obtain Accurate Positional Coordinates of Malawi Cichlids

By Rhiya Sharma
Advisor: Patrick McGrath

Abstract

Pose-estimation of the cichlid fishes of Lake Malawi is an excellent way to study variations in social behavioral and disease. Machine learning strategies can be employed to automatically quantify animal behavior which will have the potential to reveal unforeseen features and observe patterns which might otherwise go unnoticed by the human eye. I will be using various machine learning models to accurately predict the position of the cichlids while they are engaged in building bowers.

Background

Quantification of behavior of animals is important to understand how their brains work. It also helps understand the causes behind behavioral evolution [1] and its effects on the study of neural mechanisms underlying social behaviors. Extracting the pose of animals is one way of analyzing animal behavior, especially its courtship behavior. This can be done by employing a software system that combines video tracking, computer vision and machine learning strategies to automatically track the body parts and movements of animals in their environment [2]. Tracking the spatiotemporal coordinates of animals can be used to deduce their stereotypical behavior.

Large aquarium tanks, in the McGrath lab, containing a sand tray (approx. 14"x17") are used to record the movements of the Malawi cichlids and allow the male cichlids to build bowers. These sand-dwelling species have a unique innate behavioral trait of building bowers in the sand that serve as extended phenotypes and territories during courtship and spawning. The two bower types – “castles” and “pits” are distributed widely across the species *Mchenga conophorous* (castle), *Copadichromis virginalis* (pit) and *Tramitichromis intermedius* (pit).

The males court the females in parallel when constructing the bowers and also get involved in aggressive encounters with conspecific males. To build the bower, the males are required to repeatedly perform hundreds of scoops–spit bouts with their mouths per hour. Digging pits requires the males to collect sand from the center of the pit and spit it elsewhere as opposed to building castles, which requires the males to gather sand from elsewhere and spit it in a targeted location [3].

Tracking the species' behavior by a human observer is tedious and time consuming as building bowers and attracting females is a week-long process. Moreover, the similarity between the fish and the sand color can make it difficult to identify the fish with an untrained eye.

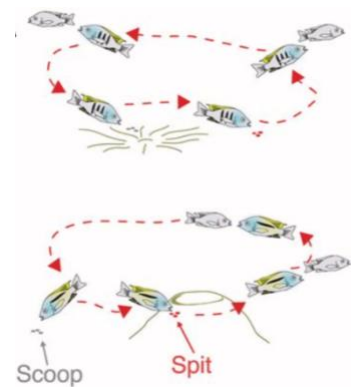


Figure 1: Pit digging & castle bower

For these reasons, the McGrath and Streelman labs are interested in automating this process to study behavioral evolution.

Work done in Spring 2020

Some of the ways to improve the accuracy of an R-CNN model is to add more data and to check the consistency and accuracy of the labelling across the input data. In order to do this, two users are made to annotate the same images and an analysis of these annotations were conducted. High agreement levels between the annotators suggest that the annotations are reliable. The agreement levels are calculated based on an IOU score (Intersection Over Union).

IOU is the overlapping areas of intersection between two bounding boxes (annotations made by the two users), divided by the total area of both bounding boxes. This produced an accuracy score that was used to measure how close two bounding boxes match. This process helped determine the annotation strategy to follow for all future datasets (extent of fish image to be included) and choose a threshold value to determine poorly annotated datasets. The threshold value taken for all the image datasets was 0.5, i.e. an accuracy score of below 0.5 is an incorrect match.

I annotated 50 frames each from 28 different trials containing only male or female fishes from the MC, CV, TI and MCxCV species. These annotations would help the model distinguish between the males and females. I then annotated 500 frames each from 7 different trials (MC, CV, TI and MCxCV) containing one male and multiple females in one tank. To do this, I had to have a clear understanding of the differences between the male and female cichlid features.

Species	Fish	Trials	No. of Annotations	Species	Fish	Trials	No. of Annotations
Copadichromis virginalis (CV)	4 females	CV_fem_con1	50	Mchenga conophorous (MC)	4 females	MC_fem_con1	150
	4 females	CV_fem_con2	50		4 females	MC_fem_con2	150
	4 females	CV_fem_con3	50		4 females	MC_fem_con3	50
	1 male	CV_male_con1	50		1 male	MC_male_con1	50
	1 male	CV_male_con2	50		1 male	MC_male_con2	50
	1 male	CV_male_con3	50		1 male	MC_male_con3	50
	1 male	CV_male_con4	50		1 male	MC_male_con4	50
	4 males	CV_social_male_con1	50		4 males	MC_social_male_con1	50
	4 males	CV_social_male_con1_2	50		4 males	MC_social_male_con1_2	50
	4 males	CV_social_male_con2	50		4 males	MC_social_male_con2	50
	4 males	CV_social_male_con3	50		4 males	MC_social_male_con2	50
	4 males	CV_social_male_con3_2	50		4 males	MC_social_male_con3	50
	1 male + 4 females	CV10_3	500		1 male + 4 females	MC6_5	500
					1 male + 4 females	MC16_2	500
Tramitichromis intermedius (TI)	1 male	TI_male_con1	50	Mchenga conophorous × Copadichromis virginalis (MCxCVF1)	1 male + 4 females	MCxCVF1_12a_1	500
	1 male	TI_male_con2	50		1 male + 4 females	MCxCVF1_12b_1	500
	4 females	TI_social_fem_con1	50				
	4 males	TI_social_male_con1	50				
	4 males	TI_social_male_con2	50				
	1 male + 4 females	TI2_4	500				
	1 male + 4 females	TI3_3	500				

Figure 2: Summary of annotations done

Analyzing the annotations:

I wrote a script to calculate the IOU score and to also observe the disagreements between the two annotations. Based on these results, an annotating strategy was developed between the two users to increase uniformity and the trials with a high disagreement were re-annotated.

ProjectID x	Framefile	Nfish x	Sex x	Box x	User x	ProjectID y	Nfish y	Sex y	Box y	User y	Same Nfish	IOU Score_calculated
MC_fem_con1	MC_fem_con1_0001_vid_10048.jpg	1	f	(1041, 258, 22, 190)	rhiyasharma	MC_fem_con1	0	(0,)	(0,)	user2	FALSE	0
MC_fem_con1	MC_fem_con1_0001_vid_103730.jpg	1	f	(643, 541, 105, 53)	rhiyasharma	MC_fem_con1	1	m	(641, 527, 123, 81)	user2	TRUE	0.558566697
MC_fem_con1	MC_fem_con1_0001_vid_108127.jpg	1	f	(831, 25, 106, 38)	rhiyasharma	MC_fem_con1	2	m	(836, 23, 100, 50)	user2	FALSE	0.726855394
MC_fem_con1	MC_fem_con1_0001_vid_108127.jpg	1	f	(831, 25, 106, 38)	rhiyasharma	MC_fem_con1	2	m	(836, 23, 100, 50)	user2	FALSE	0.726855394
MC_fem_con1	MC_fem_con1_0001_vid_112744.jpg	1	f	(511, 903, 109, 49)	rhiyasharma	MC_fem_con1	2	m	(506, 890, 106, 62)	user2	FALSE	0.710654796
MC_fem_con1	MC_fem_con1_0001_vid_112744.jpg	1	f	(511, 903, 109, 49)	rhiyasharma	MC_fem_con1	2	m	(506, 890, 106, 62)	user2	FALSE	0.710654796
MC_fem_con1	MC_fem_con1_0001_vid_118159.jpg	0	(0,)	(0,)	rhiyasharma	MC_fem_con1	0	(0,)	(0,)	user2	TRUE	1
MC_fem_con1	MC_fem_con1_0001_vid_122266.jpg	2	f	(60, 859, 85, 91)	rhiyasharma	MC_fem_con1	2	m	(49, 853, 81, 90)	user2	TRUE	0.642974303
MC_fem_con1	MC_fem_con1_0001_vid_122266.jpg	2	f	(60, 859, 85, 91)	rhiyasharma	MC_fem_con1	2	m	(462, 893, 68, 71)	user2	TRUE	0
MC_fem_con1	MC_fem_con1_0001_vid_122266.jpg	2	f	(477, 908, 42, 61)	rhiyasharma	MC_fem_con1	2	m	(49, 853, 81, 90)	user2	TRUE	0
MC_fem_con1	MC_fem_con1_0001_vid_122266.jpg	2	f	(477, 908, 42, 61)	rhiyasharma	MC_fem_con1	2	m	(462, 893, 68, 71)	user2	TRUE	0.466851925
MC_fem_con1	MC_fem_con1_0001_vid_142263.jpg	0	(0,)	(0,)	rhiyasharma	MC_fem_con1	0	(0,)	(0,)	user2	TRUE	1
MC_fem_con1	MC_fem_con1_0001_vid_148312.jpg	0	(0,)	(0,)	rhiyasharma	MC_fem_con1	0	(0,)	(0,)	user2	TRUE	1
MC_fem_con1	MC_fem_con1_0001_vid_149016.jpg	0	(0,)	(0,)	rhiyasharma	MC_fem_con1	0	(0,)	(0,)	user2	TRUE	1

Figure 3: Score Report for User 1 & User 2 for the MC_fem_con1 trial

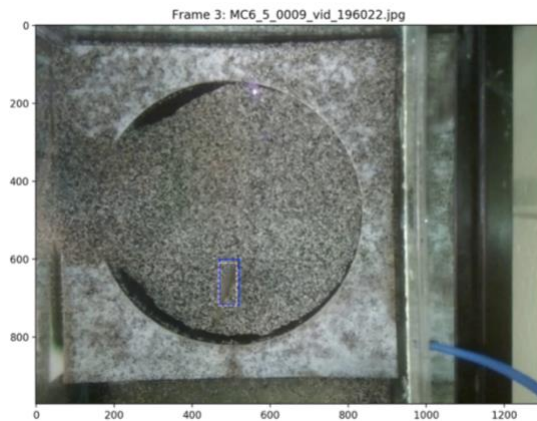


Figure 4:

- User 1 (solid line) & User 2 (dotted line) annotations for MC6_5. This image shows a disagreement between the two annotations.
- Blue bounding box – male annotation, pink bounding box – female annotation.

Proposal for Summer 2020

Convolutional networks like Fast R-CNN, YOLO and Mobilenet will be used to efficiently trace the movements of the fishes.

Fast R-CNN builds upon the existing R-CNN (Region-Based Convolutional Neural Network). It comprises of a single model to learn and output regions and classifications directly. The model takes a set of region proposals as input and passes them through a deep convolutional neural network. A pre-trained CNN is then used for feature extraction. The RoI (Region of Interest) Pooling layer at the end of the deep CNN extracts features specific for a given input candidate region. The output of the CNN is then interpreted by a fully connected layer [4].

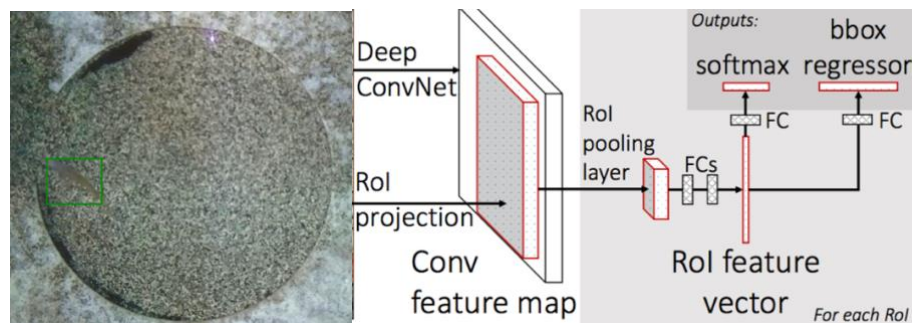


Figure 5: Fast R-CNN Model Architecture

Another neural network is the YOLO model, You Only Look Once. It is a single neural network, which is trained end to end, and predicts bounding boxes and class probabilities directly from full images in one evaluation. It works by first splitting the input image into a grid of cells and each grid predicts N bounding boxes and confidence (reflects the accuracy of the bounding box and whether the bounding box actually contains an object). YOLO also predicts the classification score for each box for every class in training. The class probabilities map and the bounding boxes with confidences are then combined into a final set of bounding boxes and class labels [5].

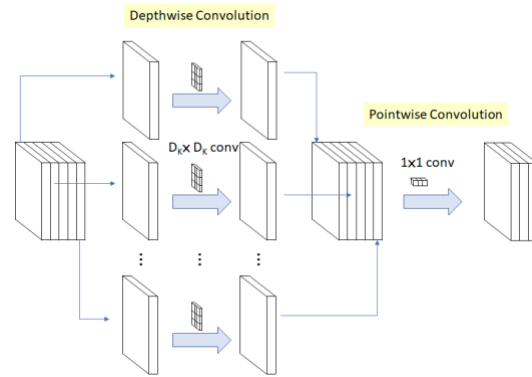


Figure 6: Summary of predictions made by YOLO model.

Finally, Mobilenet is a light weight deep neural network as it uses depth-wise separable convolutions. The model performs a single convolution on each color channel rather than combining all three and flattening it. This has the effect of filtering the input channels. The pointwise convolution then applies a 1×1 convolution to combine the outputs of the depth-wise convolution. A standard convolution both filters and combines inputs into a new set of outputs in one step. The depth-wise separable convolution splits this into two layers, a separate layer for filtering and a separate layer for combining. This factorization has the effect of drastically reducing computation and model size [6].

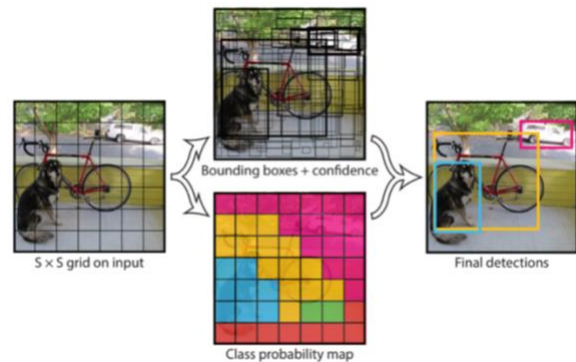


Figure 7: Depth-wise Separable Convolutions

Each model has its pros and cons. In general, Fast R-CNN models are known to be more accurate, however YOLO and Mobilenet are much faster as they trade off accuracy for good speed return. Training these models with a good balance between speed and accuracy might provide us with good results. During the summer, I will be working on making these models work to accurately predict the fish position before and after the bower-building spits and scoops.

I would like to thank the committee for taking the time to read my proposal, I hope to be funded for the Summer semester so that I can complete the proposed work.

References

- [1] York, Ryan A., et al. "Evolution of bower building in Lake Malawi cichlid fish: phylogeny, morphology, and behavior." *Frontiers in Ecology and Evolution* 3 (2015): 18
- [2] W. Hong et al., "Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning", 2019.
- [3] York, R., Patil, C., Abdilleh, K., Johnson, Z., Conte, M., Genner, M., McGrath, P., Fraser, H., Fernald, R. and Streelman, J., 2018. Behavior-dependent cis regulation reveals genes and pathways associated with bower building in cichlid fishes. *Proceedings of the National Academy of Sciences*
- [4] Girshick, R., 2020. Fast R-CNN.
- [5] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2020. You Only Look Once: Unified, Real-Time Object Detection.
- [6] Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2020. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications.