Revealing sea turtle behavior in relation to fishing gear using color-coded spatiotemporal motion patterns with deep neural networks

Janie L. Reavis[1,*], H. Seckin Demir[2], Blair Witherington[3], Michael Bresette[3], Jennifer Blain Christen[2], Jesse Senko[4] and Sule Ozev[2]

[1]*School of Life Sciences, Arizona State University, Tempe, AZ, USA*

[2]*Department of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA*

[3]*Inwater Research Group, Jensen Beach, FL, USA*

[4]*School for the Future of Innovation in Society, Arizona State University, Tempe, AZ, USA*

Correspondence*: Janie L. Reavis
jreavis3@asu.edu

ABSTRACT

Incidental capture, or bycatch, of marine species is a global conservation concern. Interactions with fishing gear can cause mortality in air-breathing marine megafauna, including sea turtles. Despite this, interactions between sea turtles and fishing gear - from a behavior standpoint - are not sufficiently documented or described in the literature. Understanding sea turtle behavior in relation to fishing gear is key to discovering how they become entangled or entrapped in gear. This information can also be used to reduce fisheries interactions. However, recording and analyzing these behaviors is difficult and time intensive. In this study, we present a machine learning-based sea turtle behavior recognition scheme. The proposed method utilizes visual object tracking and orientation estimation tasks to extract important features that are used for recognizing behaviors of interest with green turtles (*Chelonia mydas*) as the study subject. Then, these features are combined in a color-coded feature image that represents the turtle behaviors occurring in a limited time frame. These spatiotemporal feature images are used along a deep convolutional neural network model to recognize the desired behaviors, specifically evasive behaviors which we have labelled “reversal” and “U-turn.” Experimental results show that the proposed method achieves an average F1 score of 85% in recognizing the target behavior patterns. This method is intended to be a tool for discovering why sea turtles become entangled in gillnet fishing gear.

Keywords: green turtle, chelonia mydas, behavior recognition, color-coding, spatiotemporal features

This material is based upon work supported by the National Science Foundation under Grant No. 1837473. The research is also supported by Inwater Research Group and Florida Power and Light Company. The work on protected species was conducted under Florida FWC Marine Turtle Permit 20-125.

1 INTRODUCTION

Incidental capture of non-target animal species, termed bycatch, in fisheries is a global ecological threat to marine wildlife (Estes et al., 2011). Fisheries bycatch poses a threat to air-breathing animals such as sea turtles. One such gear, gillnets, can create an ecological barrier that does not naturally occur, so there is likely no evolutionary mechanism that causes avoidance (Casale, 2011). Various approaches have been proposed to reduce bycatch rates of sea turtles and other marine megafauna (Lucchetti et al., 2019; Wang et al., 2010a; Demir et al., 2020). Attempted solutions include: marine policy that sets bycatch limits for fisheries (Moore et al., 2009); acoustic deterrents similar to pingers used to prevent dolphin bycatch; buoyless nets and illuminated nets, which have shown promising results for reducing bycatch in coastal net fisheries (Peckham et al., 2016; Wang et al., 2010b). These bycatch reduction approaches can involve changing the technical design of gear or introducing novel visual or acoustic stimuli, which also changes gear configuration. However, as a part of the design process, effectiveness of different types of stimuli must be analyzed by observing the associated behavioral response of sea turtles, which has not been clearly documented in previous studies. Analyzing sea turtle interactions with fishing gear and bycatch reduction technologies (BRTs) is not an easy task, since it requires the researcher to monitor the experiment underwater for long periods while identifying and recording sea turtle behaviors and ensuring the study subject's safety. Even when experiments are recorded with GoPros, short battery life requires constant monitoring of each camera view, and the subsequent manual behavioral analysis is time-intensive for researchers. Fortunately, with the developments in computer vision-based approaches, recognition of certain behaviors can be performed automatically after training this convolutional neural network with behavioral data.

Various approaches have been proposed to complete the behavior recognition task for different applications involving humans or animals (Chakravarty et al., 2019; Nweke et al., 2018; Porto et al., 2013; Yang et al., 2018; Bodor et al., 2003; Ijjina and Chalavadi, 2017). Some of the recognition algorithms analyze the data captured using wearable sensors (Chakravarty et al., 2019; Nweke et al., 2018). While the sensors used in these type of experiments provide valuable information about the activities of interest, they are not applicable in our context as they need to be located on the subject's body in a controlled environment. Various methods use vision based approaches for the behavior recognition task (Porto et al., 2013; Yang et al., 2018; Bodor et al., 2003; Ijjina and Chalavadi, 2017). Earlier examples of the vision based methods employ hand-crafted features for analyzing the activities (Porto et al., 2013; Bodor et al., 2003). While these approaches can perform well for differentiating basic behaviors, they are not very efficient in recognizing complex activities. With the advancements in the machine learning field, recent studies employ deep neural networks (DNN) successfully for the activity recognition task (Yang et al., 2018; Ijjina and Chalavadi, 2017). Although DNNs provide powerful representations to analyze complex data sets, end-to-end training approaches usually require large amounts of data samples and a large number of network coefficients. In this study, we propose a hybrid approach for the sea turtle behavior recognition task. We use domain knowledge for determining base features to recognize certain behaviors and convert them into color-coded spatiotemporal 3-D images to train deep convolutional neural networks (CNN).

In our application, we are specifically interested in recognizing ”U-Turn” and ”Reversal” behaviors of turtles, since they are important indicators of effectiveness of the given stimuli. In order to recognize these behaviors, we combine turtle location, velocity, and orientation information in spatiotemporal images and use these images as inputs to a CNN architecture.

In the U-turn behavior, the turtle makes a u-shaped maneuver in a short amount of time possibly due to an external visual stimulus. In Reversal behavior, the turtle moves backwards while facing forward rather than changing its orientation. These are avoidance behaviors exhibited by sea turtles when faced with a barrier or other deterrent. To differentiate these behaviors from each other and from other motion patterns, we use turtle location, speed, and orientation information. In order to extract those features and combine them as an input to a deep neural network based architecture, we propose the recognition system shown in Figure 1.

Here, we explain how we conduct the physical experiments and provide an overview of the proposed behavior recognition framework and explain the functional blocks. We then present the results of our comparative study for the object tracking task on the turtle dataset that we collected. We then report the performance of the proposed orientation estimation network and behavior recognition network followed by an explanation of the anticipated results and utility for conservation purposes.

2 MATERIALS AND EXPERIMENTAL SET UP

2.1 Animal acquisition and facility maintenance

All sea turtles used in this study were captured by Inwater Research Group (IRG) via dip net, entangling net, or hand capture after entrainment in the intake canal at the St. Lucie Nuclear Power Plant in Jensen Beach, FL. Capture of these turtles is necessary for returning them to the open ocean. For our choice trials, we included healthy juvenile and subadult green (*C. mydas*) turtles with a standard carapace length of less than 78 cm. After the IRG team removed turtles from the canal and collected biometric data, all turtles were kept in separate 6 ft diameter holding tanks with circulating seawater from the canal. Turtles were not held for more than 72 hours.

2.2 Computer Setup

Captured data have been processed using a computer with Intel(R) Core(TM) i7-9750H processor and NVIDIA RTX 2070 GPU unit and 16GB RAM. For deep neural network architectures, we have used Keras Libraries[1].

3 METHODS

3.1 Animal behavior experiments and analysis

We conducted all tank experiments in a 13.9 x 2.3 x 1.5 m concrete tank beside the intake canal at the St. Lucie Nuclear Power Plant in Jensen Beach, Florida (Figure 2). The two treatments we used in the development of this method consisted of a gillnet vs. no gillnet set up during the day and at night, meaning a turtle was given the choice between a pathway with a gillnet fully blocking it or a pathway with nothing in it (see Figure 2). The variable being changed is time of day with darkness being the most important factor in nighttime experiments. Each turtle was used in three consecutive 15- minute trials with the same treatment. All trials were recorded using GoPro Hero8 cameras from 4 different viewpoints, although this

[1] https://keras.io

study focused on behaviors recorded from the primary overhead view, as shown in Figure 2. Turtle behavior was analyzed from the recordings rather than in real-time due to the need to monitor turtle safety. Here, we specifically focus on the novel turtle avoidance behavior identified in relation to the gillnet deployed in the treatment sector: Reversal and U-turn. A Reversal occured when a turtle made contact with the gillnet and then escaped by moving backward with its rear flippers and maintaining a forward-facing orientation. A U-turn involved a 180 degree turn within a 3-second period. Here, we only classify U-turns that occur near the barrier of interest (i.e. the gillnet or treatment area containing the gillnet).

3.2 Related Work

Our behavior recognition approach requires the turtle location information in every frame. Thus, we included an object tracking method as part of the design. The visual object tracking problem has long been studied in the computer vision field. Early methods have commonly used correlation based approaches and hand-crafted features for the tracking task. In (Ross et al., 2007), the authors proposed a method (IVT) that employs an incremental principal component analysis algorithm to achieve low dimensional subspace representations of the target object for tracking purposes. In (Babenko et al., 2009), a multiple instance learning (MILTrack) framework was used for object tracking where Haar-like features were used for discriminating the positive and negative image sets. In (Bolme et al., 2010), an adaptive correlation based algorithm (MOSSE) that calculates the optimal filter for the desired Gauss-shaped correlation output was proposed. In another approach (Bao et al., 2012), Bao et al. modeled the target by using a sparse approximation over a template set (L1APG). In this method, an `-1 norm related minimization problem was solved iteratively to achieve the sparse representation. In (Gundogdu et al., 2015), an adaptive ensemble of simple correlation filters (TBOOST) was used to generate tracking decisions by switching among the individual correlators in a computationally efficient manner. (Henriques et al., 2015) presents a method to use Kernelized Correlation Filters (KCF) operating on histogram of oriented gradients, where the key idea is to include all the cyclic shift versions of the target patch in the sample set, and train the network in Fourier Domain efficiently. In (Danelljan et al., 2015), authors propose a discriminative correlation filter based approach (SRDCF) where they use a spatial regularization function that penalizes filter coefficients residing outside the target region. In (Demir and Cetin, 2016), authors propose a ”co-difference” feature-based tracking algorithm (CODIFF) to efficiently represent and match image parts. This idea is further extended in (Demir and Adil, 2018) by including a part based approach (P-CODIFF) to achieve robustness against rotations and shape deformations. In (Bertinetto et al., 2016), the authors propose a method (STAPLE) to combine both correlation based and color based representations to construct a model that is robust to intensity changes and deformations. More recent methods use CNNs for the tracking task. Siamese network based methods have achieved remarkable results for the object tracking benchmarks (Kristan et al., 2019, 2020; Li et al., 2019). In our experiments, we compared the performance of various state-of-the-art tracking algorithms on our dataset and used the best performing method for our application. Detailed results are given in Section 4.1.

As a part of our design, we also estimated turtle orientation to differentiate some of the behavior patterns.

Various methods have been proposed to estimate the orientation of animals (Wagner et al., 2013), humans (Raza et al., 2018), and other objects (Hara et al., 2017). Similar to the tracking and behavior recognition problems, deep CNNs have successfully been used for the orientation estimation problem as well. In our method, a lightweight CNN architecture is employed to estimate the turtle orientation.

3.3 Proposed Method

In this study, we intended to successfully recognize *U-turn* and *Reversal* behaviors of sea turtles. To differentiate these behaviors from each other and from other motion patterns, we use turtle location, speed, and orientation information. In order to extract those features and combine them as an input to a deep neural network based architecture, we propose the recognition system shown in Figure 1.

The turtle location and speed were calculated by the visual object tracker and the turtle orientation calculated by the angle estimation network are combined to generate color-coded spatiotemporal images. The images are used by another network as the input for the behavior recognition task. Details of these building blocks are given in the subsections below.

3.3.1 Visual Object Tracker

The purpose of the visual object tracking block is to find the object location and size in every frame based on a given initial bounding box. Object location found by the visual object tracker is used to calculate the motion velocity vector ($\mathbf{v}$). Bounding box output is also used to crop the object region from the image for the angle estimation network. $\mathbf{v_n}$ is calculated from the current object location $\mathbf{p_n}$ and the previous object location $\mathbf{p_{n-1}}$ as shown in Eqn. 1.

$$\mathbf{v_n} = \begin{bmatrix} v_{x_n} \\ v_{y_n} \end{bmatrix} = \begin{bmatrix} p_{x_n} \\ p_{y_n} \end{bmatrix} - \begin{bmatrix} p_{x_{n-1}} \\ p_{y_{n-1}} \end{bmatrix} \tag{1}$$

In order to employ a successful object tracking algorithm in the proposed framework, we performed a comparison between the state-of-the-art visual object trackers IVT (Ross et al., 2007), MILTrack (Babenko et al., 2009), MOSSE (Bolme et al., 2010), L1APG (Bao et al., 2012), TBOOST (Gundogdu et al., 2015), KCF (Henriques et al., 2015), SRDCF (Danelljan et al., 2015), P-CODIFF (Demir and Adil, 2018), Staple (Bertinetto et al., 2016) and SiamMargin (Kristan et al., 2019) on our turtle dataset. Based on the quantitative metrics, we used the best performing tracker. Details of the experimental results are given in Section 4.

3.3.2 Orientation Estimation Network

We built a relatively small CNN architecture for detecting the orientation of the turtle. The network topology is summarized in Table 1. Note that we use two outputs for representing the angle values on the unit circle so that we can use the MSE loss function without any modifications. We could use a single output for the angle value. However, we would need to redefine the loss function to prevent penalizing the jumps between 0° and 360°. For training the network coefficients, we annotated nearly 25,000 turtle images with bounding box and orientation labels. We extended this number by rotating the turtle images by 30 to 330 degrees with 30 degree steps and included associated orientation labels based on the rotation angle.

3.3.3 Color Coding

This block generates spatiotemporal feature images based on the visual object tracking output and estimates the turtle orientation. We basically aim to represent the turtle behavior occurring over a time period as an RGB image. In order to generate this image, we draw the path of the turtle using the visual object tracking result. However, we also include the orientation and speed information using hue and value channels of the hue-saturation-value (HSV) color space. The angular difference between the velocity vector (v_n) direction and the turtle orientation (θ_n) is used for determining the hue channel, while the magnitude of the velocity vector is used for value channel. An example output of the color coding block is given in Figure 3.

3.3.4 Behavior Recognition Network

This block aims to recognize the target turtle behaviors using the color-coded spatiotemporal feature images. Since we formulate the behavior recognition task as a vision based classification problem, we adopt a widely used network architecture, ResNet50 (He et al., 2016), for this task. In order to train and test the network, we used a dataset consisting of 172 sequences with U-turn, Reversal, and random motions.

This dataset is further extended with rotated, shifted, and symmetric versions of the sequences. Since we have a relatively small dataset, we employed the transfer learning approach where we use the coefficients pre-trained on the ImageNet (Deng et al., 2009) dataset. We modified the last two fully-connected layers for our behavior recognition task so that the network gives a decision between three behavior classes. The coefficients in the last two layers are trained using our training set.

4 EXPERIMENTAL RESULTS

4.1 Object Tracking and Behavior Recognition

For our visual object tracking experiments, we compared several state-of-the-art algorithms on a dataset consisting of 59 sequences with nearly 25,000 frames. We use Center Location Error (CLE) and Overlap Ratio (OR) as two base metrics which are widely used in object tracking problems (Wu et al., 2013). CLE is the Euclidean distance between the ground truth location and the predicted location, while OR denotes the overlap ratio of predicted bounding box and ground truth bounding box. Based on these metrics, we generated the success and precision plots. The precision plot shows the ratio of frames where CLE is smaller than a certain threshold. The success plot shows the ratio of frames where OR is higher than a given threshold. Figure 4 shows the performance results of the compared algorithms.

Based on this comparative analysis, we determined that the SiamMargin (Kristan et al., 2019) algorithm achieved the highest success and precision graphs among the compared algorithms on the turtle dataset. Therefore, we used this algorithm in our visual object tracking block.

For the orientation estimation experiments, we used 70 percent of the images as training samples, and the rest for the validation and test samples. We used the batch size as 64, initial learning rate as 1e-3, and the number of epochs as 50. In every 20 epochs, we dropped the learning rate by using the drop factor value of 0.1. With these parameters, the model achieved a mean error value of 12.4 degrees on the test set.

In our final set of experiments, we used color-coded spatiotemporal feature images to recognize turtle behaviors. For these experiments, we similarly used 70 percent of the behavior sequences in our dataset to create spatiotemporal motion patterns and trained the last two fully connected layers of the ResNet50 architecture. Then, we used the test sequences to create similar spatiotemporal motion patterns using the outputs of SiamMargin tracker and orientation estimation network that we trained in the previous step. Based on the behavior recognition network outputs, we achieved the prediction results given in Table 2. Corresponding Precision, Recall, and F1 Scores for each behavior are presented in Table 3.

4.2 Anticipated Behavioral Results Using This Method

Studying the effectiveness of bycatch reduction technologies (BRTs) is a difficult task when conditions are less than ideal for recording sea turtle interactions with fishing gear and BRTs in the field and behavioral data requires intensive analysis by researchers even when it can be obtained. Therefore, using behavioral data from controlled experiments to train this convolutional neural network improves the process. We intend to use this initial study to discover if sea turtles do, in fact, recognize fishing nets as a barrier, in which case they would likely avoid the net with a U-turn when they can see them (presumably during the day). We expect to identify more Reversal behaviors during night trials when sea turtles most likely cannot see the net before them. These behaviors can last as little as 3 to 5 seconds, so in one 15-minute trial a sea turtle can perform dozens to hundreds of behaviors that require recording by a researcher. With most treatments involving at least 15 sea turtles at 3 trials each, it becomes a time-intensive project with natural human error that comes along with watching hours of behavior videos. This algorithm can identify these behaviors and enable a comparison between U-turn and Reversal behaviors in daytime and nighttime trials.

5 DISCUSSION

5.1 Future Uses and Related Behaviors

While this method has been created and tested exclusively on behavioral data in a controlled setting, we intend to use this method on field trials in the future. Given that most gillnet fisheries operate at night (Wang et al., 2010b), obtaining high resolution footage of sea turtle interactions is challenging. In particular, we plan on assessing video footage of *in situ* sea turtle interactions with gillnet fisheries as a future step of this research project.

We also recognize that the reversal and U-turn behaviors observed here are likely not exclusive to gillnet avoidance. While we were unable to find literature outlining these specific behaviors, we suspect that reversals and U-turns are evident in other common sea turtle interactions, such as mating (e.g. avoidance behavior by females during courtship)(Frick et al., 2000), predator avoidance (Wirsing et al., 2008), and competition over food or habitat resources (Gaos et al., 2021). Additionally, because this method was created for overhead video, drone footage of sea turtle interactions would be an ideal way to collect behavioral data in the field and subsequently detect the behaviors of interest in other contexts, which has become a common technique for capturing sea turtle behavior (Schofield et al., 2019). For example, studies have captured overhead drone footage of sea turtle courtship behavior (Bevan et al., 2016; Rees et al., 2018). In the future, our machine learning method could be used to detect these behaviors in relation to intraspecific aggression, predator avoidance, and other important interactions captured by drone footage.

5.2 Conclusion

In this study, we developed a behavior recognition framework for sea turtles using color-coded spatiotemporal motion patterns. Our approach uses visual object tracking and CNN based orientation estimation blocks to generate spatiotemporal feature images and processes them to recognize certain behaviors. Our experiments demonstrate that the proposed method achieves an average F1 score of 85% on recognizing the behaviors of interest.

REFERENCES

Babenko, B., Yang, M.-H., and Belongie, S. (2009). Visual tracking with online multiple instance learning. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 983–990. doi:10.1109/CVPR.2009.5206737

Bao, C., Wu, Y., Ling, H. L., and Ji, H. (2012). Real time robust l1 tracker using accelerated proximal gradient approach. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. 1830–1837. doi:10.1109/CVPR.2012.6247881

Bertinetto, L., Valmadre, J., Golodetz, S., Miksik, O., and Torr, P. H. S. (2016). Staple: Complementary learners for real-time tracking. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*

Bevan, E., Wibbels, T., Navarro, E., Rosas, M., Najera, B., L.A., S., et al. (2016). Using unmanned aerial vehicle (uav) technology for locating, identifying, and monitoring courtship and mating behaviour in the green turtle (chelonia mydas). *Herpetological Review* 47, 27–32

Bodor, R., Jackson, B., and Papanikolopoulos, N. (2003). Vision-based human tracking and activity recognition. In *Proc. of the 11th Mediterranean Conf. on Control and Automation*. vol. 1

Bolme, D., Beveridge, J., Draper, B., and Lui, Y. M. (2010). Visual object tracking using adaptive correlation filters. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2544–2550. doi:10.1109/CVPR.2010.5539960

Casale, P. (2011). Sea turtle by-catch in the mediterranean. *Fish and Fisheries* 12, 299–316

Chakravarty, P., Cozzi, G., Ozgul, A., and Aminian, K. (2019). A novel biomechanical approach for animal behaviour recognition using accelerometers. *Methods in Ecology and Evolution* 10, 802–814

Danelljan, M., Hager, G., Khan, F. S., and Felsberg, M. (2015). Learning spatially regularized correlation¨ filters for visual tracking. In *2015 IEEE International Conference on Computer Vision (ICCV)*. 4310–4318. doi:10.1109/ICCV.2015.490

Demir, H. S. and Adil, O. F. (2018). Part-based co-difference object tracking algorithm for infrared videos. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. 3723–3727. doi:10.1109/ICIP.2018.8451294

Demir, H. S., Blain Christen, J., and Ozev, S. (2020). Energy-efficient image recognition system for marine life. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 39, 3458–3466

Demir, H. S. and Cetin, A. E. (2016). Co-difference based object tracking algorithm for infrared videos. In *2016 IEEE International Conference on Image Processing (ICIP)*. 434–438. doi:10.1109/ICIP.2016.

7532394

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*

Estes, J. A., Terborgh, J., Brashares, J. S., Power, M. E., Berger, J., et al. (2011). Trophic downgrading of planet earth. *science* 333, 301–306

Frick, M., Slay, C., Quinn, C., Windham-Reid, A., Duley, P., Ryder, C., et al. (2000). Aerial observations of courtship behavior in loggerhead sea turtles (caretta caretta) from southeastern georgia and northeastern florida. *Journal of Herpetology* 34, 153–158

Gaos, A., Johnson, C., McLeish, D., King, C., and Senko, J. (2021). Interactions among hawaiian hawksbills suggest prevalence of social behaviors in marine turtles. *Chelonian Conservation and Biology.*

Gundogdu, E., Ozkan, H., Demir, H. S., Ergezer, H., Akagunduz, E., and Pakin, S. K. (2015). Comparison of infrared and visible imagery for object tracking: Toward trackers with superior ir performance. In *Computer Vision and Pattern Recognition Workshops, 2015 IEEE Conference on*. 1–9. doi:10.1109/ CVPRW.2015.7301290

Hara, K., Vemulapalli, R., and Chellappa, R. (2017). Designing deep convolutional neural networks for continuous object orientation estimation. *arXiv preprint arXiv:1702.01499*

He, K., Zhang, X., Ren, S., and Sun, J. (2016).Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778

Henriques, J. F., Caseiro, R., Martins, P., and Batista, J. (2015). High-speed tracking with kernelized correlation filters. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* doi:10.1109/TPAMI. 2014.2345390

Ijjina, E. P. and Chalavadi, K. M. (2017). Human action recognition in rgb-d videos using motion sequence information and deep learning. *Pattern Recognition* 72, 504–516

[Dataset] Kristan, M., Leonardis, A., Matas, J., Felsberg, M., Pflugfelder, R., and et al. (2020). The eighth visual object tracking vot2020 challenge results

Kristan, M., Matas, J., Leonardis, A., Felsberg, M., Pflugfelder, R., and et al. (2019). The seventh visual object tracking vot2019 challenge results. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*

Li, B., Wu, W., Wang, Q., Zhang, F., Xing, J., and Yan, J. (2019). Siamrpn++: Evolution of siamese visual tracking with very deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4282–4291

Lucchetti, A., Bargione, G., Petetta, A., Vasapollo, C., and Virgili, M. (2019). Reducing sea turtle bycatch in the mediterranean mixed demersal fisheries. *Frontiers in Marine Science* 6, 387

Moore, J., Wallace, B., Lewison, R., Zydelis, R., Cox, T., and Crowder, L. (2009). A review of marine mammal, sea turtle and seabird bycatch in usa fisheries and the role of policy in shaping management. *Marine Policy* 33, 435–451. doi:https://doi.org/10.1016/j.marpol.2008.09.003

Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., and Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications* 105, 233–261

Peckham, S., Lucero-Romero, J., Maldonado-D´ıaz, D., Rodr´ıguez-Sanchez, A., Senko, J., Wojakowski,´ M., et al. (2016). Buoyless nets reduce sea turtle bycatch in coastal net fisheries. *Conservation Letters* 9, 114–121. doi:https://doi.org/10.1111/conl.12176

Porto, S. M., Arcidiacono, C., Anguzza, U., and Cascone, G. (2013). A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns. *Biosystems Engineering* 115, 184–194

Raza, M., Chen, Z., Rehman, S.-U., Wang, P., and Bao, P. (2018). Appearance based pedestrians' head pose and body orientation estimation using deep learning. *Neurocomputing* 272, 647–659

Rees, A., Avens, L., Ballorain, K., Bevan, E., Broderick, A., Carthy, R., et al. (2018). The potential of unmanned aerial systems for sea turtle research and conservation: a review and future directions. *Endangered Species Research* , 81–100

Ross, D. A., Lim, J., Lin, R.-S., and Yang, M.-H. (2007). Incremental learning for robust visual tracking. *International Journal of Computer Vision* 77, 125–141. doi:10.1007/s11263-007-0075-7

Schofield, G., Esteban, N., Katselidis, K. A., and Hays, G. C. (2019). Drones for research on sea turtles and other marine vertebrates–a review. *Biological Conservation* 238. doi:10.1016/j.biocon.2019.108214

Wagner, R., Thom, M., Gabb, M., Limmer, M., Schweiger, R., and Rothermel, A. (2013). Convolutional neural networks for night-time animal orientation estimation. In *2013 IEEE Intelligent Vehicles Symposium (IV)*. 316–321. doi:10.1109/IVS.2013.6629488

Wang, J., Fisler, S., and Swimmer, Y. (2010a). Developing visual deterrents to reduce sea turtle bycatch in gill net fisheries. *Marine Ecology Progress Series* 408, 241–250

Wang, J., Fisler, S., and Swimmer, Y. (2010b). Developing visual deterrents to reduce sea turtle bycatch in gill net fisheries. *Marine Ecology Progress Series* 408, 241–250. doi:https://doi.org/10.3354/meps08577

Wirsing, A., Abernethy, R., and Heithaus, M. (2008). Speed and maneuverability of adult loggerhead turtles (caretta caretta) under simulated predatory attack: do the sexes differ? *Journal of Herpetology* 42, 411–413

Wu, Y., Lim, J., and Yang, M.-H. (2013). Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2411–2418

Yang, Q., Xiao, D., and Lin, S. (2018). Feeding behavior recognition for group-housed pigs with the faster r-cnn. *Computers and Electronics in Agriculture* 155, 453–460

FIGURE CAPTIONS

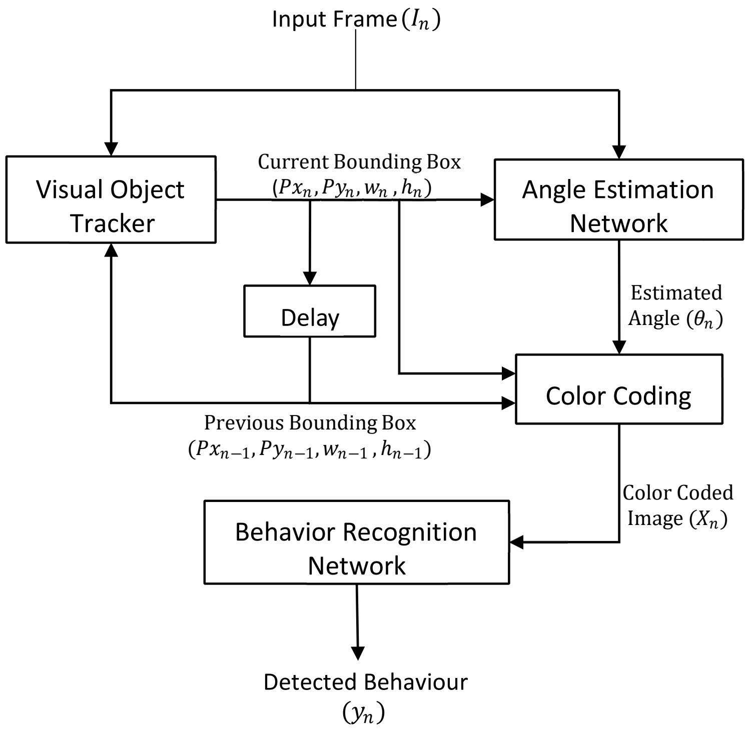


Figure 1. Overview of the proposed approach

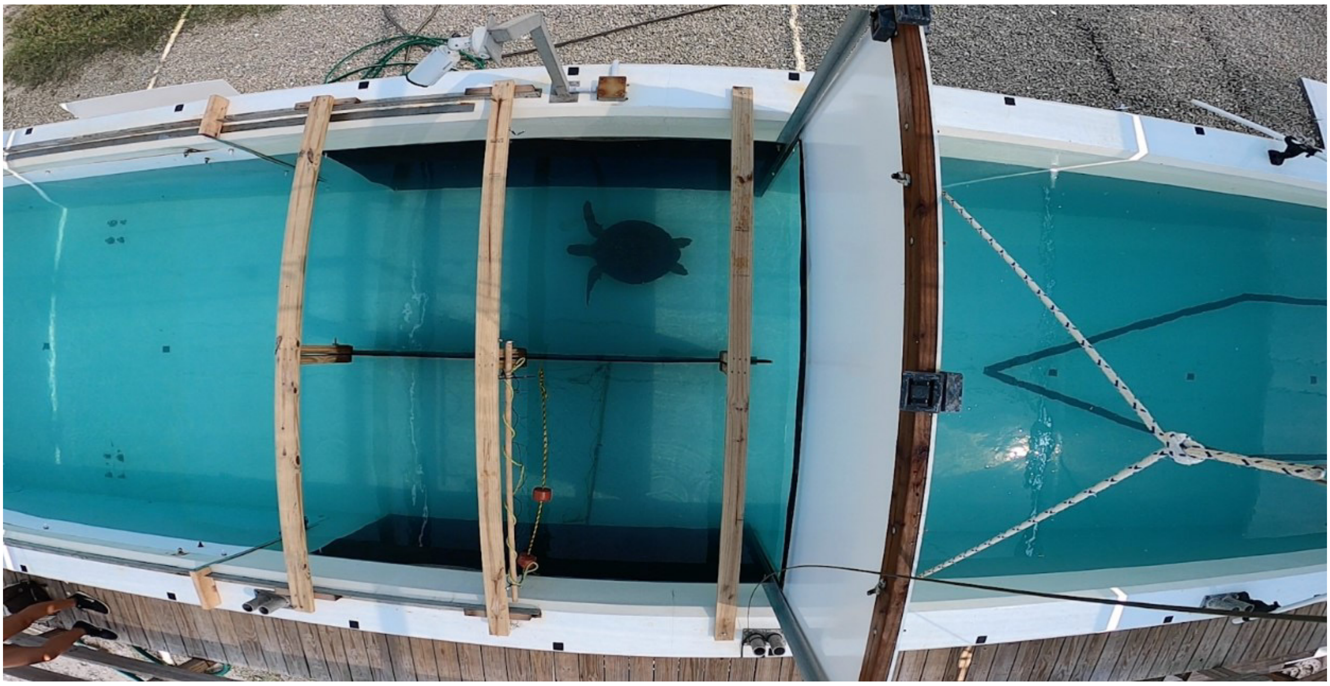


Figure 2. Experimental tank at the St. Nuclear Power Plant in Jensen Beach, FL. A juvenile green turtle (*C. mydas*) is participating in a daytime net vs. no net trial. Image captured from the primary overhead camera used to record all experiments.

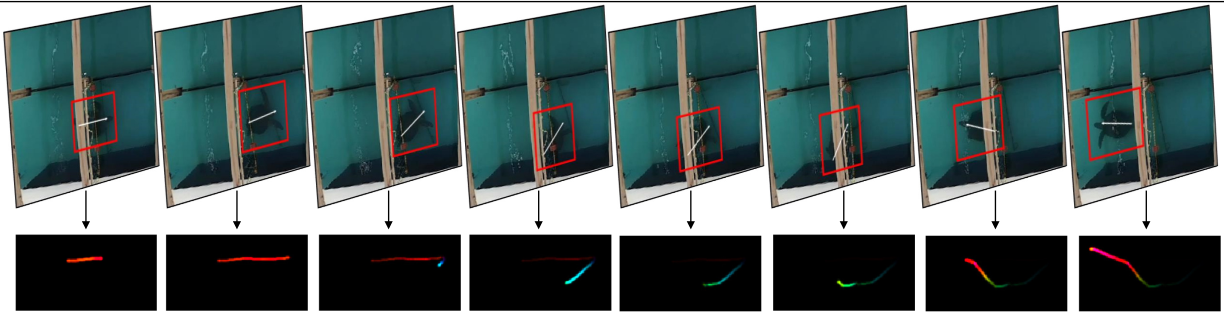
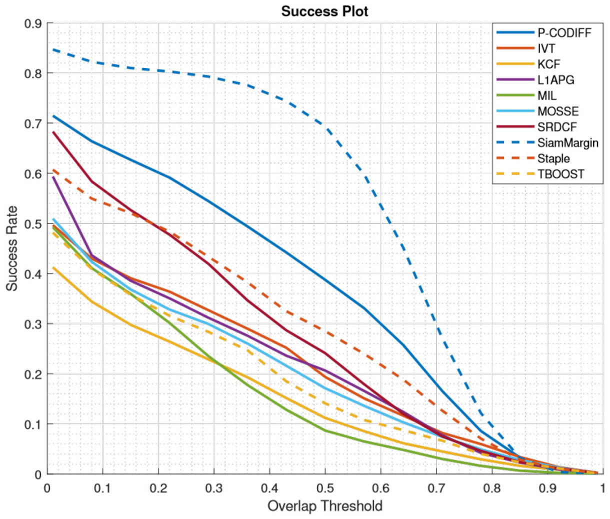
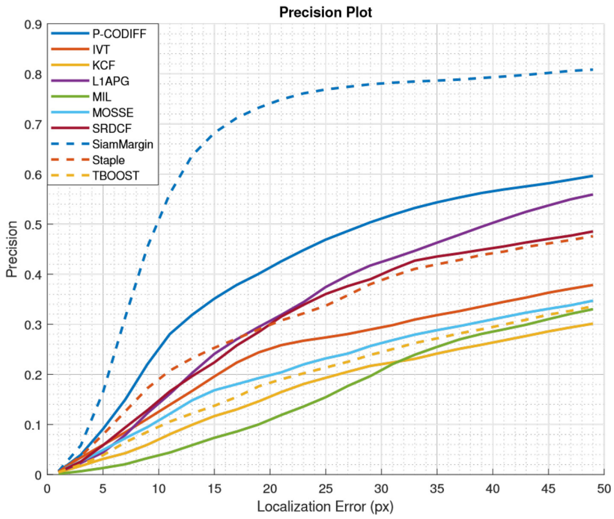


Figure 3. Color coded spatiotemporal feature images generated using turtle velocity vector and orientation information



(a) Succes vs overlap threshold plots of various methods



(b) Precision vs. localization error plots of various methods

Figure 4. Success and precision plots of various methods.

Table 1. Topology of Orientation Estimation Network

Name	Explanation
imageinput	64x64x1 input images
conv 1	8 3x3x1 convolutions, stride 1

relu 1	ReLU layer
avgpool2d 1	2x2 average pooling with stride 2
conv 2	16 3x3x8 convolutions, stride 1
relu 2	ReLU layer
avgpool2d 2	2x2 average pooling with stride 2
conv 3	32 3x3x16 convolutions, stride 1
relu 3	ReLU layer
avgpool2d 3	2x2 average pooling with stride 2
conv 4	64 3x3x32 convolutions, stride 1
relu 4	ReLU layer
avgpool2d 4	2x2 average pooling with stride 2
conv 5	128 3x3x64 convolutions, stride 1
relu 5	ReLU layer
avgpool2d 5	2x2 average pooling with stride 2
conv 6	128 3x3x128 convolutions, stride 1
relu 6	ReLU layer
dropout	20% dropout
fc	2 Fully connected layers
regressionoutput	Regression with MSE loss function

Table 2. Normalized confusion matrix showing the percentage of actual and predicted classes for 3 different behaviors

Actual	Predicted: U-Turn	Predicted: Reversal	Predicted: Random
U-Turn	83.3	3.7	13.0
Reversal	0	82.8	17.2
Random	6.7	5.0	88.3

Table 3. Precision, Recall and F1 Scores of defined behaviors

	Precision	Recall	F1 Score
U-Turn	0.833	0.925	0.877
Reversal	0.828	0.905	0.865
Random	0.883	0.745	0.808
Macro-Avg	0.848	0.858	0.85