Automatic identification of the endangered Hawksbill sea turtle behavior using deep learning and cross-species transfer

Lorène Jeantet[1,2,3,*], Kukhanya Zondo[1], Cyrielle Delvenne[4], Jordan Martin[4], Damien Chevallier[4], Emmanuel Dufourq[1,2,3]

[1]African Institute for Mathematical Sciences, South Africa

[2]African Institute for Mathematical Sciences, Research and Innovation Centre, Kigali, Rwanda

[3]Stellenbosch University, South Africa.

[4]Unité de Recherche BOREA , MNHN, CNRS 8067, SU, IRD 207, UCN, UA, Station de Recherche Marine de Martinique, Quartier Degras, Petite Anse, 97217 Les Anses d'Arlet, Martinique, France

[*]Corresponding author: Email : lorene@aims.ac.za

Keywords: Accelerometer, Behavioral Classification, Bio-logging, Deep Learning, Sea Turtles, Transfer Learning

Abstract

The accelerometer, an onboard sensor, enables remote monitoring of animal posture and movement, allowing researchers to deduce behaviors. Despite the automated analysis capabilities provided by deep learning, data scarcity remains a challenge in ecology. We explored transfer learning to classify behaviors from acceleration data of critically endangered hawksbill sea turtles (*Eretmochelys imbricata*). Transfer learning reuses a model trained on one task from a large dataset to solve a related task. We applied this method using a model trained on green turtles (*Chelonia mydas*) and adapted it to identify hawksbill behaviors like swimming, resting, and feeding. We also compared this to a model trained on human activity data. Results showed an 8% and 4% F1-score improvement with transfer learning from green turtle and human datasets, respectively. Transfer learning allows researchers to adapt existing models to their study species, leveraging deep learning and expanding the use of accelerometers for wildlife monitoring.

Introduction

Monitoring animal behavior offers a non-invasive method for studying wildlife in their natural habitats, providing valuable insights into their ecology and physiology. Through behavioral analysis, researchers can deepen their understanding of how animals interact with their environment, elucidating key aspects of their energetic strategies, survival mechanisms, and reproductive success (McHuron et al., 2018; Sansom et al., 2009; Zhang et al., 2004). Behavior monitoring also remains a critical tool for evaluating species' adaptations to extreme conditions (Abernathy et al., 2019; Sergio et al., 2018) and rapid environmental changes (Beever et al., 2017; Buchholz et al., 2019). Furthermore, it enables the measurement of remarkable physiological capabilities, such as long-distance migration and deep breath-hold diving, facilitating a more comprehensive understanding of the mechanisms driving these behaviors (Boel et al., 2014; Fossette et al., 2010; Pedersen et al., 2018).

Bio-logging has emerged as a valuable tool for studying animal behavior, enabling continuous monitoring of animals in their natural habitats. It involves deploying on-board sensors that collect high-resolution data on animal behavior, physiology, and environmental conditions. Among these sensors, the accelerometer is particularly effective for recording animal posture and movement. It operates as a piezoelectric sensor and translates the forces exerted on a mass into wave-like voltage signals. These forces include the gravitational force and the force of inertia induced by the posture and movement of the animal, respectively (Brown et al., 2013). By positioning three accelerometers orthogonally, researchers can identify behaviors characterized by specific postures and movements, such as locomotion, feeding, and inactivity (Shepard et al., 2008).

Automating data processing is a key challenge in acceleration-based behavior identification, with the aim of facilitating the analysis of large time-series datasets. Recently, deep learning has emerged as a promising tool to automatically analyze acceleration data for identifying the behavior of wild animals (Aulsebrook et al., 2024; Jeantet et al., 2021; Otsuka et al., 2024; W. Zhang et al., 2019). Deep learning algorithms, also known as neural networks, are computational models that process data through multiple layers, performing both linear and non-linear transformations. They can automatically detect complex, highly discriminative features and patterns in the data, making them capable of classifying acceleration signals into distinct behavioral categories. Deep learning, in particular, has been at the forefront of advancements in behavior identification using accelerometers in livestock monitoring and Human Activity Recognition (HAR) (Bao and Xie, 2022; Kumar et al., 2023). In livestock, many studies have developed deep learning models to track behaviors like

feeding, drinking, walking, resting, and detecting lameness in cattle, sheep, and pigs, aiming to monitor animal welfare and productivity (Kleanthous et al., 2022b; Mao et al., 2023). In humans, accelerometers are widely used to track physical activity, identifying behaviors such as running, walking, standing, sitting, and are also employed to detect falls, which is especially important for elderly care (Choi et al., 2022; Nunavath et al., 2021; Ramanujam et al., 2021). Conversely, deep learning is still underutilized in the field of ecology, with limited research dedicated to developing new methods to facilitate its use or enhance its performance.

The limited use of deep learning in ecology, compared to in livestock monitoring or HAR, may be attributed to the additional challenges of acquiring large, representative, and accurately labeled datasets in ecological research. Deep learning algorithms are generally large models with numerous successive operational layers. Each operation relies on real-valued numbers, known as weights, which are optimized, or fine-tuned, during the training phase using labeled data to tailor the model to the specific task. As the number of layers in the model increases, its ability to discriminate features improves, but this also necessitates a larger labeled dataset for effective training (He et al., 2016). In contrast, studying wild animals presents challenges in data collection due to the need to minimize disturbance, environmental constraints, and limited access to endangered species, resulting in smaller datasets. Additionally, the need for simultaneous observation to create a labeled dataset further complicates the process, leading to even smaller training datasets.

One of the primary risks of using a dataset that is too small for training deep learning models is overfitting. Overfitting occurs when the model's weights become overly tailored to the specific examples in the training data. Consequently, the model performs well on the training dataset but poorly on new, unseen data. To address data scarcity in other fields of ecology, researchers have turned to transfer learning techniques across a range of ecological tasks, including bioacoustics, camera traps, drone surveys, and animal pose estimation from video recordings (the latter three falling under computer vision) (Dufourq et al., 2022; Gray et al., 2019; Liao et al., 2023; Lu et al., 2021; Schneider et al., 2018). Transfer learning is a technique in which a model trained on one task is adapted and fine-tuned to a different, but related task. This method enables the use of deep learning models, with numerous layers, that have been pre-trained on large datasets, while mitigating the risk of overfitting on smaller, more limited datasets. For example, in bioacoustics, Batist et al. (2024) employed a pre-trained model, originally designed for classifying object categories from colour images and trained on ImageNet, a dataset comprising over 15 million annotated images, to identify sounds emitted by Black-and-white ruffed lemurs (*Varecia variegata*) using spectrograms, the

visual representations of sound. Even though the datasets and tasks are not identical (classifying images vs classifying spectrograms), a neural network pre-trained on the image dataset can learn abstract patterns that are transferrable to the spectrogram dataset. While transfer learning has recently been utilized in acceleration-based behavior identification studies of livestock (Bloch et al., 2023; Kleanthous et al., 2022a), and has long been employed in the study of human behavior (Cook et al., 2013; Hernandez et al., 2020), it has not yet been applied to the study of wild animals.

The issue of data scarcity to train deep learning algorithm in ecology is especially notable for marine species. In this study, we explore the use of transfer learning to study hawksbill sea turtles (*Eretmochelys imbricata*). Among the six endangered sea turtle species, the hawksbill turtle is the most endangered along with the Kemp's ridley turtle (*Lepidochelys kempii*), both classified as critically endangered (International Union for Conservation of Nature (IUCN), 2024). Martinique is an island in the Caribbean where a population of hawksbill turtles and juvenile green turtles (*Chelonia mydas*) reside (Lelong et al., 2024; Nivière et al., 2018). Understanding their behaviors is essential to elucidate how these endangered species utilize resources and adapt their energy strategies in a changing environment (e.g., rising water temperatures, coral bleaching and eutrophication, Li and Reidenbach., 2014), as well as under significant anthropogenic pressures (e.g., accidental bycatch in fisheries, disturbances related to tourism, Hayes et al., 2017; Louis-Jean et al., 2009; Siegwalt et al., 2022). Accelerometers are an effective tool for studying behaviors such as swimming, resting, breathing and feeding in marine turtles, thereby facilitating a deeper investigation into their energy strategies (Fossette et al., 2012; Jeantet et al., 2020a).

The green turtle population in Martinique is subject to significant monitoring (Bonola et al., 2019; Charrier et al., 2022; Siegwalt et al., 2022, 2020), and studies utilizing accelerometers have been conducted to automatically monitor their behaviors (Jeantet et al., 2020a, 2021). Conversely, studying the population of hawksbill turtles presents greater challenges, as individuals are more elusive than green turtles, making it difficult to equip them with bio-loggers. Furthermore, very few studies have employed accelerometers to investigate hawksbill turtles; to our knowledge, only two studies have explored the use of this technology to study their behavior (Jeantet et al., 2018; Okuyama et al., 2012). Consequently, there is a substantial lack of data and studies on this species, and the challenges in data collection make it difficult to apply deep learning algorithms to study their behaviors.

The aim of this study is to investigate the application of transfer learning for the automatic identification of behaviors in endangered species using accelerometer data,

addressing the common challenge of small datasets in ecological research when training deep learning algorithms. We focus on a particularly challenging species to monitor, the hawksbill sea turtle, for which limited research and data exist. Firstly, we explore the use of a pre-trained model on a closely related species, the green turtle, whose postures and movement patterns are supposed to be similar to the hawksbill turtle in the expression of behaviors of interest, such as resting, feeding, swimming and breathing. For this species, a publicly available dataset and trained model exist. Secondly, we assess the feasibility of applying transfer learning from a species with a noticeably different morphology from the hawksbill turtle, from which behavioral identification should rely on different features than those used for sea turtles. This approach enables us to test whether the method is species-specific or can be applied to a broader range of species, with minimal dependence between the species used in the pre-trained model and the species being studied. For this, we focus on humans, as data collection is more accessible and numerous datasets and pre-trained models are available. This study is the first to explore transfer learning for monitoring wild animals, aiming to propose a method that can be generalized across a broader range of species, thereby facilitating the wider application of deep learning in the study of animal behavior.

Material and Methods

Hawksbill dataset

Ethical note

This study meets the legal requirements of the countries in which the work was carried out and follows all institutional guidelines. The protocol was approved by the 'Conseil National de la Protection de la Nature' (http://www.avis-biodiversite.developpement-durable.gouv.fr/bienvenue-sur-le-site-du-cnpn-et-du-cnb-a1.html), and the French Ministry for Ecology, Sustainable Development and Energy (permit no. 2013154-0037), which acts as an ethics committee in Martinique. The fieldwork was carried out in strict accordance with the recommendations of the Prefecture of Martinique in order to minimize the disturbance of animals (authorization no. 201710-0005).

Data collection

The hawksbill turtle dataset was collected through fieldwork conducted in Martinique, Caribbean Island, France, from November 2022 to May 2023. Six free-ranging hawksbill turtles were manually caught, measured and identified as described in Nivière et al. (2018) and Jeantet et al., (2020a). We equipped them with CATS (Customized Animal Tracking Solutions, Germany) devices over a two-day period using four suction cups attached to their carapace and an automatic release system (Fig. 1, see Jeantet et al., 2020a for details). Each CATS device comprised a video recorder combined with a tri-axial accelerometer, a tri-axial gyroscope, and a pressure sensor. The devices were configured to continuously record acceleration and angular velocity (gyroscope) at a rate of 20 Hz and pressure at 1 Hz. The video recorders were programmed to record until nightfall (6 pm) and resume at daybreak (6 am). The maximum battery capacity was considered to provide a recording capacity of 18 h of video footage and 48 h for the other sensors.

Video annotation and data preprocessing

Video recordings captured by the CATS devices were analyzed through visual inspection using a custom-made software TurtleCap (https://github.com/Vadym-Hadetskyi/TurtleCap) to identify behaviors. The corresponding start and end times of each observed behavior were annotated, with precision to the nearest tenth of a second. Six main categories of behavior were identified and used to label the video recordings: *Breathing, Feeding, Gliding, Resting, Scratching* and *Swimming* (Fig. 1). Any other observed behavior was categorized as *Other*. These seven categories are similar to those used to describe the behavior of green turtles in the previous studies in Martinique (see Jeantet et al. 2020a for a precise description of the behaviors). The acceleration data, gyroscope data, and pressure data associated with the corresponding observed behaviors were visualized and checked using R software (version 3.5.3) and the *rblt* package (Geiger, 2019).

Data preprocessing mainly involved mapping the behavior labels to the multi-sensor data and removing the unlabeled night sequences for which the video recording was not available. No pre-processing was performed on the acceleration and gyroscopic data. For the pressure data, the difference was calculated between each data point (1 Hz) and a linear interpolation technique was then used to increase the sampling rate to 20 Hz.

In total, 69.7 hours of multi-sensor recordings were labeled from six different hawksbill turtles (approximately 11.6 hours of recording per individual, max = 17.8 hours, min = 6.3

hours, standard deviation = 3.6 hours). The predominant behavior observed in the videos was *Feeding*, totaling over 38.6 hours, followed by *Resting* and *Swimming*, with 19.1 hours and 7.9 hours, respectively (Table S1). The other behaviors were expressed in minority (*Breathing*: 2.2 hours, *Gliding*: 1 hour, *Scratching*: 0.8 hour and *Other*: 0.1 hour).

Datasets of the pre-trained models

Green turtle dataset

The green turtle dataset used in this study was collected in Martinique between February 2018 and May 2019 in a similar way to the hawksbill dataset. The CATS devices were deployed on 13 individuals over one or two-day period. The dataset is annotated, freely available on the Dryad digital repository (https://doi.org/10.5061/dryad.hhmgqnkd9}, Jeantet et al. 2020b) and has been precisely described in (Jeantet et al., 2020a).

In total, the green turtle dataset contained 68.6 hours of recordings from 13 individuals (approximately 5.29 hours per individual, max = 14.67 hours, min = 0.96 hours, standard deviation = 3.39 hours), labeled with the same seven behaviors as the hawksbill dataset. The predominant behavior observed in the videos was *Resting*, totaling over 34.3 hours, followed by Swimming and *Breathing*, with 22.3 hours and 5.7 hours, respectively (Table S1). The other behaviors were expressed in minority (*Gliding*: 2.3 hours, *Feeding*: 1.8 hours, *Scratching*: 1.2 hours and Other: 1 hour).

Human dataset

To evaluate the significance of the species used for pre-training, we also assessed the potential for transfer learning from other taxa, where data collection is more feasible. We focused on humans, since collecting data from them is significantly easier compared to animals, resulting in numerous open-access datasets available. In this case, we used the publicly available Intensive Care Unit (ICU) HAR dataset (Reyes-Ortiz et al., 2013), which is a dataset commonly used in the development of deep learning models for HAR (Ramanujam et al., 2021; Zhang et al., 2022).

The ICU HAR comprised of 30 individuals wearing a smartphone on the waist recording 3D acceleration and 3D angular speed sampled at 50 Hz. Each person performed six

activities (*Walking, Walking_upstairs, Walking_downstairs, Sitting, Standing*, and *Laying*) for durations ranging from one to two minutes each, amounting to approximately 220 hours of labeled sequences (see Table S1 for the specific duration available per activity).

Architecture of the deep learning model

In this transfer learning study, we tested V-Net (Jeantet et al., 2021), a model originally adapted to identify the underwater behaviors of green turtles in Martinique. The V-Net model has the advantage of generating an output with the same length as the input, allowing for predictions at each data point, or time step, from the recorded data. It can directly predict behaviors from raw accelerometer data without requiring preprocessing steps such as segmentation, filtering, or the calculation of descriptive variables (Jeantet et al., 2020a). Moreover, the V-Net was specifically designed to handle small and imbalanced datasets (Milletari et al., 2016). Beyond its application to green turtles in Martinique, the model has demonstrated satisfactory results in identifying terrestrial behaviors in green turtles in French Guiana (Jeantet et al., 2022) and prey capture events in chinstrap penguins (*Pygocelis antarctica*) in Antarctica (Schoombie et al., 2024). This evidence suggests its ability to generalize across species and contexts. For the human dataset, we tested a closely related model, U-Net, which shares a similar architecture and has been widely used in human studies (Meena and Sarawadekar, 2023; Y. Zhang et al., 2019).

The V-Net and U-Net are models also known as fully convolutional neural networks given that their primary operation is convolution. In machine learning, particularly in convolutional neural networks, convolution is a mathematical operation used to detect and extract hierarchical features from input data, such as edges, textures, and patterns in images or signals. It operates by sliding a filter over an input, multiplying the filter values with the signal values, and summing them at each position to generate a new signal that highlights specific features based on the filter. The values of the filter involved in the multiplications and summations represent the model's weights. The training phase of a neural network optimizes these weights to enable accurate identification of the behaviors of the studied species based on the labeled data provided to the model. A convolutional layer applies a set number of filters to the input data. Each filter produces a transformed version of the input signal, referred to as a feature map. A fully convolutional neural network is characterized by the sequential arrangement of multiple convolutional layers (see Fig. 2 for a schematic visualization of the V-Net and U-Net architecture). The sequential application of multiple filters enables the

extraction of various features from the signal, which subsequently enables behavior classification.

The V-Net and U-Net are encoder-decoder architectures (Fig. 2). In these models, a series of convolutional layers first reduces the dimensionality of the input, mapping it to a lower-dimensional space. This process forms the encoder, and it generates a large number of discriminative features based on the number of layers and filters in the encoder. Following that, the second part of the model, known as the decoder, synthesizes these features to predict the outcome. This results in an output matrix with the same length as the input. The final layer is a convolutional layer with seven channels, each corresponding to one of the behavioral classes. At this stage, the outcome is a matrix containing a value for each behavior at every data point. These seven values are then passed through a softmax function (Eqn 1) to obtain the probability of each behavior for each data point.

$$softmax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_i}} \qquad (1)$$

with z_i representing the i-th element of the array of size n, where n is the number of classes. The softmax function produces an array of non-negative values that sum to 1, making it interpretable as a probability distribution. For each data point, the behavior with the highest probability is then selected as the predicted behavior. A detailed description of the V-Net and U-Net architectures is provided in the supplementary materials.

Evaluation of the models

To evaluate the performance of the various models employed in this study, we tested them on a dataset distinct from the training set. This approach enables us to assess the models' ability to make predictions on data considered unseen as not used within model training. For the three datasets previously described, the data were divided into training, validation, and testing sets. The training and validation sets are used simultaneously during the iterative training process, with the validation set employed to assess the performance trends. The testing set is reserved for the final analysis of the model, simulating the conditions under which the model will be applied.

For the human dataset, the dataset is originally provided in two datasets, with 70% of the individuals randomly allocated to the training set and the remaining 30% to the test set (Reyes-Ortiz et al., 2013). From the human training dataset, 30% of the individuals were

randomly allocated to the validation dataset. For green turtles, seven individuals were used for training, three for the validation and three for testing, following the same distribution as in Jeantet et al. (2021). For the hawksbill turtles, three individuals were randomly selected for the training dataset, one for validation and two for testing. This data splitting approach was maintained consistently across all tested models.

We evaluated the model's generalizability using the testing dataset. To do so, we divided the entire recording for each test individual into consecutive windows with a 10\% overlap. We ran the trained model on all windows and then reassembled the predictions to reconstruct the signal. Based on the predictions and the manually verified labels, we generated the confusion matrix and calculated the global accuracy (Eqn 2), precision (Eqn 3), recall (Eqn 4) and F1-score (Eqn 5).

$$Global\ accuracy = \frac{TP+TN}{TP+TN+FN+FP} \qquad (2)$$

$$Precision = \frac{TP}{TP+FP} \qquad (3)$$

$$Recall = \frac{TP}{TP+FN} \qquad (4)$$

$$F1-score = \frac{2\times Precision\times Recall}{Precision+Recall} \qquad (5)$$

In these equations, TP (True Positives) and TN (True Negatives) refer to the number of correctly identified behaviors, while FN (False Negatives) and FP (False Positives) refer to the number of misclassified behaviors. For imbalanced datasets, the F1-score is considered a more reliable measure of model performance than the global accuracy (Saito and Rehmsmeier, 2015).

Transfer Learning

Fundamentals of transfer learning

As a neural network undergoes training, the model learns from annotated data and the weights are iteratively updated to minimize the difference between the outputs to the annotated values. Initially, these weights are randomly generated and then optimized for the specific classification task. Overfitting arises when these weights become overly tailored to the labeled dataset, leading to poor generalization and diminished model performance on new data (Ying, 2019). This issue is particularly prevalent when the model possesses a large number of weights and the training dataset's size is small.

Transfer learning presents a solution by leveraging pre-trained models, where the weights from a model *A* are employed to initialize a model *B*. Typically, model *A* is trained on a similar task as model *B*, but with a larger dataset. In this process, the weights from model *A* are employed as the starting point for model *B*, as opposed to randomly initialized weights. During the training phase, the weights of model *B* are then fine-tuned specifically for task *B*, using the smaller labeled dataset. In practice, the initial layers of the model are often ``frozen'', meaning their weights remain unchanged during training, while only the weights of the layers closer to the output are fine-tuned. Freezing certain layers offers computational advantages by reducing the number of weights that need to be optimized. The rationale for freezing the early layers of a model is that these layers capture general features of the data, while the deeper layers progressively learn more abstract representations (Gu et al., 2018; Yosinski et al., 2015). As a result, the deeper layers are more specialized for the specific task and must be adapted during transfer learning.

Pre-training on green turtle and human dataset

The first step of transfer learning involves training model *A* on the original dataset. To obtain the pre-trained model from the green turtle dataset, we trained the V-Net using the same parameters as in Jeantet et al. (2021). As input, we used the raw acceleration (X, Y, Z), the gyroscope (X, Y, Z) and the difference of pressure, sampled at 20 Hz over a window size of 40 seconds, resulting in a 7 x 800 matrix (Fig. 2). We employed the Generalized Dice loss (Sudre et al., 2017) and we trained the V-Net on 30 epochs with a batch size of 32, using the Adam optimizer and a learning rate of 0.0001. We obtained a F1-score of 81.1% and a global accuracy of 97.2% on the green turtle dataset.

For the human dataset, we employed the U-Net model proposed by Zhang et al. (2019), which is publicly available on GitHub. The U-Net model from Zhang et al. (2019) consists of 23 convolutional layers, while the V-Net model used for the green turtle comprises 15 convolutional layers (Fig. 2, see supplementary materials for a detailed description of the V-Net and U-Net architectures). In their study, the authors evaluated the model on the ICU HAR dataset, using a fixed-length window of size 224 to segment the data. Following this approach, we trained the U-Net on the ICU HAR dataset using raw acceleration (X, Y, Z) and gyroscope (X, Y, Z) data over a 224-length window, resulting in a 6 x 224 matrix. The model was trained for 50 epochs with a batch size of 32, using the Adam optimizer, a cross-entropy

loss and a learning rate of 0.001, consistent with Zhang et al., (2019). We obtained an F1-score of 85.6% and a global accuracy of 95.7% on the human dataset.

Transfer learning on the hawksbill dataset

To assess the benefits of using transfer learning to identify the behavior of hawksbill turtles from a small dataset, we explored various scenarios. First, to determine whether a new model is necessary for identifying the behavior of hawksbill turtles, we predicted their behaviors using the V-Net trained on green turtles. In this case, no transfer learning was employed, and the model was not fine-tuned on the hawksbill dataset, but instead solely applied on the six individuals to assess the predictions. Secondly, we established a baseline model (Model-Hawksbill) by training the V-Net from randomly initialized weights on the hawksbill dataset. In this case we did not use the pre-trained model on green turtles, but randomly initialized all the layers of the V-Net and trained it using the hawksbill dataset. Third, we used transfer learning on the V-Net pre-trained on the green turtle dataset (Model-Green_turtle). Finally, we used transfer learning on the U-Net pre-trained on the human dataset (Model-Human). We adapted the pre-trained U-Net architecture for our case study by replacing the final convolutional layer with a convolutional layer featuring seven channels (corresponding to the seven behavioral categories in the hawksbill dataset vs six in the human dataset) and fine-tuned all the layers (Fig. 2).

Additionally, since transfer learning involves freezing and fine-tuning different layers, we explored four additional scenarios for the Model-Green_turtle to determine the optimal configuration (Fig. 2). In the first scenario, we fine-tuned all the model layers. In the second, we froze the encoder weights and fine-tuned only the decoder weights. In the third scenario, we tested the opposite approach by freezing the decoder and fine-tuning the encoder weights. Finally, in the fourth scenario, we fine-tuned only the weights of the last convolutional layer.

To train or fine-tune the V-Net (Model-Hawksbill, Model-Green_turtle) on the hawksbill dataset, we fed the algorithm with a window size of 800 (40 seconds), encompassing data from the three accelerometer and three gyroscope axes along with the difference of pressure (matrix size 7 x 800). For Model-Human, we used a window of size 224 and only the three accelerometer and the three gyroscope axes (matrix size 224 x 6), reproducing the input shape used with the human dataset. To generate these windows, we

reused the method described in Jeantet et al. (2021), allowing us to produce a predetermined number of windows while fostering specific behaviors to balance the dataset (see Jeantet et al. 2021 for detailed information). We generated 6 000 windows from the training dataset and 3 000 from the validation dataset at each epoch (see Fig. S1 for the distribution of the behaviors based on this method). We ran each execution 20 times with the same hyperparameters (epoch = 20, learning rate = 0.0001, batch size = 32, Adam optimizer, the Generalized Dice loss function for the V-Net and cross-entropy loss function for the U-Net). At each epoch, the model's performance was evaluated on the validation dataset, and throughout the 20 epochs, the model weights were saved when the best performance was achieved.

Models were implemented in Python3 using the Tensorflow2 and Keras libraries (Abadi et al., 2015, ; Chollet, 2015). Model training and testing were performed on a Dell G15 with a 13th generation Intel Core i7-13650HX processor as Central Processing Unit (CPU) and an NVIDIA GeForce RTX 3050 Graphics Processing Unit (GPU). Pre-trained models and analysis scripts are available online on GitHub (https://github.com/AIMS-Research/research_za/tree/main/biologging_transferlearning_hawksbill), and the hawksbill dataset to run the scripts is available on Zenodo (https://doi.org/10.5281/zenodo.11402241).

Results and Discussion

In this study, we explored the use of transfer learning across species and taxa, employing fully convolutional neural networks to predict the behaviors of an endangered species and tackle the challenge of data scarcity. Initially, we demonstrate that while behaviors may seem similar between green and hawksbill turtles, a model trained solely on green turtles cannot accurately predict hawksbill behavior (Fig. 3, F1-score: 41.17%), highlighting the necessity for species-specific model training. Conversely, utilizing a pre-trained model (Model-Green_turtle) significantly improves predictions compared to a model trained from randomly initialized weights (Model-Hawksbill) by 8% points (Fig. 3, Table 1, Model-Hawksbill, F1-score = 69.11%; Model-Green_turtle with all layers fine-tuned, F1-score = 77.12%). Comparing the time budget predicted for the testing dataset from the Model-Hawksbill and the Model-Green_turtle with camera observations, the predictions from the Model-Green_turtle show a closer alignment with the observed data for both individuals (Fig. 4). The time budget, representing the proportion of time allocated to each behavior, was calculated by dividing the number of predicted data points for each behavior by the total number of labeled data points. For both individuals, particularly individual #2, a closer inspection of the predictions revealed

Feeding sequences that were predicted as *Resting* (Fig. S2). These *Feeding* sequences correspond to accelerometer and gyroscope signals with minimal variation, likely reflecting moments when the turtle was chewing its food. Although these chewing sequences were labeled as *Feeding* based on the video, it was difficult to observe the actual chewing, as the turtle typically had its head down. Consequently, some errors may result from our annotation of the accelerometer data—either when the turtle is not actually chewing or when the chewing is not detected by the logger—rather than from a model error. Despite being trained on only three individuals, the model demonstrates high accuracy in behavior identification and shows promising potential for broader applications.

This study presents the first application of transfer learning for wildlife monitoring using acceleration data. To our knowledge, only two studies used transfer learning for acceleration-based behavior identification in livestock. Kleanthous et al., (2022a) evaluated the effectiveness of transfer learning in identifying sheep behavior using two datasets collected simultaneously from accelerometers positioned differently on the collars of the same individuals. Bloch et al. (2023) applied transfer learning to study cow behavior, using a pre-trained model from an open-source dataset of different cows in a different environment. Likewise, a review on transfer learning applied to HAR reveals that in most studies, models are pre-trained either on the same individuals with sensors positioned at different locations or on different subjects with loggers in the same location and/or in different environments (Hernandez et al., 2020). In these studies, and across the literature, cross-species transfer learning has not been attempted. In our case, since similar datasets for hawksbills are not available, we demonstrate that transfer learning can also be applied across different species. We show that predictive accuracy improves when the model is trained on species with similar behaviors and environment, but also when trained on species with entirely different behaviors such as humans (3.8% points increase in the F1-score, Model-Hawksbill, F1-score = 69.11%; Model-Human, F1-score = 72.87%, Table 1).

Transfer learning from human data yields better results than a model trained from randomly initialized weights, suggesting that the species used for pre-training can be entirely different of the species under study while still benefiting the model. This occurs because deep learning models consist of multiple successive layers, with each layer progressively representing data in increasingly abstract forms from the input to deeper layers (Goodfellow, 2016). Since the model has already been trained to recognize human behaviors, its weights are optimized to extract signal features that distinguish between behaviors, with the earlier layers of the encoder capturing general features and the deeper layers identifying finer, more

detailed aspects of the signal (LeCun et al., 2015; Zeiler and Fergus, 2014). Although these finer features differ between turtle and human behaviors, the fine-tuning phase enables it to adapt and optimize its weights to capture fine-grained features of the new species. In their survey on transfer learning, Weiss et al. (2016) draw an analogy, suggesting that a person with guitar-playing experience and a musical background is likely to learn to play the piano more quickly than an individual lacking prior musical experience. These findings are particularly promising for the broader application of deep learning and transfer learning in acceleration-based behavior identification for wildlife monitoring, independent of the species under study. The results presented in this study suggest that the species used to pre-train the model can differ from the species being studied, yet this approach can still produce better outcomes than a model trained solely on the available data. From an applied perspective, practitioners using accelerometer data can use a model pre-trained on human data as a starting point for developing species-specific models. This could help reduce the amount of data needed for training models, which is a significant advantage for species that are difficult to observe in their natural habitats.

The primary risk of training a deep learning model on a small dataset is overfitting. The significance of a machine learning model's ability to generalize lies in its capacity to effectively handle variations in data. A well-generalized model avoids learning noise or overly specific patterns from the training data, instead capturing the essential, generalizable features that are critical for accurately recognizing relevant behaviors across diverse datasets or individuals. However, achieving good generalization is particularly challenging when the training dataset is small, as the model is more prone to overfitting to the limited data available (Ying, 2019). In our study, we tested the model on a set of new, unseen individuals who were not included in the training data to assess its generalization performance. The observed improvement in the F1-score following the application of transfer learning suggests that the model has enhanced its ability to generalize to new data. In addition, a common method to evaluate overfitting is to visualize the model's performance on both the training and validation datasets during training. If the model performs well on the training data but poorly on the validation data, it indicates overfitting. In this study, the visualization of performance trends during training for both a randomly initialized model and a model using transfer learning suggests that the transfer learning model exhibited less overfitting (Fig. S3). Training the model on a larger dataset allows it to learn from a broader range of examples, enabling it to capture more diverse patterns and variability among individuals. This helps the model identify key features that differentiate behaviors in a way that is independent of specific individuals,

thus enhancing its ability to generalize. The model's ability to avoid overfitting and generalize effectively is promising for its application to new, unseen individuals, suggesting its robustness and potential for real-world applications.

Another advantage of transfer learning highlighted in this study is the reduced variation in the F1-scores across repeated independent executions, indicating a more robust model with transfer learning (Model-Hawksbill, sd = 2.81%; Model-Green_turtle with all layers fine-tuned, sd = 0.35%, Table 1). Even using transfer learning with human data reduced this variability (Model-Human, sd = 1.17%). Our study shows that the model performance can vary by more than 10% points when starting from a random initialization of weights (Fig. 3). This is likely because the initial values of the weights affect the ability to achieve the suitable weights during training. Depending on the initial weights, training may not reach the suitable values, resulting in lower performance in some cases. Another explanation could be that the model is less sensitive to individual variations and more consistent in its ability to predict behavior for new, unseen individuals. From an applied perspective, having a robust model is particularly important since, in most cases, the model is trained only once before being applied to new data.

Deep learning remains relatively underutilized in wildlife monitoring based on acceleration data, possibly because deep learning models are often considered black boxes with the difficulty in identifying the information used by the models (Castelvecchi, 2016). In this study, we aimed to enhance our comprehension of transfer learning and deep learning models by fine-tuning different layers and evaluating the outcomes (Fig. 3, Table 1). We demonstrate that, with an encoder decoder network, better results are attained when fine-tuning either the entire model (F1-score: 77.12%) or the first layers, the encoder (F1-score = 76.55%), highlighting the significance of these early layers in adapting the model to a specific task, or in this case, a specific species. The weaker results observed when the encoder is "frozen" corroborate this finding (F1-score = 53.21%). In transfer learning applications within computer vision and bioacoustics, it is typically the final layers of the pre-trained model that undergo fine-tuning (Iman et al., 2023). In our study, the encoder emerges as the pivotal component that requires customization to suit the specific task or species under study. Based on our results, we can speculate that the decoder, responsible for signal reconstruction, is less species-specific as we obtained comparable results whether the decoder was fine-tuned along with the encoder or maintained "frozen". These results also suggest a low dependency between the encoder, responsible for extracting information, and the decoder. Lastly, while models in computer vision or bioacoustics yield favorable results with softmax fine-tuning

alone (Batist et al., 2024; Dufourq et al., 2022), this approach yields remarkably poor results in our study (F1-score = 46.90%). These results provide a clearer insight into the layers of the V-Net responsible for acquiring problem-specific knowledge. While transfer learning has become a commonly used approach in both bioacoustics and computer vision, we demonstrate that its implementation can vary significantly depending on the model architecture and the specific classification task involved. Although our study primarily focuses on developing a robust method for wildlife monitoring, our findings also enhance comprehension of the implicated models, thereby offering potential benefits to the wider deep learning and ecology communities.

One of the primary advantages of freezing layers in deep learning models is the reduction of computational cost during training. In the context of transfer learning, only the weights of unfrozen layers are fine-tuned, reducing the number of calculations required to optimize the model. For example, in bioacoustics, researchers have used the pre-trained model ResNet152V2, with 58 331 648 weights, and have fine-tuned only the softmax layers with 81 922 weights (Batist et al., 2024). As a result, with fewer weights to optimize, the training time is significantly reduced. In our case, achieving the best results requires either fine-tuning all the weights, which does not reduce training time compared to a model with randomly initialized weights (3.92 minutes with GPU for the Model-Green_turtle vs 3.96 minutes for the Model-Hawksbill), or fine-tuning only the encoder. In the latter scenario, we observe only a slight improvement in training time, with a reduction of 59 seconds without a CPU and 19 seconds with a GPU compared to the Model-Hawksbill (Table 1). The reduction in training time becomes more significant when only the decoder or softmax layers are fine-tuned, but this comes with a decrease in performance. Nevertheless, while we do not observe a significant improvement in training times due to transfer learning, the training durations remain reasonable, averaging around 10 minutes without a GPU.

In conclusion, we demonstrate that transfer learning improves behavior classification using accelerometer data of an endangered species, particularly in situations where datasets are scarce. By leveraging transfer learning, we developed a more robust model with reduced variability in predictive performance and better generalization, evidenced by a higher F1 score on new individuals compared to a model trained from randomly initialized weights. We also demonstrate that transfer learning can be successfully applied using data from entirely different species—such as humans and hawksbill turtles—though better results are obtained when the species share similar morphology, as seen with the green turtle and hawksbill turtle. From a practical perspective, it remains essential to collect data and label it through direct

observations, which can be particularly challenging for endangered species that are difficult to monitor. Nevertheless, the amount of labeled data required for training can be significantly reduced through transfer learning. While estimating the minimum number of individuals needed is difficult due to task-specific issues—such as the difficulty of distinguishing the associated accelerometer signals—transfer learning generally decreases the number of individuals required. It is therefore feasible to pre-train models on larger datasets available online that may be similar to the species of interest or on human data when specific data are lacking. Since transfer learning can be applied to all deep learning architectures, it is also possible to reuse existing models available online, usually provided with the data they were trained on (Jeantet et al., 2021; Otsuka et al., 2024; Schoombie et al., 2024; Y. Zhang et al., 2019). Fine-tuning can be applied to all layers, particularly in models with relatively few parameters, or to specific layers to adapt the model to the species under study. We demonstrate that the choice of layers for fine-tuning may vary depending on the architecture, highlighting the need for a review of common practices from other fields according to the chosen model architecture. Therefore, transfer learning serves as a valuable resource for researchers, enabling them to adapt existing models to their study species and saving time by eliminating the need to develop species tailored models, which is a common practice. In this context, it is crucial for practitioners to share their accelerometer data and associated models, allowing the broader community to benefit and ultimately facilitate the utilization of accelerometers for wildlife monitoring and expanding its scope of application.

Author Contributions
LJ, DC and ED designed the research project. DC, CD and JM collected the hawksbill dataset; LJ and ED designed the methodology; LJ, CD and JM analyzed and labeled the dataset; LJ and KZ experimented the transfer learning; LJ, KZ and ED led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data Availability Statement

Pre-trained models and analysis scripts are available online on GitHub (https://github.com/AIMS-Research/research_za/tree/main/biologging_transferlearning_hawksbill). The hawksbill dataset to run the scripts is available on Zenodo (https://doi.org/10.5281/zenodo.11402241).

Conflict of Interest Statement

The authors declare that they have no conflicts of interest.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.

Abernathy, H.N., Crawford, D.A., Garrison, E.P., Chandler, R., Conner, M., Miller, K., Cherry, M.J., 2019. Deer movement and resource selection during Hurricane Irma: implications for extreme climatic events and wildlife. Proc. R. Soc. B 286, 20192230.

Aulsebrook, A.E., Jacques-Hamilton, R., Kempenaers, B., 2024. Quantifying mating behaviour using accelerometry and machine learning: challenges and opportunities. Anim. Behav. 207, 55–76. https://doi.org/10.1016/j.anbehav.2023.10.013

Bao, J., Xie, Q., 2022. Artificial intelligence in animal farming : A systematic literature review. J. Clean. Prod. 331, 129956. https://doi.org/10.1016/j.jclepro.2021.129956

Batist, C.H., Dufourq, E., Jeantet, L., Razafindraibe, M.N., Randriamanantena, F., Baden, A.L., 2024. An integrated passive acoustic monitoring and deep learning pipeline for black-and-white ruffed lemurs (Varecia variegata) in Ranomafana National Park, Madagascar. Am. J. Primatol. 1–18. https://doi.org/10.1002/ajp.23599

Beever, E.A., Hall, L.E., Varner, J., Loosen, A.E., Dunham, J.B., Gahl, M.K., Smith, F.A., Lawler, J.J., 2017. Behavioral flexibility as a mechanism for coping with climate change. Front. Ecol. Environ. 15, 299–308. https://doi.org/10.1002/fee.1502

Bloch, V., Frondelius, L., Arcidiacono, C., Mancino, M., Pastell, M., 2023. Development and Analysis of a CNN- and Transfer-Learning-Based Classification Model for Automated Dairy Cow Feeding Behavior Recognition from Accelerometer Data. Sensors 23. https://doi.org/10.3390/s23052611

Boel, M., Aarestrup, K., Baktoft, H., Larsen, T., Søndergaard Madsen, S., Malte, H., Skov, C., Svendsen, J.C., Koed, A., 2014. The physiological basis of the migration continuum in brown trout (Salmo trutta). Physiol. Biochem. Zool. 87, 334–345.

Bonola, M., Girondot, M., Robin, J.P., Martin, J., Siegwalt, F., Jeantet, L., Lelong, P., Grand, C., Chambault, P., Etienne, D., Gresser, J., Hielard, G., Arqué, A., Régis, S., Lecerf, N., Frouin, C., Lefebvre, F., Sutter, E., Vedie, F., Barnerias, C., Thieulle, L., Bordes, R., Guimera, C., Aubert, N., Bouaziz, M., Pinson, A., Flora, F., Duru, M., Benhalilou, A., Murgale, C., Maillet, T., Andreani, L., Campistron, G., Sikora, M., Rateau, F., George, F., Eggenspieler, J., Woignier, T., Allenou, J.P., Louis-Jean, L., Chanteur, B., Béranger, C., Crillon, J., Brador, A., Habold, C., Le Maho, Y., Chevallier, D., 2019. Fine scale geographic residence and annual primary production drive body condition of wild immature green turtles (Chelonia mydas) in Martinique Island (Lesser Antilles). Biol. Open 8. https://doi.org/10.1242/bio.048058

Brown, D.D., Kays, R., Wikelski, M., Wilson, R., Klimley, A.P., 2013. Observing the unwatchable through acceleration logging of animal behavior. Anim. Biotelemetry 1, 20. https://doi.org/10.1186/2050-3385-1-20

Buchholz, R., Banusiewicz, J.D., Burgess, S., Crocker-Buta, S., Eveland, L., Fuller, L., 2019. Behavioural research priorities for the study of animal response to climate change. Anim. Behav. 150, 127–137.

Castelvecchi, D., 2016. Can we open the black box of AI? Nat. News 538, 20.

Charrier, I., Jeantet, L., Maucourt, L., Régis, S., Lecerf, N., Benhalilou, A., Chevallier, D., 2022. First evidence of underwater vocalizations in green sea turtles Chelonia mydas. Endanger. Species Res. 48, 31–41. https://doi.org/10.3354/esr01185

Choi, A., Kim, T.H., Yuhai, O., Jeong, S., Kim, K., Kim, H., Mun, J.H., 2022. Deep learning-based near-fall detection algorithm for fall risk monitoring system using a single inertial measurement unit. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 2385–2394.

Chollet, F., 2015. Keras.

Cook, D., Feuz, K.D., Krishnan, N.C., 2013. Transfer learning for activity recognition: a survey. Knowl. Inf. Syst. 36, 537–556. https://doi.org/10.1007/s10115-013-0665-3

Dufourq, E., Batist, C., Foquet, R., Durbach, I., 2022. Passive acoustic monitoring of animal populations with transfer learning. Ecol. Inform. 70, 101688. https://doi.org/10.1016/j.ecoinf.2022.101688

Fossette, S., Gleiss, A.C., Myers, A.E., Garner, S., Liebsch, N., Whitney, N.M., Hays, G.C., Wilson, R.P., Lutcavage, M.E., 2010. Behaviour and buoyancy regulation in the deepest-diving reptile: the leatherback turtle. J. Exp. Biol. 213, 4074–4083. https://doi.org/10.1242/jeb.048207

Fossette, S., Schofield, G., Lilley, M.K.S., Gleiss, A.C., Hays, G.C., 2012. Acceleration data reveal the energy management strategy of a marine ectotherm during reproduction. Funct. Ecol. 26, 324–333. https://doi.org/10.1111/j.1365-2435.2011.01960.x

Geiger, S., 2019. Package "rblt" 1–12.

Goodfellow, I., 2016. Deep learning.

Gray, P.C., Bierlich, K.C., Mantell, S.A., Friedlaender, A.S., Goldbogen, J.A., Johnston, D.W., 2019. Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. Methods Ecol. Evol. 10, 1490–1500.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2018. Recent advances in convolutional neural networks. Pattern Recognit. 77, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Hayes, C.T., Baumbach, D.S., Juma, D., Dunbar, S.G., 2017. Impacts of recreational diving on hawksbill sea turtle (Eretmochelys imbricata) behaviour in a marine protected area. J. Sustain. Tour. 25, 79–95.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2016.90

Hernandez, N., Lundström, J., Favela, J., McChesney, I., Arnrich, B., 2020. Literature Review on Transfer Learning for Human Activity Recognition Using Mobile and Wearable Devices with Environmental Technology. SN Comput. Sci. 1, 1–16. https://doi.org/10.1007/s42979-020-0070-4

Iman, M., Arabnia, H.R., Rasheed, K., 2023. A Review of Deep Transfer Learning and Recent Advancements. Technologies 11, 40. https://doi.org/10.3390/technologies11020040

International Union for Conservation of Nature (IUCN), 2024. The IUCN Red List of Threatened Species.

Jeantet, L., Dell'Amico, F., Forin-Wiart, M.-A., Coutant, M., Bonola, M., Etienne, D., Gresser, J., Regis, S., Lecerf, N., Lefebvre, F., de Thoisy, B., Le Maho, Y., Brucker, M., Châtelain, N., Laesser, R., Crenner, F., Handrich, Y., Wilson, R., Chevallier, D., 2018. Combined use of two supervised learning algorithms to model sea turtle behaviours from tri-axial acceleration data. J. Exp. Biol. 221. https://doi.org/10.1242/jeb.177378

Jeantet, L., Hadetskyi, V., Vigon, V., Korysko, F., Paranthoen, N., Chevallier, D., 2022. Estimation of the Maternal Investment of Sea Turtles by Automatic Identification of Nesting Behavior and Number of Eggs Laid from a Tri-Axial Accelerometer. Animals 12. https://doi.org/10.3390/ani12040520

Jeantet, L., Planas-Bielsa, V., Benhamou, S., Geiger, S., Martin, J., Siegwalt, F., Lelong, P., Gresser, J., Etienne, D., Hiélard, G., Arque, A., Regis, S., Lecerf, N., Frouin, C., Benhalilou, A., Murgale, C., Maillet, T., Andreani, L., Campistron, G., Delvaux, H., Guyon, C., Richard, S., Lefebvre, F., Aubert, N., Habold, C., le Maho, Y., Chevallier, D., 2020a. Behavioural inference from signal processing using animal-borne multi-sensor loggers: a novel solution to extend the knowledge of sea turtle ecology. R. Soc. Open Sci. 7, 200139. https://doi.org/10.1098/rsos.200139

Jeantet, L., Planas-Bielsa, V., Benhamou, S., Geiger, S., Martin, J., Siegwalt, F., Lelong, P., Gresser, J., Etienne, D., Hielard, G., Arque, A., Regis, S., Lecerf, N., Frouin, C., Benhalilou, A., Murgale, C., Maillet, T., Andreani, L., Campistron, G., Delvaux, H., Guyon, C., Richard, S., Lefebvre, F., Aubert, N., Habold, C., Le Maho, Y., Chevallier, D., 2020b. Raw acceleration, gyroscope and depth profiles associated with the observed behaviours of free-ranging immature green turtles in Martinique.

Jeantet, L., Vigon, V., Geiger, S., Chevallier, D., 2021. Fully convolutional neural network : a solution to infer animal behaviours from multi-sensor data. Ecol. Model. 450. https://doi.org/10.1016/j.ecolmodel.2021.109555

Kleanthous, N., Hussain, A., Khan, W., Sneddon, J., Liatsis, P., 2022a. Deep transfer learning in sheep activity recognition using accelerometer data. Expert Syst. Appl. 207, 117925. https://doi.org/10.1016/j.eswa.2022.117925

Kleanthous, N., Hussain, A.J., Khan, W., Sneddon, J., Al-Shamma'a, A., Liatsis, P., 2022b. A survey of machine learning approaches in animal behaviour. Neurocomputing 491, 442–463.

Kumar, P., Chauhan, S., Awasthi, L.K., 2023. Human Activity Recognition (HAR) Using Deep Learning : Review , Methodologies , Progress and Future Research Directions. Arch. Comput. Methods Eng. 31, 179–219. https://doi.org/10.1007/s11831-023-09986-x

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Lelong, P., Besnard, A., Girondot, M., Habold, C., Priam, F., Giraudeau, M., Le Loc'h, G., Le Loc'h, A., Fournier, P., Fournier-Chambrillon, C., Bustamante, P., Dupont, S.M., Vincze, O., Gros-Desormeaux, J.-R., Martin, J., Bourgeois, O., Lepori, M., Régis, S., Lecerf, Nicolas, Lefebvre, F., Aubert, N., Frouin, C., Flora, F., Pimentel, E., Pimentel, M., Siegwalt, F., Jeantet, L., Chambault, P., Hielard, G., Arqué, A., Arthus, M., Louis-Jean, L., Brador, A., Giannasi, P., Etienne, D., Lecerf, Nathaël, Chevallier, P., Chevallier, T., Meslier, S., Landreau, A., Maceno, M., Larcher, E., Le Maho, Y., Chevallier, D., 2024. Demography of endangered juvenile green turtles in face of environmental changes: 10 years of capture-mark-recapture efforts in Martinique. Biol. Conserv. 291, 110471. https://doi.org/10.1016/j.biocon.2024.110471

Li, A., Reidenbach, M.A., 2014. Forecasting decadal changes in sea surface temperatures and coral bleaching within a Caribbean coral reef. Coral Reefs 33, 847–861.

Liao, J., Xu, J., Shen, Y., Lin, S., 2023. THANet: Transferring Human Pose Estimation to Animal Pose Estimation. Electronics 12, 4210.

Louis-Jean, L., Lenfant, P., Galvin, R., Marechal, J.-P., 2009. Crossing fishery statistics with marine turtle bycatch data and habitat mapping in Martinique, FW.

Lu, T., Han, B., Yu, F., 2021. Detection and classification of marine mammal sounds using AlexNet with transfer learning. Ecol. Inform. 62, 101277. https://doi.org/10.1016/j.ecoinf.2021.101277

Mao, A., Huang, E., Wang, X., Liu, K., 2023. Deep learning-based animal activity recognition with wearable sensors: Overview, challenges, and future directions. Comput. Electron. Agric. 211, 108043.

McHuron, E.A., Peterson, S.H., Hückstädt, L.A., Melin, S.R., Harris, J.D., Costa, D.P., 2018. The energetic consequences of behavioral variation in a marine carnivore. Ecol. Evol. 8, 4340–4351.

Meena, T., Sarawadekar, K., 2023. Seq2Dense U-Net: Analysing Sequential Inertial Sensor data for Human Activity Recognition using Dense Segmentation Model. IEEE Sens. J.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. https://doi.org/10.1109/3DV.2016.79

Nivière, M., Chambault, P., Pérez, T., Etienne, D., Bonola, M., Martin, J., Barnérias, C., Védie, F., Mailles, J., Dumont-dayot, É., Gresser, J., Hiélard, G., Régis, S., Lecerf, N., Thieulle, L., Duru, M., Lefebvre, F., Milet, G., Guillemot, B., Bildan, B., Montgol, B. De, Benhalilou, A., Murgale, C., Maillet, T., Queneherve, P., Woignier, T., Sa, M., Le, Y., Petit, O., Chevallier, D., 2018. Identification of marine key areas across the Caribbean to ensure the conservation of the critically endangered hawksbill turtle. Biol. Conserv. 223, 170–180. https://doi.org/10.1016/j.biocon.2018.05.002

Nunavath, V., Johansen, S., Johannessen, T.S., Jiao, L., Hansen, B.H., Berntsen, S., Goodwin, M., 2021. Deep learning for classifying physical activities from accelerometer data. Sensors 21, 5564.

Okuyama, J., Kataoka, K., Kobayashi, M., Abe, O., Yoseda, K., Arai, N., 2012. The regularity of dive performance in sea turtles: A new perspective from precise activity data. Anim. Behav. 84, 349–359. https://doi.org/10.1016/j.anbehav.2012.04.033

Otsuka, R., Yoshimura, N., Tanigaki, K., Koyama, S., Mizutani, Y., Yoda, K., Maekawa, T., 2024. Exploring deep learning techniques for wild animal behaviour classification using animal-borne accelerometers. Methods Ecol. Evol. https://doi.org/10.1111/2041-210X.14294

Pedersen, L., Jackson, K., Thorup, K., Tøttrup, A.P., 2018. Full-year tracking suggests endogenous control of migration timing in a long-distance migratory songbird. Behav. Ecol. Sociobiol. 72, 1–10.

Ramanujam, E., Perumal, T., Padmavathi, S., 2021. Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. IEEE Sens. J. 21, 13029–13040.

Reyes-Ortiz, J., Anguita, D., Ghio, A., Oneto, L., Parra, X., 2013. Human Activity Recognition Using Smartphones. https://doi.org/10.24432/C54S4K

Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS One 10, e0118432.

Sansom, A., Lind, J., Cresswell, W., 2009. Individual behavior and survival: the roles of predator avoidance, foraging success, and vigilance. Behav. Ecol. 20, 1168–1174.

Schneider, S., Taylor, G.W., Kremer, S., 2018. Deep Learning Object Detection Methods for Ecological Camera Trap Data, in: 2018 15th Conference on Computer and Robot Vision (CRV). Presented at the 2018 15th Conference on Computer and Robot Vision (CRV), IEEE, Toronto, ON, Canada, pp. 321–328. https://doi.org/10.1109/CRV.2018.00052

Schoombie, S., Jeantet, L., Chimienti, M., Sutton, G., Pistorius, P., Dufourq, E., Lowther, A., Oosthuizen, W., 2024. Identifying prey capture events of a free-ranging marine predator using bio-logger data and deep learning. R. Soc. Open Sci. 11, 240271. https://doi.org/10.1098/rsos.240271

Sergio, F., Blas, J., Hiraldo, F., 2018. Animal responses to natural disturbance and climate extremes: a review. Glob. Planet. Change 161, 28–40.

Shepard, E.L.C., Wilson, R.P., Quintana, F., Laich, A.G., Liebsch, N., Albareda, D.A., Halsey, L.G., Gleiss, A., Morgan, D.T., Myers, A.E., Newman, C., Macdonald, D.W., 2008. Identification of animal movement patterns using tri-axial accelerometry. Endanger. Species Res. 10, 47–60. https://doi.org/10.3354/esr00084

Siegwalt, F., Benhamou, S., Girondot, M., Jeantet, L., Martin, J., Bonola, M., Lelong, P., Grand, C., Chambault, P., Benhalilou, A., Murgale, C., Maillet, T., Andreani, L., Campistron, G., Jacaria, F., Hielard, G., Arqué, A., Etienne, D., Gresser, J., Régis, S., Lecerf, N., Frouin, C., Lefebvre, F., Aubert, N., Vedie, F., Barnerias, C., Thieulle, L., Guimera, C., Bouaziz, M., Pinson, A., Flora, F., George, F., Eggenspieler, J., Woignier, T., Allenou, J.-P., Louis-Jean, L., Chanteur, B., Béranger, C., Crillon, J.,

Brador, A., Habold, C., Maho, Y.L., Robin, J.-P., Chevallier, D., 2020. High fidelity of sea turtles to their foraging grounds revealed by satellite tracking and capture-mark-recapture: New insights for the establishment of key marine conservation areas. Biol. Conserv. 250, 108742. https://doi.org/10.1016/j.biocon.2020.108742

Siegwalt, F., Jeantet, L., Lelong, P., Martin, J., Girondot, M., Bustamante, P., Benhalilou, A., Murgale, C., Andreani, L., Jacaria, F., Campistron, G., Lathière, A., Barotin, C., Buret-Rochas, G., Barre, P., Hielard, G., Arqué, A., Régis, S., Lecerf, N., Frouin, C., Lefebvre, F., Aubert, N., Arthus, M., Etienne, D., Allenou, J.-P., Delnatte, C., Lafolle, R., Thobor, F., Chevallier, P., Chevallier, T., Lepori, M., Assio, C., Grand, C., Bonola, M., Tursi, Y., Varkala, P.-W., Meslier, S., Landreau, A., Le Maho, Y., Habold, C., Robin, J.-P., Chevallier, D., 2022. Food selection and habitat use patterns of immature green turtles (Chelonia mydas) on Caribbean seagrass beds dominated by the alien species Halophila stipulacea. Glob. Ecol. Conserv. 37, e02169. https://doi.org/10.1016/j.gecco.2022.e02169

Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M., 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, in: Springer (Ed.), Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham, pp. 240–248. https://doi.org/10.1007/978-3-319-67558-9_28

Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. J. Big Data 3, 9. https://doi.org/10.1186/s40537-016-0043-6

Ying, X., 2019. An overview of overfitting and its solutions, in: Journal of Physics: Conference Series. IOP Publishing, p. 022022.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H., 2015. Understanding Neural Networks Through Deep Visualization. https://doi.org/10.48550/arXiv.1506.06579

Zeiler, M.D., Fergus, R., 2014. Visualizing and Understanding Convolutional Networks, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 818–833.

Zhang, G., Swaisgood, R.R., Zhang, H., 2004. Evaluation of behavioral factors influencing reproductive success and failure in captive giant pandas. Zoo Biol. Publ. Affil. Am. Zoo Aquar. Assoc. 23, 15–31.

Zhang, Shibo, Li, Y., Zhang, Shen, Shahabi, F., Xia, S., Deng, Y., Alshurafa, N., 2022. Deep learning in human activity recognition with wearable sensors: A review on advances. Sensors 22, 1476.

Zhang, W., Martinez, A., Meese, E.N., Lowe, C.G., Yang, Y., Yeh, H.G.H., 2019. Deep Convolutional Neural Networks for Shark Behavior Analysis. 2019 IEEE Green Energy Smart Syst. Conf. IGESSC 2019 2019-Janua. https://doi.org/10.1109/IGESSC47875.2019.9042394

Zhang, Y., Zhang, Z., Zhang, Yu, Bao, J., Zhang, Yifan, Deng, H., 2019. Human Activity Recognition Based on Motion Sensor Using U-Net. IEEE Access 7, 75213–75226. https://doi.org/10.1109/ACCESS.2019.2920969
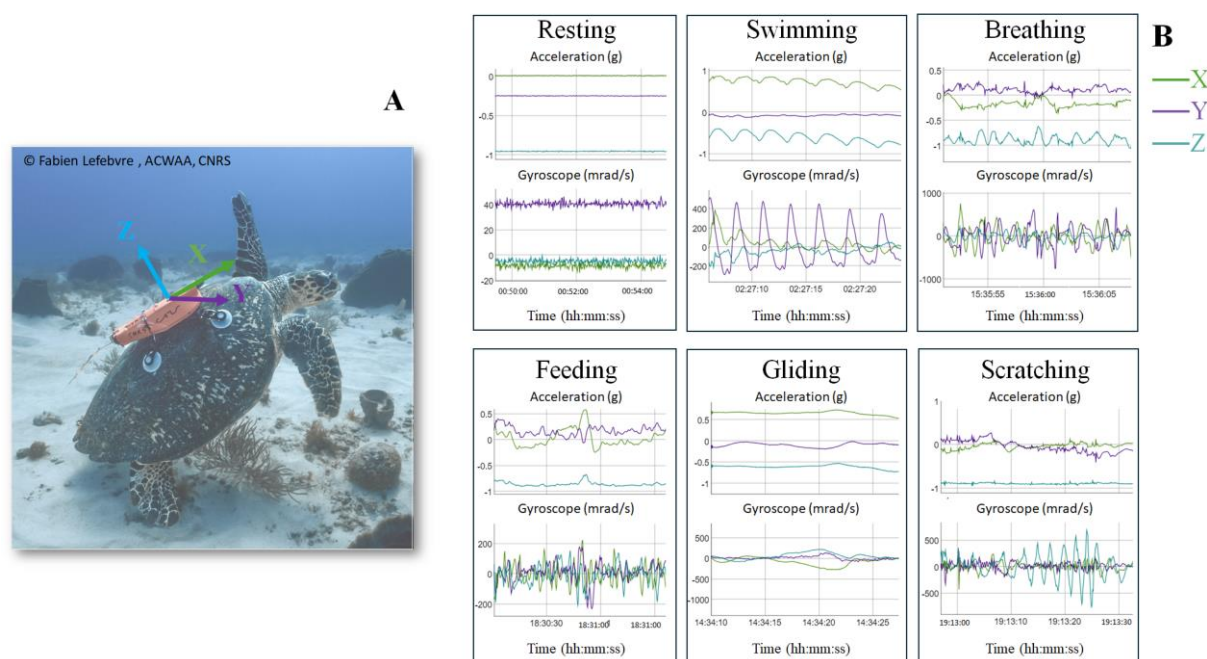
Figures and Tables

Fig. 1. On-board camera combined with tri-axial accelerometer and tri-axial gyroscope deployed on a hawksbill sea turtle in Martinique (credit photo: Fabien Lefebvre, ACWAA, CNRS). The three arrows indicate the orientation of the accelerometer and gyroscope axes: X represents the front-to-back axis, Y the lateral axis, and Z the top-to-bottom axis (**A**). The plots on the right show the raw acceleration and gyroscope signals corresponding to each observed behavior of the hawksbill sea turtle (**B**).
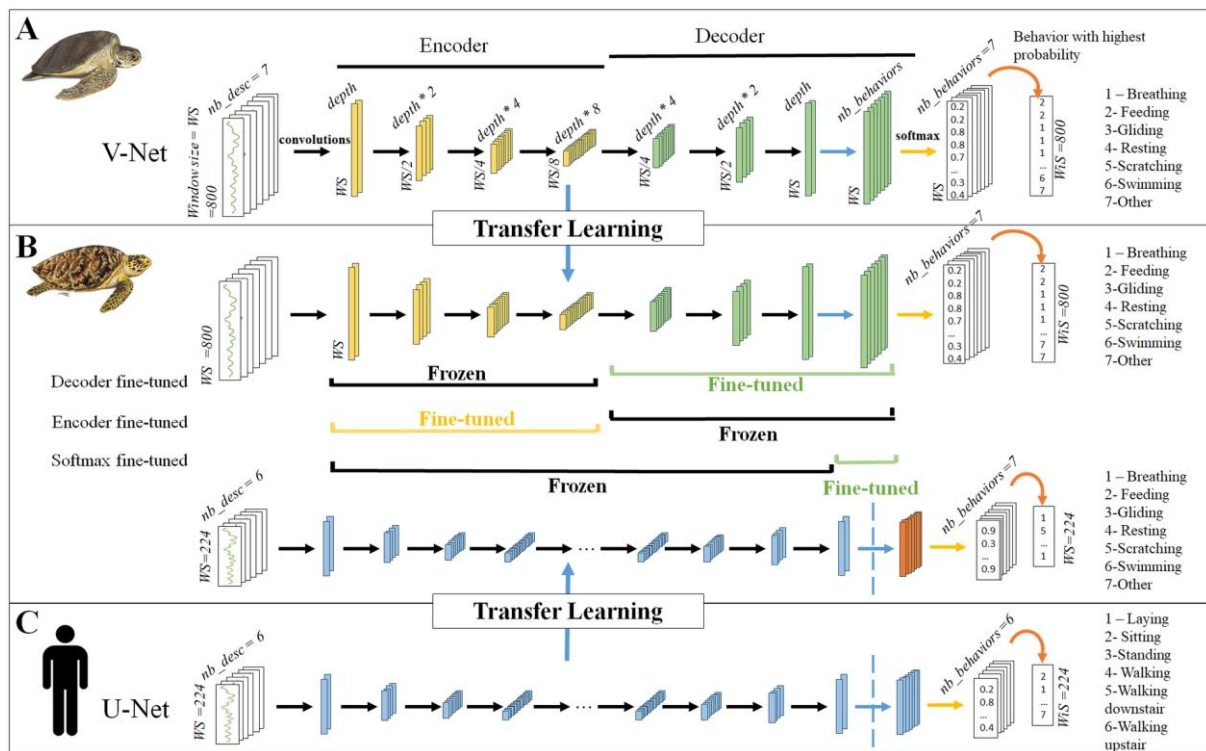
Fig. 2. Representation of the V-Net and U-Net architectures. Panel **A** depicts the V-Net pre-trained on the green turtle dataset, while Panel **C** illustrates the U-Net pre-trained on the human dataset. Panel **B** presents the transfer learning applications tested on the hawksbill dataset, highlighting the different frozen layer configurations for the model pre-trained on the green turtle dataset and the modified layer, shown in orange, for the model pre-trained on the human dataset. The black arrows represent two consecutive convolutional layers, while the blue arrows represent a single convolutional layer. *depth* corresponds to the number of features to generate (*depth*=32), *nb_desc* refers to the number of input data, and *nb_behaviors* denotes the number of behaviors. *WS* represents the window size.
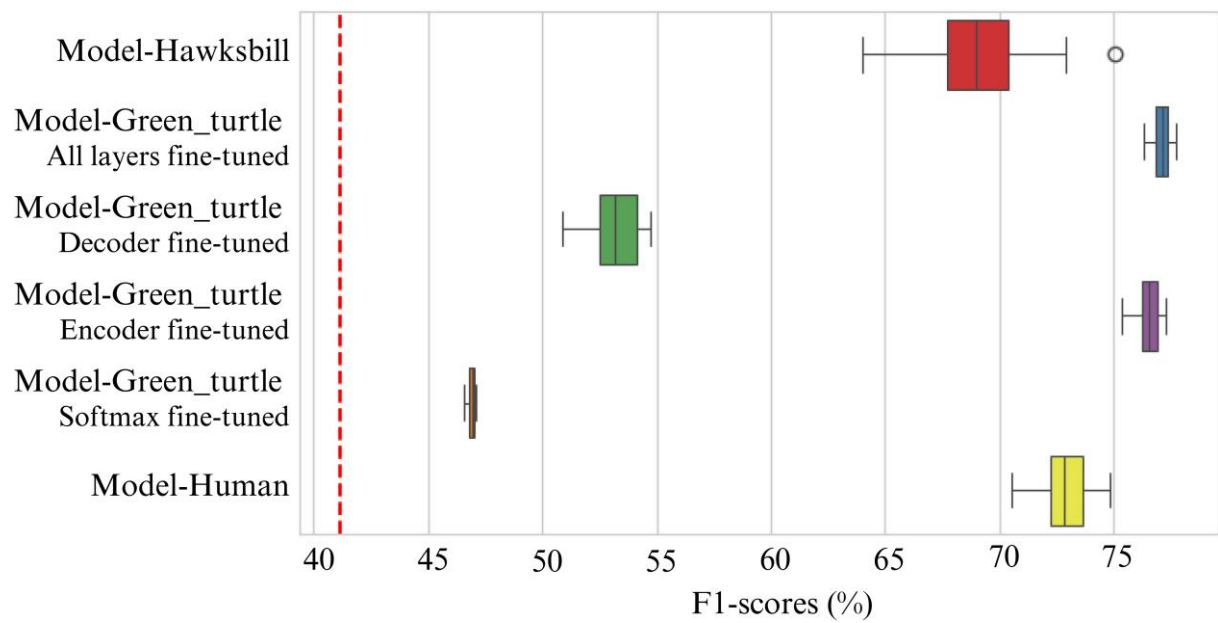
Fig. 3. Boxplots of the F1-scores obtained for each run (n=20) across all experiments. The red line indicates the F1-score of model trained on green turtles and applied to the hawksbill dataset without using transfer learning (F1-score = 41.17%).
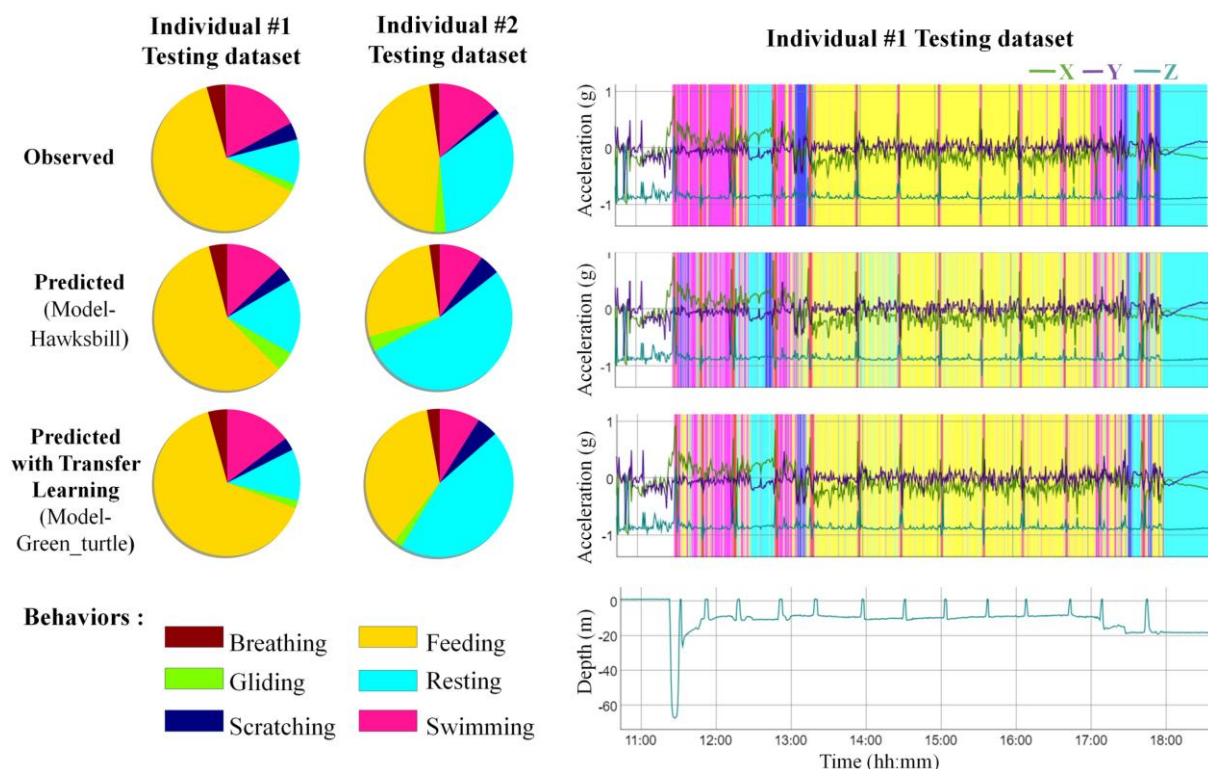
Fig. 4. Comparison of time budgets for two individuals from the testing dataset: observed behaviors from video analysis, predictions from a model with randomly initialized weights, and predictions from transfer learning using a model pre-trained on green turtles with all layers fine-tuned. The plots on the left show the time budget as a pie chart, while the plot on the right visualizes the predictions over time for one individual, using the *rblt* package.

Table 1. Average F1-score (%) and training execution time, in minutes, using either a Graphical Processing Unit (GPU) or Central Processing Unit (CPU) for each experiment across 20 independent executions. We display the pre-trained dataset used (data), which layers were fine-tuned (layers), and the number of trainable parameters (params).

Data	Layers	Params	F1-score ±sd	GPU time	CPU time
Hawksbill	No transfer learning	1,256,487	69.11 ±2.81	3.96	9.44
Green Turtle	All	1,256,487	77.12 ±0.35	3.92	9.36
Green Turtle	Decoder	302,183	53.21 ±1.02	3.03	6.47
Green Turtle	Encoder	954,304	76.55 ±0.47	3.64	8.45
Green Turtle	Softmax	231	46.90 ±0.13	2.19	4.68
Human	All	10,497,414	72.87 ±1.17	3.24	9.70

A   Appendix

Architecture of the V-Net and the U-Net

The U-Net was introduced by Ronneberger, Fischer, and Brox (2015) for detecting organs of interest in 2D biomedical images. The V-Net, introduced by Milletari, Navab, and Ahmadi (2016), is an adaptation of the U-Net designed for 3D biomedical images. Both share the same architecture with an encoder and decoder and the novelty of the V-Net lies in its use of 3D convolutions instead of 2D. Both models consist of layers, where each layer includes two convolutional layers followed by downsampling in the encoder or upsampling in the decoder. Another innovation of the V-Net lies in the addition of a "shortcut," where the input of the first convolution is added to the output of the second convolution at each layer. Milletari et al. (2016).

The V-Net developed on the green turtle dataset is therefore a fully convolutional network, with successive sequence of 1D convolutions based on the time dimension (Jeantet, Vigon, Geiger, & Chevallier, 2021). Each convolutional layer is followed by a ReLU activation. Downsampling is achieved by performing a convolution with a kernel size of 2 and a stride of 2. Upsampling is achieved by doubling the data size, effectively creating a copy of the data along the time dimension. The model consists of 3 layers corresponding to 8 convolutional layers and 3 downsampling operations in the encoder, as well as 6 convolutional layers and 3 upsampling operations in the decoder. The last layer is a 1D convolution with softmax activation to obtain the probability for each behavior.

The U-Net used in Zhang et al. (2019) is similar to the one proposed by Ronneberger et al. (2015), with the addition of a sequence of double convolutions. The model comprises 5 layers, consisting of 12 convolutional layers and 5 downsampling operations in the encoder, along with 10 convolutional layers and 5 upsampling operations in the decoder. Downsampling is performed using max pooling. In Zhang et al. (2019), the authors applied 2D convolutions to acceleration-gyroscopic time series data. The final layer is a 2D convolution with a softmax activation function, which produces probabilities for each behavior.

References

Jeantet, L., Vigon, V., Geiger, S., & Chevallier, D. (2021). Fully convolutional neural network : a solution to infer animal behaviours from multi-sensor data. *Ecological Modelling*, *450*(109555). doi: 10.1016/j.ecolmodel.2021.109555

Milletari, F., Navab, N., & Ahmadi, S. A. (2016). *V-net: Fully convolutional neural networks for volumetric medical image segmentation* (No. arXiv:1606.04797v1). arXiv. (ISBN: 9781509054077) doi: 10.1109/3DV.2016.79

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation* (No. arXiv:1505.04597). arXiv. Retrieved 2024-04-24, from `http://arxiv.org/abs/1505.04597` doi: 10.48550/arXiv.1505.04597

Zhang, Y., Zhang, Z., Zhang, Y., Bao, J., Zhang, Y., & Deng, H. (2019). Human activity recognition based on motion sensor using u-net. *IEEE Access*, *7*, 75213-75226. doi: 10.1109/ACCESS .2019.2920969
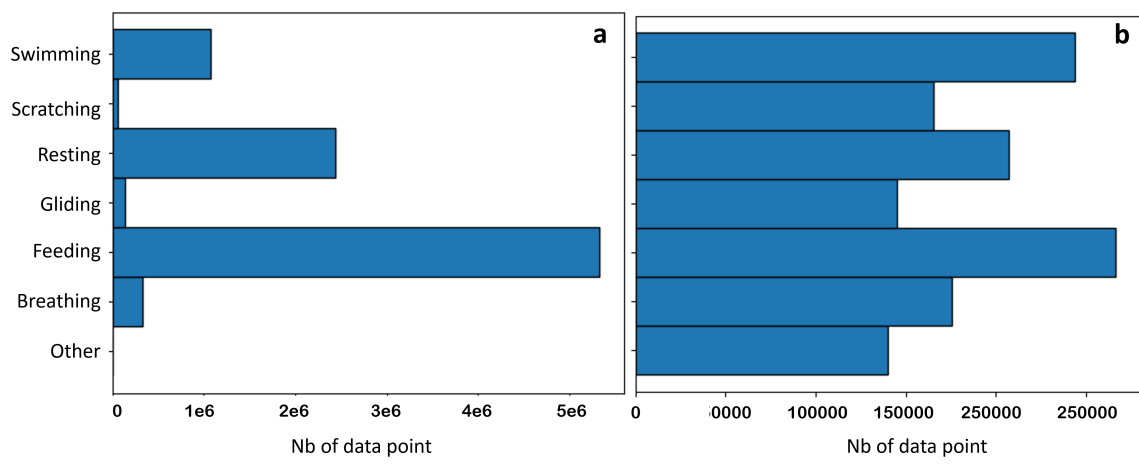
Fig. S1. Histogram of the distribution of behaviors in the hawksbill training dataset when using a sliding window (a) or the global generator described in Jeantet et al. (2021) (b). Figure (b) shows the number of data points for each behavior when 6000 windows of size 800 (40 sec) are drawn. Figure (a) represents the number of data for each behavior when using a sliding window of size 800 on the entire training data with an overlap of 200 (10 sec).
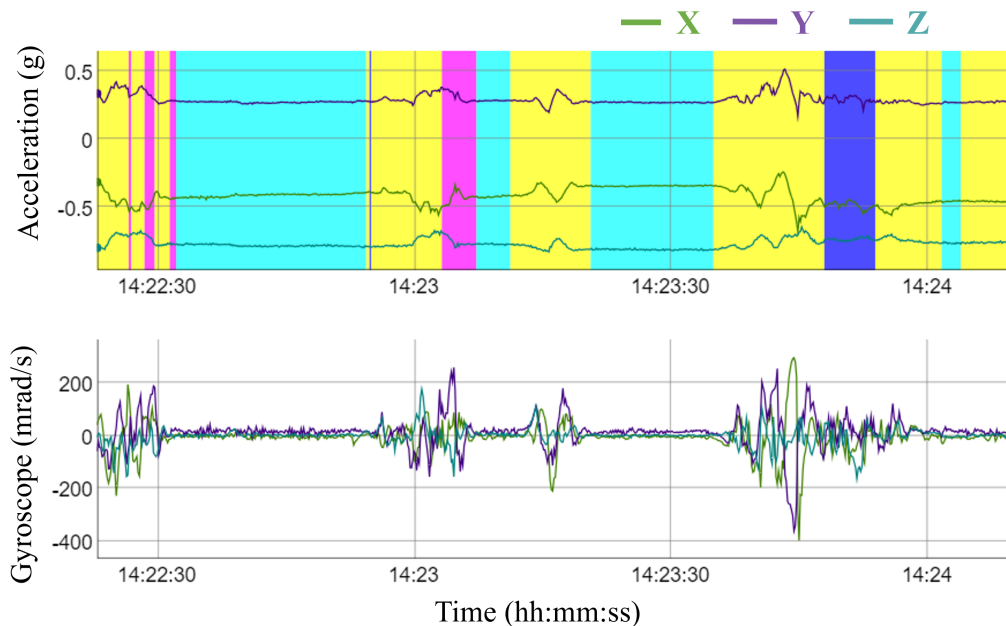

Fig. S2. Predictions from a model pre-trained on the green turtle dataset with all its layers fine-tuned on the hawksbill dataset for a sequence labelled as Feeding of a hawksbill turtle from the testing dataset (individual #2)."
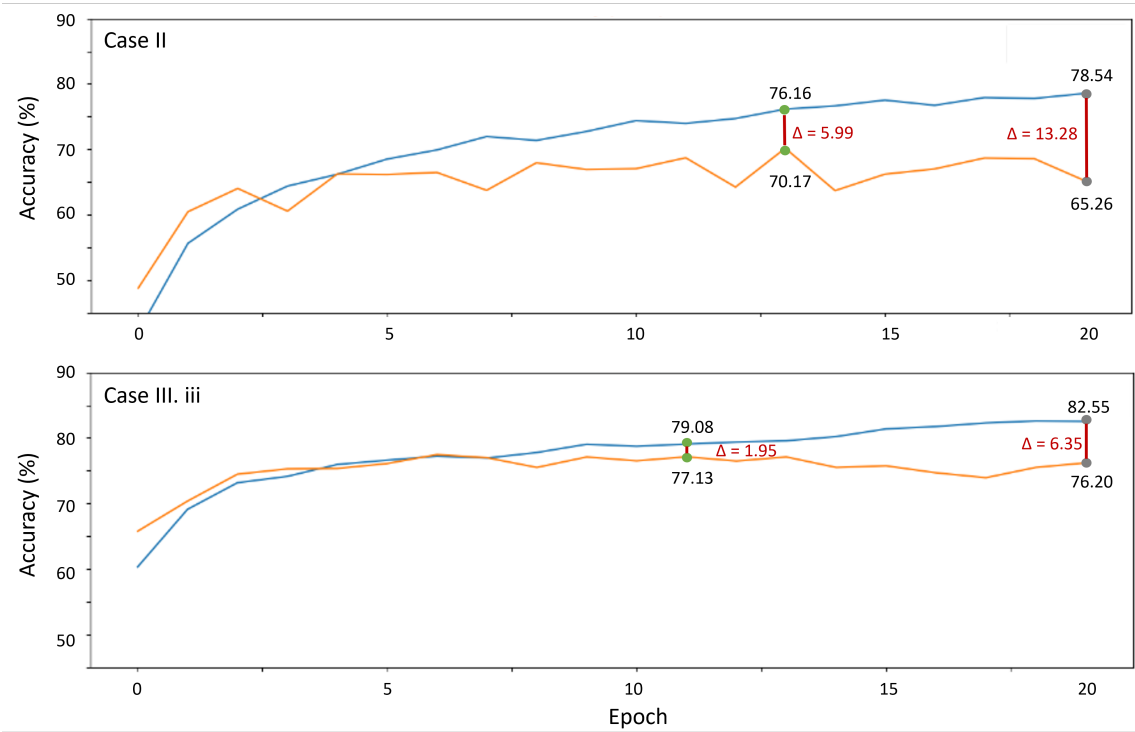
Fig. S3. Accuracy at each epoch on training and validation Hawksbill datasets: comparison between a V-Net with randomly initialized weights and a pre-trained V-Net on green turtle dataset with frozen decoder. The green dots represent the epoch at which the model was saved and the black dots the last epoch. A significant difference between the validation and training data indicates that the model cannot generalize to new data, suggesting that the weights are overly adjusted to the training dataset (overfitting).

Table S1. Presentation of the datasets used in the study, showing the different behaviors and the corresponding recording time available.

	Hawksbill turtle dataset	Green turtle dataset		Human dataset
Number of individuals	6	13		30
Breathing	2.2 h	5.7 h	Walking downstairs	29.9 min
Feeding	38.6 h	1.8 h	Standing	40.6 min
Gliding	1.0 h	2.3 h	Sitting	38.6 min
Resting	19.1 h	34.3 h	Walking	36.9 min
Scratching	0.8 h	1.2 h	Laying	41.2 min
Swimming	7.9 h	22.3 h	Walking upstairs	33.0 min
Other	0.1 h	1.0 h		