animals

MDPI

Communication

Application of Machine Learning for Automating Behavioral Tracking of Captive Bornean Orangutans (*Pongo Pygmaeus*)

Frej Gammelgård [1,†], Jonas Nielsen [1,†], Malthe G. Hansen [1], Emilia J. Nielsen [1], Aage K. Olsen Alstrup [2], Juan O. Perea-García [3], Trine H. Jensen [1,4] and Cino Pertoldi [1,4]

[1] Department of Chemistry and Bioscience, Aalborg University, Frederik Bajers Vej 7H, 9220 Aalborg, Denmark.
[2] Department of Nuclear Medicine & PET, Aarhus University Hospital and Department of Clinical Medicine, Aarhus University, Palle Juul Jensens Boulevard 99, 8000 Aarhus N, Denmark.
[3] Faculty of Social and Behavioural Sciences, Leiden University, Leiden, the Netherlands.
[4] Aalborg Zoo, Mølleparkvej 63, DK-9000 Denmark
† These authors contributed equally to this work.

Simple Summary: This study investigates the application of machine learning in the form of image classification and object detection to video material, to automate behavior recognition in orangutans. The use of machine learning showed potential for future expansion, with the goal of automating behavioral studies, but also notable issues that should be considered when using such methods.

Abstract: This article applies object detection to video material to investigate the potential of using machine learning to automate behavior tracking. The study includes video tapings of two captive bornean orangutans and their behavior. The use of object detection was shown to have potential for automating the tracking of especially locomotion, whilst filtering out false positives. Potential improvements regarding this tool were addressed and future implementation should take these into consideration. These improvements include using adequately diverse training material and limiting iterations to avoid overfitting the model.

Keywords: orangutans; object detection; image classification; zoological institution; artificial intelligence; automatic behavior tracking

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The orangutan (*Pongo pygmaeus*) is partly endemic to Borneo. The Bornean forests have experienced great reductions because of fires, deforestation, logging, and the establishment of plantations. Due to illegal hunting and a reduction in suitable habitats, the species is now listed as critically endangered, protected by law, and part of conservation programs all over the world [9]. Zoological associations such as DAZA, EAZA and WAZA also uphold requirements of their members such as prioritizing research to continuously uphold good welfare standards [2, 3, 1].

To properly understand the implications of orangutan behavior and welfare; it must first be understood what constitutes a healthy behavior. This is not easily done in captive animals, since the natural behavior of wild orangutans do not inherently constitute healthy behavior for the captive ones. This is both because not all wild behaviors are desirable in a captive setting, and because not all wild behaviors are reproducible in captivity [10]. Stereotyped behaviors in captive individuals may be assessed and analyzed in a behavioral study as indicators of welfare [11]. Continuous focal sampling is useful for analyzing behavior and understanding which behaviors are healthy, such as spending

much time foraging or interacting with enrichment, and which behaviors are considered pathological, such as stereotypies [14, 16].

In recent years, artificial intelligence (AI) has been a useful tool for improving video and image analysis, through applying machine learning to automatically identify and differentiate between species, individuals and/or behaviors [23, 24, 25]. Relevant parts of machine learning for video and image analysis are image classification and object detection [23, 26, 27]. Image classification may identify which orangutan is shown in a frame, based on a trained model [24, 27]. Object detection may be used for identifying and labeling certain behaviors executed by an orangutan, whilst giving information about the spatial coordinates and time (Figure 1) [24, 28].



Figure 1. Image of the female bornean orangutan at Aalborg Zoo, Denmark. The yellow label box specifies its location and labels the specific behavior (Moving forward/Locomotion) with the exact coordinates (at two corners of the rectangle).

Knowledge of the coordinates as a function of time may allow for further analysis, such as determining whether the subject is lying down, standing up, moving, etc. Creating a model capable of tracking certain behaviors would thus make behavioral analysis more effective, by allowing for automatization and thus reducing the workhours required for behavioral coding [23, 29].

The aim of this study was to investigate the applicability of machine learning techniques by automating the tracking of certain behaviors and recognizing individuals, from video material to evaluate the potential for greater efficiency in this type of research in the future.

2. Materials and Methods

2.1. Subjects and Enclosure

The behavior of two adult ~~b~~bornean orangutans in Aalborg Zoo (Denmark) was examined. The male and female orangutan were born in other European zoos in 2010 and 2013 respectively. They were transferred to Aalborg Zoo in June 2022 and immediately ~~after~~ placed in a recently renovated enclosure as the first two orangutans to inhabit the new enclosure. In August 2022, two Asian small-clawed otters (*Aonyx cinereus)* were introduced to the enclosure. The orangutans were separated in the enclosure until March 2023, after which they were introduced to each other and mostly kept together.

2.2. Data Collection

An ethogram was constructed based on a previous behavioral study of the same orangutans [30], conversations with the zookeepers, and personal observations made in Aalborg Zoo. The ethogram was modified with the intent of only studying recognizable behaviors with an object detection model (Table 1).

Table 1. Ethogram with the selected behaviors. The ethogram was modified from [30].

Behavior	Description
Locomotion	Walking, crawling, or climbing
Inactive	Stationary, e.g., sitting, lying down, or standing still
Covered Inactive	Stationary under cover, such as blankets
Foraging / Feeding	Searching for food, moving with food, or ingesting food or drinking
Out of View / Not labeled	When it is not possible to see the orangutan because they are out of the camera's view. Includes images that were not labeled by the object detection model
Other	Urinating, interaction with zookeepers etc.

The ~~Video~~ material was collected from two periods: June 12th, 2023, to June 16th, 2023, and July 17th, 2023, to July 21st, 2023, in the visitor opening hours (10 AM to 7 PM). The average temperatures and precipitation for the periods were 16.4 °C with 27.0 mm rain for June and 15.9 °C with 140.8 mm rain in July [31]. The outdoor areas were monitored by two Milesight Mini PTZ dome 2.0 MP Starlight Cameras placed outside the surrounding net. A manual ethogram was conducted where the video material was studied in GridPlayer version 0.5.3 and noted in Excel version 16.65 (22091101). This was done to allow for comparison between manual observations and an automatic object detection model. A total of 12 hours of video material was collected for this comparison.

Selected behaviors, as shown in the ethogram (Table 1), were tracked using object detection to demonstrate the capabilities of such a model in future behavioral studies. Video material provided for image classification was collected on October 10th, October 17th and October 19th with 12 MP cameras from an iPhone 11, in order to get high resolution images of the animals' bodies and faces.

To demonstrate the capabilities of machine learning, an object detection model trained to recognize selected behaviors was constructed. The model was constructed in Rectlabel by extracting frames from a 2-minute video. The 2-minute video consisted of 5-10 second individual clips, so each behavior (Locomotion, Inactive, Covered Inactive and Foraging/Feeding) was displayed ~~a total of~~ 30 seconds each, in different places and positions. Frames from the video were extracted in Rectlabel which gave a total output of 3331 image frames. Every tenth frame was labeled in Rectlabel with a specific behavior as seen in Figure 1. The resulting 334 labeled images were used to train a model in Create ML with 9000 iterations. The model constructed in Create ML was used to auto label the remaining image frames. The object detection model could automatically label 3281 images frames (98.5%). All unlabeled images frames were labeled manually. The behavior Inactive was labeled 887 times, Covered Inactive was labeled 835 times, Foraging/Feeding was labeled 814 times and Locomotion was labeled 795 times. The new model with a total of 3331 labeled image frames was trained in Create ML with 18000 iterations. This final model was used to analyze 12 hours of video material.

2.3. Data Analysis

The collected data from the manual observations and automated observations were analyzed using Excel version 2403 (Build 16.0.17425.20124). The machine learning computations were conducted in RectLabel version 2023.10.22 (2023.10.22) and Create ML version 5.0 (121.1).

Firstly, ~~it was~~ examined whether an image classification model could effectively distinguish between the male and female orangutan. The video material was divided into short videos of 5-10 seconds in QuickTime Player version 10.5 (1170.4.13.1) and

subsequently converted into image frames in RectLabel, that either maintained the background or had the background removed with Preview version 11.0 (1056.2.4). The entire methodological pipeline is shown in Figure 2.

It was furthermore examined whether an object detection model could recognize different behaviors for the female orangutan. In total, 12 hours of footage were analyzed with the object detection model, to somewhat replicate the focal sampling used in behavioral studies. To compare these results with manual observations of the same material, time budgets were made for both methods. The methodological pipeline is shown in Figure 3.

To further analyze the data from the object detection model, the coordinates, and the dimensions of the boxes were collected from the CSV-file.

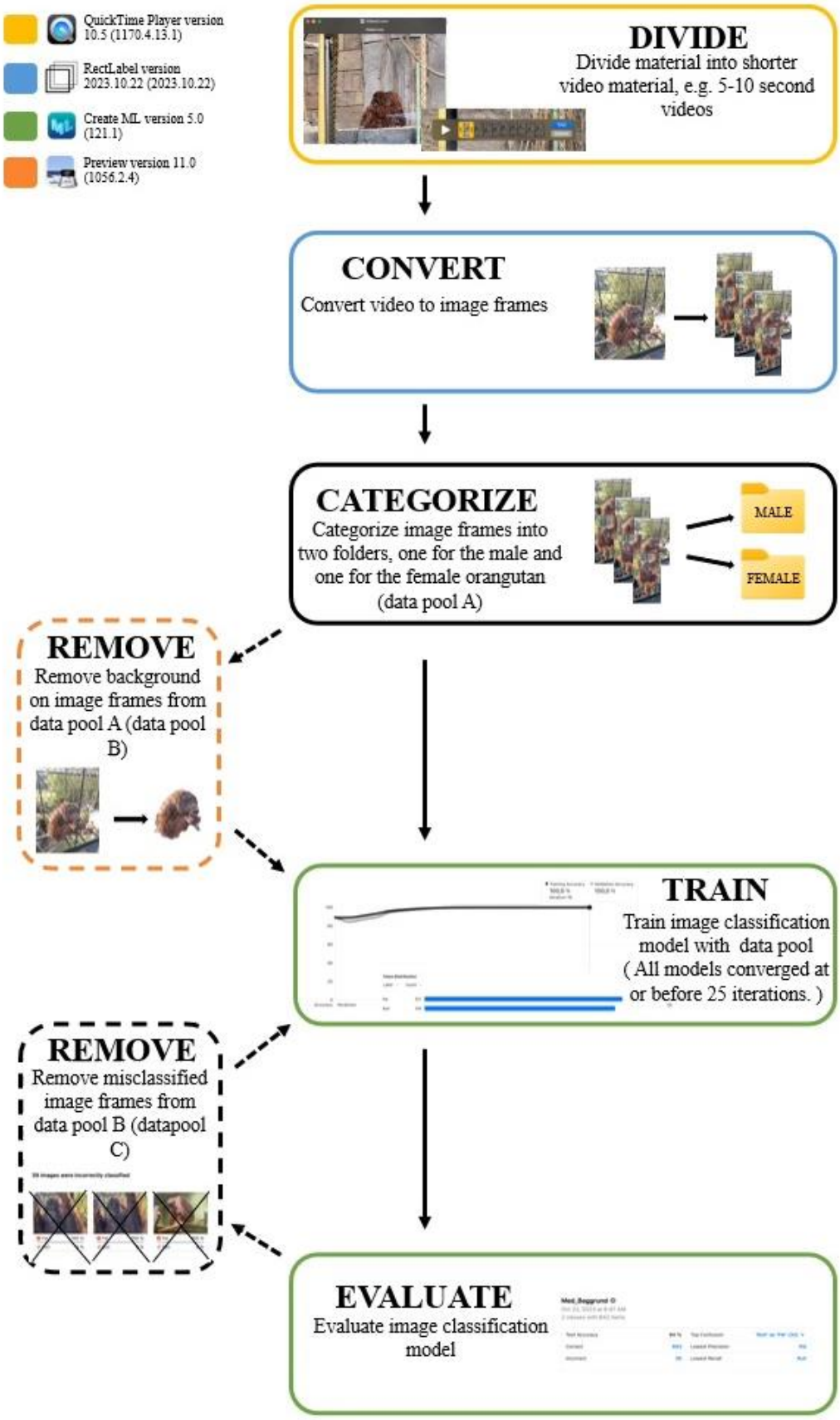


Figure 2. Pipeline showing the process of creating and evaluating an image classification model. The different colored boxes represent the software used.

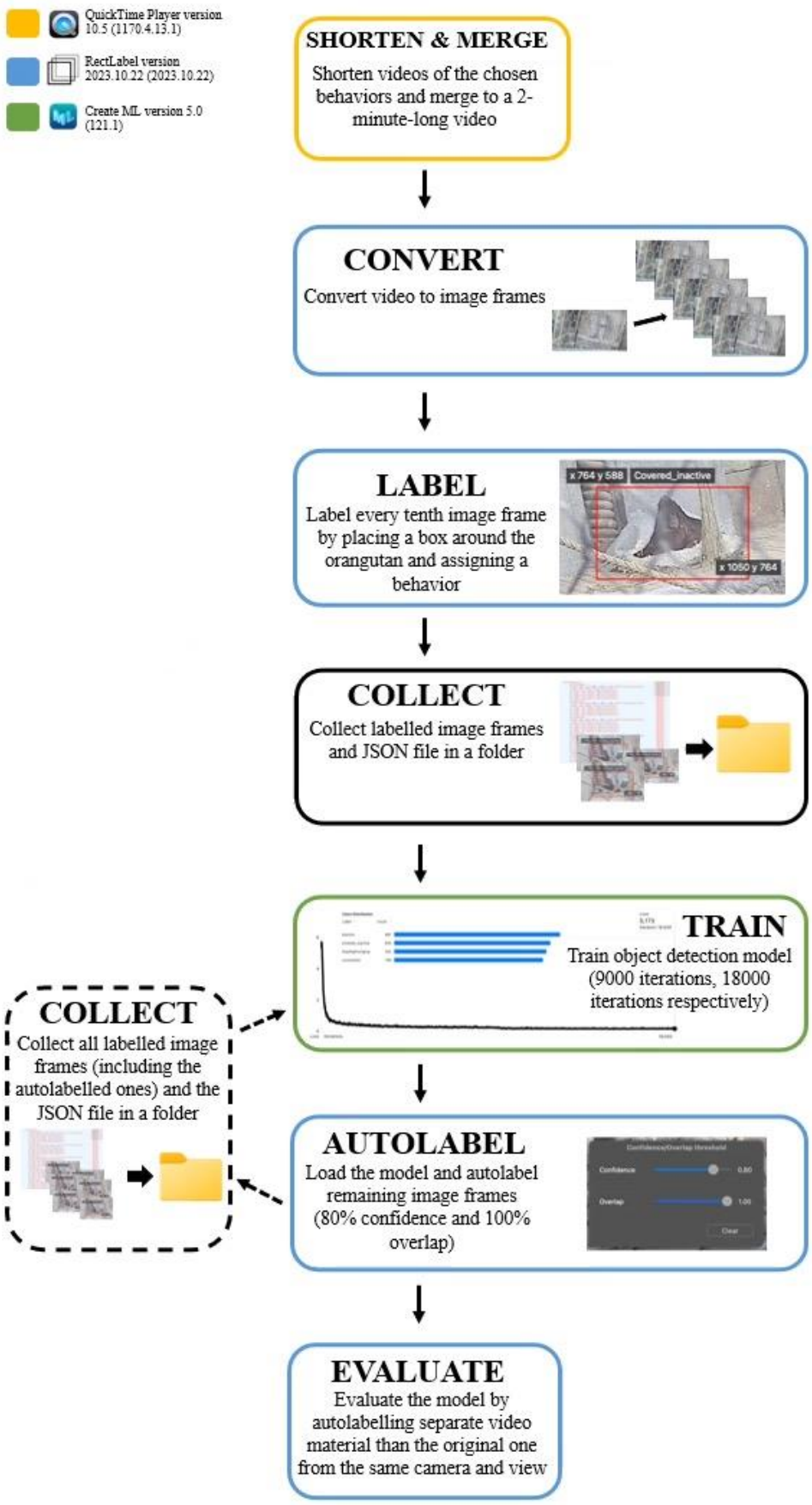

QuickTime Player version 10.5 (1170.4.13.1)
RectLabel version 2023.10.22 (2023.10.22)
Create ML version 5.0 (121.1)
SHORTEN & MERGE
Shorten videos of the chosen behaviors and merge to a 2-minute-long video
CONVERT
Convert video to image frames
LABEL
Label every tenth image frame by placing a box around the orangutan and assigning a behavior
x 764 y 588
Covered_inactive
x 1050 y 764
COLLECT
Collect labelled image frames and JSON file in a folder
TRAIN
Train object detection model (9000 iterations, 18000 iterations respectively)
COLLECT
Collect all labelled image frames (including the autolabelled ones) and the JSON file in a folder
AUTOLABEL
Load the model and autolabel remaining image frames (80% confidence and 100% overlap)
EVALUATE
Evaluate the model by autolabelling separate video material than the original one from the same camera and view

Figure 3. Pipeline showing the process of creating and evaluating an object detection model. The different colored boxes represent the software used.

3. Results

3.1. Using Image Classification to Recognize Individual Orangutans

When the image classification model was trained with data pool A, consisting of 642 images (327 and 315 images of the male and female orangutan respectively), the test accuracy was 94% when evaluated with data pool A. The image classification model trained with data pool B had a test accuracy of 95% when evaluated with data pool B. After correction the image classification model was trained with data pool C and did not improve when evaluated with data pool B. The test accuracy however increased to 99% when the corrected image classification model was evaluated with data pool C. In all cases, the female was more often misclassified as the male, and further correction failed to improve the test accuracy.

3.2. Using Object Detection to Automate Behavioral Tracking

The object detection model (Figure 4) ~~labelled~~ 13.0% of the 43483 image frames in the 12-hour video, which was approximately split into 1 frame per second. 6.9% of the image frames were labeled with 'Locomotion', 2.0% with 'Foraging/Feeding', 1.4% with 'Inactive', and 2.7% with 'Covered Inactive'. 87% of the image frames were left unlabeled due to the orangutan being out of view (41.3% of the time) or because of lack of confidence by the model. The model was set to 80% confidence. When the 12-hour video material was observed manually, 3.6% was noted as 'Locomotion', 18.1% with 'Foraging/Feeding', 13.9% with 'Inactive', 15.6% with 'Covered Inactive', and 7.5% of the behavior was categorized as 'Other'. The orangutan was out of view 41.3% of the time.

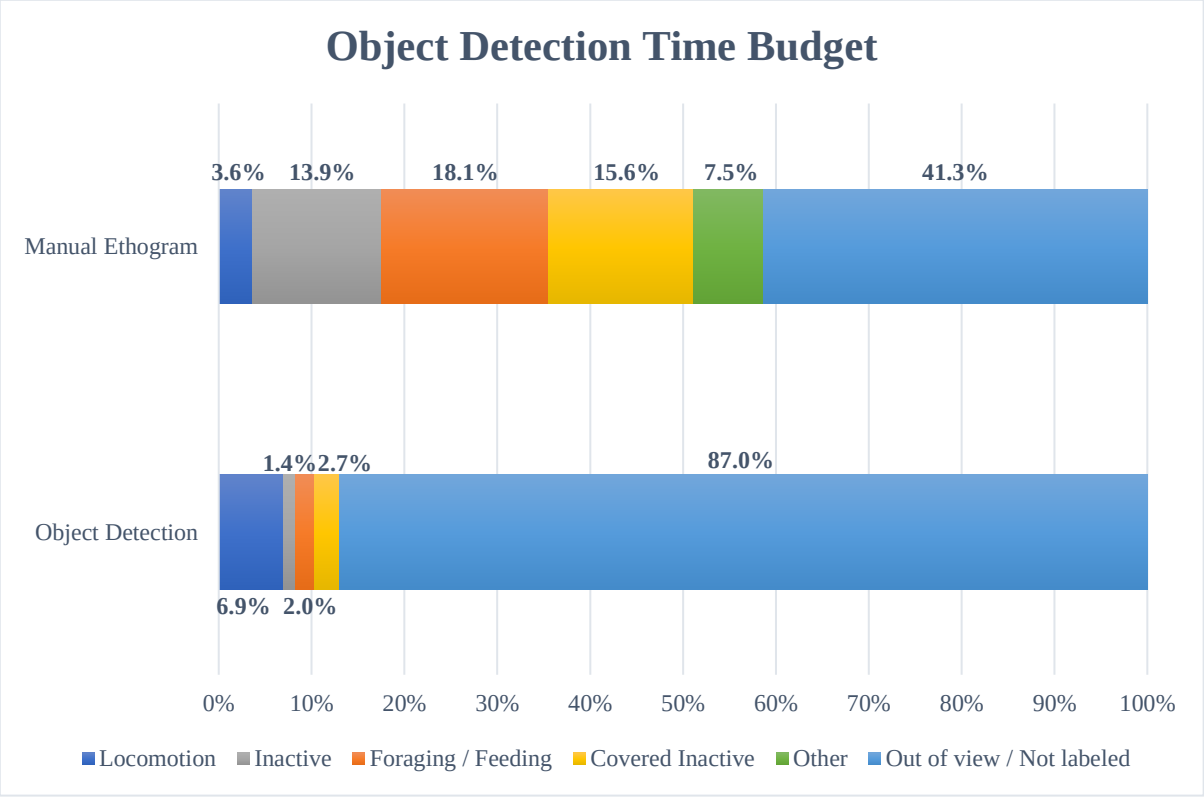


Figure 4. Time budget for the 12-hour video when observed manually and with the object detection model. Each of the behaviors are represented by a different color and show percentage of how long the female orangutan does the behavior.

3.3. Correcting Object Detection of Locomotion

It appears from the object detection time budget (Figure 4) that locomotion has almost double the value when using the object detection model as it does in the manual ethogram. This is found to potentially be caused by false positives from the model. To filter out most of the false positives of locomotion, the coordinates of the CSV file were utilized. The coordinates for each frame the model labeled as 'Locomotion' were plotted as shown below (Figure 5).

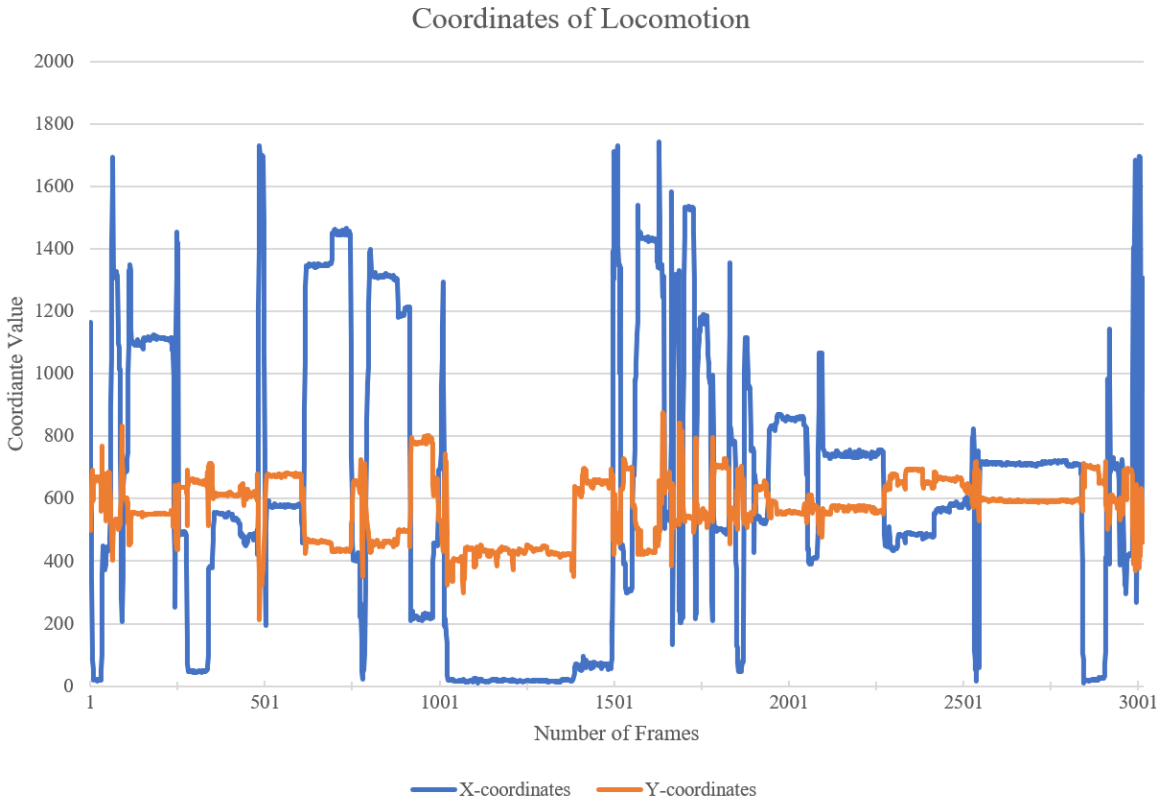


Figure 5. Plot of coordinates of the bottom corner of the label box. The blue line shows x-coordinates, and the orange line shows y-coordinates.

The frames where movement did not occur, displayed by changes in both the x-coordinate and the y-coordinate were labeled as false positives, instead of 'Locomotion'. Thus, the plateaus throughout Figure 5 were removed manually. The renewed time budget including false positives for locomotion can be seen below (Figure 6). It appears that 3.4% of the frames labeled 'Locomotion' by the model were false positives and 3.5% of the frames were left as 'Locomotion'.

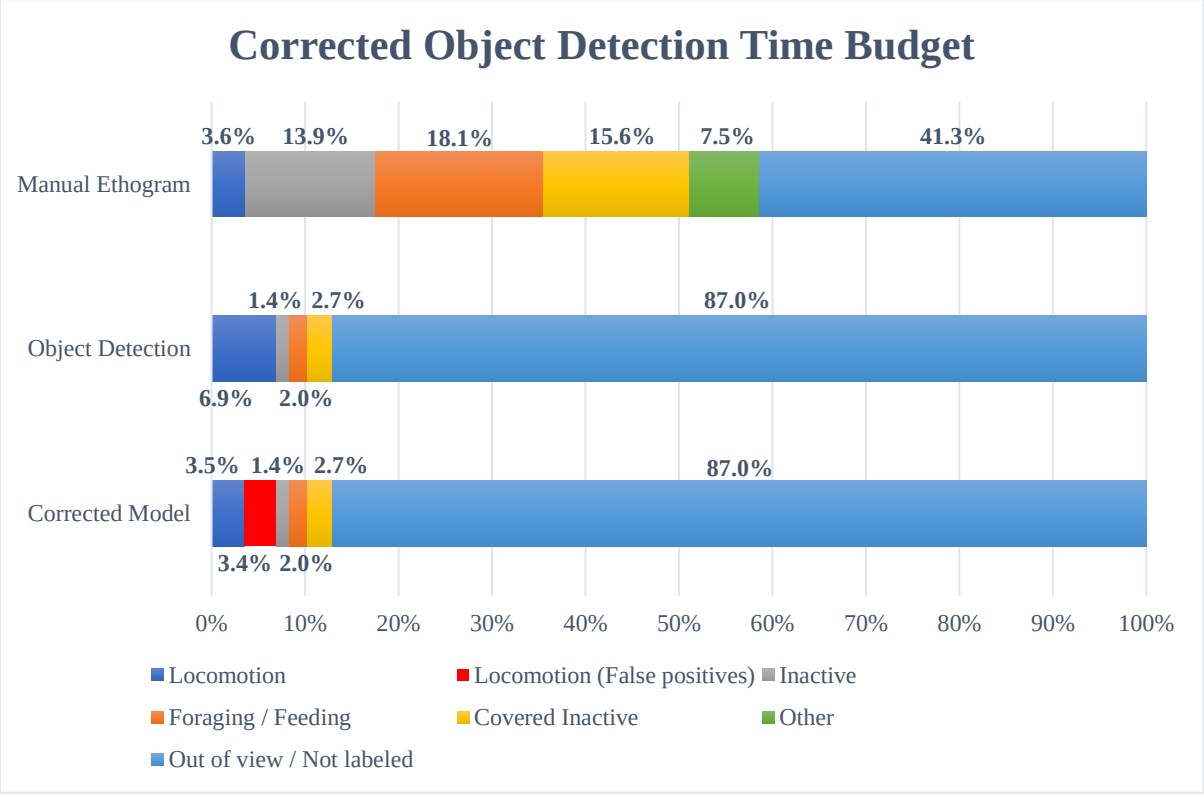


Figure 6. Time budget for the 12-hour video when observed manually and with the object detection model, as well as data corrected for false positives in 'Locomotion' (shown in red). Each of the behaviors are represented by a different color and show percentage of how long the female orangutan does the behavior.

4. Discussion

4.1. Subject Recognition with Image Classification

The image classification model with the best test accuracy in this study had a test accuracy of 99%. However, the test accuracy decreased when the evaluation data did not equal the training data which for practical use would often be the case. The model should be trained with a larger data pool containing pictures of the male and female orangutan from different angles and in different lighting ~~to improve the test accuracy~~. The models which were trained with data pools without background (data pool B and C) ~~did~~ more effectively distinguish the male and female orangutan, but an improved model trained with images with background would probably be preferable t[illegible]mplement. Additionally, it would minimize the work effort in the training and evaluation process. Occasionally it can be challenging for the model to identify the subject on the images, and therefore an object detection model where the subjects are manually encapsulated by boxes could possibly be more effective at distinguishing the orangutans.

4.2. Potential for Utilizing Label Box Dimensions to Determine Behavior

Besides using the coordinates of the label box, created by the object detection model, for correcting locomotion, there is a variety of potential uses for the information given. An example of this is using the height/width ratio of the label box to determine simple behaviors, such as the orangutan raising its arms or lying down. This is achieved through observing frames of these behaviors and setting limits that determine when a label box' ratio exceeds these, and thus communicates the presence of that behavior. This is exemplified

below (Figure 7), where the height/width ratio of 100 frames are shown, and limits for lying down and raising arms are implemented, to visualize when these behaviors happen throughout the frames.

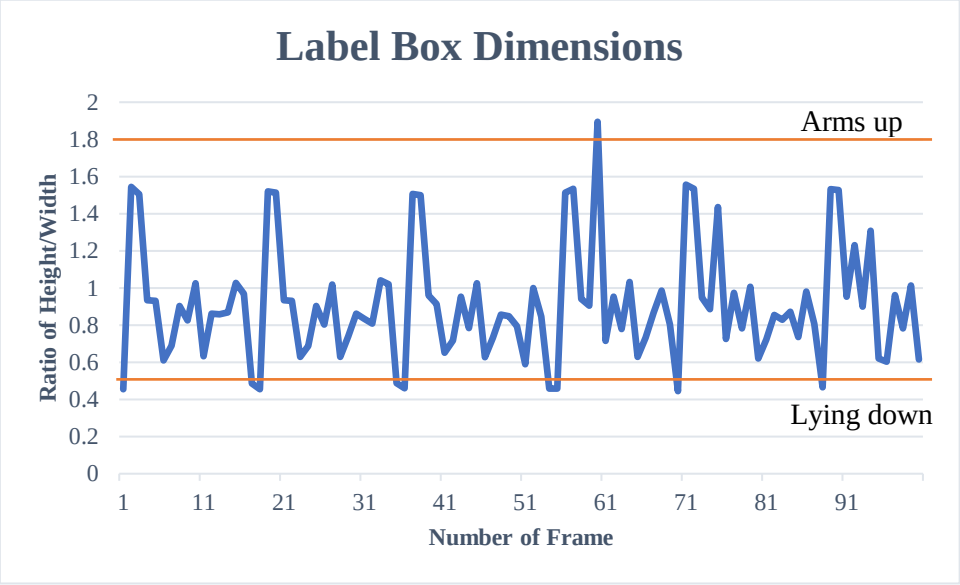


Figure 7. The height/width ratio of label boxes over 100 frames of footage. The upper line shows the limit of the ratio, where the behavior can be considered 'Arms up'. The lower line shows the lower limit of the ratio, where the frame is considered 'Lying down'.

4.3. Behavior Tracking with Object Detection

The object detection model constructed to recognize the orangutan behaviors ('Locomotion', 'Foraging/Feeding', 'Inactive', and 'Covered Inactive') could autolabel 98.5% of the training material. Again, the success rate was greater when the evaluation material equaled the training material but for practical use this is roughly never the case, and the model should be trained with more material. The extended object detection model did not label 87.0% of the image frames which indicates that the model found it difficult to recognize the four behaviors. For the behaviors 'Foraging/Feeding', 'Inactive' and 'Covered Inactive', the percentages vary when comparing the results obtained manually and from the model. For the behavior 'Locomotion', approximately double the percentage was obtained when using the model. This value was therefore corrected to exclude the number of false positives. Such corrections may be necessary to fully evaluate the model, and may be implemented in a multitude of ways, such as using the label box' height and width to determine occurrence and false positives for other behaviors. These false positives might be caused by the box-labelling of the behaviors, which leads to the model mistaking a certain background used in the training, as recognition of that behavior. A solution might be to label the subject with polygons rather than boxes to better encapsulate the subject and minimize the influence from the background. Another issue may be a lack of diversity in the training material, which may lead to overfitting of the model. This might also be exacerbated by using too many iterations when constructing the model. The extended object detection model showed the greatest potential for optimization when tracking 'Locomotion'. Contrastingly, tracking the other behaviors of 'Foraging/Feeding', 'Inactive', and 'Covered Inactive' might be harder due to the lack of uniformity. Another challenge emerges in the autolabeling process when two or more different behaviors overlap, e.g. when the orangutan feeds and performs locomotion simultaneously. Potentially the model could be trained with image frames with overlapping behaviors, and the behavior of interest can be labeled on these frames to overcome this challenge. An object detection model however can be useful for tracking certain behaviors or tendencies. Further work

is necessary to assess the limitations and challenges of an object detection model in automatization and optimization of behavioral studies.

The continued development of such models may prove a useful tool in the future for researchers and zookeepers alike. This tool may be used to quickly assess the general behavior of an animal without requiring much effort. For this reason, studying problems and possible improvements whilst constructing such models is an important and necessary first step, that provides insight into the process and helps avoid similar mistakes in the future.

5. Conclusion

This study demonstrated that the automatization of behavioral analysis with machine learning, using Image Classification to recognize individual Bornean orangutans, was possible and reliable. Tracking behaviors with Object Detection showed potential and was somewhat capable of labeling the behavior 'Locomotion'. However, the Object Detection model needs further optimization.

Author Contributions: Conceptualization, F.G., M.G.H., E.J.N., J.N. T.H.J. and C.P.; methodology, F.G., M.G.H., E.J.N., J.N. and C.P; validation, F.G., M.G.H., E.J.N., and J.N.; formal analysis, F.G., M.G.H., E.J.N., and J.N.; investigation, F.G., M.G.H., E.J.N., and J.N.; data curation, F.G., M.G.H., E.J.N., and J.N.; writing—original draft preparation, F.G., M.G.H., E.J.N., and J.N.; writing—review and editing, F.G., M.G.H., E.J.N., J.N, A.K.O.A., J.O.P-G., T.H.J and C.P.; visualization, F.G., M.G.H., E.J.N., and J.N.; supervision, A.K.O.A, T.H.J., J.O.P-G. and C.P. All authors have read and agreed to the published version of the manuscript.

Funding: Aalborg Zoo Conservation Foundation (AZCF grant number 04-2023).

Institutional Review Board Statement: Ethical review and approval were waived for this study as behavior observations of the animals were the only procedure included.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to thank the staff at Aalborg Zoo for allowing us to conduct this study. Thank you to Kenneth Madsen, Zookeeper at Aalborg Zoo tending the orangutans, for discussing the aim of the paper with us. Thank you, Kasper Kystol Andersen, Technician at Aalborg Zoo, for helping us with accessing and using the video material.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. M. Ancrenaz, M. Gumal, A. J. Marshall, E. Meijaard, S. A. Wich and S. Husson, "Pongo pygmaeus, Bornean Orangutan (errata version published in 2018)," The IUCN Red List of Threatened Species 2016: e.T17975A123809220, 2016.
2. DAZA, "Etiske retningslinjer," Danish Association of Zoos and Aquaria, DAZA, 2022.
3. EAZA, "Standards for the Accommodation and Care of Animals in Zoos and Aquaria," European Association of Zoos and Aquaria, EAZA, 2014.
4. R. Barongi, J. P. Bonner, P. Boyle, S. Burns, O. Byers, G. Dick, L. Dickie, F. A. Fisken, S. Gendron, J. Gray, M. Gusset, H. Hofer, S. Hunt, S. Luz, M. Parker and P. Riger, "Committing to Conservation: The World Zoo and Aquarium Conservation Strategy," World Association of Zoos and Aquariums (WAZA), Gland, Switzerland, 2015.
5. S. Wolfensohn, J. Shotton, H. Bowley, S. Davies, S. Thompson and W. S. M. Justice, "Assessment of Welfare in Zoo Animals: Towards Optimum Quality of Life," *Animals*, vol. 8, no. 7, 4 July 2018.
6. G. J. Mason, "Stereotypies and suffering," *Behavioral Processes*, vol. 25, no. 2-3, pp. 103-115, 23 May 1991.
7. C. K. Lutz and M. A. Novak, "Environmental Enrichment for Nonhuman Primates: Theory and Application," *ILAR journal*, vol. 46, no. 2, pp. 178-191, 1 April 2005.
8. C. Y. Ting, "Orangutan behaviour in captivity: activity budgets, enclosure use and the visitor effect," MSc thesis, Singapore, 2011.

9. Z.-H. Zhou, "Application Status," in *Machine Learning*, Singapore, Springer Nature, 2016, pp. 15-18.
10. Z.-Q. Zhao, P. Zheng, S.-T. Xu and X. Wu, "Object Detection With Deep Learning: A Review," *IEEE Transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 1-21, 11 November 2019.
11. M. Zuerl, P. Stoll, I. Brehm, R. Raab, D. Zanca, S. Kabri, J. Happold, H. Nille, K. Prechtel, S. Wuensch, M. Krause, S. Seegerer, L. von Fersen and B. Eskofier, "Automated Video-Based Analysis Framework for Behavior Monitoring of Individual Animals in Zoos Using Deep Learning - A Study on Polar Bears," *Animals*, vol. 12, no. 6, 10 March 2022.
12. Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 27 May 2015.
13. P. Povlsen, "Wildlife monitoring using drones and artificial intelligence," MSc Thesis, Aalborg, 2023.
14. F. Sakib and D. T. Burghardt, *Visual Recognition of Great Ape Behaviours in the Wild, ArXiv*, 2020.
15. L. Wang and H. Liu, "Behaviour Recognition Method Based on Object Detection for Power Operating Series," in *Proceedings of 2023 Chinese Intelligent Automation Conference*, Singapore, Springer Nature, 2023, pp. 484-499.
16. J. F. Larsen, K. K. D. Andersen, J. Cuprys, T. B. Fosgaard, J. H. Jacobsen, D. Krysztofiak, S. [illegible]. Lund, B. Nielsen, M. E. B. Pedersen, M. J. Pedersen , A. Trige-Esbensen, E. M. Walther, C. Pertoldi, T. H. Jensen, A. K. O. Alstrup and J. O. Perea-García, "Behavioral analysis of captive male Bornean orangutan (Pongo pygmaeus) when exposed to environmental changes," *Archives of Biological Sciences*, vol. 74, no. 4, pp. 443-458, 16 November 2023.
17. F. Rubek, "Sammendrag af sommer 2023," Danmarks Meterologiske Institut, 2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual authors and contributors and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.