Automating Recognition of Wild Boar Behaviors in Video Footage

Khaled Awad[1], Lior Koren[1], Achiad Davidson[1], Dan Malkinson[1[0000−0001−6172−0253]], and Ilan Shimshoni[1[0000−0002−5276−0242]]

University of Haifa, Haifa 31905, Israel

1 Introduction

Understanding the behavior of wild boars is crucial for various reasons, including conservation efforts, wildlife management, and mitigating potential conflicts with humans [14]. These resilient and adaptable animals are widely distributed across different habitats, exhibiting complex social structures and intriguing behavioral patterns. By delving into the research on wild boar behavior, scientists and conservationists can gain valuable insights into their ecological roles, reproductive strategies, foraging habits, and responses to changing environmental conditions [6]. This knowledge serves as a foundation for developing effective management strategies, minimizing human-wildlife conflicts, and ensuring the long-term viability of both wild boar populations and their ecosystems. The study of animal behavior has long been a challenging and time-consuming endeavor, requiring extensive field observations and manual data analysis. However, recent advancements in artificial intelligence (AI) have opened up new avenues for recognizing and understanding the behavior of wild boars. AI-powered technologies, such as computer vision and machine learning algorithms, offer promising opportunities to automate the process of behavioral recognition, providing researchers with faster and more accurate insights into the complex behaviors of these fascinating animals. By harnessing the power of AI, scientists can analyze vast amounts of data collected through cameras and sensors, enabling them to uncover patterns, social interactions, and behavioral responses that were previously difficult to observe.

In the extant literature, there is a notable paucity of research focusing on the recognition of behaviors in wild boars, a gap that stands in stark contrast to the more developed studies on pigs[15],[10] and [7] , as well as research in human activity recognition [13], [9] and [11]. This study introduces a novel model that automates the process of recognizing specific behaviors in wild boars within video footage.

In analyzing activities and behaviors currently the main framework consists of analyzing raw image data. Usually, each frame is encoded by some representation and then the activity representation is recovered from these representations. This is an end-to-end procedure. This framework yields high quality results but requires a relatively large training set and in most cases is not explainable.

In our algorithm, we exploit the fact that we are able to use tools which can detect and track animals automatically. For each frame there are tools that

are able to extract the animal's articulated pose. Thus, in each frame a set of landmarks of the boar's anatomy are extracted.

In the second phase, the pose data serves as input for a feature extraction process. Here, we construct a set of feature vectors by using sliding windows across each set of 11 frames and calculate our features using the output landmarks from the pose algorithm, which encapsulate critical aspects of the boar's posture and movements. These feature vectors are then used in the training of a Random Forest classifier, which classifies the sliding window as belonging to one of several behavior classes. In our case the behaviors are: walking, eating, vigilance, and no action, when the boar is not seen in the video. The framewise accuracy results are quite promising but the results are fragmented, i.e., from time to time a sliding window is misclassified. This yields a low segmental F1 score of the video to behavior segments [8]. We overcome this problem by a applying a novel method based on a decision tree of the classification results. Thus, the relatively rare classification errors are eliminated, improving considerably the segmentation accuracy.

The method presented here was applied to videos of wild boars but could be easily applied to species and other behaviors. Algorithms of this type can be very useful for ecological study of animal behavior and for monitoring their behavior in the wild.

2 Dataset

The video content in our dataset was captured using stationary cameras strategically placed in various locations and towns throughout Israel, including but not limited to Beit Oren, Ramat Hanadiv, Haifa, and other geographical areas. Subsequently, certain preprocessing steps were applied, such as ensuring that each video contains at least one behavior and verifying that the video quality is sufficient for use with algorithms like pose estimation. From 52 selected videos we extracted a total of 13,266 frames derived. Among these frames, a subset of 570 frames contained no action, indicating instances where no wild boar was detected within the frames. Furthermore, we identified 2,841 frames illustrating vigilance behavior, 5,704 frames capturing eating behavior, and 4,151 frames depicting walking behavior. Examples of them can be seen in Figure 1.

((a)) Vigilance ((b)) Eating ((c)) Walking

Fig. 1: Imaging displaying the wild boar with predefined behaviors.

3 Proposed Model

In this study, we introduce a novel model designed to identify specific behaviors of wild boars from video footage. This model is engineered to discern wild boar behaviors in videos, categorizing them into predefined labels such as walking, eating, vigilance, and a 'no action' label for frames that do not exhibit any of the aforementioned behaviors. The architecture of our model is underpinned by the integration of multiple computer vision algorithms, particularly during the feature extraction and classification stages. A model diagram is displayed in Figure 2.

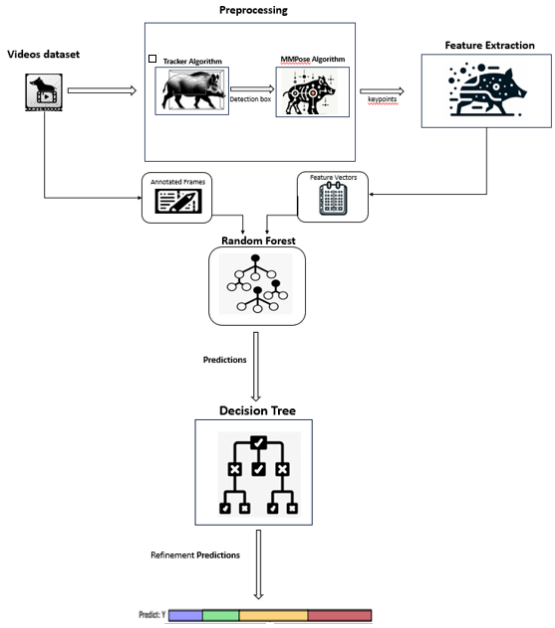


Fig. 2: proposed model component

We have compiled a dataset by accessing 52 videos from an internal repository established and maintained by Achiad Davidson and Dan Malkinson who specialize in the field of animal behavior recognition [5,3,4]. The frames of the videos where classified by them as belonging to one of the four classes mentioned above.

The preprocessing stage includes detection [12], tracking [1] and articulated pose estimation (using MMPose [2]) steps, which yields for each frame a set of 20 body 2D landmarks for each boar. In the second step a sliding window classification algorithm is applied. For each sliding window of a specific length (11 in our case), we compute a set of 20 body 2D landmarks for each boar. Initially, we identified key body parts of the boar, we used most of the generated key points

(15 of 20) to create our feature vectors, namely the nose, feet, elbows, knees, ears, eyes, chin and nose. To determine whether the boar remained stationary or changed its location, we calculated the standard deviation of all points for the x and y coordinates, creating a total of 30 features. We also calculated the distance between the first and last frame of the 11 frames, which generated another 30 features. Furthermore, we introduced additional features, specifically the differences between the y-coordinates of key points on the boar's face and the y-coordinates of the boar's paws, this process resulted in additional 20 features (80 overall in our experiments).

These features should be indicative of the targeted behavioral patterns under investigation. For instance, to accurately interpret instances where a boar exhibits eating behavior, it is imperative to identify sequences within the video data, where the boar's head is oriented downwards towards the ground. Additionally, the lack of movement of the boar's legs during such instances serves as a critical corroborative feature for this behavior. Moreover, in the assessment of vigilance behavior, it is essential to observe not only the animal's immobilization but also the specific posture where its head is elevated and remains static. We also have to be able to estimate the boars motion.

Thus, the features we compute, measure the average distance between different landmarks of a frame and for a specific landmark we measure the distance between the first and last landmark in the sub-window. For instance, an analysis of foot landmark behavior examining the differences between the first and last frames can reveal specific patterns of movement or inactivity, as illustrated in Figure 3.

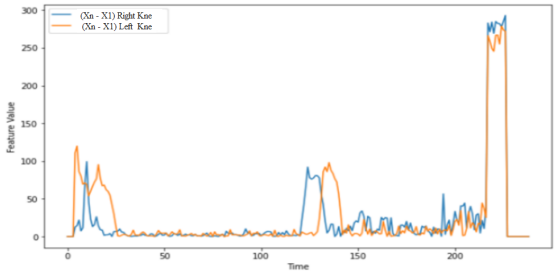
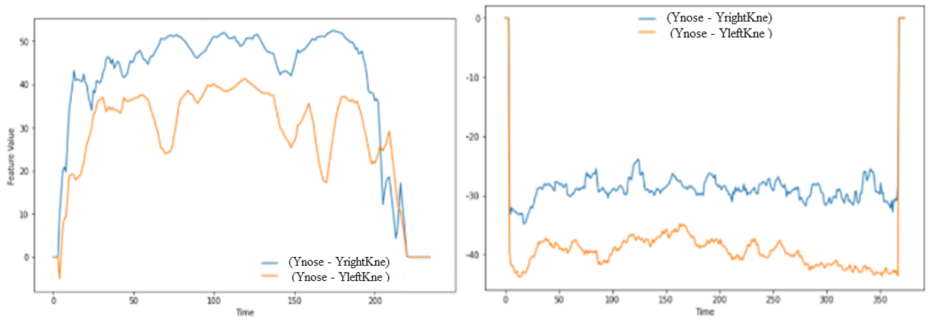


Fig. 3: An instance for boar movement in a video, showing the difference in position of a landmark between the first and last frame

In the referenced video, observations of a wild boar transitioning from walking to stopping for a brief eating period before resuming movement are shown in the figure, where initial high values diminish to near zero before increasing once more.

In Figure 4, a sequence of frames illustrates the behavior of a wild boar, showcasing vigilance in (a) and eating in (b). This is discernible through the



(a) Vigilance (b) Eating

Fig. 4: (a) The wild boar, exhibiting vigilance, holds its head high. (b) The wild boar lowers its head to eat.

analysis of specific features, namely, the differences in the y-coordinate of the nose or chin relative to the y-coordinate of the boar's legs.

For the evaluative component of our study, an exhaustive cross-validation scheme was employed, leveraging the Leave-One-Out (LOO) cross validation method, to assess the efficacy of the chosen Random Forest classifier. Here the algorithm was trained on $N - 1$ videos and tested on the remaining video. In our case the dataset consists of 52 videos.

Employing the Random Forest algorithm to our dataset observed a quite high accuracy levels; where it exceeds the 80%. Analyzing the results, there are two types of errors. The first are classification errors, where specific segments were misclassified. This problem can be dealt with by increasing the training set. The second problem is fragmentation. While the large majority of the frames in a segment were correctly classified from time to time a small number of sliding windows were incorrectly classified. This phenomenon is natural, since each sub-window is independently classified and thus, from time to time a sub-window can be mis-classified. This phenomenon has a small effect on the accuracy but has a severe effect on the segmental F1 score. In order to compute this score, a segment is considered true if the IOU between the detected and the ground truth segments is above a certain threshold. This score is used to evaluate the performance of action recognition algorithms. We tested it at several IOU thresholds 10%, 25% and 50%, which are denoted by F1@10, 25, 50.

For instance, Figure 5 illustrates an instance within a continuous segment of vigilance behavior, where certain frames were erroneously classified as eating by the model. It is implausible for the boar to exhibit two or more distinct behavioral segments within a span of 50 frames, equating to only a few seconds, which indicates that these mis-classifications are noise errors that the model was unable to resolve effectively.

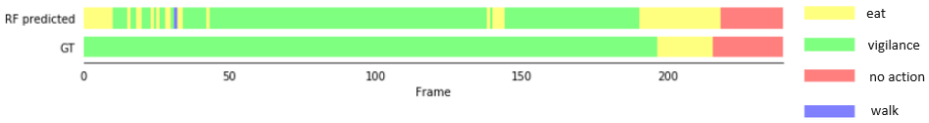


Fig. 5: Fragmentation errors within behavior segments being analyzed by the model

3.1 Addressing Fragmentation Errors.

The problem of fragmentation is a natural consequence of the fact that each sub-window is classified independently. This could be addressed by using a more complex classification method. We however, decided to address this issue in a simple way. We thus added an extra stage to our model, which is based on adapting a Decision Tree algorithm approach. In this stage, a simple decision tree was constructed whose input for frame number i is $x_i = i$ and $y_i = RF_i$ the classification result of the Random Forest classifier. If the maximal depth of the tree is h there are at most 2^h leaves in the tree. Thus, the video is divided into at most 2^h segments and all the frames belonging to a segment (leaf) are classified according to the majority class. Since in nearly all of the videos the number of ground truth segments is not more than four, the tree depth of 2 was chosen. For cases where there were more than four ground truth segments, this step could not improve the results. Naturally, this step deals with fragmentation and not with mis-classified segments.

Leveraging these two steps enabled us to develop a fully automated, novel model capable of accurately identifying behaviors of wild boars within video footage. Through these enhancements, we achieved accuracy rates approximately between 81% to 84%, alongside F1 scores of 73.24, 71.83, and 60.56 at the respective thresholds of 10, 25, and 50. This is in comparison the F1 scores 24.0, 22.14, and 18.05, which were achieved when only the RF algorithm was run.

4 Results

Our algorithm underwent a training and validation process with a dataset that included 52 videos, summing up to 13,266 frames. Each frame was categorized into one of four possible behaviors: 'Vigilance', 'Walk', 'Eat', and 'No action'. To measure the performance of our algorithm, we employed two main metrics: accuracy and the F1 score. We also used a confusion matrix to provide a detailed view of the algorithm's classification accuracy.

The results show that our RF algorithm achieves an overall accuracy of 82%. The F1 scores for different thresholds—10, 25, and 50—were 24.0, 22.14, and 18.047, respectively. These findings suggest that our model is quite adept at predicting the correct behaviors in most frame windows. The confusion matrix in Figure 6 shows that the Random Forest classifier demonstrates a proficient

capability in correctly classifying the majority of instances across all categories. Better results could be obtained by larger training sets.

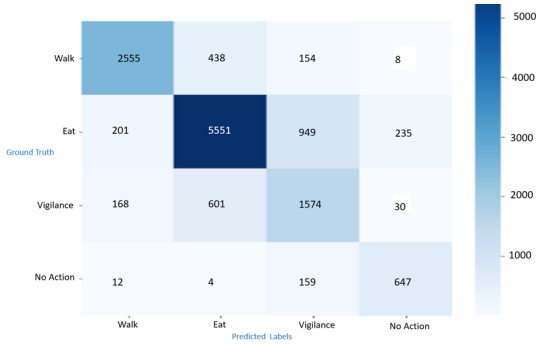


Fig. 6: Random Forest - Testing Confusion Matrix

In several instances, we encountered fragmentation and noise within segments of continuous behavior, resulting in a lower F1 score. For instance, by examining Figure 7, it is clear that within just a few seconds and across a small number of frames, there was significant fragmentation, indicating that the boar was rapidly switching between eating, walking, and vigilance actions, this rapid switching is illogical. We thus, added the decision tree step to reduce the fragmentation of the class segments.

As displayed in Figure 7 employing this strategy aimed at reducing fragmentation and the noise in behavior segments, also enhances the continuity and accuracy of the recognized behaviors. Still, in the beginning of the video the classifier mis-classifies several frames as eating instead of vigilance.

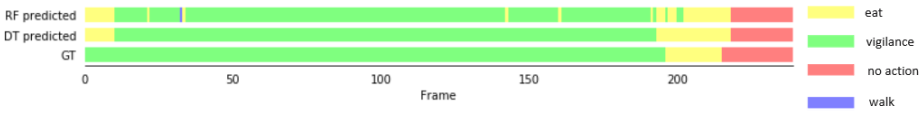


Fig. 7: Comparison graph for the predicted labels of the Random Forest and the Decision Tree with the ground truth labels.

In another example shown in Figure 8, the efficacy of the Decision Tree stage in error rectification, particularly in the initial segment of the video sequence, is markedly evident. This graph illustrates a scenario wherein the wild boar exhibits vigilance behavior denoted as first segment followed by a transition to walking until it exits the frame. The Decision Tree model demonstrates a pronounced reduction in predictive errors during the initial vigilance phase, as compared to the Random Forest model.

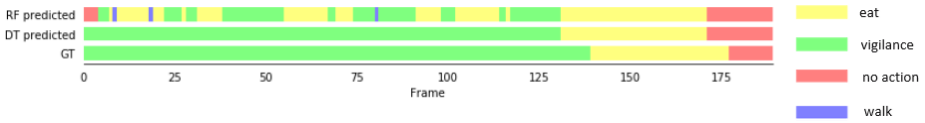


Fig. 8: Comparison graph for the Decision Tree correction.

	F1@10	F1@25	F1@50	Test Acc
Random Forest	24.0	22.14	18.047	0.818
Decision Tree	73.427	72.727	60.839	0.836

Table 1: F1 Score and testing accuracy results for Random Forest and Decision Tree algorithms

As displayed in Table 1, the integration of this approach based on a Decision Tree notably enhanced the accuracy metric; however, the most significant impact was observed in the improvement of the F1 Score across various thresholds. This enhancement represents a substantial shift in model performance, with the Decision Tree algorithm achieving an improved accuracy rate of 0.836. Furthermore, there is a remarkable improvement of the F1 scores. The scores at distinct thresholds—10, 25, and 50—registered are 73.43, 72.73, and 60.84, respectively.

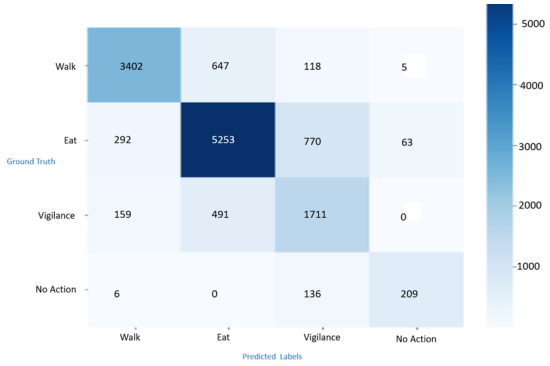


Fig. 9: Decision Tree testing confusion matrix

The results in confusion matrix in Figure 9 suggest that the added refinement stage provides competitive, if not superior, performance in certain aspects. This improvement in the results was due to the Decision Tree's ability to eliminate errors and reduce noise in the segments, as well as to prevent fragmentation.

4.1 Limitations

As mentioned above, the method we suggested for dealing with fragmentation deals with the problem effectively. Still, in cases where there is a large misclassified segment, it is not able to overcome this problem. Consider for example the video whose graph is shown in Figure 10. In that video there is a large vigilance segment, which is misclassified as walking and eating. Applying the method to this video yields a large segment of walking instead of vigilance. This problem can be dealt with by a larger more versatile training set.

Fig. 10: Miss classifying segments between 'Vigilance' and 'Eating' in both the Decision Tree and the Random Forest.

In our experiments, we decided to use a decision tree of depth 2, with at most 4 possible segments. There are however several videos with more segments. Consider for example Figure 11. In this video there exist 5 segments. Applying the DT with depth 2 naturally yields only 4 segments, while when using a depth 3 DT, the video is correctly segmented into 5 segments. It is of course possible to use more than one DT and choose between the results the best result.

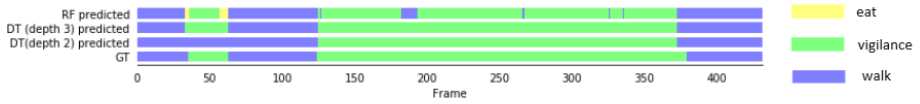


Fig. 11: Comparison between the performance of DT (depth 3) and DT (depth 2) algorithms.

Our approach is global, and requires as input the maximal number of segments. This could be addressed by local methods such as Windowed Smoothing or the median filter. Another approach is using Hidden Markov Models (HMMs), which are designed to work with sequences. HMMs smooth the transitions between hidden states because they take into account that behaviors are often dependent on previous states.

References

1. Bergmann, P., Meinhardt, T., Leal-Taixé, L.: Tracking without bells and whistles. In: The IEEE International Conference on Computer Vision (ICCV) (10 2019)

2. Contributors, M.: Openmmlab pose estimation toolbox and benchmark. https://github.com/open-mmlab/mmpose (2020)
3. Davidson, A., Malkinson, D., Schonblum, A., Koren, L., Shanas, U.: Do boars compensate for hunting with higher reproductive hormones? Conservation Physiology **9**(1), coab068 (2021)
4. Davidson, A., Malkinson, D., Shanas, U.: Wild boar foraging and risk perception—variation among urban, natural, and agricultural areas. Journal of Mammalogy **103**(4), 945–955 (2022)
5. Davidson, A., Shanas, U., Malkinson, D.: Age-and sex-dependent vigilance behaviour modifies social structure of hunted wild boar populations. Wildlife Research **49**(4), 303–313 (2021)
6. Genov, P.V., Focardi, S., Morimando, F., Scillitani, L., Ahmed, A.: Ecological impact of wild boar in natural ecosystems. In: Melletti, M., Meijaard, E. (eds.) Ecology, Conservation and Management of Wild Pigs and Peccaries. pp. 404–419. Cambridge University Press (2017)
7. Kashiha, M., Bahr, C., Ott, S., Moons, C., Niewold, T., Odberg, F., Berckmans, D.: Automatic identification of marked pigs in a pen using image pattern recognition. Computers and Electronics in Agriculture **93**, 111–120 (2013)
8. Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D.: Temporal convolutional networks for action segmentation and detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
9. Liu, J., Luo, J., Shah, M.: Recognizing realistic actions from videos "in the wild". In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1996–2003. IEEE (2009)
10. Nasirahmadi, A., Richter, U., Hensel, O., Edwards, S., Sturm, B.: Using machine vision for investigation of changes in pig group lying patterns. Computers and Electronics in Agriculture **119**, 184–190 (2015)
11. Reddy, K.K., Shah, M.: Recognizing 50 human action categories of web videos. Machine Vision and Applications **24**(5), 971–981 (2013)
12. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)
13. Soomro, K., Zamir, A.R., Shah, M.: Ucf101: A dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:1212.0402 (2012)
14. Stillfried, M., Fickel, J., Börner, K., Wittstatt, U., Heddergott, M., Ortmann, S., et al.: Secrets of success in a landscape of fear: urban wild boar adjust risk perception and tolerate disturbance. Frontiers in Ecology and Evolution **5**, 157 (2017). https://doi.org/10.3389/fevo.2017.00157
15. Yang, Q., Xiao, D., Lin, S.: Feeding behavior recognition for group-housed pigs with the faster r-cnn. Comput. Electron. Agric. **155**, 453–460 (2018), https://api.semanticscholar.org/CorpusID:53744668