JOURNAL OF
EMERGING INVESTIGATORS

Article

SmartZoo: A Deep Learning Framework for an IoT Platform in Animal Care

Siheon Ji[1], Sangyoon Bae[2]
[1]Mount Pisgah Christian School, Johns Creek, GA
[2]Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, South Korea

SUMMARY

Zoos offer educational and scientific advantages but face high maintenance costs and challenges in animal care due to diverse species' habits. Challenges include tracking animals, detecting illnesses, and creating suitable habitats. Despite the potential benefits, data-driven approaches like those in digital agriculture are rarely used in zoos due to cost and technical limitations. We developed a deep learning framework called SmartZoo to address these issues and enable efficient animal monitoring, condition alerts, and data aggregation. Animal movement data was treated as a time sequence; the time sequence was predicted using a transformer-based model; and the range of the cage was predicted by plotting the predicted animal movement based on the generated time sequence. We used K-means clustering to evaluate whether the data generated by the SmartZoo model would be more different than a real dataset when each were compared to random data from a Gaussian distribution. We discovered that the data generated by our model is closer to real data than random data, and we were able to demonstrate that the model excels at generating data that resembles real-world data. In the future, we hope our framework may assist zoological experts in caring for animals, enabling them to support the important educational missions of zoos.

INTRODUCTION

Zoos are important for children and adults alike. Zoos entertain and educate the public and are also critical to scientific research and animal conservation. There are many animals that are currently in danger of extinction, such as the golden lion tamarins, with about 2500-3000 individuals left, and the Arabian oryx, with about 1000 individuals left, but zoos keep these animals in a safe habitat so that they can live peacefully, thrive, and even reproduce (1,2,3). Zoos also educate the public about animals, sometimes about species that are very rare. Zoos can provide a highly important infrastructure for education and conservation, as well as inspire people to become better conservationists and scientists.

However, some animals like polar bears, cheetahs, and lions face challenges being kept in small and compact areas (4). They often do not successfully adapt to zoo-like environments and, as a result, suffer shorter lifespans and reproductive failures (4). Many animals suffer from zoochosis, experiencing stress and depression-like symptoms such as isolation from other animals and lethargy, which negatively impacts their quality of life without proper care (5).

With the advent of modern Internet of Things (IoT), there are many technologies that can be used to improve animals' lives in zoos. IoT technology includes gadgets and devices that function through the internet, often communicating with one another to achieve a goal such as communication, motion detection, or home security. There are two main reasons why IoT is beneficial for zoos. First, using IoT can reduce the cost of zoo maintenance. Animal behavior data that is collected by various algorithms can be used to track abnormal animal behavior/movements and identify whether an animal is ill and allows fast meditation. Second, using IoT can enhance the welfare of the animals. For example, zookeepers can find the most active locations where animals tend to spend their time and reflect on the information to design an improved layout of the cage. A better zoo layout or design can be used to increase positivity in animals so that they can be more active (6). By changing the current design with these technologies, zoos may have a much better chance of keeping animals happier and decreasing annual spending on maintenance and other tasks (7).

For an animal caring IoT platform, we first developed a deep learning framework to detect abnormal animal movement and propose an appropriate animal cage layout. We found that the prediction of animal activities can be cast as a sequence generation problem.

To address this, we will utilize duck movement data to train and test our model. Hence, we developed a machine learning-based method using a sequence model to predict the ordinary future movements of ducks. Since a cage is the usual roaming area for the animal, the future movement pattern was used to determine the shape and size of an ideal cage layout. We also used ordinary movement patterns of ducks, which is generated by our model for abnormal movement detection. Our method, SmartZoo, shows promise for benefiting animals that need a specific area of space for a high standard of living.

RESULTS

We predicted animal movement by building a transformer-based model and evaluated the effectiveness of our model by comparing whether the predicted movement was closer to ground truth movement than Gaussian random movement. We defined 'animal care' as detection of abnormal movement and prediction of suitable range of cage. Animal care is accomplished by a deep learning model that generates animal movement sequence. Our goal was to predict the movement patterns of an animal, compare these predictions to the recent movement of the animal to detect abnormal activities, and draw the expected cage range of the animal using our algorithm.

https://doi.org/10.59720/22-276

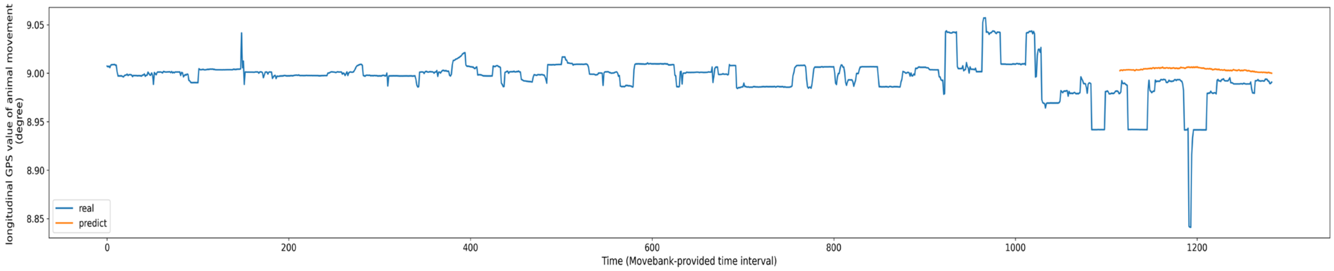

Figure 1: Longitudinal sequence prediction achieved by SmartZoo. The blue graph shows the empirical sequence, and the orange graph shows the sequence generated by the SmartZoo transformer model. The length of the empirical sequence is 1283 observed timepoints. Time is in units of hours, but the intervals between timepoints are irregular. The MAPE score is 0.3325; the RMSE score is 0.0415; and the MAE score is 0.03.

We used the duck movement dataset from Movebanks since it has more timepoints than other animal datasets (8). The duck dataset contains the longitudinal and latitudinal GPS value of duck's location, obtained with variable time intervals (1 to 12 hours) between datapoints. We trained our model using movement timeseries data to predict the next movement timeseries data. A section of movement timeseries data is defined as a sequence. The predicted sequence was regarded as future movement of an individual duck. We used mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE) to measure the performance of the model. MAPE was used for its ease of interpretation and straightforward comparison between different models or datasets. RMSE was used for its ability to weigh larger errors more heavily, providing a conservative assessment of model performance. The strength of MAE is its simplicity in interpretation. The closer all three metrics were to 0, the better the performance of the model.

We generated predicted longitudinal and latitudinal sequences with our model SmartZoo, a transformer-based model (**Figure 1-2**). The ratio of training data to test data was 4:1. But, since there were many outliers whose z-score beyond the threshold (±3) near the end of training dataset, the model learned this pattern of outliers and the sequence after outlier timepoints were not properly fitted. The model trained with longitudinal sequences showed a 0.3325 MAPE score, 0.0415 RMSE score, 0.03 MAE score, and the model trained with latitudinal sequences showed a 0.1180 MAPE score, 0.0615 RMSE score, 0.0563 MAE score. Overall, these scores indicate that the model predicts time series well. The fact that MAPE scores show significant differences while RMSE scores are similar indicates that the longitudinal sequence is more susceptible to outliers.

After using the algorithm to predict duck movement patterns, we wanted to test the feasibility of using the algorithm for cage prediction. We plotted the generated latitudinal and longitudinal sequence and drew a range of a cage that would correspond with the sequences (**Figure 3**). The animal cage shape that is drawn in this image should support the ducks' movements well because it is suggested by their movements themselves. The cage should reduce the probability of blocking the ducks in any way and allow the ducks to move freely in every direction that they were predicted to move in.

To evaluate how well our generated sequences modeled normal behavior, we wanted to test whether the sequences generated by our normal-behavior-trained algorithm were more similar to normal behavior sequences or abnormal behavior sequences. Abnormal duck movements, arising from a different distribution than that of the acquired, normal data, should be easily distinguishable from normal data. In contrast, movements generated by a model trained on acquired, normal behavior will likely be challenging to differentiate from acquired, normal data. Duck movement data do not follow a Gaussian distribution, and Gaussian distributions are widely used to generate noisy data from natural observations. Therefore, we simulated abnormal movement data by extracting sequences from a Gaussian distribution with a mean and variance equal to that of the acquired, normal behavior sequences (11).

We did K-means clustering on a mixture of model-generated sequences and ground truth sequences (from the acquired data of real duck movements) and on a mixture of model-generated sequences and Gaussian random sequences. We determined how well model-generated

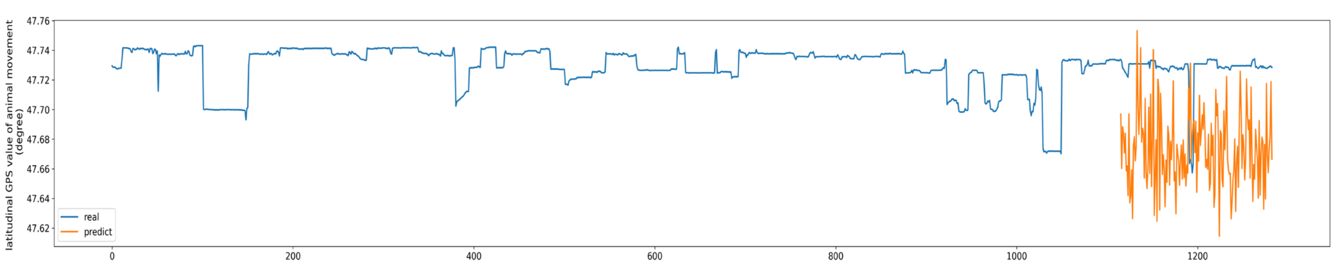

Figure 2: Latitudinal sequence prediction achieved by Smartzoo. The blue graph shows the empirical sequence, and the orange graph shows the sequence generated by the SmartZoo transformer model. The length of the empirical sequence is 1283. Time is in units of hours, but the intervals between timepoints are irregular. The MAPE score is 0.1180; the RMSE score is 0.0615; and the MAE score is 0.0563.

JOURNAL OF
EMERGING INVESTIGATORS

https://doi.org/10.59720/22-276

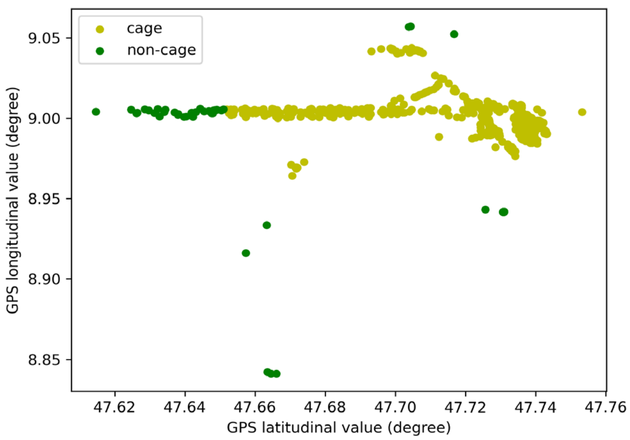

Figure 3: Range of cage prediction by using SmartZoo-generated and ground truth movement data of ducks. Real and computer-generated datapoints on movement of ducks are plotted in 2D Euclidean space. The latitudinal and longitudinal sequences were scaled using scikit-learn's MinMax scaler to fit between 0 and 1, maximizing the differences between values. Then, z-scores were calculated for each sequence, and data points falling outside the commonly used outlier detection threshold of z-score 3 were designated as outliers. Non-outlier data points in both latitudinal and longitudinal sequences were designated as the cage.

sequences, ground truth, and Gaussian random data are distinguished by calculating the alignment of clusters generated through K-means clustering with the original classes. Distinguishing between ground truth sequences and model-generated sequences showed lower accuracy (0.2054) than distinguishing between Gaussian random sequences and model-generated sequences (0.4732) (**Figure 4-5**). These results indicate that the sequences generated by our model are more similar to the ground truth sequences than the Gaussian random sequences.

DISCUSSION

SmartZoo is a modeling algorithm with the potential to save zookeepers' time and effort by supporting improved

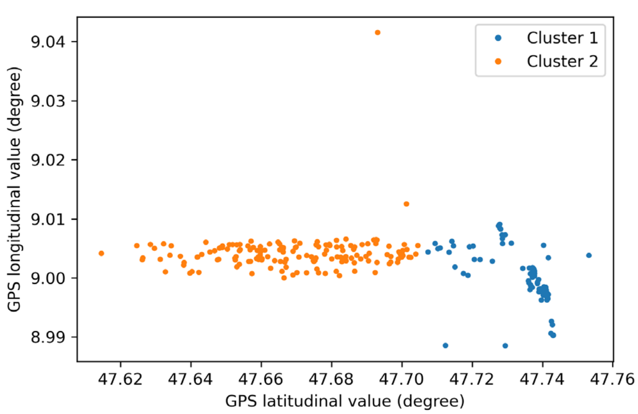

Figure 4: K-means clustering (k=2) results for mixture of SmartZoo-generated data and ground truth duck movement data. We mixed ground truth data and SmartZoo-generated data with same ratio. Data points are clustered into two groups via K-means clustering (k=2). Clustering accuracy is 0.2054, which means it is hard to distinguish generated data and ground truth data.

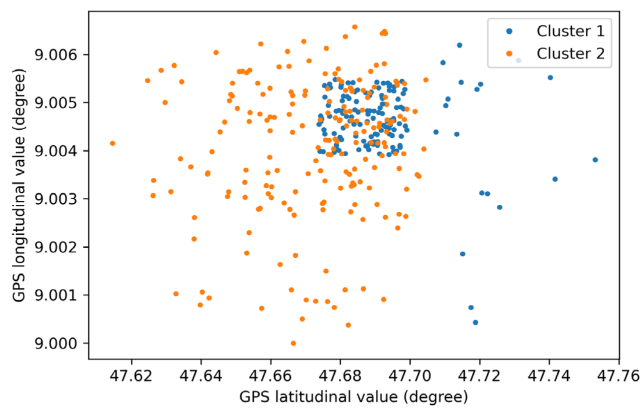

Figure 5: K-means clustering (k=2) results for mixture of SmartZoo-generated data and Gaussian random data. We mixed Gaussian random data and SmartZoo-generated data with same ratio. Data points are clustered into two groups via K-means clustering (k=2). Clustering accuracy is 0.4732, which means it is easier to distinguish SmartZoo-generated data and Gaussian random data compared to same task with ground truth data (**Figure 4**).

cage design and detecting abnormal movement through a transformer-based movement data generation model. We achieved successful movement data generation, which we then use to produce appropriate cage shape prediction and abnormal movement detection. We found that our generated movement data was closer to real movement data than random movement data from a Gaussian distribution.

Improper cages can negatively impact animals (12). Creating a cage based on the model-generated sequences together with the ground truth sequences takes into consideration the potential future activities of the animals, thus fully accommodating their movements. SmartZoo-generated cages remove the inconvenience and time consumption that comes with determining the right cage size and make the process robust from human errors like setting the size of cage too small for animal whose movement range is very large.

SmartZoo is currently capable of generating realistic movement data, but in the future, we plan to utilize its data generating power for anomaly detection as well. Since K-means cluster of model-generated sequence is similar to ground truth cluster and disimlar to Gaussian random sequence cluster, SmartZoo has the potential to be used identify unusual animal behaviors, e.g. if new data point significantly deviates from the value predicted by SmartZoo. Abnormal movement detection may allow a zookeeper to see that the animal is not in a normal state and may need assistance (13). After executing K-means clustering, a centroid for the normal and abnormal clusters can be obtained. Then, these centroid locations can be used to determine camera placement; to capture animal movements efficiently a camera should be aimed at each of the centroids of the clusters.

However, since we used only one species of animal (ducks), it is hard to generalize our model to all species of animals, and our data is not sufficient to capture the whole movement pattern of ducks because the data only span one month. Also, our analysis did not account for confounding variables like natural habitat, diet, and seasonal behavior. Our approach may be somewhat simple, but we think it will serve

https://doi.org/10.59720/22-276

as the basis for further research. For further research, we are planning to use data that cover longer timespans, to account for confounding variables, and to adapt our model to as many species as possible.

In addition to movement data, we can also use closed-circuit television (CCTV), a form of video surveillance, to add video data to the movement data. Deep learning through video analysis, especially in the future, could potentially play a key role in future versions of SmartZoo. Current deep learning models for video can detect objects in the video, verify the class of the object, classify the type of video, and even do question-answering tasks on videos (14,15,16,17). In the case of SmartZoo, deep learning models for video could detect an animal and identify its behavior, unveiling where it usually does specific behaviors like drinking, eating, and sleeping. Finally, a deep learning model for video could help derive insights into where to put food supplies, enrichment, and structures such as ponds depending on the usual movement of the animals. On a broader note, SmartZoo has potential to help zoos decide the optimal location to place the cages within the zoo grounds for people to view the animals in their comfortable enclosures.

MATERIALS AND METHODS

Data

Data were obtained from the open source Movebanks, a large dataset of tracked movements of animals across the world (8). We chose Lake Constance duck data, which was collected from 18th December 2008 to 4th January 2009 in Lake Constance, Germany. The dataset consisted of GPS information of animal movements in longitude and latitude. Both sequences have 1283 timepoints. Each timepoint also contains the observation time, and the time interval between each observation varies within the range of 1 to 12 hours.

Programming details

We executed our experiments on Google Colab. We used python version 3.7, scikit-learn version 1.2.2, numpy version 1.22.4, matplotlib version 3.7.0, and transformers version 4.7.0. We used transformers with basic settings provided by Huggingface.

Transformer model

Using the transformer model (**Figure 6**), time series data were generated. The transformer model uses a self-attention algorithm, which connects positions within a single sequence, making learning dependency and fastened training possible. Moreover, it uses positional information of the sequence via positional encoding, where it encodes sequence data efficiently and shows powerful performance. The transformer model is appropriate for capturing dependence of long-range data. Since it is not possible to describe an animal's overall behavioral pattern based on its behavior in a short period of time, we adopted the transformer model for our animal movement modeling. We used only the encoder part of the transformer and added two linear layers for sequence generation.

Model training details and metrics

The data was split into training and testing sets to avoid overfitting and to increase generalizability. A Transformer model was trained to predict the immediate subsequent

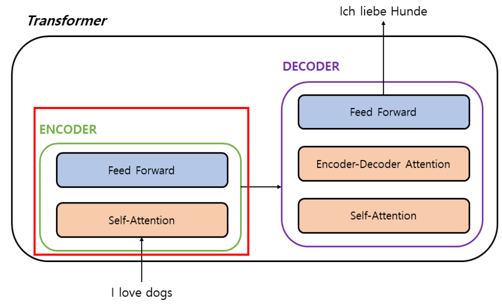

Figure 6: Transformer model architecture. The transformer model architecture represents a significant shift in the approach to sequence-to-sequence tasks in deep learning (9). Transformers consist of an encoder and a decoder, each made up of multiple layers containing self-attention and feed-forward neural networks. This design enables the model to handle a wide range of tasks, including language translation, text summarization, and even image recognition, with high parallelization and efficiency. We used only encoder part to encode time-series data and predict future sequences.

sequence, taking 336 timepoints as input and the successive 168 timepoints as output. To evaluate the model, the last 336 timepoints of the training data were used, which the model had never seen as input, to predict the final 168 timepoints corresponding to the test data. The training set was former 80% of entire dataset and test set was latter 20% of entire dataset. The training set was min-max scaled using the Scikit-learn package for efficient learning. Data was fed into the model by a sliding window technique. Using 336 timepoints of input data, the output data of 168 timepoints immediately following the input data was predicted.

MAPE, RMSE, and MAE were calculated to evaluate the output data. The equations of each metric are as follows:

$$MAPE = \frac{100}{n} \times \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n}}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{Y}_i|$$

The MAPE, RMSE, and MAE scores were reported at test set. The performance metric was computed by the MAPE, score on the test set.

K-means clustering

K-means clustering, which identifies cluster of a particular data point, was performed to prove that the sequence created by our model belongs to a ground truth rather than a Gaussian random sequence.

Random sequence generating

The random sequence is generated by python package 'numpy'. We used 'randn' function from numpy random module, which returns random numbers sampled from a

https://doi.org/10.59720/22-276

standard normal distribution.

All the codes and data are available at https://github.com/DeanJii/SmartZoo.

Received: November 11, 2022
Accepted: August 28, 2023
Published: August 7, 2024

REFERENCES

1. "Golden Lion Tamarin." *Smithsonian's* National Zoo. www.nationalzoo.si.edu/animals/golden-lion-tamarin. Accessed 20 Nov. 2022.
2. Linda Lombardi, "8 Zoos Helping Animals Edge Out of Extinction". *Vetstreet*. www.vetstreet.com/our-pet-experts/8-zoos-helping-animals-edge-out-of-extinction. Accessed 22 Nov. 2022.
3. "The Conservation Mission of Zoos." *Wild Welfare*. wildwelfare.org/the-conservation-mission-of-zoos-nabila-aziz/#:~:text=Zoos%20engage%20in%20research%2C%20preserve,conservation%20projects%20across%20the%20world. Accessed 10 Oct. 2022.
4. Edna Francisco. "Zoo Carnivores Need More Space." *Science and AAAS*. www.science.org/content/article/zoo-carnivores-need-more-space. Accessed 2 Jan. 2023.
5. "What is zoochosis and how do animals get it?" *In Defense of Animals*. www.idausa.org/campaign/elephants/what-is-zoochosis/. Accessed 3 Apr. 2023.
6. R.Scott Nolen. "Designing zoo habitats that promote animal well-being." *American Veterinary Medical Association*. www.avma.org/javma-news/2002-12-01/designing-zoo-habitats-promote-animal-well-being. Accessed 3 Apr. 2023.
7. Wijeyakulasuriya, Dhanushi A., et al. "Machine learning for modeling animal movement." *PloS one*, vol. 15, no. 19, Jul. 2020, https://doi.org/10.1371/journal.pone.0235750.
8. "Movebanks." *Max Planck Institute for Animal Behavior*, www.movebank.org/cms/webapp?gwt_fragment=page. Accessed 20 Oct. 2023
9. Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems*, vol. 30, Dec. 2017, https://doi.org/10.48550/arXiv.1706.03762.
10. Bock, Hans-Hermann. "Clustering methods: a history of k-means algorithms." *Selected contributions in data analysis and classification*, 1st ed., Springer, 2007, https://doi.org/10.1007/978-3-540-73560-1_15.
11. Tian, Chunwei, et al. "Deep learning on image denoising: An overview." *Neural Networks*, vol. 131, Nov. 2020, https://doi.org/10.1016/j.neunet.2020.07.025.
12. Hartcher, Jones, et al. "The welfare of layer hens in cage and cage-free housing systems." *World's Poultry Science Journal*, vol. 73, 2017, https://doi.org/10.1017/S0043933917000812.
13. Fowler, M. E., "Behavioral clues for detection of illness in wild animals: models in camelids and elephants." *In Zoo and Wild Animal Medicine,* 2008, pp. 33-cp3. WB Saunders, https://doi.org/10.1016/B978-141604047-7.50007-5.
14. Zhao, Zhong-Qiu, et al. "Object detection with deep learning: A review." *IEEE transactions on neural networks and learning systems*, vol. 30.11, 2019, pp. 3212-3232, https://doi.org/10.48550/arXiv.1807.05511.
15. Wang, Xiaogang. "Deep learning in object recognition, detection, and segmentation." *Foundations and Trends in Signal Processing*, vol 8.4, 2016, pp. 217-382, http://dx.doi.org/10.1561/2000000071.
16. Rehman, et al. "Deep learning for video classification: A review." *Authorea Preprints*, 2023, https://doi.org/10.36227/techrxiv.15172920.v1.
17. Kim, et al. "Video question answering using language-guided deep compressed-domain video feature." *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, https://doi.org/10.1109/ICCV48922.2021.00173.