SiamBOMB: A Real-time AI-based System for Home-cage Animal Tracking, Segmentation and Behavioral Analysis

Xi Chen[1] , **Hao Zhai**[1,2] , **Danqian Liu**[3,5] , **Weifu Li**[1,4] , **Chaoyue Ding**[1] , **Qiwei Xie**[1] and **Hua Han**[1,5]

[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[2]School of Automation and Electrical Engineering, University of Science and Technology Beijing, China
[3]Howard Hughes Medical Institute, University of California, Berkeley, CA, USA
[4]College of Science, Huazhong Agricultural University, Wuhan, China
[5]Center of Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences
{xi.chen, zhaihao2020, qiwei.xie, hua.han}@ia.ac.cn, dqliu@ion.ac.cn, wfli@stu.hubu.edu.cn

Abstract

Biologists often need to handle numerous video-based home-cage animal behavior analysis tasks that require massive workloads. Therefore, we develop an AI-based multi-species tracking and segmentation system, SiamBOMB[1], for real-time and automatic home-cage animal behavioral analysis. In this system, a background-enhanced Siamese-based network with replaceable modular design ensures the flexibility and generalizability of the system, and a user-friendly interface makes it convenient to use for biologists. This real-time AI system will effectively reduce the burden on biologists.

1 Introduction

It is of great significance to explore the relationship between the brain and animal behaviors. Recording home-cage animals with video cameras means a lot for investigation, and neuro-behavioral analysis of phenotypes requires the monitoring of animal behavior over long periods of time [Jhuang et al., 2010]. Therefore, automatic home-cage animal behavioral analysis has attached enough attention in recent years. Most of previous methods [Dietrich et al., 2015; Yu et al., 2017; Singh et al., 2019; Rao et al., 2019] implemented traditional machine learning models for object localization and detection. However, tracking-based systems cannot provide enough accuracy for fine posture and state of animals. From this perspective, Liu et al. [2020] proposed a deep learning model U-Net for image segmentation and motor activity recognition. Although achieving high accuracy, it has to run offline and retrain for another different species.

Deep learning methodology for image segmentation have thrived these years and many methods have been proposed, such as encoder-decoder architecture [Ronneberger et al., 2015] networks, region proposal networks [He et al., 2017], multi-scale networks [Lin et al., 2017], Siamese architecture [Bertinetto et al., 2016; Li et al., 2018], etc. Our tracking
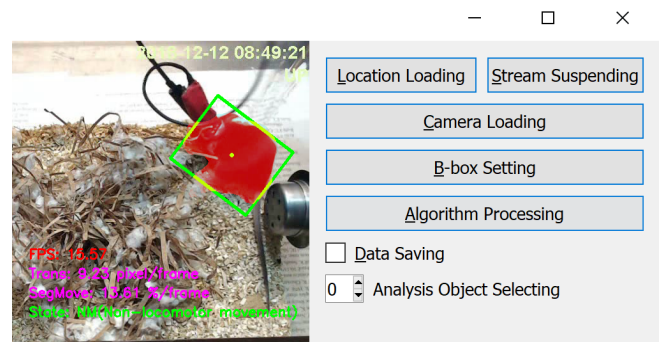

Figure 1: A part of our software interface. It is tracking and segmenting a mouse while doing fundamental behavioral analysis.

and segmentation method is inspired from SiamMask [Wang et al., 2019]. By implementing this kind of structure, we can obtain approximate species-specific feature at the first frame (from one-shot learning), instead of with extra labels of this species (from transfer learning [Mathis et al., 2018]).

Note that the backgrounds in such home-cage tasks usually remain relatively fixed, we introduce the background image b for template frame and enhance negative information in the region proposal subnetwork [Ren et al., 2015]. Our network also has additional advantages for online multi-species tracking. It can run nearly real-time just with a single laptop GPU. Furthermore, we build a user-friendly interface (Figure 1), embedding a fundamental behavioral analysis system. It contributes to the home-cage segmentation of various species, and also develops automatic biological experiments.

2 Algorithm

Our methodology is concluded as Siamese network using Background information for Online Multi-species home-cage Behavioral analysis (SiamBOMB), shown in Figure 2.

2.1 Using Background Information to Enhance One-shot Learning

Using one-shot learning from the first frame ***bounding box*** labels, the Siamese subnetwork is able to adapt features of

---

[1]Our demonstration video: https://youtu.be/x4tjOdrtQ98, open source project: https://github.com/JackieZhai/SiamBOMB.
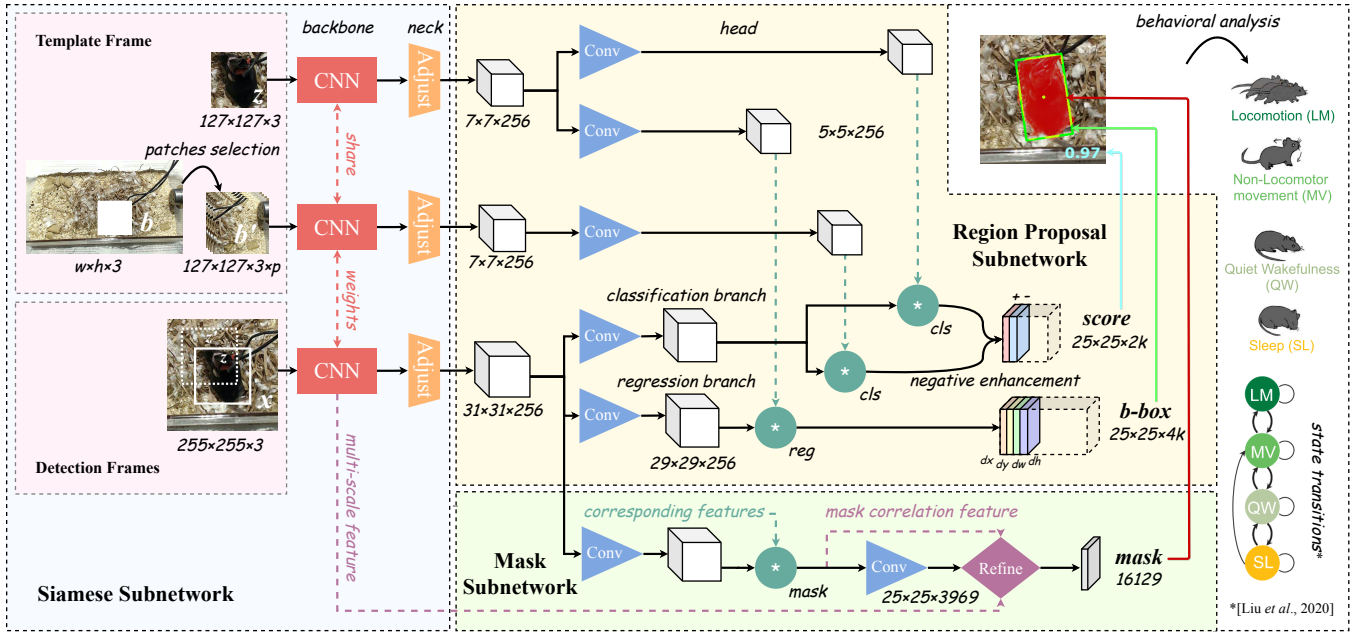
Figure 2: The architecture of our network (SiamBOMB): left side is Siamese subnetwork for feature extraction, top-right side is region proposal subnetwork generating two output feature maps for **score**s and **bounding box**es, bottom-right side is mask subnetwork for refined **mask** results. z denotes exemplar image, b denotes background image, x denotes search images, p denotes the number of patches seleted from b, k denotes the number of anchor, $*$ denotes depth-wise cross correlation, the refinement module are omitted here for simplicity.

different species. Meanwhile, it can also obtain the background texture, color or intensities information from b' due to the approximate invariance of b (Figure 2). To adapt to the CNN feature extractors, we select p dissimilar patches from b appropriately according to the scale and variance of b. In the region proposal subnetwork, the additional background features enhance the negative side of k pairs **score** map. In the mask subnetwork, the refinement model acquires extra multi-scale background features during upsampling layers.

In order to evaluate the effectiveness of proposed network, we have labeled 5000 continuous frames of home-cage mice movement **mask**s by two manuals, and divided the training and testing sets by a ratio of 7:3. The results are shown in Table 1. $*$ Metrics from DAVIS 2016 dataset [Perazzi *et al.*, 2016]. $\dagger$ Manual variability (discrimination between different sets of manual labeling). [2] [Liu *et al.*, 2020]. [3] [Sauer *et al.*, 2019]. Our proposed network has higher accuracy by implementing rotated bounding boxes [Chen and Tsotsos, 2019] and update models [Zhang *et al.*, 2019], and has faster speed by using AlexNet or ResNet-50 [Zhang *et al.*, 2017] as CNN feature extractors (Figure 2).

2.2 Achieving Real-time Multi-species Monitoring

With modular design, we can simply substitute the backbone, neck or head of our network to suit various species, environments and purposes. Then, we duplicate CNNs in the same tracker to accelerate multi-object speed. As the number of objects increases, CNNs can adaptively share weights or networks to reduce GPU spatial complexity. However, if objects are from other species (or with huge varieties), CNNs terminate sharing to keep multi-species synchronous accuracy.

Method	$\mathcal{J}^*(\%)$	$\mathcal{F}^*(\%)$	$\mathcal{T}^*(\%)$	$\mathcal{S}$ (fps)
Another GT†	87.92	89.58	3.75	–
U-Net[2]	**81.32**	**84.73**	**3.22**	0.33
SiamMask[3]	77.31	76.82	6.61	**3.74**
SiamBOMB (acc.)	**84.35**	**85.17**	**4.31**	3.19
SiamBOMB (spe.)	77.04	74.52	4.66	**14.65**

Table 1: Comparsion for mouse segmentation under the accuracy, robustness and speed. $\mathcal{J}$ denotes Jaccard index (IoU) for region similarity, $\mathcal{F}$ denotes F-measure for contour accuracy, $\mathcal{T}$ denotes Temporal stability and $\mathcal{S}$ denotes Speed (including fundamental behavior analysis and data storage strategy).

SiamBOMB can achieve laptop-level portability. During the experiments, we use a single GPU (NVIDIA GeForce GTX 1050 Ti, Max-Q, 4 GB memory) for evaluation. As shown in Table 1, U-Net as an offline method [Liu *et al.*, 2020] is well-performed but quite slow. In contrast, our accuracy-priority network increases speed while reaching the same level accuracy. Moreover, our speed-priority network for online multi-species monitoring achieves 20 fps (maximum) monitoring. Accordingly, it is able to monitor 9 objects of various species simultaneously with just about 2.2 GB GPU memory usage, and suitable for any biological home-cage animal experiment.

3 Innovations of The System

SiamBOMB dominates home-cage segmentation of various species by using background enhanced one-shot learning. It

also achieves nearly real-time speed and manual accuracy (Table 1) by running our modular Siamese-based network just on a single laptop-level GPU. We supplement a user-friendly interface developed by PyQt5, making it convenient for biologists and researchers in other fields to get started. The entire system provides capacities to monitor the trajectories, movements and states of multiple experimental objects automatically and concurrently. It contributes for biologist to get rid of cumbersome workloads, which may mean dozens of videos or millions of frames [Dugatkin, 2020].

Acknowledgments

References

[Bertinetto *et al.*, 2016] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr. Fully-convolutional siamese networks for object tracking. In *European conference on computer vision*, pages 850–865. Springer, 2016.

[Chen and Tsotsos, 2019] Baoxin Chen and John K Tsotsos. Fast visual object tracking with rotated bounding boxes. *arXiv preprint arXiv:1907.03892*, 2019.

[Dietrich *et al.*, 2015] Marcelo O Dietrich, Marcelo R Zimmer, Jeremy Bober, and Tamas L Horvath. Hypothalamic agrp neurons drive stereotypic behaviors beyond feeding. *Cell*, 160(6):1222–1232, 2015.

[Dugatkin, 2020] Lee Alan Dugatkin. *Principles of animal behavior*. University of Chicago Press, 2020.

[He *et al.*, 2017] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[Jhuang *et al.*, 2010] Hueihan Jhuang, Estibaliz Garrote, Xinlin Yu, Vinita Khilnani, Tomaso Poggio, Andrew D Steele, and Thomas Serre. Automated home-cage behavioural phenotyping of mice. *Nature communications*, 1(1):1–10, 2010.

[Li *et al.*, 2018] Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, and Xiaolin Hu. High performance visual tracking with siamese region proposal network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8971–8980, 2018.

[Lin *et al.*, 2017] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1925–1934, 2017.

[Liu *et al.*, 2020] Danqian Liu, Weifu Li, Chenyan Ma, Weitong Zheng, Yuanyuan Yao, Chak Foon Tso, Peng Zhong, Xi Chen, Jun Ho Song, Woochul Choi, et al. A common hub for sleep and motor control in the substantia nigra. *Science*, 367(6476):440–445, 2020.

[Mathis *et al.*, 2018] Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe, Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience*, 21(9):1281, 2018.

[Perazzi *et al.*, 2016] Federico Perazzi, Jordi Pont-Tuset, Brian McWilliams, Luc Van Gool, Markus Gross, and Alexander Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 724–732, 2016.

[Rao *et al.*, 2019] Srinivasa R Rao, Sam WZ Olechnowicz, Patrick Krätschmer, James EC Jepson, Claire M Edwards, and James R Edwards. Small animal video tracking for activity and path analysis using a novel open-source multi-platform application (animapp). *Scientific reports*, 9(1):1–9, 2019.

[Ren *et al.*, 2015] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[Sauer *et al.*, 2019] Axel Sauer, Elie Aljalbout, and Sami Haddadin. Tracking holistic object representations. *arXiv preprint arXiv:1907.12920*, 2019.

[Singh *et al.*, 2019] Surjeet Singh, Edgar Bermudez-Contreras, Mojtaba Nazari, Robert J Sutherland, and Majid H Mohajerani. Low-cost solution for rodent home-cage behaviour monitoring. *PloS one*, 14(8), 2019.

[Wang *et al.*, 2019] Qiang Wang, Li Zhang, Luca Bertinetto, Weiming Hu, and Philip HS Torr. Fast online object tracking and segmentation: A unifying approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1328–1338, 2019.

[Yu *et al.*, 2017] Yao-Qing Yu, Devin M Barry, Yan Hao, Xue-Ting Liu, and Zhou-Feng Chen. Molecular and neural basis of contagious itch behavior in mice. *Science*, 355(6329):1072–1076, 2017.

[Zhang *et al.*, 2017] Qiao Zhang, Zhipeng Cui, Xiaoguang Niu, Shijie Geng, and Yu Qiao. Image segmentation with pyramid dilated convolution based on resnet and u-net. In *International Conference on Neural Information Processing*, pages 364–372. Springer, 2017.

[Zhang *et al.*, 2019] Lichao Zhang, Abel Gonzalez-Garcia, Joost van de Weijer, Martin Danelljan, and Fahad Shahbaz Khan. Learning the model update for siamese trackers. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4010–4019, 2019.