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

Xi Chen¹, Hao Zhai¹,², Danqian Liu³,⁵, Weifu Li¹,⁴, Chaoyue Ding¹, Qiwei Xie¹ and Hua Han¹,⁵

¹National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
²School of Automation and Electrical Engineering, University of Science and Technology Beijing, China
³Howard Hughes Medical Institute, University of California, Berkeley, CA, USA
⁴College of Science, Huazhong Agricultural University, Wuhan, China
⁵Center 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, SiamBOMB¹, for real-time and automatic home-cage animal behavioral analysis. In this system, a background-enhanced Siamese 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 attracted 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 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
CNN feature extractors, we select the approximate invariance of Table 1. The results are shown in manual labeling).

2016].

movement we have labeled 5000 continuous frames of home-cage mice scale background features during upsampling layers.

mask subnetwork, the refinement model acquires extra multi-features enhance the negative side of

appropriately according to the scale and variance of

ground texture, color or intensities information from

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 masks 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]. † 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.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\mathcal{J}$ (%)</th>
<th>$\mathcal{F}$ (%)</th>
<th>$\mathcal{P}$ (%)</th>
<th>$S$ (fps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Another GT†</td>
<td>87.92</td>
<td>89.58</td>
<td>3.75</td>
<td>—</td>
</tr>
<tr>
<td>U-Net²</td>
<td><strong>81.32</strong></td>
<td><strong>84.73</strong></td>
<td><strong>3.22</strong></td>
<td>0.33</td>
</tr>
<tr>
<td>SiamMask³</td>
<td>77.31</td>
<td>76.82</td>
<td>6.61</td>
<td>3.74</td>
</tr>
<tr>
<td><strong>SiamBOMB</strong> (acc.)</td>
<td><strong>84.35</strong></td>
<td><strong>85.17</strong></td>
<td><strong>4.31</strong></td>
<td>3.19</td>
</tr>
<tr>
<td><strong>SiamBOMB</strong> (spe.)</td>
<td>77.04</td>
<td>74.52</td>
<td>6.61</td>
<td><strong>14.65</strong></td>
</tr>
</tbody>
</table>

Table 1: Comparison 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, $\mathcal{P}$ denotes Temporal stability and $S$ 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

![Diagram of network architecture](image-url)
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

This research is supported by the National Natural Science Foundation of China (61673381), Special Program of Beijing Municipal Science & Technology Commission (Z181100003818001, Z181100000118002), the Strategic Priority Research Program, CAS (XDB32030200), Bureau of International Cooperation, CAS (153D31KYSB20170059).

References


