先端科学技術研究科　修士論文要旨

所属研究室 (主指導教員)	数理情報学 (池田　和司　(教授))		
学籍番号	2411271	提出日	令和 8年 1月 17日
学生氏名	宮井　菜名子		
論文題目	Hierarchical Representation Learning of Dog Behavior via Single-View 3D Pose Estimation		

要旨

Dogs exhibit diverse behaviors that function as important signals in human-dog communication. Automatic analysis of such behaviors is increasingly needed in both scientific and applied contexts. However, conventional methods for behavior analysis face two major challenges: (i) 3D pose estimation typically requires multi-camera setups or prior training with complex calibration, and (ii) behavior classification relies heavily on predefined labels, limiting the ability to detect previously unseen behaviors. To address these limitations, we combine D-Pose, a model that estimates 3D dog poses from a single camera by learning pose representations, with h/BehaveMAE, a self-supervised framework that learns hierarchical behavior representations from pose sequences without predefined labels. Using a dataset of annotated dog behaviors, we perform evaluation by applying linear probing on the learned embeddings. To evaluate the effectiveness of our approach, we conduct behavior classification experiments comparing embeddings learned from 3D pose trajectories obtained via D-Pose with those learned from 2D pose trajectories extracted using DeepLabCut (DLC), where h/BehaveMAE is applied in both settings. In addition, we compare h/BehaveMAE with a supervised Transformer encoder baseline to assess the benefit of self-supervised and hierarchical representation learning, particularly under limited annotation conditions. Furthermore, to examine the general applicability of the learned embeddings beyond supervised classification, we evaluate their use in an anomaly detection task. The results show that embeddings learned from 3D pose trajectories consistently outperform those based on 2D poses. Moreover, under limited annotation conditions, our approach achieves higher performance than the supervised Transformer encoder baseline. We also demonstrate that the hierarchical representations learned by h/BehaveMAE enable the detection of rare behaviors at multiple levels of granularity. Our results suggest that this approach provides a flexible and generalizable pipeline for behavior analysis, enabling promising representation learning from videos. While this study focuses on dog behavior, the proposed framework may serve as a step toward uncovering the mechanisms of animal communication in the future.