

課題名：深層学習技術を用いた映像解析による動物園動物の行動分析

実施時期：2025/4～2026/3

所属機関名：公益財団法人京都高度技術研究所

代表者氏名：吉田 信明

成果概要：動物園で飼育されている動物の行動分析におけるVision-Language Modelの応用可能性の評価を行った。京都市動物園で飼育されているワオキツネザルの行動を記録した監視カメラ2台の映像と、それらに基づく人手での行動観察結果をデータセットとし、映像から行動(31分類)への分類モデル (ActionCLIP(Wang et al., 2021)) をファインチューニングにより作成した。その結果57.8%の正解率となり、一定の分類精度はえられた一方で、より精度の高い分類モデルには、特に、頻度の少ない行動などが今後の課題となった。

成果のポイント：動物園の複雑な飼育環境で撮影した映像を用いて、撮影された個体の行動認識の評価をVLMを用いて行い、課題を抽出した。

データセット

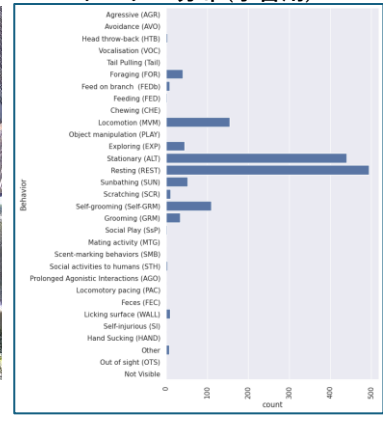
2視点の監視カメラ映像 + 観察データ

入力画像の例



映像：
- カメラ2台の映像を連結
- 20秒・16フレーム分
観察データ：
- エングラム 31分類
- 偏り大(右グラフ)

データの分布(学習用)



データ件数

| 用途 | 件数 |
|-----|------|
| 学習用 | 1419 |
| 評価用 | 251 |

モデルの学習

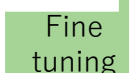
分類タスク向けVLMを選定、
本データセットで学習

使用モデル：ActionCLIP(Wang et al., 2021)
画像向け分類モデルCLIPの映像版

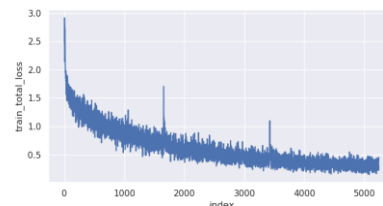


ヒトの行動での
事前学習済みモデル

HMDB51データセット
(Kuehne et al., 2011)



評価用モデル



学習経過

評価・今後の課題

分類タスク向けVLMを選定、
本データセットで学習

主要な行動の分類結果

| 行動 | 件数 | 精度 | 再現率 | F値 |
|---------------|----|------|------|------|
| Resting | 88 | 0.84 | 0.75 | 0.79 |
| Stationary | 78 | 0.56 | 0.60 | 0.58 |
| Locomotion | 28 | 0.48 | 0.39 | 0.43 |
| Self-grooming | 20 | 0.36 | 0.40 | 0.38 |
| Others | 37 | 0.40 | 0.46 | 0.42 |

結果

- ・全体正解率：57.8%
- ・データ件数が多いと精度は相対的に高くなる
- ・誤分類
 - 移動を伴わない行動間 (RestingとStationary)
 - 移動の有無に違いがあっても誤分類あり

今後の課題

- ・件数の少ない行動の精度低下：従来手法でも課題
- ・「行動」の表現の多様性：同じ行動でも、実際の見え方は千差万別
⇒ 行動ごとのデータ件数の増加と、各行動の見え方のバリエーションの増加等に向けた取り組みが必要

混同行列

| | Resting | Stationary | Locomotion | Self-grooming | Others |
|---------------|---------|------------|------------|---------------|--------|
| Resting | 66 | 16 | 1 | 3 | 2 |
| Stationary | 4 | 47 | 8 | 5 | 14 |
| Locomotion | 2 | 6 | 11 | 3 | 6 |
| Self-grooming | 4 | 4 | 0 | 8 | 4 |
| Others | 3 | 11 | 3 | 3 | 17 |

成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：

吉田信明, 山梨裕美, 松阪智子. ワオキツネザルの複数視点映像からの行動分類へのマルチモーダルAIの利用の評価. 動物の行動と管理学会 2025年度研究発表会, 2025年9月18日.