

令和7年度 修士論文

動物種および環境の多様性に頑健な  
行動認識のためのマルチモーダル表現空間の構築

北海道大学 大学院 情報科学研究科

複合情報学専攻 自律系工学研究室

菊地 真優

令和8年2月 xx 日

# 概要

本研究では、動物種および撮影環境の多様性に頑健な動物行動認識を実現するため、マルチモーダル表現学習に基づく共有表現空間の構築手法を提案する。動物行動認識は、動物園における個体管理や福祉評価を支援する重要な技術である。しかし、従来手法の多くは膨大な行動ラベル付きデータを必要とし、未知の動物種や環境への適用が困難であるという課題があった。

そこで本研究では、RGB 映像から得られる外観情報と、オプティカルフロー（光フロー）に基づく運動情報を統合するマルチモーダル表現学習手法を設計した。各モダリティから抽出された特徴を Gated Fusion に基づく統合機構により適応的に統合することで、行動に対して判別的な表現空間を形成する。さらに、獲得される表現から動物種固有のバイアスを除去するため、種分類器を用いた敵対的学習を導入し、動物種に依存しにくい汎用的な行動特徴の獲得を目指す。

提案手法の有効性を検証するため、大規模動物行動データセットを用いて学習を行い、未知環境で撮影された動物園のシロクマ映像に対して評価を行った。クラスタリング指標を用いた評価の結果、提案手法は単一モダリティに基づく手法と比較して、動物種や環境の変化に対して頑健な行動表現を獲得できることを示した。

本研究の成果は、アノテーションコストを抑制しつつ、多様な実環境へ適用可能な動物行動認識システムの実現に寄与するものである。

# 目次

|                             |    |
|-----------------------------|----|
| 概要                          | i  |
| 第 1 章 序論                    | 1  |
| 1.1 研究背景                    | 1  |
| 1.2 研究目的                    | 2  |
| 1.3 本論文の構成                  | 2  |
| 第 2 章 関連研究                  | 3  |
| 2.1 動物行動認識                  | 3  |
| 2.2 実環境における動物行動解析の事例        | 4  |
| 2.3 表現学習と行動埋め込み             | 4  |
| 2.4 敵対的学習による不変表現学習          | 4  |
| 第 3 章 提案手法                  | 6  |
| 3.1 入力動画の前処理                | 6  |
| 3.2 RGB 映像に基づく外観特徴抽出        | 7  |
| 3.3 オプティカルフローに基づく運動特徴抽出     | 7  |
| 3.4 Gated Fusion による適応的特徴統合 | 7  |
| 3.5 敵対的学習による種情報の分離          | 8  |
| 3.6 目的関数と最適化                | 8  |
| 第 4 章 実験と評価                 | 10 |
| 4.1 実験設定                    | 10 |
| 4.1.1 データセット                | 10 |
| 4.1.2 実装詳細とハイパーパラメータ        | 11 |
| 4.2 評価指標                    | 11 |

|              |                              |           |
|--------------|------------------------------|-----------|
| 4.3          | 実験結果 . . . . .               | 12        |
| 4.3.1        | 定量評価 . . . . .               | 12        |
| 4.3.2        | 定性評価 (可視化) . . . . .         | 13        |
| 4.4          | 考察と限界 . . . . .              | 13        |
| 4.4.1        | ドメイン適応における動作情報の重要性 . . . . . | 13        |
| 4.4.2        | 敵対的学習による種不変性の獲得 . . . . .    | 13        |
| 4.4.3        | 本手法の限界と課題 . . . . .          | 13        |
| <b>第 5 章</b> | <b>実験結果</b>                  | <b>14</b> |
| 5.1          | 大規模データセットにおける学習結果 . . . . .  | 14        |
| 5.2          | シロクマ映像に対する評価結果 . . . . .     | 14        |
| 5.3          | モダリティ別比較 . . . . .           | 14        |
| 5.4          | 敵対的学習の効果分析 . . . . .         | 14        |
| 5.5          | 表現空間の可視化 . . . . .           | 14        |
| <b>第 6 章</b> | <b>考察</b>                    | <b>15</b> |
| 6.1          | マルチモーダル統合の有効性 . . . . .      | 15        |
| 6.2          | 種に依存しない表現の獲得について . . . . .   | 15        |
| 6.3          | 動物園実環境への適用可能性 . . . . .      | 15        |
| 6.4          | 本手法の限界 . . . . .             | 15        |
| <b>第 7 章</b> | <b>結論</b>                    | <b>16</b> |
| 7.1          | 本研究のまとめ . . . . .            | 16        |
| 7.2          | 今後の課題 . . . . .              | 16        |
|              | <b>謝辞</b>                    | <b>17</b> |
|              | <b>参考文献</b>                  | <b>18</b> |

# 第 1 章

## 序論

### 1.1 研究背景

近年、動物行動の定量的な解析は、動物園における個体管理や行動モニタリング、さらには動物福祉の向上を目的として重要性を増している。従来、動物行動の評価は飼育員や研究者による目視観察に基づいて行われてきたが、長時間にわたる継続的な観察には多大な人手と時間を要するという課題がある。このような背景から、映像データを用いたコンピュータビジョン技術による動物行動認識の自動化が強く望まれている。

深層学習に基づく行動認識技術は、特定のデータセットにおいて高い認識性能を達成している。しかし、これらの手法の多くは膨大な行動ラベル付きデータを必要とするため、新たな動物種や撮影環境へ適用する際には、再度大規模なアノテーション作業が必要となる。特に動物園環境では、動物種ごとの外観差や行動様式の多様性に加え、照明条件や檻による遮蔽といった撮影環境の影響が大きく、特定の環境で学習されたモデルをそのまま他環境へ適用（汎化）することが困難であるという問題がある。

このような課題に対し、特定のドメインに依存しない行動に関する本質的な特徴を抽出し、動物種や環境の変化に頑健な表現を獲得することが重要である。その有効なアプローチとして、映像から得られる外観情報に加え、オプティカルフローに基づく運動情報を併用するマルチモーダル表現学習が注目されている。静的な外観特徴と動的な運動特徴を統合することで、単一のモダリティでは捉えきれない複雑な行動パターンを表現可能になると期待される。

さらに、表現学習の過程において、行動認識に不要な属性情報を抑制する手法として、敵対的学習を用いた不変表現学習 (Invariant Representation Learning) が提案されている。これらの手法は、人や物体の認識分野において、ドメイン間の差異を克服した特徴表現の獲得に成功しており、動物行動認識への応用も期待されている。しかし、マルチモー

ダル表現学習と敵対的学習を組み合わせ、動物種および環境の多様性を同時に克服することを目的とした研究は、未だ十分に検討されていない。

## 1.2 研究目的

本研究の目的は、動物種および撮影環境の多様性に頑健な動物行動認識を実現するため、マルチモーダル表現学習に基づく共有表現空間の構築手法を提案することである。具体的には、RGB 映像から得られる外観情報と、オプティカルフローに基づく運動情報を統合することで、行動に関する判別的な特徴表現を学習する。

本研究の核心的な提案として、獲得される表現から動物種固有のバイアスを排除することを目的とした、種分類器との敵対的学習を導入する。これにより、種の違いというドメインの壁を越え、未知の動物種や環境に対しても高い汎化性能を持つ、汎用的な行動表現の獲得を目指す。

本研究では、大規模動物行動データセットを用いて提案手法の学習を行い、その有効性を検証する。さらに、未知環境で撮影された動物園のシロクマ映像を用いた評価実験を通じ、実環境における適用可能性と提案手法の頑健性を明らかにする。

## 1.3 本論文の構成

本論文の構成を以下に示す。第 2 章では、動物行動認識およびマルチモーダル表現学習、ならびに敵対的学習に関する関連研究について述べる。第 3 章では、本研究で提案するマルチモーダル表現学習および敵対的学習に基づく手法の詳細を説明する。第 4 章では、実験に用いたデータセットや実装条件、評価指標について述べる。第 5 章では、提案手法の実験結果を示し、単一モダリティ手法との比較や分析を行う。第 6 章では、実験結果に基づく考察を行い、提案手法の有効性や限界について議論する。最後に、第 7 章では本研究のまとめと今後の課題について述べる。

## 第 2 章

# 関連研究

### 2.1 動物行動認識

動物行動認識は、映像データから動物の振る舞いを自動的に解析する技術であり、その応用範囲は基礎的な生態学研究から動物園における飼育管理まで多岐にわたる。動物の行動は、個体の精神的、身体的、および認知的な状態を色濃く反映することが知られており [?], 近年では特に、アニマルウェルフェア（動物福祉）を客観的に評価するための重要な指標として注目されている [?, ?].

中でも、常同行動と呼ばれる反復的な異常行動は、不適切な飼育環境や慢性的なストレスに起因して発現することが報告されており、動物の健康状態や福祉水準を評価する上で極めて重要な指標となる [?]. 従来、これらの行動評価は専門家による長時間の目視観察や記録に依存してきた。しかし、人手による観察は多大な労力と時間を要するだけでなく、観察者間の主観的ばらつきや、24 時間にわたる連続的なモニタリングが困難であるといった課題を抱えている。そのため、低コストかつ継続的な運用が可能であり、定量的な行動データを自動的に取得できるシステムの開発が強く求められている。

こうした背景のもと、近年の深層学習技術の発展は、動物行動解析の分野に大きな変革をもたらした。Convolutional Neural Networks (CNN) や Vision Transformer (ViT) の登場により、画像分類や物体検出の精度は飛躍的に向上し、特定のデータセットにおいては、人間と同等、あるいはそれ以上の認識性能が報告されている。また、DeepLabCut [?] に代表される姿勢推定技術の進展により、映像から動物の関節位置を高精度に推定し、微細な動作を定量的に解析することも可能となった。

## 2.2 実環境における動物行動解析の事例

動物園環境における代表的な解析事例として、Wang ら [?] は、監視カメラ映像を用いたシロクマの常同行動検出手法を提案している。この研究では、時間平均画像を用いた背景差分により動体を検出し、その重心軌跡の周期性を解析することで、往復歩行などの常同行動を特定している。照明変化や天候の影響を受けやすい屋外環境において、ロバストな検出を実現した点は、実運用を想定した重要な成果である。

一方で、同手法は特定の動物種（シロクマ）および移動軌跡という限定的な特徴量に依存しており、移動を伴わない行動や、外観の異なる他種への適用には課題が残る。より多様な行動を種横断的に認識するためには、軌跡情報に加え、身体の動きや外観の変化を含む高次元な表現学習が必要である。

## 2.3 表現学習と行動埋め込み

行動認識におけるアノテーションコストを削減し、未知の行動に対応するためのアプローチとして、「表現学習 (Representation Learning)」および「距離学習 (Metric Learning)」が注目されている。表現学習は、意味的に類似したサンプル（同じ行動）が特徴空間上で近接し、意味的に異なるサンプル（異なる行動）が分離されるような判別的な埋め込み空間 (Embedding Space) を構築することを目的とする [cite: 40]。

FaceNet [?] によって広く知られるようになったトリプレット損失 (Triplet Loss) などの距離学習手法は、クラス分類層を持たないため、学習データに含まれないクラスや、サンプル数の少ない希少な行動に対しても、特徴空間上の距離に基づいてクラスタリングや異常検知が可能となる。動物行動認識の文脈においても、この埋め込みに基づくアプローチは、教師なしあるいは半教師ありの行動解析を実現するための重要な基盤となる [cite: 42]。

## 2.4 敵対的学習による不変表現学習

高次元な特徴表現を学習する際、行動認識に不要な属性情報（動物種や背景など）が特徴量に強く残存すると、汎化性能を阻害する要因となる。この問題に対処する手法として、敵対的学習 (Adversarial Learning) を用いたドメイン適応が提案されている。

Domain-Adversarial Neural Networks (DANN) [?, ?] では、特徴抽出器の後段にタスク分類器とドメイン識別器を配置し、勾配反転層を用いて学習を行う。ドメイン識別器



がドメインを識別できないように特徴抽出器を更新することで、特徴空間からドメイン特有の情報を抑制し、タスクに本質的な情報のみを保持した不変表現の獲得が可能となる。人物行動認識などの分野では、この枠組みの有効性が広く示されているが動物行動認識において、動物種をドメインとして扱い、種に依存しない表現を獲得する試みは未だ十分に検討されていない

## 第 3 章

# 提案手法

本研究では，ラベル付けされたデータが乏しい環境下や，学習時に含まれない未知の動物種に対しても頑健に機能する，行動に基づく特徴表現空間を構築することを目的とする．提案手法は，RGB 映像から得られる外見情報と，オプティカルフローから得られる動作情報を適応的に統合するマルチモーダル学習フレームワークである．さらに，学習された特徴空間から動物種に依存するバイアスを除去するため，敵対的学習に基づく種情報分離モジュールを導入する．

システム全体のアーキテクチャを図??に示す．入力映像はスライディングウィンドウによって複数のクリップに分割され，それぞれ RGB ストリームおよびオプティカルフローストリームによって特徴抽出が行われる．抽出された特徴は Gated Fusion により統合された後，距離学習によって行動クラスごとのクラスタリングが促進される．同時に，敵対的種識別器を用いることで，埋め込み空間から動物種に由来する情報の除去が行われる．

### 3.1 入力動画の前処理

入力映像は，撮影時間やフレーム数が動画ごとに異なるため，可変長の映像をそのまま処理することは困難である．そこで本研究では，映像を一定長のクリップ列に分割するために Sliding Window モジュールを採用する．

具体的には，特徴抽出エンコーダの入力サイズに合わせて，映像を固定フレーム長のクリップへと分割する．例えば，エンコーダが 16 フレームを入力とする場合，フレーム 1-16，2-17，3-18 のように，1 フレームずつ重複させながらクリップを生成する．このような重なりを持つ分割により，映像中の局所的な動作の連続性を保持したまま特徴抽出を行うことが可能となる．

各クリップは独立にエンコーダへ入力され、クリップ単位の特徴ベクトルが得られる。その後、得られた複数の特徴ベクトルは、時間方向に沿って集約され、Adaptive Average Pooling により映像全体を表現する単一の特徴ベクトルへと変換される。この処理により、映像長に依存しない一定次元の埋め込み表現を安定して得ることができる。

Sliding Window を用いた特徴抽出は、フレームを等間隔にサンプリングする手法と比較して、短時間に発生する動作変化や局所的な運動パターンをより詳細に捉えることが可能であり、行動認識において有効である。

## 3.2 RGB 映像に基づく外観特徴抽出

RGB ストリームは、映像フレーム列から動物の外観や姿勢、および周囲環境の文脈情報を捉えることを目的とする。本研究では、事前学習済みの VideoMAE エンコーダを用いて、各クリップから高次元の外観特徴を抽出する。

外見情報の抽出には、事前学習済みの VideoMAE を用いる。VideoMAE は映像内の文脈的特徴や大域的な外見情報を捉える能力に優れており、本研究ではエンコーダ部分の重みを凍結した状態で使用する。これにより、動物の姿勢や周囲環境の文脈情報を 768 次元の特徴ベクトルとして抽出する。

## 3.3 オプティカルフローに基づく運動特徴抽出

オプティカルフローストリームは、動物の動作に起因する時間的变化を捉えるために設計されている。まず、隣接フレーム間の画素移動量を推定するため、RAFT に基づくオプティカルフロー推定を行う。

得られたオプティカルフローは、X3D エンコーダへ入力され、微細な運動パターンを表現する特徴ベクトルへと変換される。オプティカルフローは、被写体の色や模様依存せず、動きそのものを表現できるため、外観差の大きい動物種間においても共通性の高い特徴を抽出できる利点を持つ。

## 3.4 Gated Fusion による適応的特徴統合

外見情報と動作情報は互いに補完的であるが、行動の種類や状況によって重要度は大きく異なる。そこで本研究では、両者を単純に連結するのではなく、Gated Fusion モジュールを用いて適応的な重み付けを行う。

まず、X3D からの特徴量  $\mathbf{E}_{X3D} \in \mathbb{R}^{2048}$  と VideoMAE からの特徴量  $\mathbf{E}_{VideoMAE} \in$

$\mathbb{R}^{768}$  を、線形層を用いて共通の 512 次元空間へ射影する.

次に、以下の式によりゲーティングベクトル  $\alpha \in [0, 1]^{512}$  を算出する.

$$\alpha = \sigma(\mathbf{W}[\mathbf{E}_{X3D}; \mathbf{E}_{VideoMAE}] + \mathbf{b}) \quad (3.1)$$

ここで、 $[\cdot; \cdot]$  はベクトルの連結、 $\mathbf{W} \in \mathbb{R}^{512 \times 1024}$  および  $\mathbf{b} \in \mathbb{R}^{512}$  は学習可能なパラメータ、 $\sigma(\cdot)$  はシグモイド関数を表す.

最終的な統合特徴量  $\mathbf{F}$  は、次式により計算される.

$$\mathbf{F} = \mathbf{E}_{X3D} \odot \alpha + \mathbf{E}_{VideoMAE} \odot (1 - \alpha) \quad (3.2)$$

ここで、 $\odot$  は要素ごとの積を示す. この機構により、モデルは行動や環境条件に応じて、外見情報と動作情報の寄与度を柔軟に調整できる.

### 3.5 敵対的学習による種情報の分離

異なる動物種間での汎化性能を高めるためには、埋め込み空間から種に依存した情報を排除し、純粋な行動情報のみを保持する必要がある. そこで本研究では、Domain-Adversarial Neural Networks (DANN) に基づく敵対的学習を導入する.

統合特徴量  $\mathbf{F}$  に対して種識別器  $D_s$  を接続し、その前段に勾配反転層 (Gradient Reversal Layer: GRL) を配置する. 学習過程において、種識別器は種ラベル  $y_s$  を正しく分類するように学習する一方、特徴抽出器および統合層は、GRL により反転された勾配を受け取ることで、種識別を困難にする特徴表現を生成するように更新される.

敵対的損失  $L_{adv}$  は、以下の交差エントロピー誤差として定義される.

$$L_{adv} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{s,i}^{(k)} \log D_s(\text{GRL}(f(x_i)))^{(k)} \quad (3.3)$$

ここで、 $N$  はバッチサイズ、 $K$  は種クラス数、 $f(x_i)$  は入力映像  $x_i$  に対する特徴抽出および統合処理を表す. 距離学習によって行動判別性が維持されるため、種情報のみを抑制しつつ、行動識別能力を損なわない表現学習が可能となる.

### 3.6 目的関数と最適化

提案手法全体の目的関数  $L_{total}$  は、距離学習による損失  $L_{triplet}$  と敵対的損失  $L_{adv}$  の加重和として定義される.

$$L_{total} = L_{triplet} + \lambda L_{adv} \quad (3.4)$$

ここで、 $L_{triplet}$  は同一行動のサンプルを近づけ、異なる行動のサンプルを遠ざけるトリプレット損失であり、行動識別的なクラスタリングを実現するために不可欠である。

ハイパーパラメータであるトリプレット損失のマージン  $m$ 、および敵対的損失の重み係数  $\lambda$  は、検証データセットにおける ARI (Adjusted Rand Index) を最大化するように最適化を行った。探索範囲は  $m \in [0.05, 0.5]$ 、 $\lambda \in [0.01, 1.0]$  と設定した。

## 第 4 章

# 実験と評価

### 4.1 実験設定

#### 4.1.1 データセット

本研究では、提案手法の有効性と汎化性能を検証するために、学習用として大規模な野生動物データセット、評価用として実際の動物園で撮影された監視映像データセットの 2 種類を使用した。

##### 学習用データセット: Animal Kingdom

学習には、多様な動物種と行動を含む大規模データセットである Animal Kingdom を使用した。このデータセットは 850 種類以上の動物と 140 種類以上の行動クラスを含んでおり、多様な環境下で撮影されているため、汎用的な特徴表現の学習に適している。本研究では、実験の対象を哺乳類（Mammals）に限定し、以下の基準でデータのフィルタリングを行った。

- 100 以上の動画サンプルが存在する行動クラスのみを選定。
- 各動画には単一の動物種のみが映っていること。
- 各クリップの長さが 16 フレーム以上であること。
- 複数の個体が異なる行動をとっている動画は除外する。

結果として、「Walking」「Eating」「Running」などを含む主要な行動クラスを用いてモデルの学習を行った。

### 評価用データセット: シロクマ監視映像

提案手法の未知の環境に対する適応能力を評価するため、札幌市円山動物園のホッキョクグマ館で撮影された監視映像を使用した。データは 2020 年 8 月 30 日に固定カメラを用いて撮影されたもので、解像度は  $368 \times 640$  ピクセルである。Animal Kingdom のようなドキュメンタリー映像とは異なり、照明条件の変化やカメラのブレ、複雑な背景など、実環境特有のノイズを含んでいる。行動ラベルは飼育員の観察記録に基づき、以下のクラスについて、個体が単一の行動を継続している 30 秒間のセグメントを手動で抽出した。

- **基本行動:** Sleeping (睡眠)、Swimming (遊泳)、Walking (歩行)、Keeping still (静止)
- **常同行動 (Stereotypic behavior):** 動物福祉の観点で重要な指標となる、ペーシング (常同歩行) などの反復的な異常行動。

#### 4.1.2 実装詳細とハイパーパラメータ

外見特徴の抽出には ImageNet で事前学習された VideoMAE を、動作特徴の抽出には Kinetics-400 で事前学習された X3D を使用した。これらのバックボーンネットワークの重みは固定し、特徴抽出器として利用した。オプティカルフローの抽出には RAFT を用い、隣接フレーム間の密なフローを推定した。

学習の安定性を確保するため、ハイパーパラメータの最適化には Optuna を使用した。具体的には、Animal Kingdom の検証セットにおける ARI (Adjusted Rand Index) を最大化するように、トリプレット損失の-margin  $m$  を  $[0.05, 0.5]$  の範囲で、敵対的損失の重み係数  $\lambda$  を  $[0.01, 1.0]$  の範囲で探索した。

## 4.2 評価指標

学習された埋め込み空間において、行動クラスがどの程度適切に分離されているかを定量的に評価するため、抽出された特徴ベクトルに対して KMeans++ によるクラスタリングを適用した。評価指標には以下の 2 つを採用した。

1. **ARI (Adjusted Rand Index):** クラスタリング結果と正解ラベルの類似度を測定する指標であり、偶然による一致を補正したもの。値は  $[-1, 1]$  の範囲をとり、1 に近いほど高い性能を示す。
2. **NMI (Normalized Mutual Information):** クラスタリング結果と正解ラベ

ルの相互情報量を正規化したもの。クラスサイズが不均衡な場合でも安定した評価が可能である。

また、定性的な評価として、t-SNE を用いて特徴空間を 2 次元に圧縮・可視化し、クラス間の分離や重なりを視覚的に確認した。

## 4.3 実験結果

### 4.3.1 定量評価

札幌市円山動物園のシロクマデータセット（未知のドメイン）に対するクラスタリング性能の比較結果を表 4.1 に示す。比較対象として、RGB のみを用いたモデル、オプティカルフローのみを用いたモデル、敵対的学習を行わない Gated Fusion モデルを設定した。

表 4.1 シロクマデータセットに対する各手法のクラスタリング性能比較 (ARI および NMI)

| 手法   | ARI          | NMI          |
|--|--------------|--------------|
| RGB Only (Appearance)                      | 0.093        | 0.278        |
| Optical Flow Only (Motion)                 | 0.557        | 0.539        |
| Gated Fusion (Proposed w/o Adv)            | 0.615        | 0.622        |
| <b>Adversarial Gated Fusion (Proposed)</b> | <b>0.742</b> | <b>0.728</b> |

表 4.1 より、以下の傾向が確認された。まず、RGB 情報のみを用いた場合、ARI は 0.093 と極めて低い値にとどまった。これは、学習データ（野生動物のドキュメンタリー映像）と評価データ（動物園の檻の中）の背景や環境が大きく異なるため、モデルが行動そのものではなく背景情報に過学習してしまったことによると考えられる。一方、オプティカルフローのみを用いた場合は ARI 0.557 と大幅な性能向上が見られた。これは、動作情報が動物種や背景の色に依存しにくく、ドメイン間のギャップに対して頑健であることを示唆している。

さらに、Gated Fusion によって両者を統合することで ARI は 0.615 まで向上し、マルチモーダル化の有効性が確認された。最も重要な点として、提案手法である敵対的学習を導入したモデル（Adversarial Gated Fusion）は、ARI 0.742、NMI 0.728 という最高の性能を達成した。これは、敵対的学習によって種固有の特徴（シロクマ特有の外見など）が抑制され、純粋な行動の特徴に基づいた埋め込み空間が形成されたためであると考えられる。



### 4.3.2 定性評価 (可視化)

図?? (別途挿入) に、各モデルによって学習された特徴空間の t-SNE 可視化結果を示す。RGB のみのモデルでは、クラスごとのクラスタが崩れており、多くのサンプルが混在している様子が確認できる。これに対し、提案手法 (Adversarial Gated Fusion) では、「Swimming」や「Walking」などの各行動クラスが明確に分離されたクラスタを形成しており、クラス間の境界も明瞭である。特に、動物福祉の観点で重要となる「Stereotypic behavior (常同行動)」のクラスタも他の行動と明確に区別されており、異常検知への応用可能性を示唆している。

## 4.4 考察と限界

### 4.4.1 ドメイン適応における動作情報の重要性

実験結果より、ドメインシフト (学習データとテストデータの環境差) が大きい場合、RGB による外見情報はノイズとなりやすく、オプティカルフローによる動作情報が支配的な役割を果たすことが明らかになった。しかし、動作情報だけでは静止画的な文脈 (例: 食事中の姿勢など) を取りこぼす可能性があるため、Gated Fusion による適応的な統合が性能向上に寄与したと考えられる。

### 4.4.2 敵対的学習による種不変性の獲得

敵対的学習を導入したモデルがベースラインを大きく上回ったことは、提案手法が「種に依存しない行動表現」の学習に成功していることを裏付けている。種識別器を騙すようにエンコーダを学習させることで、特定種 (この場合はシロクマ) に特化した特徴が削ぎ落とされ、結果として未知の種に対しても汎用的な特徴空間が構築されたと言える。

### 4.4.3 本手法の限界と課題

一方で、いくつかの課題も残されている。第一に、背景の動的ノイズの影響である。オプティカルフローは動物の動きだけでなく、背景の揺れやカメラノイズも拾ってしまうため、これらが行動特徴として誤学習される場合がある。より高度な背景差分技術との併用が望まれる。第二に、複数個体の対応である。本実験では単一個体の映像に限定したが、実際の飼育環境では複数個体が同時に異なる行動をとる場合がある。個体追跡 (Tracking) と行動認識を組み合わせたフレームワークへの拡張が今後の課題である。

## 第 5 章

# 実験結果

5.1 大規模データセットにおける学習結果

5.2 シロクマ映像に対する評価結果

5.3 モダリティ別比較

5.4 敵対的学習の効果分析

5.5 表現空間の可視化

## 第 6 章

# 考察

- 6.1 マルチモーダル統合の有効性
- 6.2 種に依存しない表現の獲得について
- 6.3 動物園実環境への適用可能性
- 6.4 本手法の限界

## 第 7 章

# 結論

### 7.1 本研究のまとめ

### 7.2 今後の課題

# 謝辞

みんなありがとう.

## 参考文献