

卒業論文概要書

Summary of Bachelor's Thesis

Date of submission: 02/07/2020

学科名 Department	情報通信	氏名 Name	藤森詩織	指導 教員 Advisor	渡辺 裕 ㊞
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	1w162328-2 ^{CD}		
研究題目 Title	DeepLabCut を用いた動物の行動分類 Animal Behavior Classification Using DeepLabCut				

1 まえがき

近年、外出先からスマートフォン等でペットの様子をリアルタイムに確認できるペットの見守りカメラが普及してきている。ペットの様子を録画する機能や、ペットとコミュニケーションをとる機能など様々な機能がある。しかし、ペットがどんな行動をしているかは映像を見ないと確認ができない。

外出中のペットの行動を、動画を確認することなく知ることができれば、見守りカメラの需要はさらに高まると考えられる。さらに、行動に加えて運動量と食事量のバランスを知ることができれば、ペットの健康管理にも役立つ。

そこで本研究では、カメラ映像から猫の行動を分類し、猫の運動量と食事量を知る際に重要となる行動を検出することを目的とする。

2 DeepLabCut

DeepLabCut は、深層学習によって実験動物の動画像から、所望の関節位置などを推定し、追跡するためのツールとして Mathis らによって開発された[1]。

DeepLabCut は特定の実験動物に特化した技術ではなく、追跡対象のデータセットを用意することで動物の特徴点推定やその追跡が可能である。これは、転移学習と呼ばれる手法によって実現されている。

また、DeepLabCut は ResNet 構造を持つ。ResNet は Deconvolution 層を出力とすることで対象動物の姿勢を高精度に推定できる。

3 外れ値検知

外れ値検知とは、普段起こりえないようなデータ点を検出する手法である。

外れ値検知の手法の一つに、One Class Support Vector Machine (OneClassSVM) がある。この手法は、Support Vector Machine (SVM) を教師なしの1クラス分類に応用した手法で、ほとんどが正常データであるようなデータ分布で異常検知が可能である。「正常」の1クラスのみを有し、識別境界を設定することで正常か異常かを判定する。

4 提案手法

本研究では、猫の運動量を知る際に重要となる、「上下運動」と食事量を知る際に重要となる「食事」を含めた、「座る」「寝そべる」「食事」「歩行」「上下運動」の五つの行動を分類する。

移動物体を検出する従来手法として、背景差分法やフレーム間差分法が挙げられる。しかし、背景差分法は日照条件の変化などに弱い。また、フレーム間差分法は移動物体がしばらく静止した場合に、移動物体を見失う。

そこで、DeepLabCut による骨格情報を用いた手法を提案する。提案手法の処理の流れを図1に示す。用意したデータセットを用いて学習させた学習済み DeepLabCut を使用し、入力動画の姿勢推定を行う。姿勢推定によって得られた座標値を用いて、行動分類に必要な特徴量をフレームごとに算出する。最後に分類器を用いてクラス分類を行い、行動を分類する。

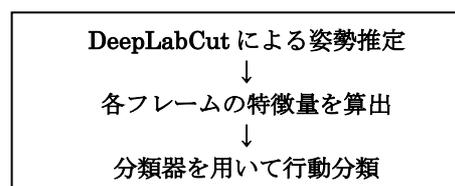


図1 提案手法

「座る」「寝そべる」「食事」は「静止中」に行う行動（姿勢）であり、「歩行」「上下運動」は「運動中」に行う行動である。提案手法では静止と運動に分けてそれぞれで分類を行う。

静止中の「座る」「寝そべる」「食事」の分類には、関節角度を特徴量として用いて、SVMによって姿勢分類を行う。運動中の「歩行」「上下運動」の分類には、前足の付け根の y 座標の変化率を特徴量として用いる。「上下運動」が起きた場合、y 座標に急激な変化が起きる。この変化率を外れ値として検出することで、「歩行」と「上下運動」を分類する。

5 実験

5.1 静止時の姿勢分類実験

静止時の姿勢分類実験では、各行動 1177 フレームで 1920×1080[pe]のサイズの動画を使用した。DeepLabCut を用いて得られた足の座標から、前足と後足の関節角度を算出した。両足の関節角度を特徴量に用いた場合と、前足のみ、後ろ足のみで SVM によるクラス分類をした。「寝そべる」「座る」の 2 クラス分類と、「食事」を加えた 3 クラス分類の分類正解率をそれぞれ表 1、表 2 に示す。

2 クラス分類では、両足を特徴量として用いた場合、100%の精度で分類に成功した。しかし、「食事」を加えた場合には、精度が低下した。頭の動きを考慮するなど、特徴点をさらに加えて精度を向上させる必要がある。また、後ろ足のみを特徴量として用いた場合には、51.3%と低い分類精度となった。

表 1 静止時の姿勢分類の正解率

使用した 関節角度	「寝そべる」 「座る」	「寝そべる」 「座る」「食事」
両足	1.000	0.937
前足	1.000	0.871
後足	0.755	0.513

5.2 運動時の行動分類実験

従来手法と提案手法の比較実験をした。総フレーム中 4 割を上下運動フレーム、残りの 6 割が歩行フレームである 1920×1080[pe]サイズの動画を使用した。使用する特徴点座標を DeepLabCut、背景差分

法、フレーム間差分法の 3 つの手法で取得した。提案手法では、DeepLabCut を用いて前足の付け根の座標を取得した。従来手法である背景差分法とフレーム間差分法では、差分画像の重心座標を取得した。

取得した y 座標からフレーム間の y 座標変化率を算出し、OneClassSVM を用いて外れ値検知をした。実験結果を表 2 に示す。表 2 中の正常標本精度は正常サンプルを正しく正と予測できた割合を表す。異常標本精度は異常サンプルを正しく異常と予測できた割合を表す。

表 2 の F 値を比較すると、提案手法が従来手法よりも高い精度で外れ値検知に成功したことが分かる。背景差分法による差分画像には影が映っていた。また、フレーム間差分法では猫の動きが小さい場合に差分画像をうまく取得できなかった。これらの原因で、提案手法の精度のほうが高くなった。

表 2 比較実験の精度

	正常標本 精度	異常標本 精度	F 値
提案手法	0.714	0.571	0.645
背景差分法	0.762	0.500	0.604
フレーム間 差分法	0.667	0.500	0.571

6 まとめ

本研究では、DeepLabCut を用いて猫の行動を分類する手法を提案した。DeepLabCut は影や明るさの影響を受けることなく、指定した特徴点を追跡し続けることができる。この点により、従来手法である背景差分法やフレーム間差分法よりも移動物体を高精度に検出できることが確認できた。

しかし、体の一部が家具に隠れた場合には特徴点を取得できなくなるため、時系列の情報を用いるなど手法を改善させる必要がある。

参考文献

- [1] A. Mathis, P. Mamidanna1, "Markerless tracking of user-defined features with deep learning," arXiv preprint arXiv: 1804.03142, 2018.

2019年度 卒業論文

DeepLabCut を用いた動物の行動分類

Animal Behavior Classification Using DeepLabCut

指導教員 渡辺 裕 教授

早稲田大学 基幹理工学部

情報通信学科

1W162328-2

藤森 詩織

目次

第1章 序論	3
1.1 研究の背景.....	3
1.2 本研究の目的.....	3
1.3 本論文の構成.....	4
第2章 関連技術	5
2.1 まえがき	5
2.2 DeepLabCut	5
2.2.1 DeepLabCut の構造	5
2.2.2 DeepLabCut の検出精度.....	6
2.3 移動物体検知.....	8
2.3.1 背景差分法.....	8
2.3.2 フレーム間差分法.....	8
2.4 SVM	9
2.5 外れ値検知.....	9
2.5.1 IsolationForest	9
2.5.2 OneClassSVM	10
2.6 むすび.....	10
第3章 提案手法	11
3.1 まえがき	11
3.2 手法の概要.....	11
3.3 特徴量の取得.....	12
3.4 行動の分類.....	13
3.5 むすび.....	15
第4章 実験	16
4.1 まえがき	16
4.2 静止時の姿勢分類実験	16
4.2.1 データセット	16
4.2.2 評価指標	16
4.2.3 2値分類の実験結果	18
4.2.4 3値分類の実験結果	19
4.2.5 考察	20
4.3 運動時の行動分類実験	21
4.3.1 データセット	21

4.3.2 評価指標	22
4.3.3 従来手法との比較	23
4.3.4 実験結果	24
4.3.5 考察	27
4.4 提案手法による行動の分類精度	28
4.5 むすび	28
第5章 結論と今後の課題	29
5.1 結論	29
5.2 今後の課題	29
謝辞	30
参考文献	31
図一覧	32
表一覧	33

第1章 序論

1.1 研究の背景

近年、外出先からスマートフォン等でペットの様子をリアルタイムに確認できるペットの見守りカメラが普及してきている。ペットの様子を撮影し録画する機能に加え、マイクやスピーカーを内蔵することで、外出先からペットとコミュニケーションをとる機能がある。また、温度センサーが搭載されていて、室内の温度が設定温度の範囲外になるとスマートフォンに通知をする機能がある。スマートフォンで遠隔操作が可能な家電が増えてきているため、外出先から空調を管理することが可能である。さらに深層学習を活用し、ペットが写っている部分の映像だけを抽出することが可能なカメラアプリも存在する。しかし、ペットがどんな行動をしているかは映像を見ないと確認することができない。

外出中のペットの行動を、動画を確認することなく知ることができれば見守りカメラの需要はさらに高まると考えられる。さらに、行動に加えて運動量と食事量のバランスを知ることができれば、ペットの健康管理にも役立つ。

1.2 本研究の目的

一般社団法人ペットフード協会の調査によると、2018年の全国の犬・猫推計飼育頭数は、犬が890万3千頭に対し、猫は946万9千頭と猫の飼育頭数は犬を上回っていた[1]。

そこで本研究では、カメラ映像から猫の行動を分類し、猫にとって運動となる行動と食事時の姿勢を検出することを目的とする。

猫は高いところを好む習性がある。また、室内だけで飼われることが多い家庭の猫は、運動不足であることが多いと言われている。高い場所への上り下りは、猫の運動不足解消のためには必要な行動である。そのため猫の運動量を知るには、「上下運動」を猫の行動から検出することが必要である。

1.3 本論文の構成

本論文の構成を以下に示す.

第 1 章は本章であり, 本研究の背景, 目的について述べる.

第 2 章では本研究に関連する技術について述べる.

第 3 章では本研究で提案する手法について述べる.

第 4 章では提案手法を用いた実験, 結果および考察について述べる.

第 5 章では本論文の結論と今後の課題について述べる.

第2章 関連技術

2.1 まえがき

本章では、本研究で用いる技術である DeepLabCut について述べ、固定式カメラの場合に移動物体を検知する手法として従来用いられていた基本的な手法である背景差分法とフレーム間差分法について述べる。また、教師あり学習を用いるパターン認識モデルの一つである SVM および機械学習を用いた外れ値検知の手法である IsolationForest と OneClassSVM について述べる。

2.2 DeepLabCut

DeepLabCut は、ディープラーニングによって実験動物の動画像から、所望の関節位置などを推定し、追跡するためのツールとしてドイツエバーハルト・カール大学テュービンゲンの Alexander Mathis らの研究者グループによって開発された[2]。GitHub からダウンロードして利用することができる[3]。

DeepLabCut は特定の実験動物に特化した技術ではなく、追跡対象のデータセットを用意することで高精度に動物の特徴点推定やその追跡が可能である。これは、転移学習と呼ばれる手法によって実現されている。DeepLabCut の詳細な構造に関しては 2. 2. 1 節で述べる。

2.2.1 DeepLabCut の構造

DeepLabCut は Deeper, Stronger, and Faster Multi-Person Pose Estimation Model (DeeperCut) を基に構築されている。

DeeperCut について簡単に説明する。DeeperCut は Elder らによって提案されたボトムアップ型の姿勢推定技術である[4]。ボトムアップ型の姿勢推定は、検出する身体部分の候補点となるキーポイントを画像から抽出し、抽出されたキーポイントを、人物ごとにクラスタリングし複数人物の姿勢推定を行う手法である。DeeperCut のキーポイント抽出には Residual Network (ResNet) とよばれる構造が使われている。ResNet は Microsoft Research によって 2015 年に提案されたニューラルネットワークの一種である[5]。画像分類問題において、ネットワークの層の深くすることでより高次元の特徴を獲得することができるが、ただネットワークの層を重ねるだけでは性能が劣化してしまった。これは勾配消失問題が原因であった。ResNet では shortcut

connection を導入することで、勾配消失問題を解決し、152 もの層を重ねることを達成した。

また、キーポイントを抽出するだけでなく、クラスタリングの補助となる出力を ResNet の出力に追加している。具体的には、肘なら肘、膝なら膝というような身体部分ごとの位置を表すベクトル場と、各キーポイントのペアに関するベクトル場の二つである。ペアに関するベクトル場というのは、あるキーポイントから他の部位への線分を出力し、関節点ペアの距離や角度など、部位ごとの関係性を表すベクトル場である。ResNet でキーポイントを抽出し、各身体部位に関する出力と部位ごとの関係性に関する出力の二つを ResNet の出力に追加することで、DeeperCut は高精度に姿勢推定を行うことを可能とした。

DeepLabCut では DeeperCut の ResNet 構造に、出力として Deconvolution 層を追加している。Deconvolution 層は各関節位置をそれぞれ異なる画像として表現し、各画像に対する信頼度を学習する。

また、DeepLabCut は ImageNet とよばれるデータセットによって事前学習されている。これは転移学習と呼ばれる手法であり、転移学習によって少数の訓練データを学習させるだけで、追跡対象の身体部分を推定し追跡することが可能となった。

2.2.2 DeepLabCut の検出精度

Alexander Mathis らは、マウスを対象として、DeepLabCut の骨格検出精度を調べる実験を実施した。

まずは、人間の手動ラベリングの変動性を定量化するため実験が行われた。全 1080 フレームから成る動画から複数のフレームを抽出し、すべてのフレームに関してマウスの鼻、耳（左右）、尻尾の付け根を手動でラベル付けした。同じデータを使用し、手動ラベリングを 2 回行い、2 回の試行間の平均平方二乗誤差 (RMSE : Root Mean Squared Error) を求めた。これを "Human Variability" とした。

次に、DeepLabCut による自動ラベリングの実験を行った。DeepLabCut による自動ラベリングの流れは、用意した動画の全フレームのうち、一部を訓練データとして使用し、手動でラベリングする。ラベル付けされた訓練データによって DeepLabCut のネットワークを学習させ、学習済み DeepLabCut で残りフレームを自動ラベリングさせる。自動ラベリング結果と正解ラベルのそれぞれのピクセル値間の RMSE を Human Variability と比較することにより、DeepLabCut による自動ラベリングと人間による手動ラベリングの精度を比較した。

人間の変動性の定量化実験と同じデータを使用し、データの 8 割を訓練データとしてネットワークを学習させ、残りのデータを DeepLabCut によって自動ラベリングさせたところ、鼻と尻尾は人間レベルで検出ができていることが確認できた。

また、訓練データの割合を変化させると検出精度にどの程度影響するかを調査する実験が行われた。訓練データの割合を1%、5%、10%、20%、50%、80%とした6種類のネットワークを学習させた。訓練データ数の変化による自動ラベリングのRMSEを図2-1に示す。自動ラベリングのRMSEを実線で示し、Human Variabilityを点線で示している。訓練データの割合が80%の場合、人間と同じレベルの検出精度となっていることがわかる。訓練データの割合が10%を超えるとRMSEは5ピクセル以下の値をとっている。実験で使用した動画中のマウスの鼻のサイズは約30ピクセルであり、5ピクセル以下の誤差になっていることは高い精度で検出ができたと言える。つまり、DeepLabCutを使用すると、動画の全フレームの10%程度を訓練データとすると、指定した特徴点を高い精度で検出できる。

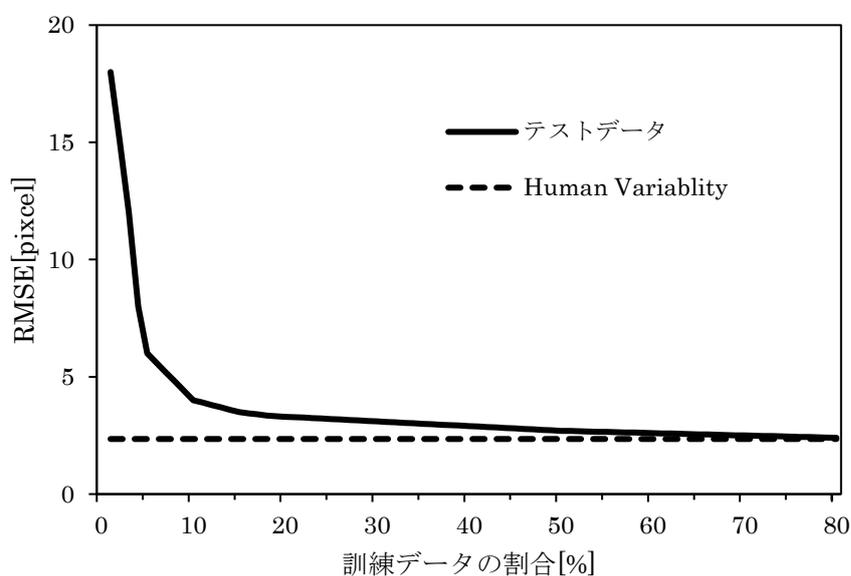


図 2-1 訓練データ数と RMSE の関係

出典：Math, PranavMamidanna1”Markerless tracking of user-defined features with deep learning” arXiv preprint arXiv:1804.03142, 2018. P. 3

マウスを追跡対象とした実験では、動き回るマウスを真上から撮影した動画が使用されており、検出する特徴点とした鼻、耳そして尻尾の付け根は常にカメラから隠れることはなかった。DeepLabCut のさらなる柔軟性を実証するために、立体環境中を自由に行動するショウジョウバエを追跡対象とした実験も行われた。固定カメラからショウジョウバエを見ると、図2-2のように体の一部がカメラから見えないフレームが存在し、また体の向きが大きく変化していることがわかる。しかし、そのような条件でも指定した身体部分を検出できている。つまり、身体部位が一時的に隠れることや体の向きの変化はDeepLabCutの精度に大きな影響を与えないことが分かった。

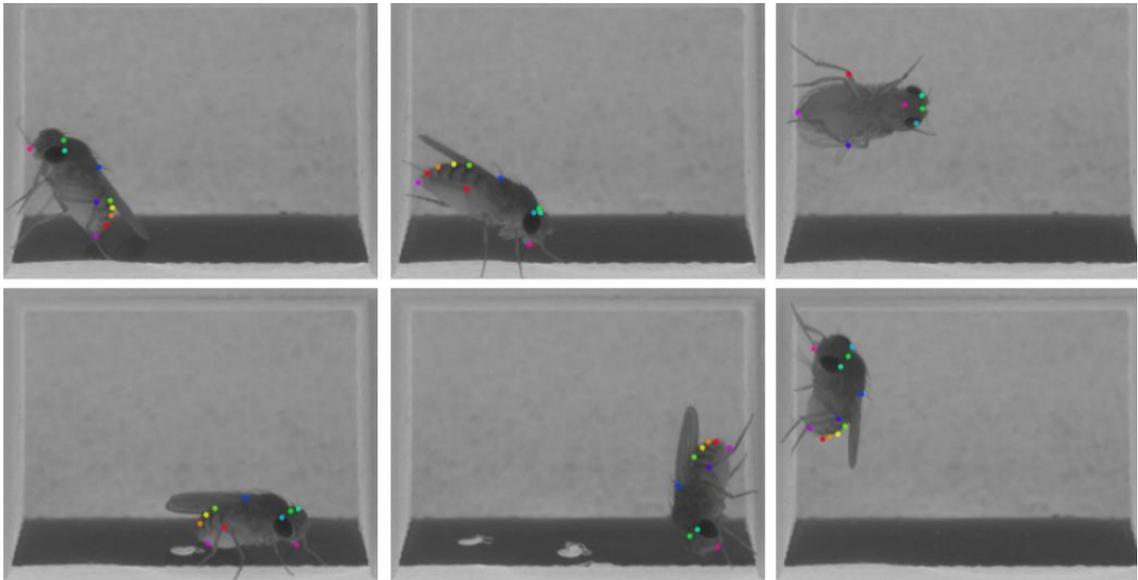


図 2-2 DeepLabCut によるショウジョウバエのトラッキング

出典 : Alexander Math, PranavMamidanna1, “Markerless tracking of user-defined features with deep learning”, arXiv preprint arXiv: 1804. 03142, 2018. p. 6

2.3 移動物体検知

固定式カメラ中に映る移動物体を検知する代表的な手法として、背景差分法とフレーム間差分法がある。それぞれの手法の原理、利点および欠点について 2.3.1 節と 2.3.2 節で述べる。

2.3.1 背景差分法

背景差分法は、移動物体の映っていない背景画像を用意し、入力画像から背景画像を引いた差分領域を移動物体とする手法である。背景差分法は、背景の変化に弱い。特に屋外の映像を使用した場合、木々の揺れや時間と共に変化する日照条件などの影響を受ける。

2.3.2 フレーム間差分法

フレーム間差分法は、連続するフレーム間の差分から移動物体を検出する手法である。背景差分法とは異なり、背景画像を用意する必要がない。また、背景の環境変化に強い。しかし、物体が一時停止した場合に、連続フレーム間の差が無くなり、移動物体を見失う。

2.4 SVM

Support Vector Machine (SVM) は、1960年代に Vapnik らによって考案された教師あり学習を用いたパターン認識手法の一つである[6]。Vapnik らによって考案された SVM は線形の識別手法であったが、Boser らによってカーネル学習法と組み合わせた非線形の識別手法へと拡張された[7]。

SVM は、学習ベクトルの中で最も他クラスと近い位置にあるデータをサポートベクトルとして、その他クラスとの距離が最大になるような位置に識別境界を引く。これはマージン最大化と呼ばれる。境界線とサポートベクトルとの距離であるマージンを大きくすることによって誤判別を防ぐ。

2.5 外れ値検知

外れ値検知とは、系列データに対する異常検知の代表的な手法のうちの一つである。外れ値検知の他には、異常部位検出と変化点検知が挙げられる。外れ値検知とは、普段起こりえないようなデータ点を検知する手法である。異常部位検出とは、異常が起きている部分時系列を検出する手法である。変化点検知とは、時系列データのパターンが急激に変化する箇所を検知する手法である。

外れ値検知の手法として、IsolationForest と OneClassSVM についてそれぞれ 2.4.1 節と 2.4.2 節で説明する。

2.5.1 IsolationForest

IsolationForest は Liu らによって提案された決定木を拡張させた手法である[8]。異常値は正常値よりも少ないであろうという前提のもとに外れ値を検知する。観測値からランダムに値を選択し、選択した値を分割する値をランダムに決め、観測値を分割していく。図 2-3 のような観測値の分布があったとき、値 a は右の木構造のように分割される。異常値は正常値よりも少ないという前提であるため、最も浅い木に分割された値が異常値であるとする。図 2-3 のような木構造を複数作成し、終端ノードまでのパス長の平均を最終的な異常値とする。

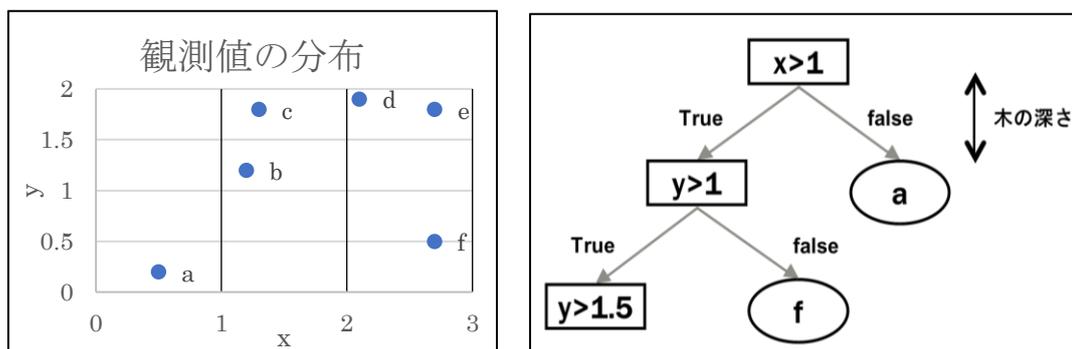


図 2-3 IsolationForest による観測値の外れ値検知

2.5.2 OneClassSVM

One Class Support Vector Machine (OneClassSVM) は、SVM を教師なしの 1 クラス分類に応用した手法で、ほとんどが正常データであるようなデータ分布で異常検知が可能である。SVM は学習データとして複数クラスのデータが用意される。識別境界を決定することで、データのクラス分類を行う。一方、OneClassSVM は「正常」の 1 クラスのみを有する。よって、識別境界を設定し、正常であるか異常であるかを判定する。

2.6 むすび

本章では、本研究で用いる技術である DeepLabCut とパターン認識モデルの一つである SVM および外れ値検知の手法である IsolationForest と OneClassSVM について述べた。また、移動物体検知の従来手法として、背景差分法とフレーム間差分法について述べた。

第3章 提案手法

3.1 まえがき

本章では、DeepLabCut を用いて、定点カメラで撮影された猫の行動を分類する手法を提案する。

3.2 手法の概要

本研究では、1.3 節で述べたように猫の運動量を知る際に重要となる、「上下運動」と食事量を知る際に重要となる「食事」を含めた、「座る」「寝そべる」「食事」「歩行」「上下運動」の五つの行動を分類する。

移動する物体を検知する方法としては、2.3 節で述べたような、背景差分法やフレーム間差分法が存在する、しかし、今回は室内で飼われる猫の行動分類を目的としている。飼い主の留守中は電気が消えており、日照条件によって背景画像が変化することから、背景差分法は適さない。また、猫がしばらく静止した場合に猫を見失う場合があることからフレーム間差分法は適さないと言える。

また、本研究で分類動作としている「上下運動」の検出には、3次元の位置情報を取得することが効果的である。従来3次元情報の取得には、ステレオカメラを用いる手法が多く挙げられている。しかし、ステレオカメラは、構造上2台以上のカメラが必須であり、事前にキャリブレーションの必要があるなど一般家庭の使用には適していない。

DeepLabCut を用いることで、室内中の猫を見失うことなく追跡し、姿勢推定によって得られる垂直方向の情報から「上下運動」を分類することが可能であると考えた。さらに得られた座標情報から関節角度を計算し、用いることも可能となる。これらの理由により、DeepLabCut を用いた手法を提案する。

提案手法での、処理の流れを図3-1に示す。用意したデータセットを用いて学習させた学習済みDeepLabCutを使用し、入力動画の姿勢推定を行う。姿勢推定によって得られた、座標値を用いて、行動分類に必要な特徴量をフレームごとに算出する。最後に分類器を用いてクラス分類を行い、行動を分類する。

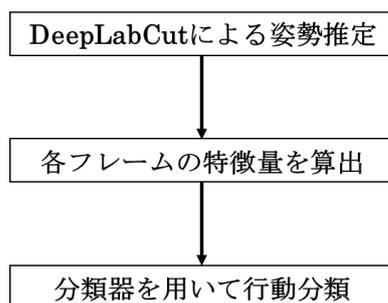


図 3-1 提案手法の流れ

分類に用いる特徴量について 3.3 節で述べ、詳しい行動の分類方法について 3.4 節で述べる。

3.3 特徴量の取得

猫の足のつま先、関節、付け根を特徴点として指定した。図 3-2 のように各足 3 箇所ずつ、合計 12 箇所の座標データを、DeepLabCut を用いて取得する。取得された座標を用いて、つま先と付け根の間の関節角度を求める。つま先、関節、付け根の座標をそれぞれ A, B, C とし、3 点の座標から次のように二つの二次元ベクトルを定義する。

$$\vec{a} = BA$$

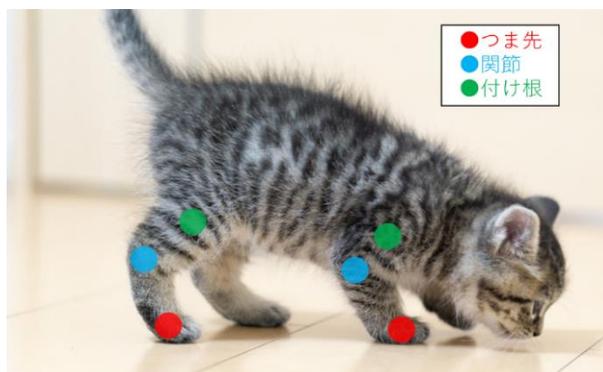
$$\vec{b} = BC$$

このとき、A, B, C のなす角 θ は、

$$\theta = \arctan\left(\frac{\vec{a} \times \vec{b}}{\vec{a} \cdot \vec{b}}\right) \quad (3.3)$$

によって求められる。

各特徴点の座標値と、式 (3.3) で求めた各足の関節角度を用いて行動分類を行う。



出典：フリー素材ぱくたそ (www.pakutaso.com)

図 3-2 取得する特徴点

3.4 行動の分類

「座る」「寝そべる」「食事」「歩行」「上下運動」の五つの行動を分類する処理の流れを図 3-3 に示す。五つの行動は「静止中」と「運動中」に行う行動に分けることができる。「座る」「寝そべる」「食事」は「静止中」に行う行動（姿勢）であり、「歩行」「上下運動」は「運動中」に行う行動である。

フレームごとに取得した座標情報を用いて、フレーム間で座標変化がある場合は「運動」している、座標変化がない場合は「静止」しているといえる。静止中の「座る」「寝そべる」「食事」の分類には、関節角度を特徴量として用いて、SVM によって姿勢分類を行う。

「座る」「寝そべる」「食事」の姿勢の例をそれぞれ図 3-4, 図 3-5, 図 3-6 に示す。図からわかるように「座る」と「寝そべる」では前足の関節角度に大きな違いがあることがわかる。「食事」は「座る」の姿勢よりも頭が下がり、前足の関節角度が小さくなっていることがわかる。

運動中の「歩行」「上下運動」の分類には、前足の付け根の y 座標の変化率を特徴量として用いる。歩行中には、垂直方向の大きな変化は起こらないが、「上下運動」が起きた場合には垂直方向に急激な変化が起きると考えられる。「上下運動」をした場合の垂直方向への座標変化の例を図 3-6 に示す。棒グラフは y 座標の前フレームとの変化率を表す。このように y 座標に大きな変化がある、つまり「上下運動」が起きると変化率は急激に大きな値をとる。閾値を設定して、閾値を超える変化率となったフレームを「上下運動」が起きていると検出することで、「歩行」と「上下運動」を分類できると考えた。異常検出手法のうち外れ値検知を用い、変化率の異常値を検出することで閾値を設定する。

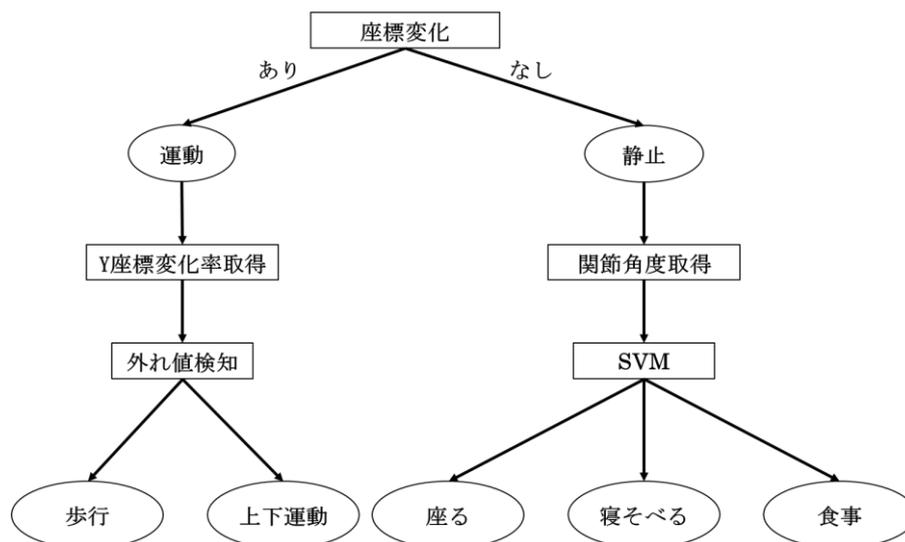


図 3-3 分類手法の処理の流れ



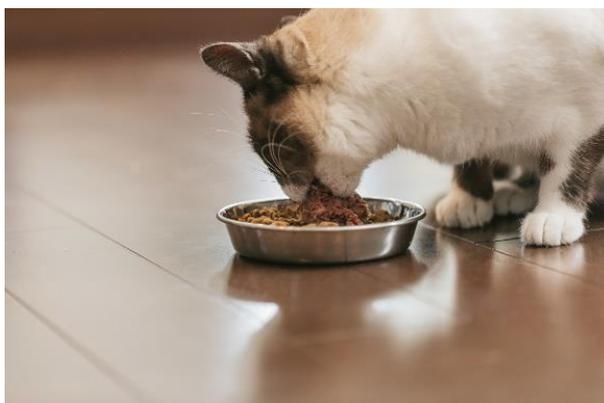
出典：フリー素材ぱくたそ (www.pakutaso.com)

図 3-4 「座る」の例



出典：フリー素材ぱくたそ (www.pakutaso.com)

図 3-5 「寝そべる」の例



出典：フリー素材ぱくたそ (www.pakutaso.com)

図 3-6 「食事」の例

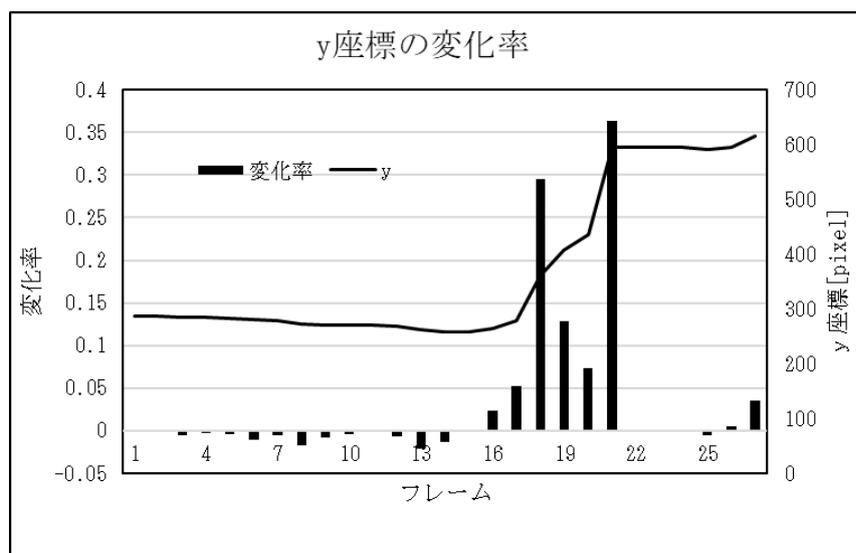


図 3-7 「上下運動」による y 座標変化の例

3.5 むすび

本章では, DeepLabCut によって行動を分類する手法について述べた.

第4章 実験

4.1 まえがき

本章では、提案手法による静止時と運動時それぞれのデータセット、評価指標、実験結果および考察について述べる。さらに提案手法を用いた場合の行動の分類精度について述べる。

4.2 静止時の姿勢分類実験

提案手法を用いて分類する行動のうち、静止時の「座る」と「寝そべる」と「食事」の分類実験に使用したデータセット、評価指標、実験結果および考察について述べる。

4.2.1 データセット

本実験では、一つの動画ではなく、複数の動画をデータセットとして使用した。また、それぞれの動画は同じ猫ではなく、異なる種類の猫が撮影された動画を使用した。各行動における動画の総フレーム数を表 4-1 に示す。使用した動画は、右の前足と右の後ろ足の 6 箇所の特徴点、または左の前足と左の後ろ足の 6 箇所の特徴点のいずれかが必ず全て映っているものとした。また、1920×1080[pe]のサイズの動画を使用した。

表 4-1 実験に使用したデータセット

行動	フレーム数
座る	1177
寝そべる	1177
食事	1177

4.2.2 評価指標

クラス分類の評価指標である、正解率(Accuracy)、適合率(Precision)、再現率(Recall)、F 値 (F-measure) について述べる。

2 値分類における予測値と真値の関係を表す混同行列を図 4-1 に示す。

真値 \ 予測値	正	負
正	True Positive (TP)	False Negative (FN)
負	False Positive (FP)	True Negative (TN)

図 4-1 2 値分類の混同行列

正解率は全サンプルのうち正解したサンプルの割合を表す。正解率は式 (1.2.1.1) によって表される。

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1.2.1.1)$$

入力サンプルに正が多いなど偏りがある場合には、正解率以外の指標が使われることが多い。

適合率は、精度と訳されることもあり。正と予測されたサンプル中の実際に正であるサンプルの割合を表す。適合率は式 (1.2.1.2) で表される。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1.2.1.2)$$

再現率は、真値が正であるサンプルを正と判定できた割合を表す。再現率は式 (1.2.1.3) で表される。

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1.2.1.3)$$

適合率と再現率はトレードオフの関係にあり、適合率が高いシステムは、再現率が低く、再現率が高いシステムは適合率が低い。そのため適合率と再現率の調和平均である、F 値で分類精度を評価することもある。F 値は式 (1.2.1.4) で表される。

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1.2.1.4)$$

多クラス分類の場合も、2値分類と同様に混同行列で表すことができる。3値分類における混同行列を図4-2に示す。ここで図中のFA(B)は真値がBであるが、予測値がAであったサンプル数を表す。

真値 \ 予測値	A	B	C
A	TA	FB (A)	FC (A)
B	FA (B)	TB	FC (B)
C	FA (C)	FB (C)	TC

図 4-2 3 値分類の混同行列

クラスAに関する適合率,再現率,F値はそれぞれ式(1.2.1.5),式(1.2.1.6),式(1.2.1.7)で表される。

$$\text{Precision (A)} = \frac{TA}{TA + FA (B) + FA (C)} \quad (1.2.1.5)$$

$$\text{Recall (A)} = \frac{TA}{TA + FB (A) + FC (A)} \quad (1.2.1.6)$$

$$\text{F - measure (A)} = \frac{2 \times \text{Precision (A)} \times \text{Recall (A)}}{\text{Precision (A)} + \text{Recall (A)}} \quad (1.2.1.7)$$

クラスB,クラスCに関しても同様に表すことができる。

各クラスのごとに求められた評価指標を平均した値はマクロ平均と呼ばれる。本実験ではマクロ平均の値を多クラス分類の評価指標に使用している。

4.2.3 2 値分類の実験結果

「寝そべる」と「座る」はそれぞれ1177フレームを実験に使用した。各フレームの前足と後ろ足の関節角度の分布を図4-3に示す。

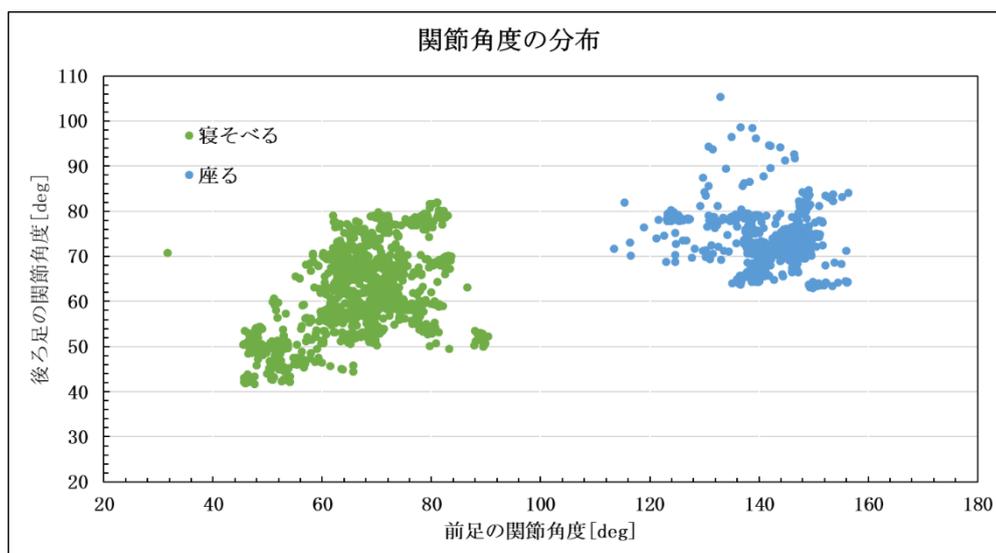


図 4-3 「寝そべる」と「座る」の関節角度分布

前足と後ろ足の関節角度の両方を特徴量として使用した場合、前足の関節角度のみを特徴量として使用した場合、後ろ足の関節角度のみを特徴量として使用した場合の 3 通りで実験した。交差検証を用いた SVM による 2 クラス分類の精度を表 4-2 に示す。

表 4-2 2 クラス分類の精度

使用した 関節角度	正解率	適合率	再現率	F 値
両足	1.0000	1.0000	1.0000	1.0000
前足	1.0000	1.0000	1.0000	1.0000
後足	0.7550	0.5716	0.4620	0.5002

4.2.4 3 値分類の実験結果

「食事」中の姿勢を分類行動に加えた実験を行った。1177 フレームの「食事」のデータを 2 値分類のデータに追加し、SVM による 3 値分類を行った。各行動の前足と後足の関節角度の分布を図 4-4 に示す。

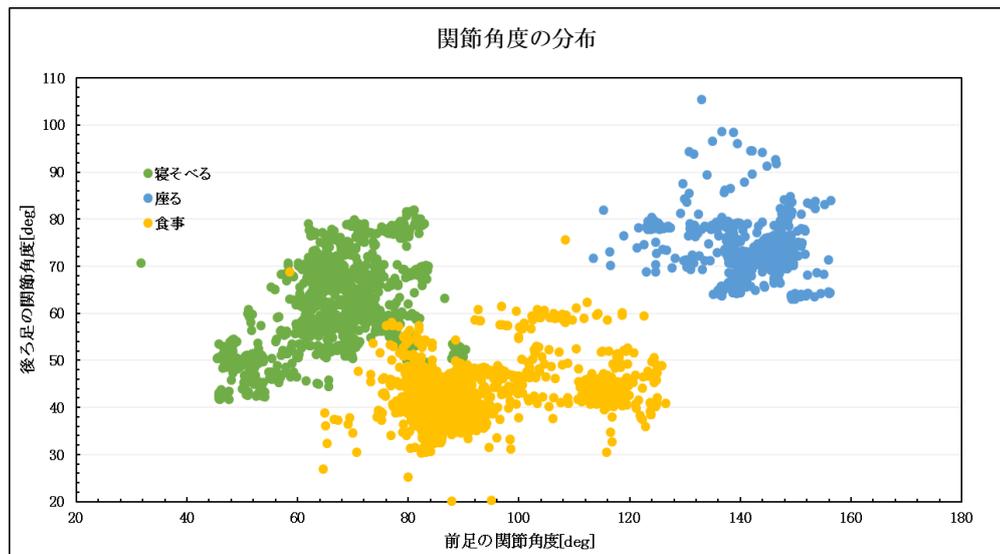


図 4-4 「寝そべる」と「座る」と「食事」の関節角度分布

交差検証による 3 値分類の精度を表 4-3 に示す。

表 4-3 3 クラス分類の精度

使用した 関節角度	正解率	適合率	再現率	F 値
両足	0.9370	0.6418	0.6394	0.6362
前足	0.8712	0.5289	0.5089	0.5059
後足	0.5125	0.4187	0.3158	0.3472

4.2.5 考察

両足または前足のみを特徴量として用いた場合と、後ろ足のみを特徴量として用いた場合を比較すると後ろ足のみを用いた場合は、他の二つの場合よりも大きく精度が低下した。つまり、前足の関節が隠れた場合に分類精度が低下することがわかる。

また、両足の関節角度を特徴量として用いた場合、「寝そべる」と「座る」は 100% の精度で分類に成功した。図 4-3 からわかるように、「寝そべる」と「座る」の姿勢は、前足の角度の分布が大きく異なる。「寝そべる」は前足の関節角度は 100° 以下に分布しているのに対し、「座る」では 100° 以上に分布している。これにより、前足の関節角度を特徴量に加えた場合、分類精度が高くなった。

しかし、「食事」を分類に加えると精度が低下した。図 4-4 の関節角度の分布からわかるように、「食事」は、「寝そべる」の分布と被っている部分が存在する。関節角度だけでなく、頭の位置や動きを特徴量に加えることで分類精度が上がると考えられる。

4.3 運動時の行動分類実験

提案手法を用いて分類する行動のうち、運動時の「歩行」と「上下運動」の分類実験に使用したデータセット、評価指標、従来手法との比較、実験結果および考察について述べる。

4.3.1 データセット

「上下運動」と「歩行」を含む四つの動画を用い行動の分類を行った。それぞれ 1920×1080 [pel]サイズの動画を使用した。

それぞれの動画を上下運動 (1) , 上下運動 (2) , 上下運動 (3) , 上下運動 (4) とする。前足の付け根を特徴点として、その y 座標と y 座標変化率をそれぞれ図 4-5, 図 4-6, 図 4-7, 図 4-8, 図 4-9, 図 4-10, 図 4-11, 図 4-12 に示す。また、四つの動画の総フレーム数と、「歩行」と「上下運動」の各フレーム数それぞれを表 4-4 に示す。

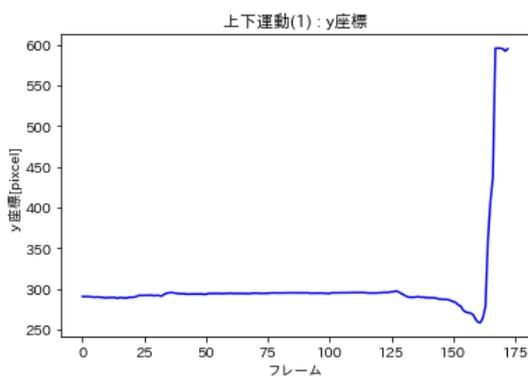


図 4-5 上下運動 (1) の y 座標

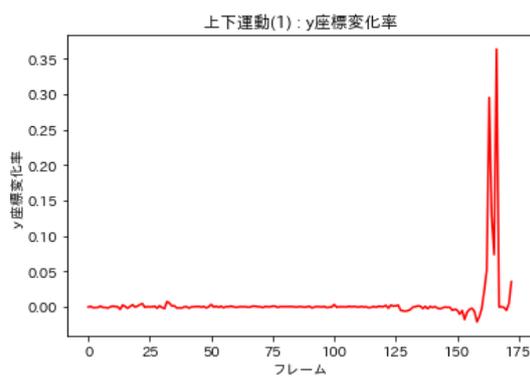


図 4-6 上下運動 (1) の y 座標変化率

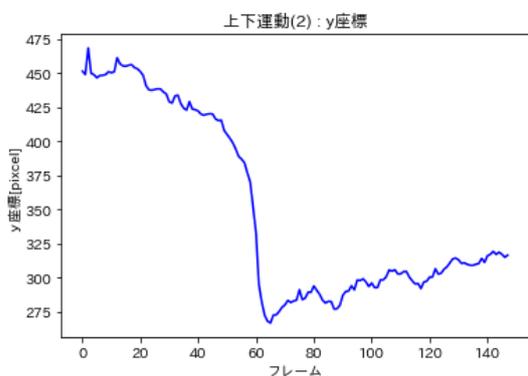


図 4-7 上下運動 (2) の y 座標

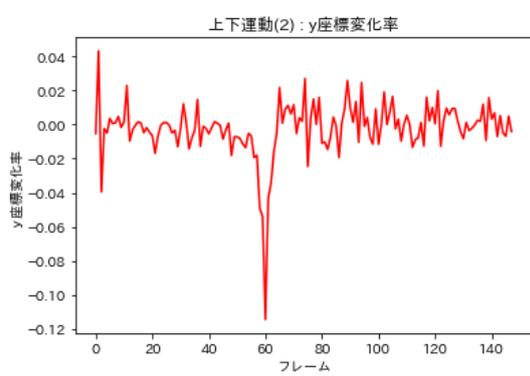


図 4-8 上下運動 (2) の y 座標変化率

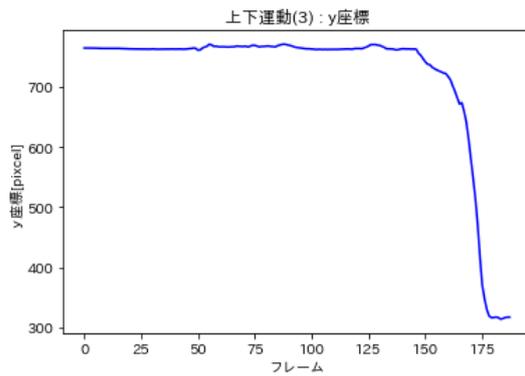


図 4-9 上下運動 (3) の y 座標

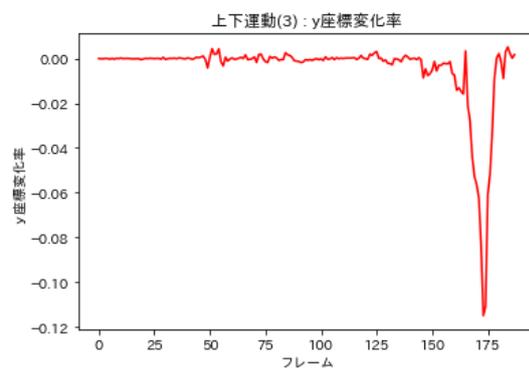


図 4-10 上下運動 (3) の y 座標変化率

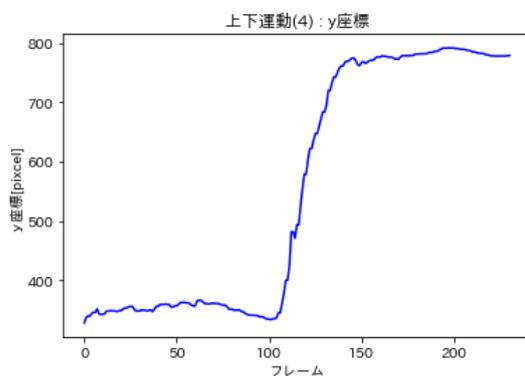


図 4-11 上下運動 (4) の y 座標

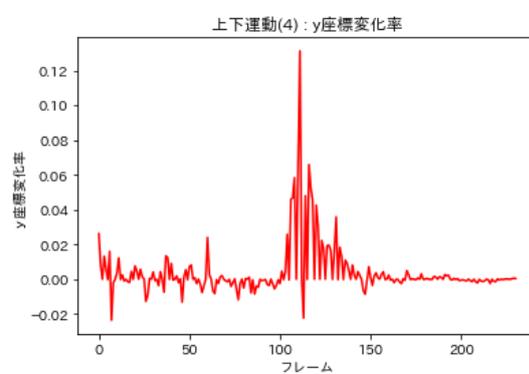


図 4-12 上下運動 (4) の y 座標変化率

表 4-4 上下運動 (1)-(4) のフレーム数

	上下運動 (1)	上下運動 (2)	上下運動 (3)	上下運動 (4)
総フレーム数	173	148	189	232
歩行フレーム数	168	142	168	196
上下運動フレーム数	5	6	13	36

4.3.2 評価指標

異常検知の評価指標には、正常標本精度と異常標本精度を用いる。それぞれ式 (1.2.1.8)、式 (1.2.1.9) のように定義される。本実験では、異常な標本が「上下運動」であり、正常な標本が「歩行」である。

正常標本精度

$$= \frac{\text{実際に正常な標本のうち、正しく正常と判定できた数}}{\text{実際に正常な標本の総数}} \quad (1.2.1.8)$$

異常標本精度

$$= \frac{\text{実際異常な標本のうち、正しく異常と判定できた数}}{\text{実際にな異常な標本の総数}} \quad (1.2.1.9)$$

ここで正常標本精度を γ_0 、異常標本精度を γ_1 とすると正常標本精度と異常標本精度の調和平均である F 値は式 (1.2.1.10) で表される。

$$F - measure = \frac{2 \times \gamma_0 \times \gamma_1}{\gamma_0 + \gamma_1} \quad (1.2.1.10)$$

4.3.3 従来手法との比較

移動物体検知の従来手法と提案手法の精度を比較する実験を行った。

提案手法では、DeepLabCut を用いて、前足の付け根の y 座標を取得し、フレーム間の y 座標変化率を特徴量として外れ値検知を行う。従来手法である背景差分法とフレーム間差分法はそれぞれ差分画像が取得できる。差分画像の重心の y 座標変化率を算出し、外れ値検知を行った。

比較実験は、全 35 フレームから成り、歩行フレームが 21 フレーム、上下運動フレームが 14 フレームである動画を使用した。DeepLabCut によって取得した前足の付け根の y 座標を図 4-13、背景差分法、フレーム間差分法によって取得した重心の y 座標をそれぞれ図 4-14、図 4-15 に示す。

取得した y 座標からフレーム間の y 座標変化率を算出し外れ値検知を行った。異常値の割合を指定可能な OneClassSVM を分類器に用いた。異常フレームの割合を 40% と指定し、外れ値検知を行った。提案手法、背景差分法、フレーム間差分法のそれぞれの外れ値検知による精度の比較を表 4-5 に示す。

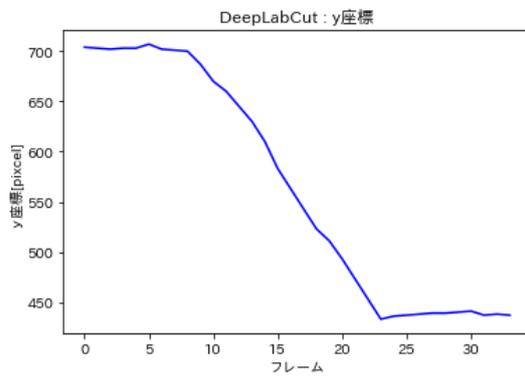


図 4-13 DeepLabCut を用いて取得した
y 座標

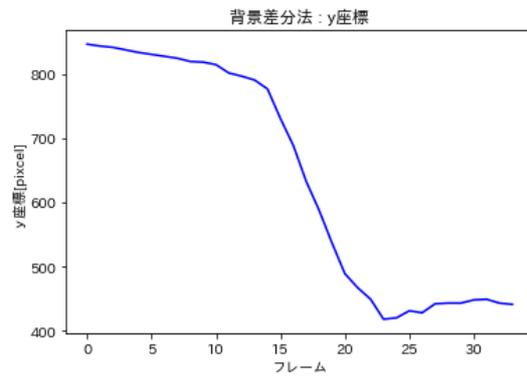


図 4-14 背景差分法を用いて取得した
y 座標

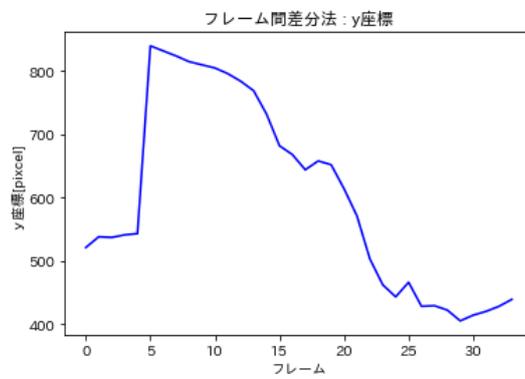


図 4-15 フレーム間差分法を用いて取得した
y 座標

表 4-5 提案手法と従来手法の比較

	正常標本精度	異常標本精度	F 値
提案手法	0.7143	0.5714	0.6349
背景差分法	0.7619	0.5000	0.6038
フレーム間差分法	0.6667	0.5000	0.5714

4.3.4 実験結果

上下運動 (1)-(4) の y 座標変化率に対し、「歩行」の変化率を正常値、「上下運動」の変化率を外れ値として、「歩行」と「上下運動」の分類を行った。

OneClassSVM による外れ値検知は、正常値と異常値を分割する超平面との距離が求まる。この距離が負の値をとるフレームが外れ値として検出される。OneClassSVM は異常値の割合を設定する必要がある。上下運動 (1)-(4) に対して、異常値の割合を変化させながら、外れ値検知の精度を求めた。異常値の割合を変化させながら求めた精度の変化を表 4-6、表 4-7、表 4-8、表 4-9 に示す。また、各動画で F 値が最大となる時の外れ値検知の結果を図 4-16、図 4-17、図 4-18、図 4-19 に示す。

IsolationForest による外れ値検知は、複数作成された木構造の終端ノードまでのパス長の平均が求まる。このパス長が負であるフレームが外れ値として検出される。上下運動 (1)-(4) に対して、IsolationForest による外れ値検知の結果を図 4-20, 図 4-21, 図 4-22, 図 4-23 に示す。また精度を表 4-10 に示す。

表 4-6 上下運動 (1) の外れ値検知精度

異常値の割合	正常標本精度	異常標本精度	F 値
0.05	0.9583	0.4000	0.5644
0.06	0.9583	0.4000	0.5644
0.07	0.9464	0.6000	0.7344
0.08	0.9345	0.6000	0.7308
0.09	0.9286	0.6000	0.7290

表 4-7 上下運動 (2) の外れ値検知精度

異常値の割合	正常標本精度	異常標本精度	F 値
0.08	0.9507	0.6667	0.7837
0.09	0.9366	0.6667	0.7789
0.10	0.9366	0.8333	0.8820
0.11	0.9225	0.8333	0.8757
0.12	0.9085	0.8333	0.8693

表 4-8 上下運動 (3) の外れ値検知精度

異常値の割合	正常標本精度	異常標本精度	F 値
0.13	0.9286	0.6923	0.7932
0.14	0.8988	0.6923	0.7822
0.15	0.8929	0.7692	0.8264
0.16	0.8750	0.7692	0.8187
0.17	0.8690	0.7692	0.8161

表 4-9 上下運動 (4) の外れ値検知精度

異常値の割合	正常標本精度	異常標本精度	F 値
0.14	0.9031	0.3889	0.5437
0.15	0.9031	0.3889	0.5437
0.16	0.8929	0.4167	0.5682
0.17	0.8724	0.4167	0.5640
0.18	0.8163	0.4167	0.5517

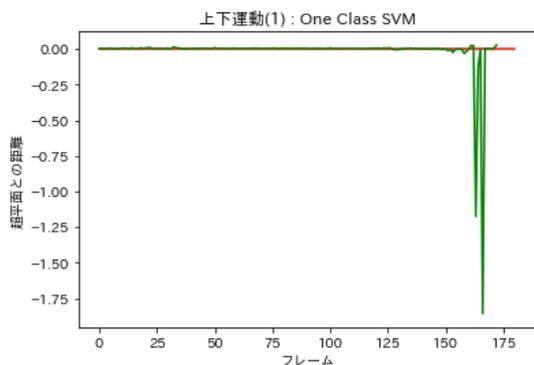


図 4-16 OneClassSVM による
上下運動 (1) の外れ値検知
(異常値の割合=0.07)

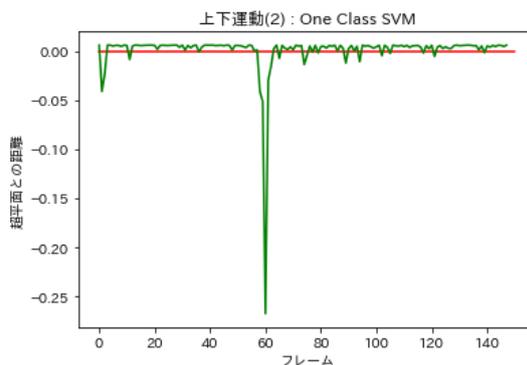


図 4-17 OneClassSVM による
上下運動 (2) の外れ値検知
(異常値の割合=0.1)

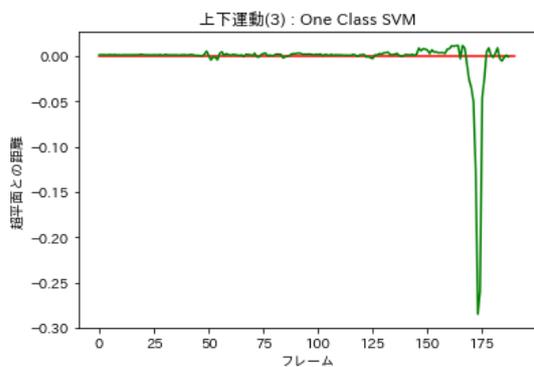


図 4-18 OneClassSVM による
上下運動 (3) の外れ値検知
(異常値の割合=0.15)

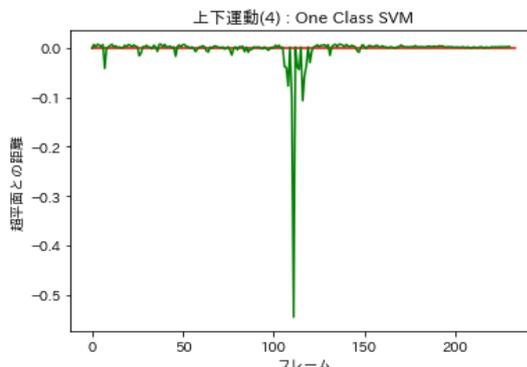


図 4-19 OneClassSVM による
上下運動 (4) の外れ値検知
(異常値の割合=0.16)

表 4-10 IsolationForest による外れ値検知精度

	正常標本精度	異常標本精度	F 値
上下運動 (1)	0.9167	0.8000	0.8544
上下運動 (2)	0.9225	0.8333	0.8757
上下運動 (3)	0.9583	0.9231	0.9404
上下運動 (4)	0.9745	0.4167	0.5837

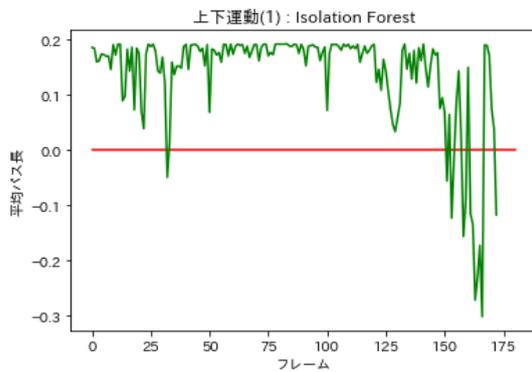


図 4-20 IsolationForest による
上下運動 (1) の外れ値検知

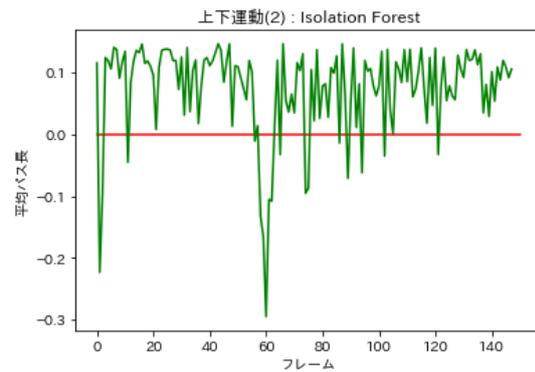


図 4-21 IsolationForest による
上下運動 (2) の外れ値検知

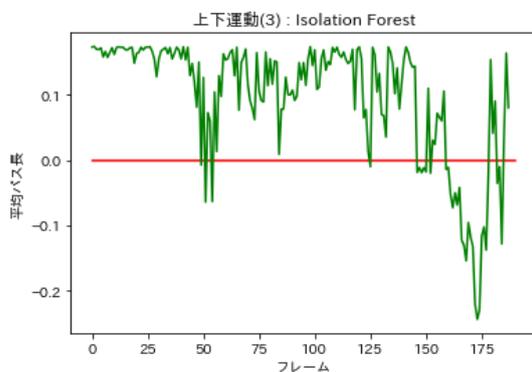


図 4-22 IsolationForest による
上下運動 (3) の外れ値検知

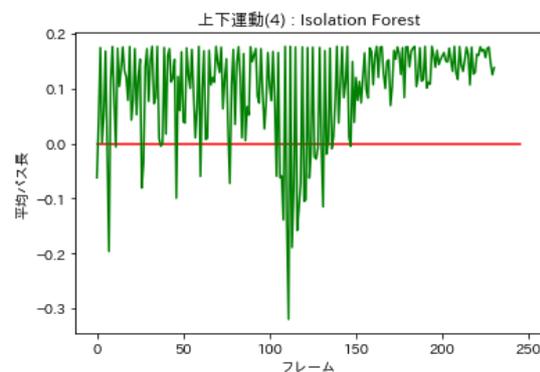


図 4-23 IsolationForest による
上下運動 (4) の外れ値検知

4.3.5 考察

表 4-5 の F 値を比較すると、提案手法は背景差分法よりも 5.2%、フレーム間差分法よりも 11.1%精度が高くなったことがわかる。これは、背景差分法を用いた場合には、差分画像に影が映ってしまうことが原因の一つに挙げられる。また、フレーム間差分法の場合は、猫の動きが小さくなるフレームで差分画像をうまく取得できないことが原因として挙げられる。

また、背景差分法とフレーム間差分法ともに、重心が足を指しているフレームや、背中を指しているフレームなどフレームごとに重心の位置が異なる。一方、DeepLabCut を用いる場合は、追跡する特徴点を指定することが可能である。これにより、猫の運動に合わせて正確に特徴点の y 座標変化を取得できる。

OneClassSVM と IsolationForest の精度を比較する。OneClassSVM の精度は、F 値が最大となるように異常値の割合を設定した場合を用いる。比較結果を表 4-11 に示す

表 4-11 各分類器による F 値の比較

	OneClassSVM	IsolationForest
上下運動 (1)	0.7344	0.8544
上下運動 (2)	0.8820	0.8757
上下運動 (3)	0.8264	0.9404
上下運動 (4)	0.5682	0.5837

IsolationForest を用いた場合は OneClassSVM を用いた場合より，上下運動 (2) では 0.71% 精度が低くなった。しかし，上下運動 (1) では 16.3%，上下運動 (3) では 13.8%，上下運動 (4) では 2.7% 精度は高かった。

OneClassSVM を用いた場合は，精度が最大になるように，設定する異常値の割合を変化させた。しかし，留守中に猫が何回上下運動を行うかは予測がつかない。つまり事前に異常値の割合はわからない。

これらのことより，見守りには IsolationForest を用いることが適していると考ええる。

IsolationForest を用いた場合，上下運動 (1)–(4) の異常と正常の分類正解率は，平均 84.3% であった。正常標本精度の平均値は 94.3%，異常標本精度は 74.3% となり，異常標本精度は低い値をとった。これは，正常であるフレームを異常であると誤検出していたことが原因の一つであった。猫がジャンプをするとき，飛び上がる前に足を曲げ，状態をかがめる準備運動のようなものをしていた。この準備運動フレームが異常フレームとして検出される場合があったため，準備運動を考慮することで検出精度が向上すると考えられる。

4.4 提案手法による行動の分類精度

提案手法による「座る」「寝そべる」「食事」の分類の正解率は 93.7% であった。また，運動時の「歩行」と「上下運動」の分類の正解率は 84.3% であった。

4.5 むすび

本章では，提案手法による静止時と運動時それぞれのデータセット，評価指標，実験結果および考察について述べる。さらに提案手法を用いた場合の行動の分類精度について述べた。

第5章 結論と今後の課題

5.1 結論

本研究は、カメラ映像から猫の行動を分類し、猫にとって運動となる行動と食事の姿勢を検出することを目的とした。まず、DeepLabCut から得られる各特徴点の座標データから、関節角度と y 座標変化率を求めた。

静止時の「寝そべる」「座る」「食事」の分類は、足の膝関節の角度を特徴量として、SVM を分類器に用いた。両足の関節角度を用いた場合、「寝そべる」「座る」の 2 値分類では、100%の精度で分類に成功したが、「食事」を加えた 3 値分類は、93.7%と精度が低下した。頭の動きなども考慮し、「食事」を加えた場合にも、さらに精度を高める必要がある。また、両足ではなく後足の関節角度のみを特徴量として用いた場合には、3 値分類の精度は 51.3%と低い値となった。

運動時の「歩行」「上下運動」の分類は、y 座標変化率を特徴量として、外れ値検知を行った。従来の移動物体検知と提案手法を比較するために、背景差分法とフレーム間差分法を用いた手法と提案手法で実験した。提案手法では背景差分法より 5.2%、フレーム間差分法よりも 11.1%精度が上がった。提案手法は背景差分法やフレーム間差分法よりも、外れ値検知の精度が上がることを確認できた。しかし運動時の分類の正解率は 84.3%であり、静止時と比較すると約 10%低い精度となった。「上下運動」の前に行う、「状態をかがめる」などの準備運動も考慮し精度を上げる必要がある。

5.2 今後の課題

本研究では、取得したい特徴点が全て映っている動画をデータセットで使用した。しかし、実際に家庭で用いる場合は、体の一部が家具などに隠れているなど悪条件の状況も考えなければならない。そのため、足だけでなく頭なども特徴点に加え、一部の特徴量が使用できない条件でも分類を可能にする必要がある。また、時系列の関係を考慮するなどの手法を改良させる必要がある。

謝辞

本研究の実験環境を与えてくださり、研究テーマ等に丁寧かつ熱心にご指導して下さった渡辺裕教授に心から感謝いたします。また、本研究に対し様々なご提案を頂きました早稲田大学国際情報通信センターの石川孝明様に深く感謝申し上げます。また、日頃から御意見やアドバイスをくださった研究室の皆様に御礼申し上げます。

最後に、学費や生活面で支えてくださり、ここまで育ててくださった家族に深謝いたします。

参考文献

- [1] 一般社団法人ペットフード協会,” 平成 30 年(2018 年)全国犬猫飼育実態調査結果”, (2018 年 12 月 25 日).
- [2] A. Mathis, P. Mamidanna1, ” Markerless tracking of user-defined features with deep learning,” arXiv preprint arXiv: 1804. 03142, 2018.
- [3] AlexEMG/DeepLabCut, <https://github.com/AlexEMG/DeepLabCut>, GitHub, Inc. (2020 年 1 月 24 日アクセス)
- [4] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, B. Schiele, “Deep Residual Learning for Image Recognition,” arXiv preprint arXiv: 1605.03170, 2016.
- [5] K. He, X. Zhang, “DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model,” arXiv preprint arXiv: 1512. 03385, 2015.
- [6] V. Vapnik, A. Lerner, “Pattern recognition using generalized portrait method,” Automation and Remote Control, vol. 24, pp. 774-780, 1963.
- [7] B. Boser, I. Guyon, “A Training Algorithm for Optimal Margin Classifiers,” Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT’92), pp. 144-152, July. 1992.
- [8] F. Liu, K. Ting, Z. Zhou, “Isolation forest”, The IEEE International Conference on Data Mining (ICDM), December. 2008.

図一覧

図 2-1 訓練データ数と RMSE の関係.....	7
図 2-2 DeepLabCut によるショウジョウバエのトラッキング.....	8
図 2-3 IsolationForest による観測値の外れ値検知.....	10
図 3-1 提案手法の流れ.....	12
図 3-2 取得する特徴点.....	12
図 3-3 分類手法の処理の流れ.....	13
図 3-4 「座る」の例.....	14
図 3-5 「寝そべる」の例.....	14
図 3-6 「食事」の例.....	14
図 3-7 「上下運動」による y 座標変化の例.....	15
図 4-1 2 値分類の混同行列.....	17
図 4-2 3 値分類の混同行列.....	18
図 4-3 「寝そべる」と「座る」の関節角度分布.....	19
図 4-4 「寝そべる」と「座る」と「食事」の関節角度分布.....	20
図 4-5 上下運動 (1) の y 座標.....	21
図 4-6 上下運動 (1) の y 座標変化率.....	21
図 4-7 上下運動 (2) の y 座標.....	21
図 4-8 上下運動 (2) の y 座標変化率.....	21
図 4-9 上下運動 (3) の y 座標.....	22
図 4-10 上下運動 (3) の y 座標変化率.....	22
図 4-11 上下運動 (4) の y 座標.....	22
図 4-12 上下運動 (4) の y 座標変化率.....	22
図 4-13 DeepLabCut を用いて取得した y 座標.....	24
図 4-14 背景差分法を用いて取得した y 座標.....	24
図 4-15 フレーム間差分法を用いて取得した y 座標.....	24
図 4-16 OneClassSVM による上下運動 (1) の外れ値検知 (異常値の割合=0.07) ..	26
図 4-17 OneClassSVM による上下運動 (2) の外れ値検知 (異常値の割合=0.1)	26
図 4-18 OneClassSVM による上下運動 (3) の外れ値検知 (異常値の割合=0.15) ..	26
図 4-19 OneClassSVM による上下運動 (4) の外れ値検知異常値の割合=0.16)	26
図 4-20 IsolationForest による上下運動 (1) の外れ値検知.....	27
図 4-21 IsolationForest による上下運動 (2) の外れ値検知.....	27
図 4-22 IsolationForest による上下運動 (3) の外れ値検知.....	27
図 4-23 IsolationForest による上下運動 (4) の外れ値検知.....	27

表一覧

表 4-1 実験に使用したデータセット.....	16
表 4-2 2 クラス分類の精度.....	19
表 4-3 3 クラス分類の精度.....	20
表 4-4 上下運動 (1)-(4) のフレーム数.....	22
表 4-5 提案手法と従来手法の比較.....	24
表 4-6 上下運動 (1) の外れ値検知精度.....	25
表 4-7 上下運動 (2) の外れ値検知精度.....	25
表 4-8 上下運動 (3) の外れ値検知精度.....	25
表 4-9 上下運動 (4) の外れ値検知精度.....	25
表 4-10 IsolationForest による外れ値検知精度.....	26
表 4-11 各分類器による F 値の比較.....	28