

## レスキュー犬の一人称動画を用いた動作分類

発表者： 情報学専攻 メディア情報学 プログラム 学籍番号 1730010 荒木 勇人  
 指導教員： 柳井啓司 教授

### 1 はじめに

被災地での救助活動を行う際に、人間の補助として訓練されたレスキュー犬（災害救助犬）が探査を行う場合がある。レスキュー犬は、犬としての特性を生かして人間と協力して被災地の探索を行う。がれきの隙間などの狭い空間、倒壊した建物など人間には踏破困難な環境でも探査可能であり、また発達した嗅覚を頼りにした救助活動が可能である。しかし、彼らは人間に向けた言語を持たないため、人間はレスキュー犬の行動から彼らが収集した情報を理解しなくてはならない。現状では、レスキュー犬を指揮するハンドラーと呼ばれる人がレスキュー犬の行動を手動でマーキングしており、その情報を消防などの指揮命令者に口頭伝達している。このレスキュー犬との共同探索の問題点として、トリアージ（緊急度に従った手当の優先順位付け）のための周辺環境情報や、要救助者情報の不足があげられる。また、ハンドラーによる記録はどうしても主観的になり客觀性が不足し、さらにそれを口頭伝達することで正確性がより不足する。

本研究では、レスキュー犬にセンサを装着して得られたデータ用いてレスキュー犬の行動をリアルタイムに分類すること目的とする。深層学習を用いた画像識別にある既存手法を予備実験として行った。予備実験をもとに、動画からのレスキュー犬行動分類を行う。本研究は映像だけでなく音声などのデータも活用したマルチモーダルな動画分類である。本研究により、レスキュー犬が今何をしているのか明示的に判断することが可能となり、トリアージに必要な情報が整理され、災害救助活動の効率化が期待される。

### 2 関連研究

レスキュー犬の行動をモニタリングするために、濱田、大野らによって装着型計測・記録装置が開発された[1]。図1にレスキュー犬に装着可能な軽量な行動計測スーツを示す。各種センサを用いた計測データを記録し、リアルタイムに映像などのデータを無線配信することが可能である。そのため、レスキュー犬が人の目の及ばない範囲で活動する際にもレスキュー犬の行動やその周辺環境などを把握するの

に役立つ。



図1 装着型計測・記録装置 [1] より引用

また、Ehsani.K らによる犬の一人称視点動画からの犬行動予測の研究がある[2]。これは、犬の行動をモデリングし、犬が次にどのような道をたどり行動するかを予測している。

しかし、これらの研究は犬の行動のモデリングであり、犬の周辺環境の推定などは行っていない。また、入力は動画像のみであり、音声などのデータは利用していない。レスキュー犬の課題には、犬の周辺環境情報や動画像からだけでは判断できない情報の取得が含まれている。例えばレスキュー犬は要救助者を発見するとその場で待機し吠え続けるように訓練されている。このように、動画像データからだけではなく、音声データ、および慣性データ・GPSデータなどの情報を複合的に用いてレスキュー犬の状態を判断しなければならない。本研究は動画像と音声からなるマルチモーダルな情報を入力とした犬の行動の分類を目的としている。

### 3 手法概要

動画からレスキュー犬の行動を推定するための手法の概要是以下の通りである。

- 動画から一定フレームと対応する音声を取り出しその音声を整形する。
- LSTM などの時系列情報を扱う手法により、データから特徴量を抽出する。
- 単位データから犬行動を分類するよう学習し、分類する。

## 4 予備実験

データセットに犬の一人称視点動画 DogCentric Activity Dataset(DCAD) [3] を用い,これを分類する予備実験を行った.これは犬の散歩を記録したデータであり,災害救助活動やその訓練データではない.災害救助活動および訓練データは現在作成中である. DCAD データセットは 10 クラス 209 クリップで構成されている. クラスはそれぞれ,横断前の待機 *Car*, 水分の摂取 *Drink*, 手渡しでの食事 *Feed*, 左を見る *Look\_at\_Left*, 右を見る *Look\_at\_Right*, 人間が犬を撫でる *Pet*, ボールで遊ぶ *Play\_with\_ball*, 体をブルブルと振る *Shake*, 何かの臭いを嗅ぐ *Sniff*, 歩く *Walk*, である.

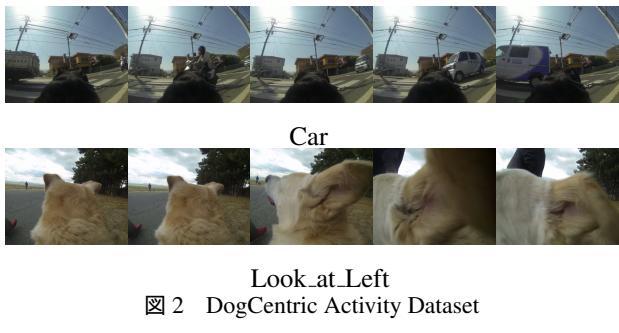


図 2 DogCentric Activity Dataset

動画ひとつにつきフレーム全体の平均を取り(式 1)) , 画像として扱い分類した(図 3). ResNet と VGG16 をそれぞれ用いた Pre-trained model の fine-tuning と二通り行った.

$$Input = \frac{\sum_{sec \times FPS} Frame}{sec \times FPS} \quad (1)$$



図 3 ネットワーク

## 5 実験結果

予備実験の結果を(図 4,5)にそれぞれ示す. 分類率は, VGG16 モデルを利用したものが 48.5%, ResNet モデルを利用したものが 59.5% であった. 全般的に, データの多いクラスは精度が高い傾向にあるが, データの少ないクラスは精度が低い傾向にある. 加えて, *Car* クラスは道路の進行方向に対して垂直に待機している 10 クラスの中で特殊なクラスであり, 車などの写ったフレームの影響で分類精度が上昇していると考えられる. *Feed* クラス, *Pet* クラス, *Play\_with\_ball* クラスは, それぞれフレーム内を人間が占める割合が多いクラスと言え, そのため混同が起こりやす

いと考えられる.

	Car	Drink	Feed	Left	Right	Pet	Ball	Shake	Sniff	Walk
Car	6	0	1	0	0	1	0	0	0	0
Drink	0	1	2	0	0	3	0	0	0	0
Feed	0	0	5	0	0	4	1	1	1	0
Left	0	0	0	4	2	1	0	2	1	0
Right	0	0	0	0	1	1	0	0	0	0
Pet	0	0	1	0	0	3	0	1	0	0
Ball	0	0	0	0	0	0	2	0	1	1
Shake	0	0	0	0	1	1	1	2	1	0
Sniff	0	0	0	0	0	1	1	1	4	0
Walk	0	0	1	0	0	1	0	1	2	4

図 4 VGG16 pretrained model による finetuning の結果

	Car	Drink	Feed	Left	Right	Pet	Ball	Shake	Sniff	Walk
Car	6	0	0	0	0	1	0	0	0	0
Drink	0	1	0	0	0	0	0	0	0	0
Feed	0	0	1	0	1	1	1	0	0	0
Left	0	0	0	1	1	0	0	1	0	2
Right	0	0	0	1	2	0	0	0	0	0
Pet	0	0	2	0	0	3	1	2	1	0
Ball	0	0	0	0	0	0	2	0	0	0
Shake	0	0	0	0	0	0	0	1	1	0
Sniff	0	0	0	0	0	0	1	0	5	0
Walk	0	0	0	0	0	0	0	0	0	3

図 5 ResNet pretrained model による finetuning の結果

## 6まとめ, 今後の課題

動画の各フレームの平均を取り, 画像として識別した. データの少ないクラスは精度が低いため, データを補う必要がある. 予備実験では簡易的な方法を用いたが, 今後は最新手法による分類を検討している. またレスキュードogの行動を認識する際には複数クラスの出力にする必要がある.

今後の課題として, 時系列情報を特徴量抽出に使う. また, 音声データから特徴量を抽出し, 動画特徴量と併せてマルチモーダルな特徴量を利用し, レスキュー犬の行動分類を行う.

## 参考文献

- [1] Y. Komori, T. Fujieda, K. Ohno, T. Suzuki, and S. Tadokoro. Detection of continuous barking actions from search and rescue dogs' activities data. In *hoge*, pp. 630–635, 2015.
- [2] K. Ehsani, H. Bagherinezhad, J. Redmon, R. Mottaghi, and A. Farhadi. Who let the dogs out? modeling dog behavior from visual data. 2018.
- [3] Iwashita, Y., Takamine, A., Kurazume, R., and M. S. Ryoo. First-person animal activity recognition from egocentric videos. In *Proc. of International Conference on Pattern Recognition (ICPR)*, Stockholm, Sweden, August 2014.