

レスキュー犬の一人称動画をを用いた動作推定

荒木勇人* 井出佑汰* 濱田龍之介** 大野和則** 柳井啓司*

*電気通信大学 **東北大学NICHe/理研AIP

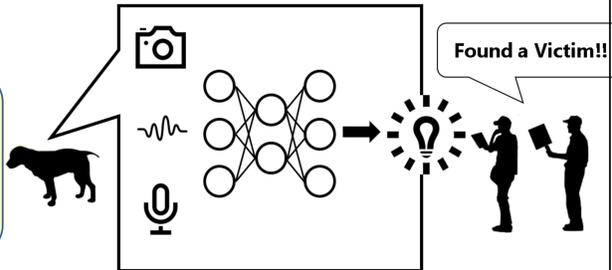
はじめに

背景

- ・ **レスキュー犬**: 被災地で救助活動を補佐する犬
- ・ 人から見えない活動も客観的に記録したい
- ・ 犬にセンサーを搭載 (サイバーレスキュー犬スーツ)

目的

犬一人称映像と音の統合によるレスキュー犬の行動認識



データセット

サイバーレスキュー犬スーツ

- ・ 東北大学の犬大野らによる犬行動計測記録装置



レスキュー犬訓練データセット

- ・ 救助訓練をしている様子



動画: 7本(2分~20分) 総時間数57分40秒 29.97FPS

利用フレーム数: 103,696枚

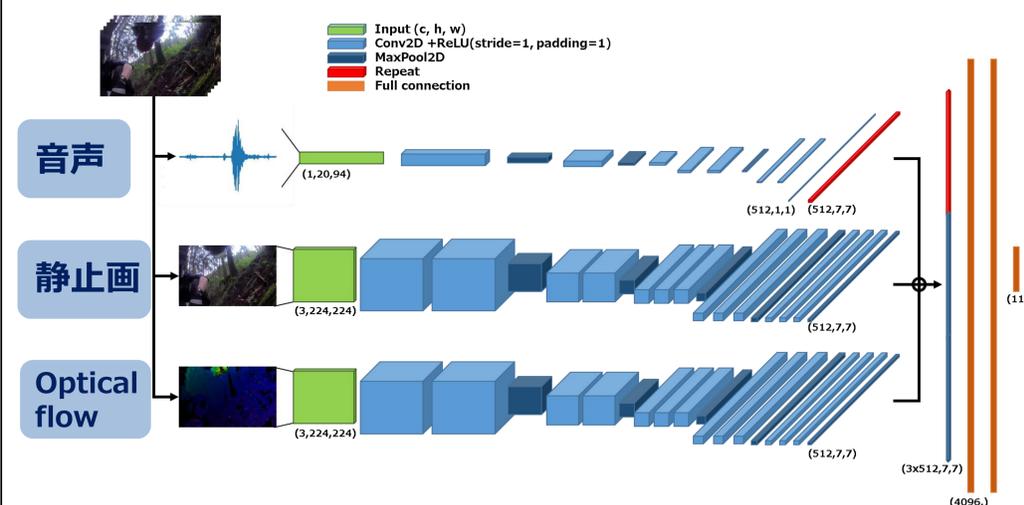
行動ラベル: 11クラス (同一時間に複数(マルチ)ラベル)

bark	Cling	Command	eat-drink	handler	run	see victim	shake	sniff	stop	walk-trot
吠える	執着	指示	飲食	見る	走る	発見	振る	嗅ぐ	停止	進行
1,744	1,127	2,439	343	2,011	98	1,549	239	7,719	6,384	8,764 (枚)

手法

Sound/image-based Three Stream Network

- ・ Two-StreamとSoundNetの組み合わせ
- ・ 動画から得られる3種類のデータから特徴抽出し、それぞれのストリームに入力



実験

比較実験

- ・ データの組み合わせ毎に比較実験を行った

	静止画像	optical flow	音声	
(1)	✓	-	-	VGG16 pre-trained
(2)	-	✓	-	VGG16 pre-trained
(3)	-	-	1D	音声ストリームのConv1D
(4)	-	-	2D	音声ストリームのConv2D
(5)	✓	✓	-	2-stream CNN
(6)	✓	-	2D	Sound/image-based 2-stream CNN
(7)	-	✓	2D	Sound/image-based 2-stream CNN
(8)	✓	✓	2D	Sound/image-based 3-stream CNN

- ・ 動画毎に、前半70%を学習に使い、後半30%を評価に用いた

実験結果

	bark	cling	command	eat-drink	look at handler	run	see victim	shake	sniff	stop	walk-trot	全体
(1)	0.244	0.066	0.0	0.024	0.057	0.0	0.204	0.0	0.0	0.588	0.510	0.436
(2)	0.141	0.0	0.0	0.0	0.017	0.0	0.017	0.0	0.0	0.586	0.476	0.406
(3)	0.669	0.078	0.22	0.023	0.138	0.0	0.274	0.44	0.502	0.745	0.704	0.512
(4)	0.563	0.04	0.188	0.001	0.059	0.0	0.201	0.304	0.524	0.744	0.74	0.512
(5)	0.110	0.018	0.043	0.0	0.155	0.0	0.259	0.0	0.426	0.705	0.668	0.435
(6)	0.662	0.031	0.195	0.018	0.115	0.002	0.308	0.402	0.498	0.726	0.694	0.5
(7)	0.667	0.054	0.234	0.014	0.123	0.01	0.223	0.356	0.487	0.759	0.692	0.493
(8)	0.577	0.135	0.186	0.066	0.183	0.026	0.433	0.409	0.530	0.779	0.725	0.518

まとめ

- ・ 提案手法で推定精度51.8%
- ・ 3種類のデータにそれぞれ得意な動作がある。

今後の課題

- ・ 精度の向上
 - 犬一人称視点用の前処理
 - データセット増強
- ・ リアルタイム処理

参考文献

[1] Aytaç, Y., Vondrick, C. and Torralba, A. A.: Soundnet: Learning sound representations from unlabeled video, Advances in Neural Information Processing Systems (2016).
 [2] Ehsani, K., Bagherinezhad, H., Redmon, J., Mottaghi, R. and Farhadi, A.: Who Let The Dogs Out? Modeling Dog Behavior From Visual Data, Proc. of IEEE Computer Vision and Pattern Recognition (2018).
 [3] Iwashita, Y., Takamine, A., Kurazume, R. and Ryoo, M. S.: First-Person Animal Activity Recognition from Egocentric Videos, Proc. of International Conference on Pattern Recognition (ICPR), Stockholm, Sweden (2014).
 [4] Komori, Y., Fujieda, T., Ohno, K., Suzuki, T. and Tadokoro, S.: Detection of Continuous Barking Actions from Search and Rescue Dogs' Activities Data, Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 630-635 (2015).
 [5] Minghuang, M., Haoqi, F. and Kris, M. K.: Going Deeper into First-Person Activity Recognition, Proc. of IEEE Computer Vision and Pattern Recognition (2016).
 [6] Simonyan, K. and Zisserman, A.: Two-stream convolutional networks for action recognition in videos, Advances in Neural Information Processing Systems, pp. 568-576 (2014).
 [7] 丹野良介, 小澤 暖, 伊藤浩二.: 危険運転シーン抽出のためのマルチモーダル深層学習技術, 第11回データ工学と情報マネジメントに関するフォーラム (DEIM) (2019).