



「次世代人工知能・ロボット中核技術開発」
(人工知能分野) 中間成果発表会
－人間と相互理解できる人工知能に向けて－

きめの細かい動作認識の研究開発

平成29年 3月29日

学校法人 千葉工業大学

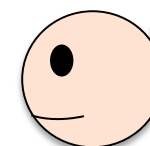
人工知能・ソフトウェア技術研究センター

竹内 彰一

国立研究開発法人 産業技術総合研究所

国立研究開発法人 新エネルギー・産業技術総合開発機構

ヒトの多様な動作を高精度で認識する技術の研究開発



サザエさんが
タラちゃんを抱いて
部屋から出て行ったよ

- ① 日常的なシーンにおけるヒトの動作100種類を定義
 - 部屋から出る、座る、ご飯を食べる、水を飲む、抱き合う、など
- ② 100種類の動作 x 動画1000本 合計100,000本の動画データセットを構築
- ③ このデータセットと深層学習で精度の高い認識エンジンを開発
日常生活シーンを理解・支援できる人工知能へ
(保育、介護、家事手伝い)



AI for Human Life

深層学習の成功の方程式： 大規模データセットによる訓練が高性能を引き出す

！ 2012年 ImageNet の100万枚超の画像データセットが深層学習による画像認識

のブレークスルーを引き起こした

予想 ヒトの動作(動画)の大規模データセットがあれば、

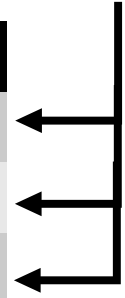
動作認識は飛躍的に進歩する

動画データセットを構築

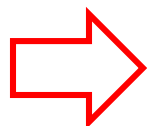
【従来のデータセット】

動作： 日常的な動作を100個選択
 動画： 一つの動作につき1000本(長さは5秒)
 規模： 100動作 x 1000本

データセット	動作の種類	全動画数
HMDB51	51	6,766
UCF-101	101	13,320
ActivityNet 200	200	23,064
Stair Actions	100	100,000



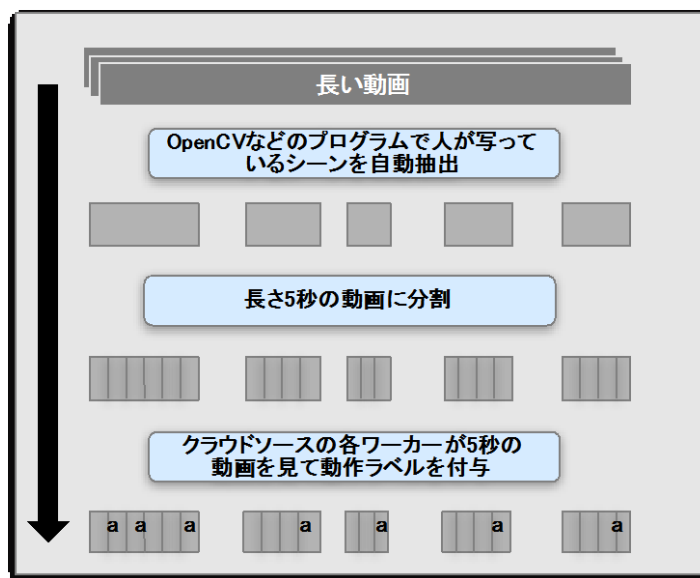
動画数は従来最大のものの 4.3倍
 動作別の動画数は 7.5倍
 日常動作の種類数は最大レベル



クラウドソーシングを活用してデータセットを構築

動画ソース

- 本案件のために新たに制作された動画
 - お題を提示して5秒の動画を制作依頼
- ネット上のCC0動画
 - 候補クリップをワーカーに提示して動作ラベルづけを依頼

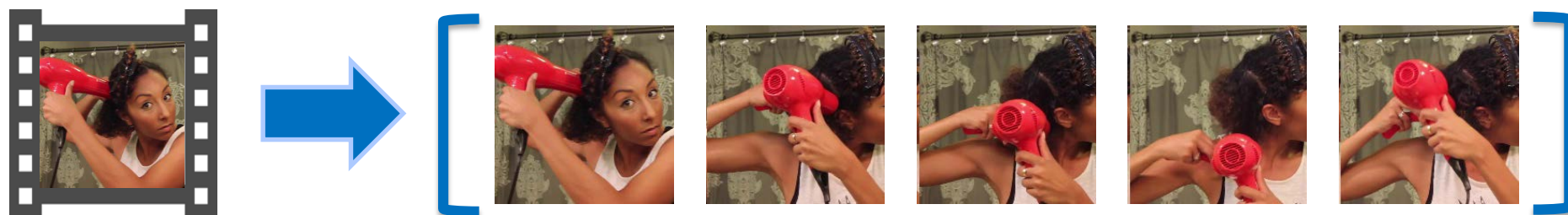


データセット提供物（案）

- 動作ラベル（日本語、英語）付き動画データセット 100,000本
 - 動画URL+タイムコードのみ（著作権への配慮）
 - オリジナル動画
- ツール
 - URL+タイムコードから動画をダウンロードするスクリプト

深層学習による動作認識

動作認識の入力データは動画（= 画像の系列データ）

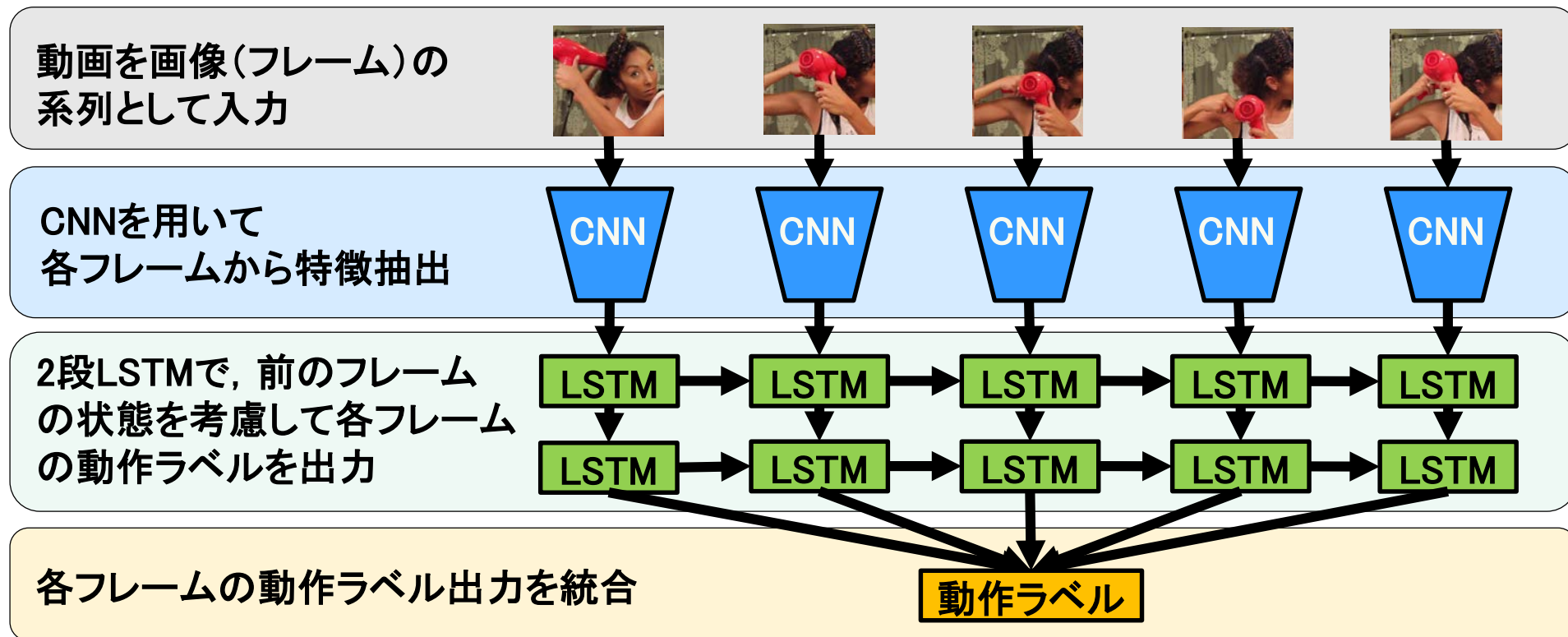


- 画像と系列データは深層学習の得意分野
 - 画像: 畳み込みニューラルネットワーク (CNN)
 - 系列データ: リカレントニューラルネットワーク (RNN)
または, Long Short-Term Memory (LSTM)

CNNとLSTMを結合したCNN+LSTM方式のネットワークで動作認識を行う

CNN+LSTM方式に基づく動作認識エンジンの開発

[Donahue et al., CVPR 2015]のLRCNモデルを参考に実装



開発中の認識エンジンの精度

構築中のデータセット	45.60%
UCF-101	86.04%
HMDB51	52.87%

従来のデータセットよりも高難易度

同設定の[Donahue et al., CVPR 2015]の結果と比べると、**4ポイント程度高い精度**