

2 値化オートエンコーダを用いたヒューマノイドロボットによる柔軟物操作のための軽量 End-to-End 学習

大原 慧[†]
大阪大学 基礎工学部情報科学科[†]

尾形 哲也[‡]
早稲田大学 理工学術院[‡]

栗野 皓光[§]
大阪大学 情報科学研究科[§]

1 はじめに

近年深層学習の活用により、動的に変化する環境に応じたロボットの動作生成が可能となり、複雑なタスクを担うロボットが実現されつつある。外部環境により変形しやすい柔軟物はロボットによる操作が困難なタスクとして一般的に知られているが、従来研究では複数の深層学習技術を応用することで柔軟物の操作に成功している [1]。しかし従来手法は推論に GPU を用いており、低消費電力化が必要である。そこで小型かつ数ワット程度の電力でも動作する Field-Programmable Gate Array(FPGA) への実装を試み、上記で示した課題の達成を含めたロボットシステム全体の性能向上を目指す。FPGA へ実装するにあたって、高々数十 MB 程度しか搭載されていないブロックメモリ (BRAM) を活用できるように、モデルサイズ的大幅な圧縮が必要となる。モデルサイズを圧縮する手法として、ニューラルネットワークの重みと活性化関数の出力を 1bit に制限する Binarized Neural Network(BNN) が報告されている [2]。

そこで、本研究では BNN を応用した 2 値化オートエンコーダと再起結合ニューラルネットワークを用いて、カメラ画像から柔軟物操作をするロボットの関節角を End-to-End に予測する軽量ニューラルネットワークを提案する。数値実験の結果、推論精度を損なうことなく、モデルサイズを 95.7% 削減できることが明らかになった。

2 柔軟物操作生成モデル

本研究で提案するモデルは図 1 に示すように (a) 画像特徴抽出モジュールと (b) ロボット動作生成モジュールで構築される。

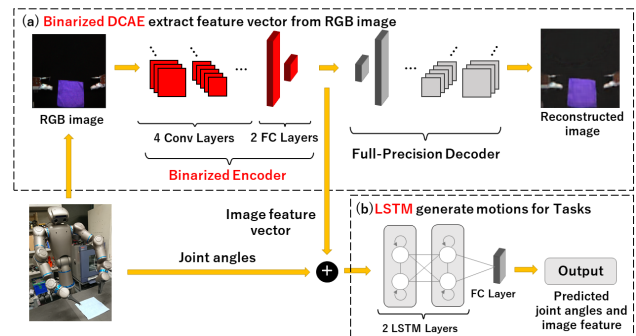


図 1 柔軟物操作生成モデルの概要図

2.1 Binarized Deep Convolutional Auto-Encoder

画像特徴抽出モジュールとして Deep Convolutional Auto-Encoder(DCAE)[3] の一部を 2 値化した Binarized DCAE を用いる。低次元の画像特徴量を抽出するエンコーダと特徴量から元の入力データを復元するデコーダで構築される DCAE 全体へ BNN を適用すると関節角度の予測精度が大幅に悪化する。本稿では DCAE のエンコーダ部分のみを 2 値化したモデルを提案する。ロボットの動作生成時はエンコーダのみを利用し、誤差伝播を容易とするためにデコーダ部分は full precision のままで学習する。

2.2 Long Short-Term Memory

ロボット動作生成モジュールとして Long short-term memory(LSTM)[4] を用いる。再帰結合ニューラルネットワークとして知られる LSTM は複数の時系列データを学習・汎化できる学習器であり、Binarized DCAE で抽出した低次元の画像特徴量とロボットの関節角度を統合して学習することでロボットの動作を生成する。また、同時に外部環境の予測も行うため DCAE のデコーダ部分と組み合わせて次状態の画像を予測することも可能である。

3 数値実験

3.1 実験設定

本手法の有効性を検証するために、テーブルのほぼ中央に配置されたタオルを折り畳む動作をシ

Light-weight End-to-End Learning of Flexible Object Manipulation for Humanoid Robot Using Binarized Auto-Encoder

[†] Satoshi Ohara, Osaka Univ.

[‡] Tetsuya Ogata, Waseda Univ.

[§] Hirimitsu Awano, Osaka Univ.

ミュレートする数値実験を行った。データセットとして各腕 6 自由度の両腕を持ったヒューマノイドロボット Nextage Open で取得した約 33000 ステップのカメラ画像と関節角データを用いた。約 22000 ステップのデータで学習を行い、約 11000 ステップのデータを評価に使用する。学習には 2 段階あり、まず $112 \times 112 \times 3\text{ch}$ のカメラ画像をもとに画像特徴量を抽出する Binarized DCAE を学習する。次に、13 次元の関節角データと抽出した 64 次元の画像特徴量をもとに折りたたみ動作を生成する LSTM を学習する。表 1 に本モデルの詳細を示す。

評価時はタオルの初期位置を撮影した画像 1 枚とロボットの初期関節角から折りたたみ動作を生成し、評価用データセットの関節角と比較する。

表 1

Layer	Output Dim.	Input Fmaps	Output Fmaps
Conv1	112×112	3	32
Max Pool	56×56	32	32
Conv2	56×56	32	64
Max Pool	28×28	64	64
Conv3	28×28	64	128
Max Pool	14×14	128	128
Conv4	14×14	128	256
Max Pool	7×7	256	256
FC1	1×1	12544	1024
FC2	1×1	1024	64
LSTM1	1×1	77	100
LSTM2	1×1	100	100
FC3	1×1	100	77

3.2 実験結果

実験の結果、本モデルは DCAE 全体を 2 値化したモデルに比べて関節角度の予測精度が向上し (図 2), DCAE のデコーダ部分を用いてタオル折りたたみ動作の予測画像を生成することに成功した (図 3)。また、表 2 に、重みサイズについて提案モデルと従来の Full-Precision モデルを比較した結果を示す。提案手法によってモデルサイズが約 95.7% 削減されていることが確認できる。

表 2 重みサイズの比較

	Full-Precision	提案モデル
DCAE	50.1MB	1.57MB
LSTM	0.611MB	0.611MB
合計	50.7MB	2.19MB

4 おわりに

本稿ではハードウェア実装に適した Binarized DCAE と LSTM を用いたヒューマノイドロボッ

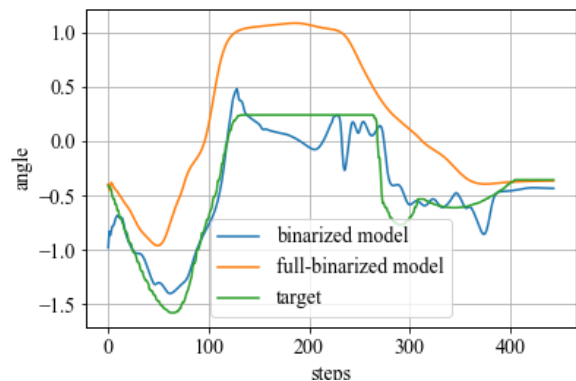


図 2 関節角度の予測

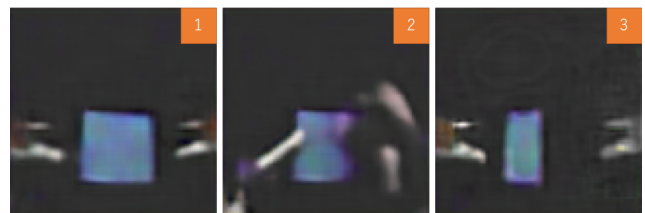


図 3 タオル折りたたみ動作の予測画像

トの動作生成について述べた。ロボットシステム全体の性能向上を FPGA によるハードウェアアクセラレータの支援によって実現する際、メモリサイズの制約を満たすためにモデルサイズの圧縮は重要である。モデルサイズの大幅な圧縮のため、画像特徴量を抽出する DCAE のエンコーダ部分の 2 値化を提案しロボットの動作生成を行った。数値実験の結果、推論精度を損なわずにモデルサイズを約 95.7% 削減できることが明らかになった。今後の展望として、実ロボットによる動作生成の実験や FPGA への実装、タスク動作の拡張を試みる。

謝辞

本研究は、JST さきがけ (JPMJPR18M1) 及び、JSPS 科研費 (18K13800) の支援を受けたものである。

参考文献

- [1] Pin-Chu Yang, Kazuma Sasaki, Kanata Suzuki, Kei Kase, Shigeki Sugano, and Tetsuya Ogata: Repeatable Folding Task by Humanoid Robot Worker using Deep Learning, IEEE Robotics and Automation Letters (RA-L), Vol.2, No.2, pp. 397-403, Nov. 2016. DOI: 10.1109/LRA.2016.2633383
- [2] M. Courbariaux, I. Hubara, D. Soudry, et al., "Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1," arXiv:1602.02830 [cs.LG].
- [3] G.E. Hinton et al., "Reducing the dimensionality of data with neural networks", Science 313 (5786)
- [4] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural computation, 9(8):17351780, 1997.