

令和 3 年 6 月 4 日現在

機関番号：82626
研究種目：若手研究(B)
研究期間：2017～2020
課題番号：17K18420
研究課題名（和文）深層学習を用いたアクション指向物体認識

研究課題名（英文）Deep learning for action

研究代表者

吉安 祐介 (Yoshiyasu, Yusuke)

国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員

研究者番号：10712234

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本課題では、深層学習を発展させ行動を導出するための認識技術確立することを目指し、（1）言語と視覚を利用したナビゲーション技術、（2）環境やものの認識とロボット軌道の学習と（3）物体位置・姿勢認識技術とロボット物体操作への応用という3つのテーマを実施した。画像、俯瞰地図、3D画像などの視覚入力をニューラルネット通して得た認識結果に基づき、モーションプランニングや強化学習ポリシーによりアクションを生成した。加えて、視覚情報と言語情報を組み合わせてナビゲーションを行う技術を構築した。最終的には、深層学習に基づく物体位置・姿勢認識技術を実装したアーム型のロボットシステムで物体把持を実現できた。

研究成果の学術的意義や社会的意義

本研究では、深層学習を用いた物体認識技術とそれに基づく物体操作やナビゲーション技術を構築した。中でも開発した視覚と言語情報を組み合わせるナビゲーション技術は、知識ベースの手法とデータ駆動の手法を組み合わせた技術であり、今後のさまざまな展開の可能性があり、学術的価値が高いと考える。一方で、本課題のようにアクションを生成するという複雑な問題においては、人と同等のレベルに達するにはさらなる改良が必要であることもわかってきている。本課題で得た成果や知見は、行動知能の研究分野を推し進めていくうえで重要な学術的意義があり、この分野におけるさらなる発展にも寄与すると考える。

研究成果の概要（英文）：The aim of this project is to develop a recognition technology for producing actions based on deep learning which has been mainly used in image recognition. We divide this research project into (1) Navigation technology using language and vision, (2) Recognition of environment and for generating robot trajectories and (3) 6-DoF object pose recognition and its application to robot object manipulation. Several different types of inputs, such as 2D image, bird-eye-view image, 3D image, were used to recognize objects and environments and then to produce actions. In addition, we devised a learning-based navigation technique by combining visual and semantic information. As a result of comparisons with the previous approaches, we showed that the proposed techniques outperform previous techniques in terms of speed and accuracy. Finally, we implemented a 6-DoF object pose recognition techniques based on deep learning on robotic arm systems and achieved object grasping.

研究分野：知能ロボティクス、人工知能、コンピュータビジョン

キーワード：深層学習 アクション

1. 研究開始当初の背景

生活の質の向上や近年の社会問題である労働力不足への対策として、住居などの環境で作業するようなサービスロボットが注目を集めている。このようなロボットは、産業用ロボットのように高精度に動きを制御する必要はないが、人の意図を理解し、周囲の環境を認識しながら作業を行うことが求められる。しかし、現状のロボットは、AR マーカーなどを貼り付けて物体を認識するなどその認識性能は十分でない。これに対し、近年、大規模データを用いて学習を行う深層学習の発展により、画像や言語の認識性能が飛躍的に高まっている。実際、2次元画像中に存在する物体の分類においては、人間をしのぐ認識性能を発揮できるようになった。しかしながら、現実世界においては人間やロボットなどの認識主体は動き、行動する。このような状況では、視点は動き、まわりのものや環境はその位置が相対的に刻々変化する。また、動作を実現していくためには、ものの名前(ラベル)に加えて、どこにどのようなものがあり、それに対してどうアクションするかといったことが重要になる。

深層学習をロボットに応用する試みは、2015年ころに始まった。国内では、早稲田大学の尾形研究室が時系列データを処理できる RNN を用いて、ビデオ映像からロボットに動作を学習させた。また、Ceatec Japan 2016 では、Preferred networks 社が、深層学習と強化学習を用いてドローンに飛行を学習させ注目を集めた。国外においても、CNN を用いた把持領域の特定や深層学習に基づく複雑な力制御モデルの構築などが行われており、ロボット実機での野菜のカットに挑戦した例もある。最近では、Google や UC Berkeley のチームが、深層学習と強化学習を組み合わせることで、end-to-end に視覚入力とロボットの制御量をつなぎ、ロボットによるピッキングを実現した。これに対して、生活支援ロボットやサービスロボットにおいては、視覚認識を直接的に動作に結びつけることよりも、言語命令を理解し、視覚情報を参照しながら運動を生成し、作業を実現するということが求められる。

2. 研究の目的

近年、深層学習の発展により、2D 静止画の認識精度が飛躍的に向上した。しかし、現実世界においては、人間やロボットなどの認識主体は動き、行動する。このような状況では、視点は動き、まわりのものや環境はその位置が相対的に刻々変化する。また、動作を実現していくためには、ものの名前やラベルなどの「what」に加えて、どこにどのようなものが存在するかなど「where」やそれに対してどうアプローチするかなどの「how」の認識も重要となる。本研究では、これまで画像や言語の認識に用いられてきた深層学習を発展させ、行動する知能を構築するための基礎として、知覚から運動につなぐ「アクション指向の物体認識技術」を研究する。具体的には、住居で作業するサービスロボットを想定し、視覚情報に基づいて周囲の環境を認識しながら、言語で受けた命令を実行するシステムの構築を目指す。

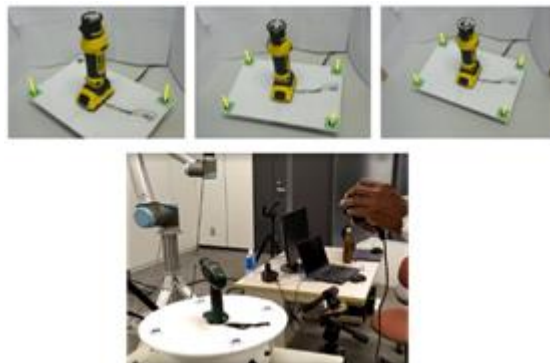
3. 研究の方法

本研究の目的は、これまで画像の認識に用いられてきた深層学習を発展させ、行動のための認識技術を確認することである。研究開始時には、このテーマを A. 言語命令の理解する方法、B. 移動しながら空間内の物体を検出する方法、C. パーツセグメンテーションに基づく道具使用時の手先軌道の推定、D. 身体動作の生成と F. 提案手法の評価の 5 つに分解して研究を遂行するという計画を立てた。研究開始後にロボットへの実装実現性等を考慮し、サブテーマ C. を「物体位置・姿勢認識技術と物体操作への応用」という課題に修正し、位置姿勢推定技術を駆使して、あらかじめ物体 CAD モデルに定義された把持点を使ってロボット手先軌道を推定する技術を開発した。また、A、B と D を組み合わせた「言語と視覚を利用したナビゲーション技術」と環境認識と障害物回避を目的とした「環境認識とロボット軌道の学習」という課題に取り組み、各課題においては既存手法との比較やロボット実験を行うことで提案技術の評価を行った。

4. 研究成果

(1) 物体位置・姿勢認識技術とロボット物体操作への応用

ロボットアームによる把持動作を実現するための基礎として、CNN を用いた姿勢推定技術を提案した。Yolo などの物体検出器を用いて物体領域を切り出した画像を入力として、CNN を用いて物体の姿勢を推定した。物体の姿勢は、CNN のネットワーク層中に、行列を直交行列に近づける制約層を組み込み、3次元姿勢を回転行列(9パラメータ)として回帰した。これにより、クォータニオンや回転ベクトルを推定す



る手法よりも姿勢推定精度を向上できた。また、CNNに基づく物体姿勢技術の学習画像データとラベルデータを収集するために、ターンテーブルとARマーカーを用いて半自動的に物体3次元姿勢をアノテーションする方法や Structure-from-motion を用いてほぼ自動的にラベル付けする技術を構築した。次に、CNNで推定した位置・姿勢を探索の初期値として、ポイントクラウドに対して物体3次元モデルの位置あわせを行うICPアルゴリズムを統合しRGBDデータに基づく物体位置姿勢認識システムを構築した。図1のように双腕ロボットBaxterと単腕ロボットSawyerを用い、これらのロボットにRGBDセンサを取り付け、ソフトウェアプラットフォームROSに実装されているモーションプランニング技術を活用することで、物体把持動作生成を可能にした。その結果、物体認識と把持動作を安定的に生成できるシステムとなり、棚への物体陳列作業も行うことができた。

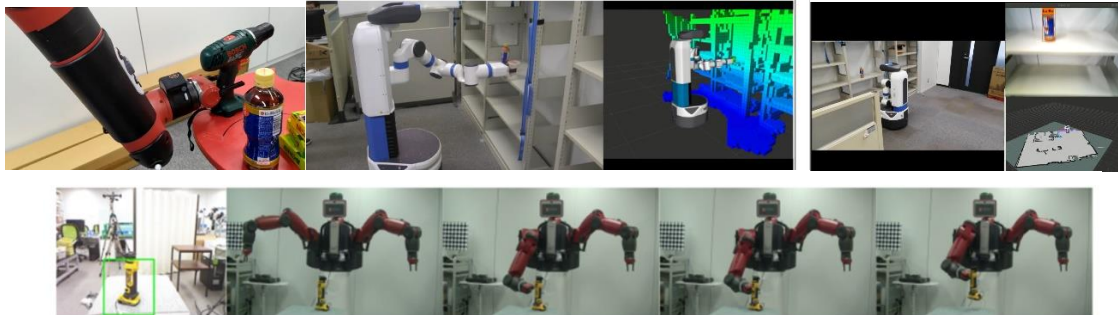


図1 ロボット実装

(2) 環境認識とロボット軌道の学習

移動ロボットやマニピュレータロボットなどへの活用を想定し、深層学習を活用して、環境の視覚情報の圧縮技術とそれに基づく軌道生成技術を構築した。2Dの環境地図をオートエンコーダで圧縮し、圧縮した潜在変数から軌道を生成する深層ニューラルネットワークを学習した。学習データの生成には、RRT*という既存の運動計画を用いた。環境データの圧縮にはDeepSDFと呼ばれる符号付距離画像を入力とする技術を用いた。ランダムに生成した障害物を含んだ未知の2D地図を使用した実験を行った結果、学習に基づく本軌道生成技術は従来よりも高速であることがわかった。

また、物体検出と環境2Dマップに基づく実空間におけるナビゲーション検証実験を行った。移動マニピュレータFetchにYOLOv3を搭載し、対象物の検出結果と深度情報を用いて目標位置を決め、あらかじめ作成した地図上でロボットをナビゲーションした。その結果、シミュレータ(Gazebo)と実環境において対象物の1m程度近くまでロボットを移動させることができた。

(3) 言語と視覚を利用したナビゲーション技術 (図2)

言語と視覚情報に基づいて物体の名前で提示したゴールに対してロボットを導く学習に基づくナビゲーション技術を構築した。具体的には、物体検出を用いて画像から得た視覚情報と、大規模な知識データから変換されたword embeddingを用いて表現されたモノの意味をニューラルネットワーク内で特徴表現として融合した。この特徴表現を入力とした動作ポリシーに基づいてアクションを生成した。これにより、大規模な知識グラフから事前に獲得した知識を利用することが可能になり、学習環境に存在しない未知物体への適用が可能になるなど汎化性が向上した。加えて、モノとモノの関係的知識を利用することで、かくれなどにより直接視認できないモノの把握を可能にし、物体探索能力の向上を図ることができた。

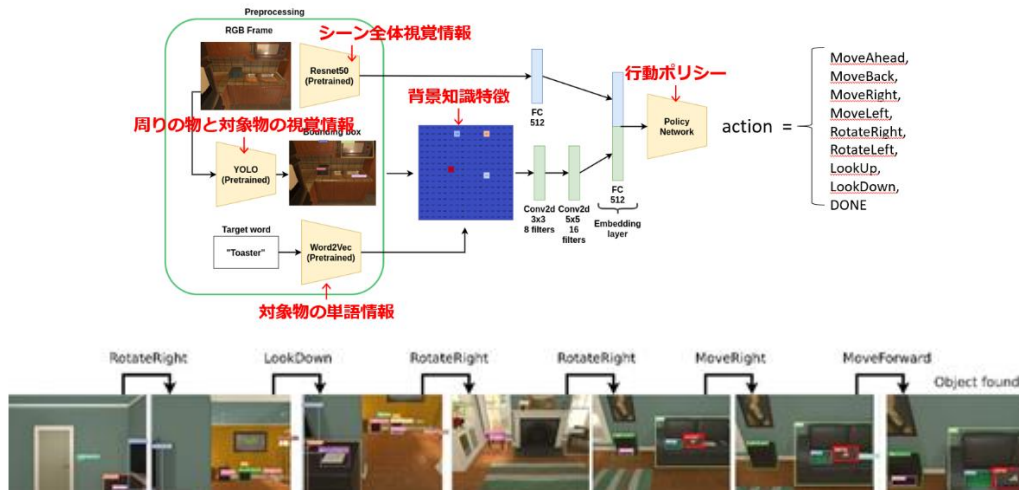


図2 言語と視覚を利用したナビゲーション技術

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 8件/うち国際共著 4件/うちオープンアクセス 0件）

1. 著者名 Watt Alassane M., Yoshiyasu Yusuke	4. 巻 -
2. 論文標題 Pathnet: Learning To Generate Trajectories Avoiding Obstacles	5. 発行年 2020年
3. 雑誌名 2020 IEEE International Conference on Image Processing (ICIP)	6. 最初と最後の頁 3194-3198
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICIP40778.2020.9191088	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Gabas Antonio, Yoshiyasu Yusuke, Singh Rohan Pratap, Sagawa Ryusuke, Yoshida Eiichi	4. 巻 -
2. 論文標題 APE: A More Practical Approach To 6-Dof Pose Estimation	5. 発行年 2020年
3. 雑誌名 2020 IEEE International Conference on Image Processing (ICIP)	6. 最初と最後の頁 3164-3168
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICIP40778.2020.9190664	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Druon Raphael, Yoshiyasu Yusuke, Kanezaki Asako, Watt Alassane	4. 巻 5
2. 論文標題 Visual Object Search by Learning Spatial Context	5. 発行年 2020年
3. 雑誌名 IEEE Robotics and Automation Letters	6. 最初と最後の頁 1279 ~ 1286
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/LRA.2020.2967677	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Kalenga-Bimpambu Tshilombo, Yusuke Yoshiyasu, Antonio Gabas, Kota Suzui	4. 巻 -
2. 論文標題 Automatic dataset generation for object pose estimation	5. 発行年 2018年
3. 雑誌名 SIGGRAPH Asia 2018 Posters	6. 最初と最後の頁 16:1--16:2
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3283289.3283326	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Yusuke Yoshiyasu, Ryusuke Sagawa, Ko Ayusawa, Akihiko Murai	4. 巻 -
2. 論文標題 Skeleton Transformer Networks: 3D Human Pose and Skinned Mesh from Single RGB Image	5. 発行年 2019年
3. 雑誌名 In proceedings of Asian Conference on Computer Vision	6. 最初と最後の頁 1-16
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-20870-7_30	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kota Suzui, Yusuke Yoshiyasu, Antonio Gabas, Fumio Kanehiro, Eiichi Yoshida	4. 巻 -
2. 論文標題 Toward 6 DOF Object Pose Estimation with Minimum Dataset	5. 発行年 2019年
3. 雑誌名 In proceedings of SII 2019	6. 最初と最後の頁 462-467
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/SII.2019.8700331	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Pirk Soren, Krs Vojtech, Hu Kaimo, Rajasekaran Suren Deepak, Kang Hao, Yoshiyasu Yusuke, Benes Bedrich, Guibas Leonidas J.	4. 巻 36
2. 論文標題 Understanding and Exploiting Object Interaction Landscapes	5. 発行年 2017年
3. 雑誌名 ACM Transactions on Graphics	6. 最初と最後の頁 1~14
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3072959.3083725	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Yoshiyasu Yusuke, Yoshida Eiichi, Pirk Soeren, Guibas Leonidas	4. 巻 -
2. 論文標題 3D convolutional neural networks by modal fusion	5. 発行年 2017年
3. 雑誌名 2017 IEEE International Conference on Image Processing (ICIP)	6. 最初と最後の頁 1777-1781
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICIP.2017.8296587	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Yoshiyasu Yusuke
2. 発表標題 Visual Object Search by Learning Spatial Context
3. 学会等名 IEEE International conference on Robotics and Automation (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------