

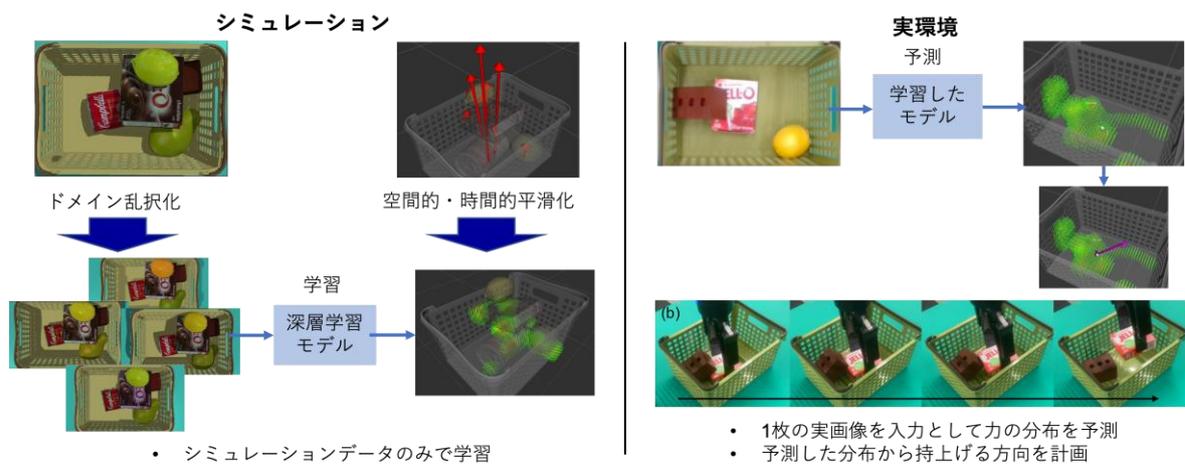
2023年9月29日

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 科学技術振興機構 (JST)

人間らしい「視覚から物体間に働く力を想起する能力」を AI が再現することに成功 安価なセンサーによりロボットの丁寧な物体操作を実現

ポイント

- 視覚情報だけから「崩しそう」「つぶしそう」と想像する力を AI が獲得
- 物体を「壊さないよう」人間らしい推論に基づく行動を AI が立案
- 店舗や工場、物流倉庫におけるロボットの作業、カメラによる事故予測などの応用に期待



視覚情報から物体間に働く力を想起する AI 技術とロボット作業への応用

概要

国立研究開発法人 産業技術総合研究所 (以下「産総研」という) インダストリアル CPS 研究センター オートメーション研究チームの花井亮主任研究員、堂前幸康研究チーム長、Ixchel Ramirez 主任研究員、牧原昂志リサーチアシスタント、原田研介特定フェロー、人工知能研究センター 尾形哲也特定フェローは、視覚情報から物体間に働く力を想起する AI 技術を開発しました。

人間は物体にかかるおおまかな力や物体の柔らかさなどを視覚のみから経験的に推論し作業することができます。例えば崩れやすそうなものや柔らかそうなものを見つけたら、崩したりつぶしたりしないように丁寧に扱わなければならないと考えます。このように人間は経験に基づき視覚から異なる感覚を呼び起こすことで、多様な行動を計画することができます。

本研究では、このような人間らしい「視覚から異なる感覚を想起する能力」を AI が再現することに成功しました。物理シミュレーター上で物体間にかかる力を可視化した仮想的な経験データを構築し、この仮想的な経験から視覚と別の感覚（力）の関係を AI が学習することでこれを実現しました。

実験では、カメラ 1 台で未知の物体間に発生するおおまかな力分布をリアルタイムで可視化することに成功しました。また力分布を理解したロボットが周辺の物体の損傷が小さくなるように指定された対象物を持ち上げるなど、

人間らしい推論に基づく行動ができることを確認しました。今後、ロボットによる丁寧な物体操作や、自動運転における事故予測などに役立つことが期待されます。

本研究開発成果は、2023年10月1日から5日まで開催のロボット分野の国際会議「IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2023」にて発表されます。

下線部は【用語解説】参照

開発の社会的背景

人間は過去のさまざまな経験に基づき、視覚から別の感覚を想起することで柔軟かつ多様な作業を行うことができます。目で見ただけで、柔らかいものや崩れやすいものを丁寧に操作するように意識します。しかしロボットにおいて、搭載される視覚センサーだけでは、このような感覚を再現することは困難でした。例えばロボットにおける運動のダイナミクスに関する推定は、一般的にロボットと物体の接触後に力覚センサーや触覚センサーを使って判断していました。しかし、この判断法では対象物が崩れてしまうなど、物体に接触してからでは遅い場合において、器用な作業をすることは困難でした。これに対して、視覚から運動のダイナミクスに関するような別の感覚の想起（クロスモーダル）を簡易に実現することができれば、安価なセンサーから人間のような行動計画を実現することができ、将来のロボットや自動運転システムなどへの貢献が見込まれます。

研究の経緯

産総研は、シミュレーションを利用したロボットの認識や動作生成技術の研究を行ってきました。本研究ではその知見を生かし、画像から物体間にかかる力の分布を想起するAI技術を開発、ロボットへの応用の可能性を探りました。

なお、本研究開発は、国立研究開発法人 科学技術振興機構（JST）ムーンショット型研究開発事業研究開発プログラム：「2050年までに、AIとロボットの共進化により、自ら学習・行動し人と共生するロボットを実現」（2020～2025年度）による支援を受けた研究開発プロジェクト「一人に一台一生寄り添うスマートロボット（PM：菅野重樹（早稲田大学）」の研究課題の一つとして取り組んでいます。

研究の内容

花井らは視覚センサーにより得られる画像に写る物体が周囲の物体と接触することで生じる力のおおまかな分布を想起し、リアルタイムに3次元で可視化することに成功しました。今回開発したAI技術では、物体同士が接触することで生じる力を画像から推定し、力の大小を視覚的に表現することができます。図1にその例を示します。図1上は視覚センサーによる入力画像、図1下はその入力画像からAIが想起した力のおおまかな分布を可視化したものです。カゴの中にある物体が（カゴの側面や底面を含む）周囲の物体と接触することで生じる力の大きさを、緑から赤の色で表しており、赤に近い色ほど相対的に大きな力がかかっていることを表しています（ただし、図を見やすくするために一定値よりも小さい力は表示していません）。さまざまなシーンにおいて物体とカゴの側面や底面、物体同士の接触位置周辺に力がかかっています。また対象物にはAIの学習に使われていない未知物体が含まれています。これは提案するAIが多様な日用品に対する汎化性能を獲得できる可能性を示しています。さらに提案手法はシミュレーション内で学習をしていますが、現実環境に適応するにあたって追加学習をしていません。これは機械

学習の研究において、ゼロショット転移といわれるものです。仮想世界で学習した AI をそのまま追加学習なしで現実世界に適用できるため、学習効率が良く、応用がしやすいことを表しています。このように画像 1 枚（カメラ 1 台）から日用品の物体間にかかる力分布をリアルタイムに推定できる AI 技術を実現しました。



図 1 提案手法により、画像から物体間に働くおおまかな力分布を想起した様子
（緑から赤になるほど相対的に大きな力がかかっている）

提案手法は、(a)寄りかかった箱とカゴとの接触位置に力を想起（カゴ側面からの接触力は小さい）、
(b)円筒の缶とカゴ底面の間に細い線状領域での接触を予測、
(c)積み重なった物体間（りんごとオレンジの間など）の接触位置にも力を想起している。

提案手法は、(d)(e)未知物品に対しても同様の想起が可能

上：入力画像、下：想起された 3 次元の力分布

※原論文の図を引用・改変したものを使用しています。

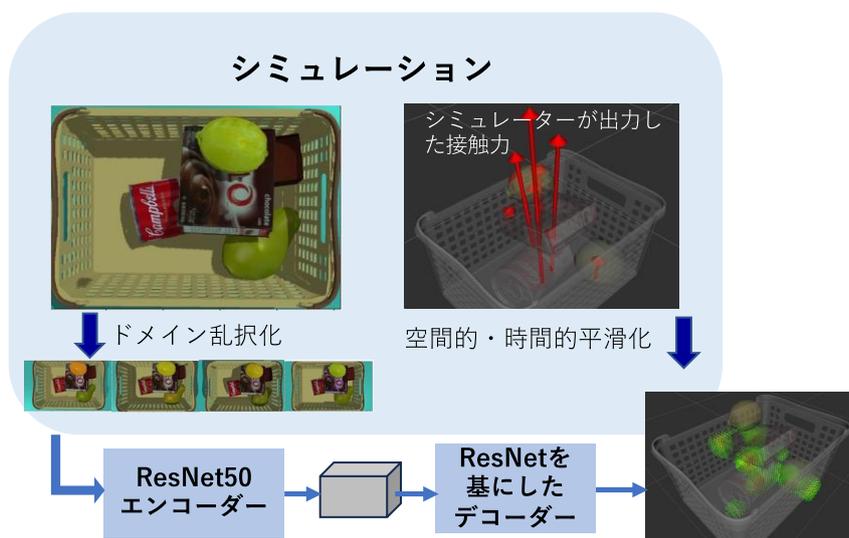


図 2 視覚から物体間の力の分布を推定する学習方法

※原論文の図を引用・改変したものを使用しています。

図 2 は提案手法による学習方法の概要です。視覚情報に対して正確な力分布を得ることは現実世界では困難です。ここでは物理シミュレーターを使い、ランダムに配置される日用品の 3D モデル間にかかる「物体間に働く力」を物理演算に基づき計算しています。しかし、物理シミュレーター上の物理演算は正確に実世界の物理現象を再現しま

せん。そのため、正確な力の量を推定しようとするればシミュレーションと現実のギャップ（機械学習の研究ではドメインギャップといわれます）を埋めるために膨大な労力を必要とします。そこで研究チームは、正確な物体の力の量ではなく、物体間のおおまかな力の分布に着目をしました。この場合、物理シミュレーションで得られる情報でも時間的・空間的な平滑化を行うことで比較的安定して現象を再現することが可能です。そのため、AIに大量の経験を積ませなくても、ドメインギャップを埋めやすいことがわかりました。この工夫に基づきシミュレーター上で仮想経験データを生成することで、図2に示すように、視覚的情報を入力し、物体間の力分布を出力する深層学習モデルを訓練しました。モデルは ResNet50 をエンコーダーとし、ResNet を基に設計したデコーダーと組み合わせることで構成しました。その結果、図1に示したようにゼロショットでの実カメラへの適用が可能となりました。通常シミュレーションを現実に似せることで学習効率を高めますが、今回は「現実では得難い経験」を積ませることで、現実世界での経験を補完した点も、技術的に重要な成果です。

またロボットによる物体操作（マニピュレーション）の計画問題に提案手法を応用しました。図3(a)、(b)に示すように、本提案手法を加えることで、視覚情報だけから、人間のような丁寧な物体操作方法を計画できました。将来器用に物体を操作するロボットへの応用が期待されます。



図3 ロボットマニピュレーションの計画問題への応用

(a) 提案手法がない場合。物体を持ち上げる際に上にある他の物体を落としてダメージを与えている。

(b) 提案手法がある場合。上の物体のダメージが少ないような丁寧な操作が実現できている。

※原論文の図を引用・改変したものを使用しています。

また、これまでに産総研の牧原昂志リサーチアシスタントらは、力の分布と同じような考え方に基づき、物体の柔らかさを推定しました[1]。少数の商品の3Dモデルに対して手動で与えた物体の柔らかさを表すマップを貼り付け、シミュレーション上で大量にデータを生成することで、深層学習モデルに視覚（距離画像センサーによる深度画像）から得られる物体形状と柔らかさの関係を学習させました（図4）。これにより、形状から物体の種類が推定できる場合にシーン中の柔らかさの分布を予測できるようになりました。例えばあるペットボトルの形と柔らかさを訓練すれば、少し形の異なる未知のペットボトルに対しても柔らかさを推定できます。このAIをロボットに適用したところ、物体把持に関する最新手法である Dex-Net と比較して、作業成功率は同等でかつ把持対象の物体のつぶれは70%以上抑えられることがわかりました。また、逆に形状だけから判断すると把持できない状況で、柔らかか

さを利用して周辺のものを変形させて押しつけて把持するという人間らしい行動をロボットがとることにも成功しました（図5）。

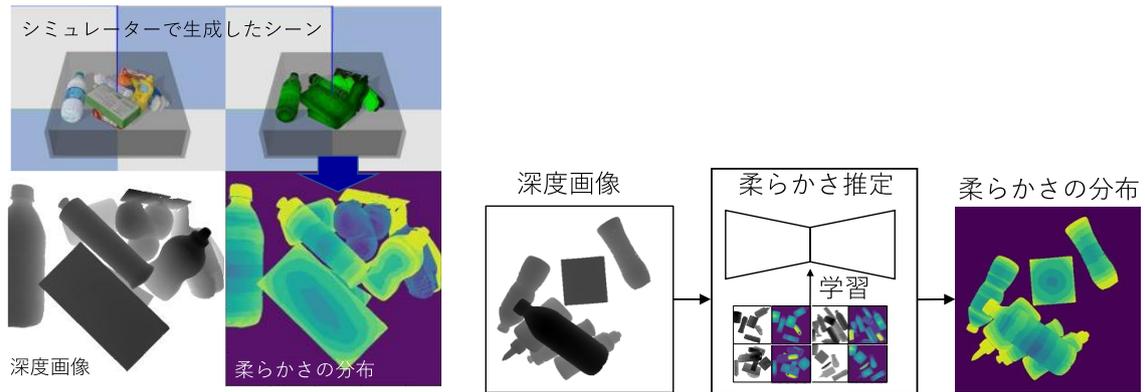


図4 視覚から柔らかさの想起

- (左) シミュレーター上で深度画像と物体の柔らかさの分布からなる仮想経験データを構築
- (右) 深度画像から柔らかさを予測するモデルを学習（黄から緑、青と変わるほど柔らかい）



図5 柔らかい周辺物体を押しつけて対象物を把持

今後の予定

このような視覚から別の感覚を想起する技術は、安価なセンサーにより人間らしい器用な行動計画を実現します。工場や物流倉庫におけるロボットによる物体操作や、今後家庭に普及するスマートロボットへの展開が期待されます。またシミュレーションによる経験データの作り方・与え方を工夫することで、自然環境において崩れやすい場所を見つけるという応用も想定されます。将来的には自動運転における危険予知や、衛星画像からの災害予知など、広い分野への適用を目指します。

学会情報

IROS2023 で発表予定（2023年10月1日から5日まで開催）

論文タイトル：“Forcemap: Learning to Predict Contact Force Distribution from Vision”

著者：Ryo Hanai, Yukiyasu Domae, Ixchel G. Ramirez-Alpizar, Bruno Leme and Tetsuya Ogata

参考文献

[1] Koshi Makihara, Yukiyasu Domae, Ixchel G. Ramirez-Alpizar, Toshio Ueshiba and Kensuke Harada, *Advanced Robotics*, 2022, vol. 36, no. 12, 600–610.

<https://doi.org/10.1080/01691864.2022.2078669>

用語解説

物理シミュレーター

物体の運動や衝突などの物理現象をコンピュータ上でシミュレーションするソフトウェアです。

ダイナミクス

ロボットや物体に働く力とその結果生じるそれらの運動との間の関係を指します。

追加学習

機械学習のコンセプトの一つで、事前に学習したモデルを新しいタスクやデータに適応させるプロセスを指します。ここではシミュレーションデータを用いて学習したモデルを、現実環境のデータを用いて再学習させることを指します。

ゼロショット転移

あるドメイン（シミュレーションデータ）で学習したモデルを別のドメイン（実環境データ）に応用することを転移学習といいます。このときに転移先ドメインのデータを用いずに転移させることをゼロショット転移と呼びます。

ResNet50

ResNet は畳み込みニューラルネットワークの代表的なアーキテクチャの一つであり、同一構造の層を積み重ねることでパラメータ数の異なるネットワークを構成することができます。ResNet50 はその中で 50 層の構成のものを指します。

Dex-Net

視覚センサーで得られたデータに対してロボットの適切な把持位置を計算する手法です。物品の 3 次元形状とハンドのモデルを使って生成した大量データを用いた深層学習を行います。

本件に関する問い合わせ先

国立研究開発法人 産業技術総合研究所

インダストリアル CPS 研究センター オートメーション研究グループ

主任研究員 花井 亮

〒135-0064 東京都江東区青海 2-4-7 産総研臨海副都心センター

029-861-2780 ryo.hanai[at]aist.go.jp

[JST 事業に関すること]

国立研究開発法人 科学技術振興機構 ムーンショット型研究開発事業部

古賀 明嗣

03-5214-8419 moonshot-info[at]jst.go.jp

機関情報

国立研究開発法人 産業技術総合研究所

<https://www.aist.go.jp/>

ブランディング・広報部 報道室 hodo-ml[at]aist.go.jp

国立研究開発法人 科学技術振興機構

<https://www.jst.go.jp/>

広報課 03-5214-8404 jstkoho[at]jst.go.jp