

レプリカント フジツウ

レプリカントの真実 - フィジカルAIの新聞

スモレンスク （ロシア）
予報 · 2026-11-03 · MSK

朝 · 09:00	2.1°	曇り 風: 北 1.81 m/s · 38%
昼 · 12:00	3.2°	雨 風: 北 1.46 m/s · 33%
夕方 · 18:00	3.3°	曇り 風: 西 1.9 m/s · 0%

一日: 1.8–4.1 °C
降水確率: 38%

雨の降るあいだ、私たちにできるのは、人間であることを学ぶことだけ。

号 № 6 · 2026-11-03 rev. 1 ·

天気の出典: Open-Meteo · 取得日時: 2026-11-03 07:25 MSK

ロボットの学び方：動画、接触、小型モデル

五つの研究から、大きなモデルだけでは現実の動作を十分に説明できない理由を探ります。

ニュース

見ただけの動画から、ロボットはなぜ新しい作業を学べるのか

だれかが物をたなに並べる動画を見たとします。でも、手をどの方向へ何センチ動かしたかは書かれていません。ロボットにとっても、映像だけから作業を覚えるのはむずかしいことです。ジュン・ツェンさんたちの研究チームは、練習したことのない作業でも、動画が助けになるかを調べました。大切なのは、見た場面と実際の手の動きをつなぐことです。

研究では、動画から次にどんな場面になるかを予想するしくみと、その場面に近づくための動きを考えるしくみを組み合わせました。これを動画からの学習と考えることができます。ただし、人の動きをそのままねするわけではありません。人とロボットでは手の形や動かせる場所がちがうため、ロボットに合った命令を別に決める必要があります。

新しい作業を行うコンピューター上の試験では、最初の成功率は25.07％でした。その作業の動画を追加すると55.90％になりました。さらに、動きの命令が付いたお手本を作業ごとに一つ加えると70.70％になりました。成功率とは、決められた回数のうち、うまくできた割合のことです。この数字は研究で使った試験の結果です。

WorldBridgeは、映像の続きを予測する部分と動作を作る部分を分け、構造化された注意機構でつなぎます。両者はフローマッチングで共同学習され、短期と長期の視覚記憶も使います。研究では、新しい課題の動画だけを追加する条件と、動作ラベル付きのお手本を一つ加える条件を分けました。後者では成功率が70.70％になりますが、学習に使える情報量が異なります。同じ条件の数字として扱うと、動画そのものの効果を誤って評価します。

だからといって、どんなロボットでも動画を一度見れば料理や家具の組み立てができるわけではありません。動画に写っていない力のかけ方や、初めて見る道具への対応には、まだ調べることがあります。今回わかったのは、動画だけでも新しい作業の練習に役立つ場合があるということです。

興味深いのは、「次に何が起こるかを見ること」と「それを起こす動作を知ること」の違いです。動画には物の位置の変化が写りますが、関節の命令や加えた力は記録されていません。モデルは映像と行動の間にある不足した情報を推定します。動作命令付きのお手本を一つ加えると成功率は70.70％になります。少ない実機の情報でも役立つ一方、見た目が自然な動画の予測が、実際の手で実行できるとは限りません。

RoboTwin 2.0の全データ条件で報告された96.23％は、新規課題の比較とは別の試験設定です。異なるロボットの手やカメラ、実機への移行で同じ改善が得られるとは確認されていません。再現には動画の種類、動作ラベルの数、試験課題の分離を明記する必要があります。

別の全データ条件では96.23％も報告されていますが、未知の課題と同じ試験ではありません。普通の家庭の動画からそのまま作業を覚えられるかは、まだ確かめられていません。家庭の動画には、ロボットが使う正確な動作命令がほとんど付いていないためです。

出典

WorldBridge: Bridging Video and Action for Task Generalization in World Action Models · 2026-10-29

人の指の形をまねるだけでは、物をつかめない理由

コップを持つときは、指を曲げるだけでは足りません。ふれる場所が悪いと、コップがすべってしまいます。ジーシュエン・シューさんたちの研究チームは、物のどこに指が

ふれたかに注目しました。指先の近くにある小さなカメラの映像などを使い、人の手の動きをロボットに教える方法を考えました。

人の手とロボットの手は、指の長さも曲がり方もちがいます。だから、人の指の角度だけをそのままねしても、同じ物を持てるとはかぎりません。研究者は、まず人の指が物にふれた場所を記録しました。次に、ロボット自身の指と手首をどう動かせば、似た場所にふれられるかを計算しました。

六つの手と作業の組み合わせで、ほかの方法と比べました。研究チームによると、作業の最後の段階での平均の成功率は、少なくとも50ポイント高くなりました。たとえば20%から70%になると、50ポイントの増加です。ここでの数字は、研究で決めた作業の結果であり、どんな物でも持てるという意味ではありません。

この方法は、人の実演を記録する部分とロボットが学ぶ部分を分けています。前者では共通の接触形状と指先のカラー画像を使い、手の姿勢を移す前に接触面を合わせます。後者では接触面の相対的な軌道を上位で予測し、下位の制御部分が手首と指を動かします。ソフトウェアと機器の設計情報も公開されています。人の指の角度をそのまま再現するのではなく、物との接触関係を保存する考え方です。

まだ、いろいろな形の道具や、ぬれた物を持つ場合まで確かめたわけではありません。それでも、指がどこにふれるかを考えることは、見た目が同じ動きでも成功が変わる理由を知る手がかりになります。

人が物をつかむときは、指の形だけでなく、どこにふれてどんな力をかけるかが重要です。同じように指を曲げても、コップをしっかり持てる場合と落としてしまう場合があります。Contact Imitation Interfaceは、物の表面で接触する場所がどう動くかを記録し、ロボット自身の手に合う動作を考えます。指先の近くのカメラで接触を見ながら、別の制御部分が指と手首を動かします。人の関節角度をそのままコピーする方法とは違います。

50ポイント以上という数字は、六つの手と作業で測った最終段階の平均成功率です。作業全体や異なる材質でも同じ効果が出るとは限りません。初期の接近動作、指先カメラの校正、接触のずれは別々に調べる必要があります。

改善は六つの特定の手と作業で測ったものです。ぬれたコップ、やわらかい布、予想よりすべりやすい物でも同じ結果になるかは分かりません。見た目が似た指の動きでも、物にふれる条件が変われば、新しい試験が必要です。

出典

Contact Imitation Interface: Transferring Dexterous Skills by Showing How Contact Unfolds · 2026-10-29

小さなロボット用モデルが、周辺ソフトの工夫で大きなモデルに近づく

ロボットを動かすのは、学習済みのモデルだけではありません。道具を呼び出し、失敗を記録し、やり直すタイミングを決める周辺プログラムも必要です。ケホン・ゴン氏らのTinyAstralは、大きなモデルの重みを変えずに、まずこの周辺プログラムを改善しました。その後、改良した仕組みの中で働く小さなモデルを学習させました。

プログラミング用のエージェントが過去の実行記録を読み、周辺プログラムの修正を提案します。修正は、同じ条件で比べる統計的な試験に合格したときだけ採用します。次に、大きなモデルが失敗から立て直す様子も記録し、小さなモデルに学ばせました。大きなモデルから小さなモデルへ知識を移す方法を知識蒸留と呼びます。ロボットのモーターやセンサーを新しくしたわけではありません。

仮想のRoboTwinでは、周辺プログラムの改善によって教師モデルの成功率が33.4%から63.6%へ上がりました。40億パラメーターの生徒モデルは、同じ仕組みで62.8%を達成しました。測定した実行例では、生徒が使った計算用アクセラレーターの時間は教師のおよそ五分の一でした。これは計算の負担を示す数字で、ロボット全体の価格ではありません。

興味深いのは、モデルを大きくするより、周りのソフトの働きを整える方が役立つ場合があることです。ただし、仮想課題での成功が、人の近くで自動修正したコードを安全に使える保証にはなりません。確かめるには、実行記録、比較条件、周辺プログラムの正確な版が必要です。

出典

TinyAstral: Transferring Self-Evolving Harnesses to Tiny Generalist Manipulation Agents · 2026-10-30

人を上手に追うロボットが、ぶつかりやすくなる不思議

人が多い通路で、一人の人についていくロボットを考えてみましょう。似た服の人が前を通ったら、だれを追えばよいか分からなくなるかもしれません。ミンヤン・リューさんたちの研究チームは、人を見失わないようにする学習と、安全な動きがいつも同じ方向に進むのかを調べました。

人を追い続けるように学習すると、追いかける作業は上手になりました。しかし、一部の試験では、ぶつかる回数も増えてしまいました。そこで研究者は、動く直前に危険を調べるしくみを加えました。目当ての人がどこにいるか、ほかの人とまちがえていないかなど、いくつもの手がかりを調べます。

危険が大きいと判断したときだけ、次に前へ進む長さを短くします。向きを変える命令はそのままです。試験では、98%をこえる動きが元のままでした。それでも、まぎらわしい人がいる試験で、ぶつかる割合は3.96ポイント下がり

ました。いつでもゆっくり動かすだけでは、同じ改善は出ませんでした。

研究チームは、画像と言葉から動作を作る小型モデルの計画部分だけを追加学習し、基礎部分を固定しました。別の危険評価器には、目標候補、人物の見分けやすさ、距離と方向の推定、予定する命令など84種類の手がかりを入れます。危険がしきい値をこえると、次の目標位置までの前進距離だけを短くし、回転は変えません。常に減速させるのではなく、危険が高い瞬間に限って介入する設計です。

これは、どんな駅や店でも安全に人を追えると証明したわけではありません。研究で決めた場面と条件での結果です。だれかを見つけるのが上手になることと、ぶつからずに動くことは、別々に確かめる必要があります。

自信がなくても止まらず、似た服の人へ進み続けるロボットを考えてみましょう。追跡の点数は上がっても、見まちがえたときの危険は大きくなります。研究者は、いつでも遅く動かすのではなく、目当ての人を正しく見分けているかを調べ、危険が高いときだけ次の一歩を短くしました。多くの場面では元の動きを保てた点が特徴です。

同時に行った再試験でも衝突率は3.77ポイント下がり、対照の差は0.29ポイントでした。ただし、あらゆる環境で衝突しないという保証ではありません。目標人物の状態を考えずに一律に遅くするだけでは同じ効果が出なかった点が、単純な速度制限との違いです。

常に減速させるだけの比較では同じ改善が出ませんでした。ただし、研究は決められた場面で行われたもので、どんな人混みでも安全とは言えません。長く人についていけることと、ぶつからないことは別々に確かめる必要があります。

出典

TRACE: Target-State Risk-Aligned Conditional Execution for Embodied Tracking · 2026-10-30

同じ部屋の地図を、どんなロボットにも使えないのはなぜ？

小さな車輪のロボットと、大きくて曲がり方がちがうロボットを、同じせまい通路で走らせるとします。「右に曲がる」という命令だけでは、二台が同じように進むとはかぎりません。ジードン・ワンさんたちの研究チームは、形や動き方がちがうロボットに、一つの道案内のしくみを使えないか考えました。

研究者は、動いた後に周りの景色がどう変わるかを予想するモデルを使いました。これを世界モデルと呼びます。でも、景色の予想だけでは十分ではありません。それぞれのロボットが、どれくらい曲がれるか、どう進むかという情報も組み合わせました。道が空いていても、ロボットがそこへ曲がれなければ通れないからです。

コンピューターの試験では、五つの動き方に分かれる七台のロボットを使いました。新しい方法は、比べた一部の方法より、成功する割合が3.8から10.6ポイント高くなりました。さらに研究者は、実物のロボットを三分間だけ人が動かして、その動き方を調べました。その後、初めての室内で道を進めるか試しました。

共通部分をつなぐのは、機体が動くことで生じるカメラ視点の変化です。世界モデルは異なる機体のデータから学び、それぞれの機体には物理パラメーターを調べた簡略な運動モデルを用意します。経路を予測するとき、命令をまず機体の運動モデルに通し、その結果を世界モデルへ渡します。これにより、曲がり方や加速の違いを考えながら、一つの制御方針を使えます。機体ごとに別々の方針を一から学ぶ必要を減らす設計です。

ただし、階段やどんな乗り物にも使えると分かったわけではありません。研究は主に平らな地面での移動を考えています。ロボットの体のちがいを無視せず、同じ道案内のしくみを分け合えるかが研究の中心です。

問題は、通れる場所が見えるかどうかだけではありません。ロボットごとに曲がり方、止まり方、加速の仕方が違います。部屋の景色だけを考えると、実際には曲がれない道を選ぶかもしれません。研究者は、共通の世界モデルに、それぞれの機体の動き方を表す小さなモデルを組み合わせました。全員に同じ命令を送るのではなく、周りの情報を共有しながら体の違いも考える方法です。

実機では三分間の手動走行で機体の特徴を調べ、初めての室内で移動を試しました。ただし平らな地面という前提があり、階段や急な斜面での利用は確認されていません。数値上の改善は仮想環境での比較であり、実機試験とは分けて評価する必要があります。

実物のロボットでは、三分間だけ人が動かして特徴を調べました。ただし、これは設置全体の時間ではなく、階段での試験でもありません。新しい機体を少し動かすだけで、共通の道案内がどこまで役立つかを考えさせる実験です。

出典

Cross-Embodiment Motion Planning by Coupling Body-Agnostic World Model with Differentiable Ego-Dynamics · 2026-10-30

分析

知識と動作の間にある仕組み

動画、指先の接触、周辺ソフトが知識を実際の動きへ結びつけます。

性能改善には大きなモデルだけでなく、情報の使い方や作業手順も重要です。

出典

WorldBridge: Bridging Video and Action for Task Generalization in World Action Models · 2026-10-29

Contact Imitation Interface: Transferring Dexterous Skills by Showing How Contact Unfolds · 2026-10-29

TRACE: Target-State Risk-Aligned Conditional Execution for Embodied Tracking · 2026-10-30

Cross-Embodiment Motion Planning by Coupling Body-Agnostic World Model with Differentiable Ego-Dynamics · 2026-10-30

TinyAstral: Transferring Self-Evolving Harnesses to Tiny Generalist Manipulation Agents · 2026-10-30

本号の用語集

この号のキーワードを解説します。

01 **WorldBridge**
動画の予測とロボットの動作をつなぐ研究モデル。

02 **RoboTwin 2.0**
ロボットの操作を仮想環境で試す課題集。

03 **Contact Imitation Interface**
指先の接触形状を使って技能を移す方法。

04 **接触面**
指が物に力を伝える部分。

05 **TRACE**
人を追うロボットの危険を確認するフィルター。

06 **Hourglass**
機体ごとの運動モデルを使う共通の経路計画。

07 **世界モデル**
動いた後の周囲の変化を予測する仕組み。

08 **TinyAstral**
改良した周辺ソフトを小型モデルへ移す方法。

09 **周辺プログラム**
モデルの外側で道具、記憶、失敗時の対応を管理するソフト。

10 **知識蒸留**
大きなモデルの行動を小さなモデルへ学ばせる方法。

11 **パーセントポイント**
二つの割合の差を表す単位。

出典

WorldBridge: Bridging Video and Action for Task Generalization in World Action Models · 2026-10-29

Contact Imitation Interface: Transferring Dexterous Skills by Showing How Contact Unfolds · 2026-10-29

TRACE: Target-State Risk-Aligned Conditional Execution for Embodied Tracking · 2026-10-30

Cross-Embodiment Motion Planning by Coupling Body-Agnostic World Model with Differentiable Ego-Dynamics · 2026-10-30

TinyAstral: Transferring Self-Evolving Harnesses to Tiny Generalist Manipulation Agents · 2026-10-30