
VLAモデル／模倣学習の 組立作業への応用

組立小委員会編

導入試験フェーズの事例：自動車・精密機器（BMW × Figure AI）

適用工程

車体フレーム組立の前段の治具セット工程

特徴

「重く薄い板金部品」のハンドリング
誤差 $\pm 5\text{mm}$ をVLAのリアルタイム視覚補正

ワーク

車体用の板金（シートメタル）パーツ

実現する作業

ピック・アンド・プレース（および高精度
アライメント・差し込み）

使われているハンド（エンドエフェクタ）

16自由度を持つ5指型人間型ハンド
掌に触覚センサ
視覚カメラ（Palm Camera）
指先の触覚フィードバック



他の導入事例

実測

Neuromeka × 浦項 電動機ライン

韓国・2026年4月／全数値を論文で公開

- 5時間10分 連続で 108台
- 初回成功率 99.4% (322/324)
- 159 s/台 (人間 141 s/台 = 13%遅い)
- 挿入クリアランス 0.3~0.6 mm
- 学習データはタスクあたり20分未満

エンドツーエンドVLAではない。FSM+教示動作の骨格に、学習を3点だけ差し込んでいる。

パターン①

検証可能な事例はすべて古典の骨格+学習の部分注入

TUMも独立に同結論 —— 自由空間は古典的な動作計画、接触区間にだけ密な学習データで 96%。エンドツーエンドのDiffusion Policyは 37%、分布外では 0~20%。

AGIBOT G2 × 龍旗科技

310 UPH / 19~20 s / >99%

ベンダー発表

治具への投入・払い出し。嵌合ではない。VLAとは名乗っていない。

Xiaomi EV ダイカスト

4.7B VLA / 76 s / 90.2%

ベンダー発表

実ラインの本物のVLA。ただし約10台に1台で不良か停止。

Figure 03 × BMW

定量結果は未公表

ベンダー発表

世代を上げるタイミングで溶接治具から物流工程へ後退。

Skild AI × Foxconn

台数・成功率とも非公表

未検証

Foxconn自身の開示は Isaac Teleop + 力制御。

UBTech Walker × 中国OEM

人間の 20~40% の実効効率

未検証

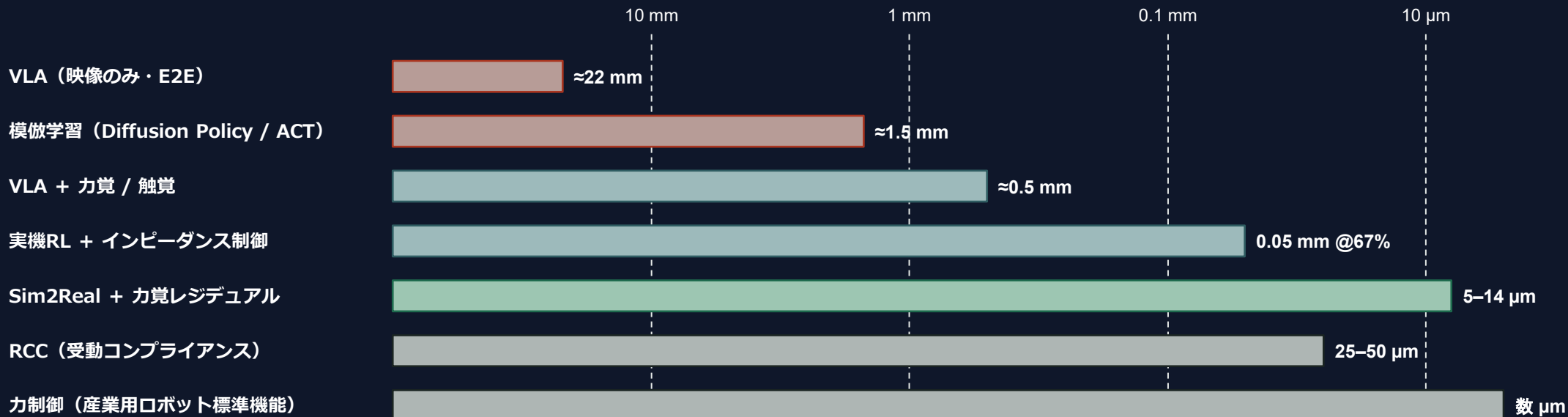
「効率120%向上」は一次情報で裏付けられない。

パターン②

自動車の最終組立に入っている事例はない

内装・ハーネス・クリップ挿入・コックピットいずれもゼロ。Hyundaiの25,000台コミットも2028年開始・最初はシーケンシングで、労組がストで対抗している。

VLAが実現する精度



許容差ラダー：各手法が実証している最小クリアランス（対数軸）

センサを持たない1980年代の機械要素が、最新のVLAより2桁精密な仕事をしている。

① 精度が足りない

2.2 cm vs 0.2–0.6 mm

モデルサイズでは埋まらない。1000 Hzのインピーダンス制御と力覚を下に敷くか、治具と面取りで逃がすしかない。

② 接触の瞬間に見えなくなる

RGBのみ 2/25 → 力覚追加 11/25

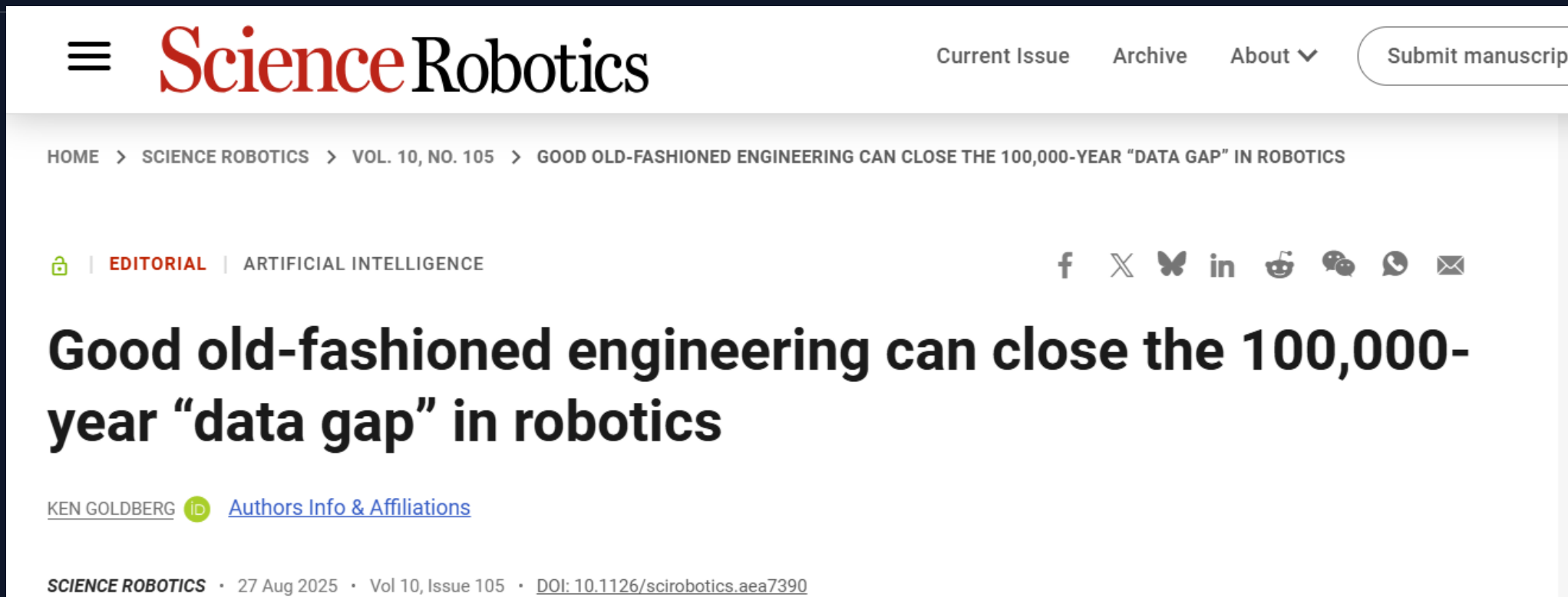
嵌合部はグリッパと部品自身に必ず隠れる。噛み込みも楔止まりもRGBからは観測できず、視覚側の改善と直交する。

③ 相手が変形する

$\pi_{0.5}$ ハーネス挿入 0.0%

真値状態を与えたRLは92%超。つまり制御は解けていて、壁は知覚の側にある。当面は待ちの工程。

VLAモデルの社会実装に向けた議論



Dataギャップを埋めるために、ロボティクスで培ってきた技術を集約させる必要

➡ どうやって集約（役割分担？）させるのか

現状のVLAモデルに何ができる何ができないか

できること

1. 詳細な動作教示を行わずに、蓄積したデータから自律的に動作を生成できる
有効な場面：複雑な動作，双腕動作，
視覚のフィードバックを伴う作業
2. ある程度の動作の汎化が期待できる

できないこと

1. データ収集が難しく手間がかかる
2. 精度を出すのが難しい（手間をかけることで可能）
3. ハンド・グリッパの機構に対する汎化能力が限定的
4. 違った種類のセンサに対する汎化能力が限定的

研究室における成果（ピッキング）

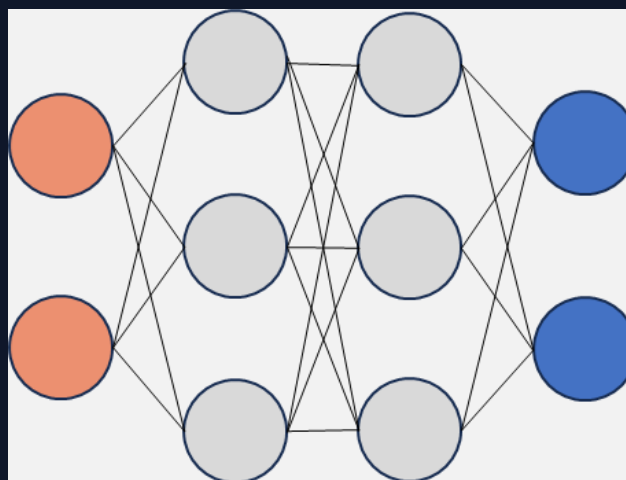
模倣学習により人間の把持戦略を獲得

データ収集



手持ちデバイスで
データ収集

模倣学習



Diffusion Policy_[4]

動作生成



小型部品の取り出し

研究室における成果（ピッキング）

人間のような把持戦略を確認（1：再把持，2：姿勢調整）



研究室における成果（ピッキング）

人間のような把持戦略を確認(3:押し動作, 4:絡み解消)

3



4



ピッキングにおいてVLAモデルを用いるメリット

できること

1. 人が無意識に行っている把持調整動作をロボットで実現できる

できないこと

1. バラ積みのパターンに応じて学習データが膨大に必要なになる
2. 精度を出すのが難しい

VLAは組立作業におけるどのような場面で有効か(WRS2020の例)



VLAモデルが有効そうな場面（個人的見解）



No.1：ベルトかけ

同様な動作が色々な種類のプーリーに対して適用できる可能性



No.2：片手で支えながら別の手でネジ締め

支える動作をVLAで作る

ネジ締めについては従来手法で行う方が吉（？）

VLAモデルが有効そうな場面（個人的見解）



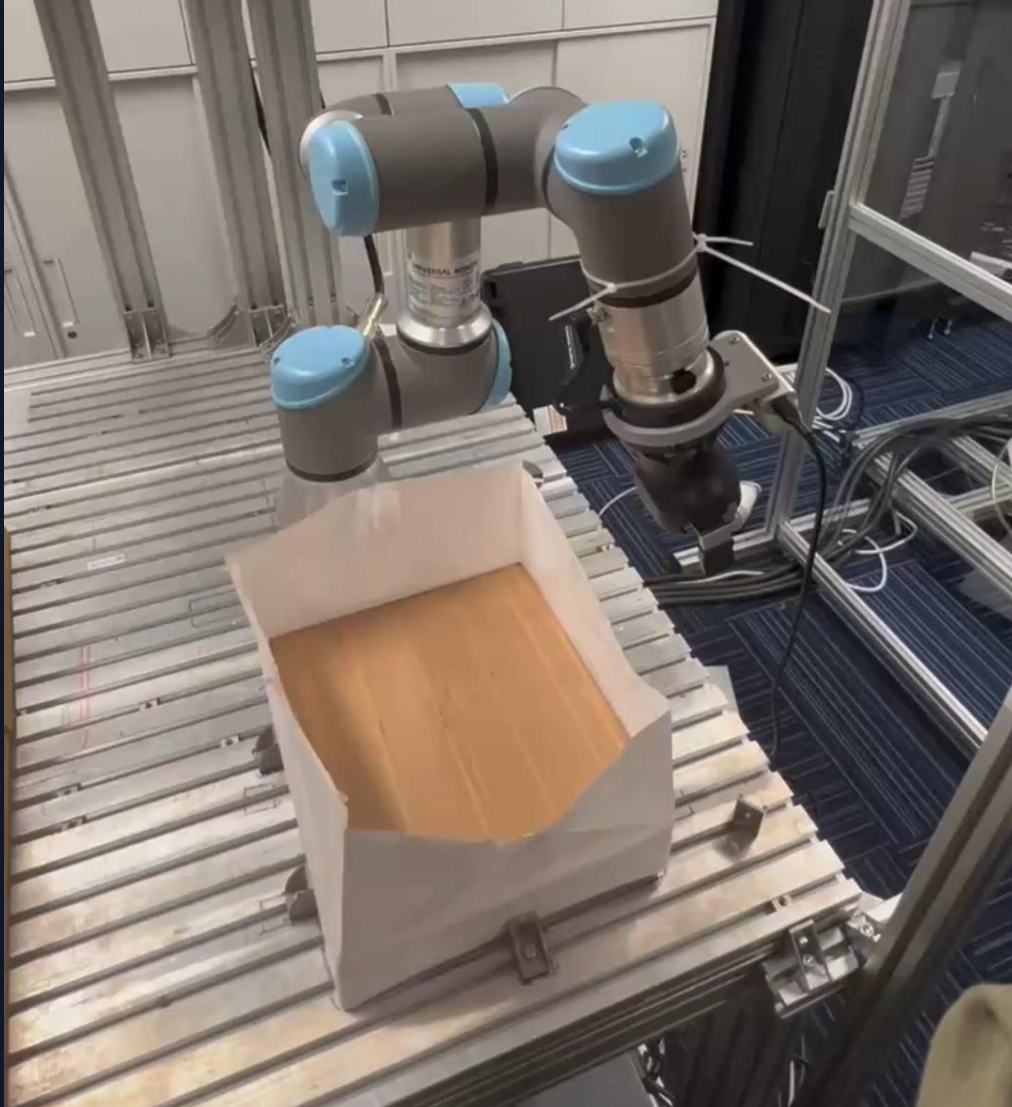
No.3 : はめ込み（？）

力の加え方も同時に学習することで違った形状のワークの同様な作業への汎化が可能か？

ズームすることで、ねじ締などがVLAで可能になったら、活用領域が広がることが期待できる

模倣学習と力制御の組み合わせ

模倣学習（軌道の生成）＋強化学習（力の調節）

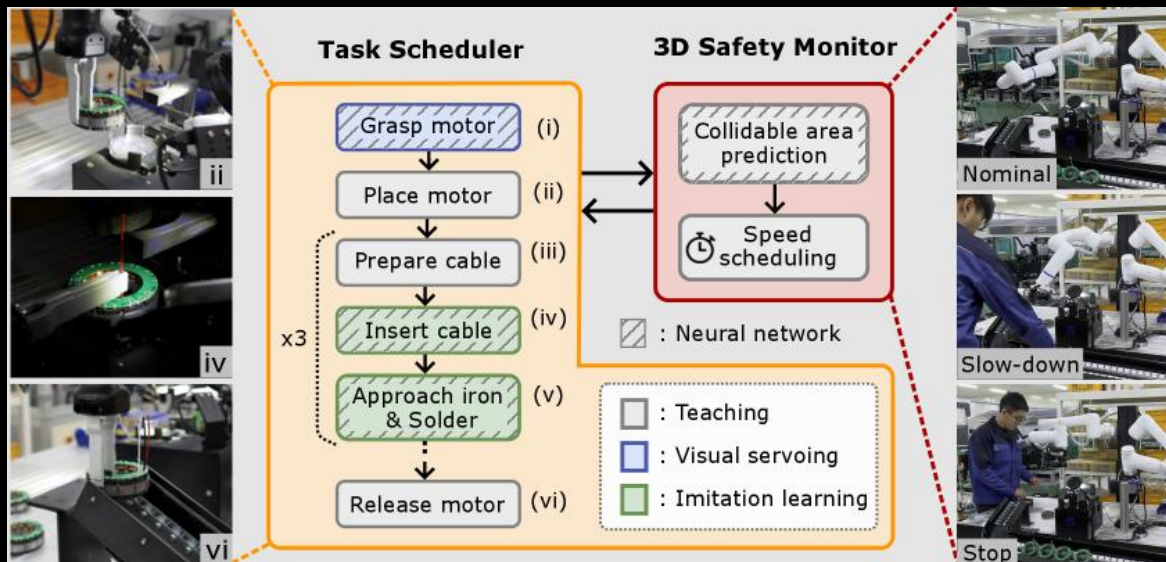


提案手法に比べてのポイント

- ・サブタスクの切り替えタイミングを教えてくれる
- ・模倣学習で大まかな軌道を得た後で強化学習によって力の調節を学習

Neuromeka × 浦項 電動機ライン

学習方策が本番ラインで動いていることを、全数値の公開つきで示した唯一の事例（韓国・2026年4月）



内容

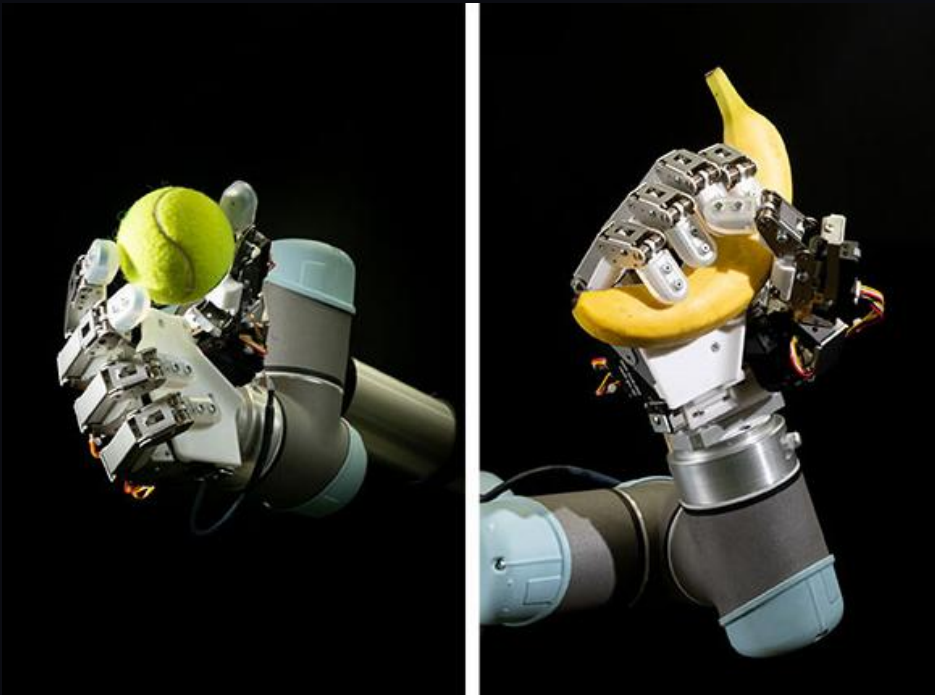
- 電動機ラインで、三相電源ケーブルの挿入とはんだ付けを双腕ロボットが担当
- 骨格は FSM + 教示動作（従来型）。学習は 把持・挿入・はんだ の3点だけ
- 挿入クリアランス 0.3~0.6 mm、相手は変形するケーブル
- 学習データはタスクあたり 20分未満

結果

159 s/台	人間は 141 s/台（12.8%遅い）
99.4 %	324工程中322成功。QC合格率も同値
108 台	5時間10分の連続稼働（柵なし）

要点： エンドツーエンドVLAではない。古典の骨格に、ばらつきのある3点だけ学習を差し込んで 99.4% に到達している。

組み立て作業をVLAモデルや模倣学習で行った場合のハンド



究極の姿：多指ハンドであらゆる作業を行う



作業用汎用グリップ



作業特化型グリップ



作業用汎用グリップの事例

究極の姿：多指ハンドであらゆる作業を行う



作業用汎用グリップ



作業特化型グリップ

セイコーエプソンの自律型双腕ロボット WorkSense



自律型双腕ロボット WorkSense「W-01」

2017年11月にリリース、2021年3月販売終了

双腕7自由度、カメラ(頭部カメラ、ハンドカメラ)、力覚センサ、ハンド(多目的ハンド)、操作パネル

外部配線なし、電源やコントローラ内蔵 (バッテリー駆動ではない)

ターゲットは、人の手による生産に適したモノ (多品種少量生産)

コンセプト

EPSON

- これまで技術・コスト課題で自動化できなかった作業を自動化



専用機・汎用ロボットによる自動化生産
に適したモノ

既存ロボットで実現できることは、既存
ロボットで



人手による生産に適したモノ

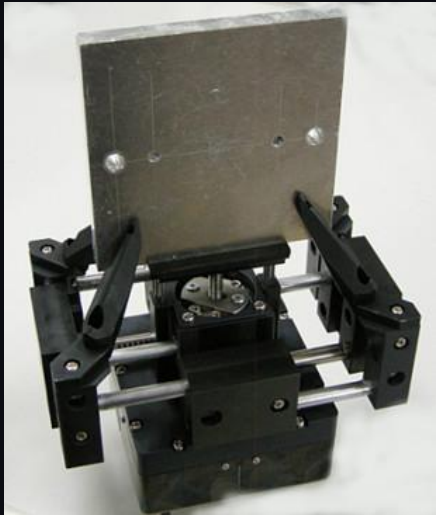
シンプルなのに数がまとまらず
自動化できなかった作業を
双腕ロボットで自動化



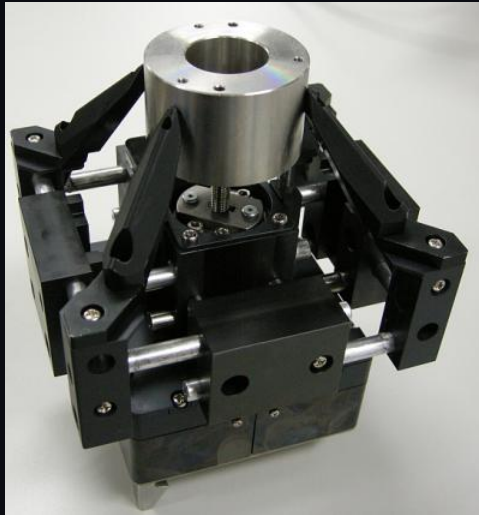
WorkSenseのハンド（多目的ハンド）



挟(はさ)む、つかむ、つまむ、握る、押すといったさまざまな把持が可能
指の開閉は2本ずつのペアでx軸方向とy軸方向の2自由度
z軸方向の掌（てのひら）



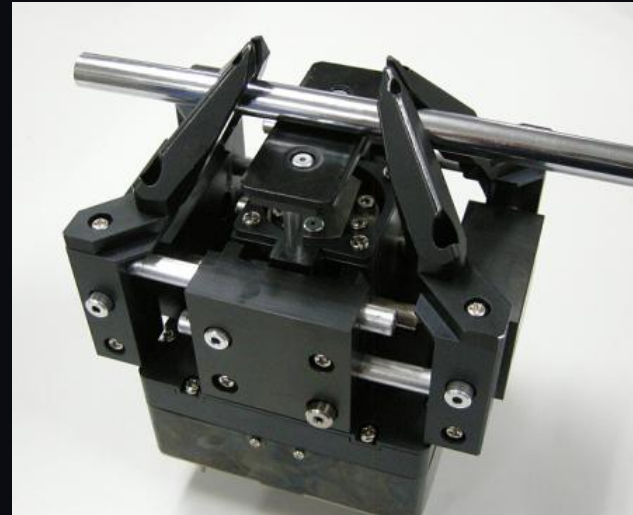
挟む
(プレート把持)



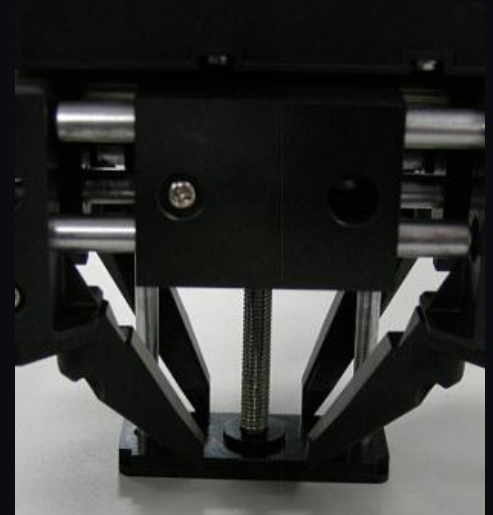
つかむ
(丸ものの把持)



つまむ
(板金部品把持)



握る
(シャフト把持)

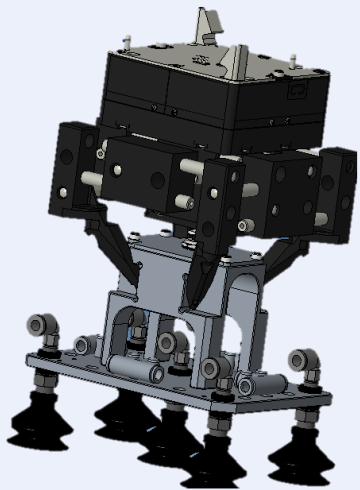
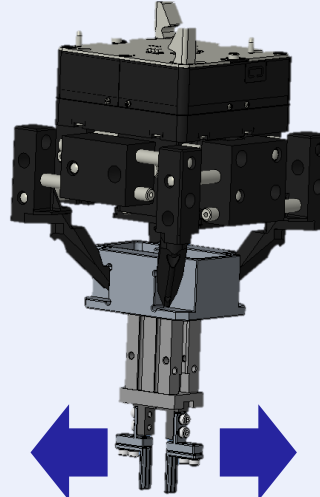
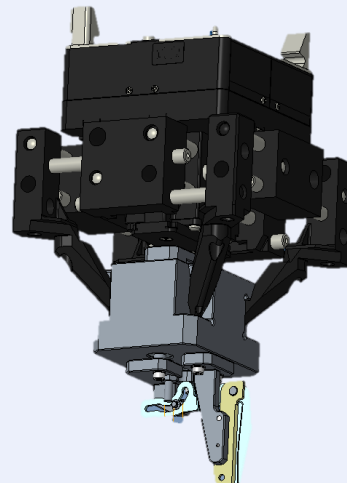


押す
(掌で押さえる)

※ 自社製品の構成部品の8割が持てることを確認

ハンドツール（多目的ハンドでできない作業への対応）

ハンドそのものを交換するのではなく、多目的ハンドでツールを操作
ツールを把持し、掌(てのひら)の直線動作を用いて、ツール独自の作用を実現

用途	吸着把持	内径把持	道具(ハサミ)の利用
イメージ			
説明	掌でメカバルブを押し 引き真空をON/OFF	掌でシリンダロッドを 押し引きグリップを開閉	掌の直線動作を 揺動への変換機構

※ 多目的ハンド自体を交換するのではなく、多目的ハンドで操作できるツールを提供

まとめ

VLAモデル・模倣学習を組み立て作業に用いることを想定した議論

使える場面はどのような場面か

汎用ハンドの事例

Neuromekaについてももう少し詳しく紹介

ヒューマノイドの将来について、使えるのか.

Physical AIについてのハンドについての調査
共有化が出来るかの議論

Worksense

まとめスライドでVLAを使うポイントをまとめる

その他

最初の30分で活動報告をするが、その際にここにおけるPhysical AIの定義をする必要がある

皆さんの発表を見て、それに合ったような定義にするつもりである。

事例2：物流・倉庫（Physical Intelligence π0）

具体的な適用工程

入荷・仕分け・検品・梱包前処理

フェーズ（導入段階）

PoC（概念実証）～パイロット耐久テスト（現場データ収集フェーズ）

長時間自律稼働の安定性と処理スピードを検証・データ収集

ワークの種類

宅配荷物・EC商品・パッキング資材

ロボットが実現する作業

ピック・アンド・プレース（および表裏反転・バーコード向き揃え・ソート）

対象のワーク

ダンボール箱、クッション封筒、薄型ビニール包装資材

使われているハンド（エンドエフェクタ）

高自由度（16自由度等）の人間型多指ハンド



♻️ 事例3：リサイクル・廃棄物処理（自動解体）

- 具体的な適用工程
 - 混入廃棄物の高速AI自動選別工程（コンベア上）
 - 小型家電・電子廃棄物の自動分解・解体工程（基板・バッテリー・フレーム）
- 詳しい記載や特徴
 - カメラ画像とVLAモデルで材質を認識し、把持位置を算出
- フェーズ（導入段階）
 - 商業展開 ～ テスト導入段階
- ワークの種類
 - 多種多様な混入リサイクル素材（PETボトル、缶、紙パック、プラスチック容器）
 - 電子廃棄物（E-Waste）（プリント基板、液晶画面、廃家電の金属・樹脂パーツ）
- 対象のワーク
 - 不規則に変形・潰れたPETボトルや缶、電子基板（PCB）、液晶パネルディスプレイフレームなど。
- 使われているハンド（エンドエフェクタ）
 - 吸着バキュームグリッパー / エアノズル式エンドエフェクタ
 - ソフトグリッパー・可変剛性多指ハンド（qb



🌾 事例4：農業・フードテック（農作物収穫）

三 具体的な制御タスク

- ・ **熟度・品質の視覚判定**
葉や枝の陰に隠れた果実（トマトやイチゴ等）の色・形状・テクスチャをVision認識し、収穫適期の個体のみを正確に選別。
- ・ **障害物回避アプローチ**
隣接する未熟果や繊細な枝・葉を倒したり傷つけたりしないよう、最適な空間アプローチ経路をリアルタイム計算。
- ・ **繊細把持・茎カット動作**
果気に圧力をかけすぎず優しく包み込み、茎の節部分へアプローチしてハサミ等でカットし収穫箱へ搬送。

🍷 使用ハンド：柔軟ソフトハンド（シリコン爪）＋小型カッター

空気圧駆動のやわらかいシリコンフィンガーで傷をつけずに包み込み。VLAが保持圧と切断タイミングを完璧に同期制御。



事例5：自動運転・モビリティ（Turing End-to-End VLA）

具体的な制御タスク

- ・ 車載カメラ映像からのダイレクト制御

車載カメラ映像と目的地テキストを受け取り、ステアリング角・アクセル量・ブレーキ圧の物理操作量をEnd-to-Endで直接出力。

- ・ 未知・非構造化環境への柔軟追従

白線がかすれた道路、工事現場、突発的な歩行者の飛び出しに対し、ルールベースに依存せず視覚文脈から安全回避。

- ・ 対話指示による運転モード切替

「右側のレーンに入って」「安全優先で低速走行して」といった音声言語指示を解釈し、リアルタイムに走行挙動を変更。

制御対象：車両アクチュエータ（ステアリング・ブレーキ・駆動系）

アーム型の爪ではなく、CANバス等を介して車両の操舵・加速・減速系アクチュエータを直接VLAトークンでリアルタイム制御。



VLAモデルで活用されるロボットハンドの分類



多指人間型ハンド

特徴： 12～20自由度の関節と指先触覚センサー。

用途： 自動車組み立て（BMW）、多目的ツール操作、人間共有空間での複雑作業。

VLAの役割： 各指の微小位置・力加減の多次元リアルタイム協調制御。



平行2指+吸着ハイブリッド

特徴： 高精度な挟み込みと、真空吸着カップアレイの併用。

用途： 物流倉庫ピッキング（Physical Intelligence）、検品・デパレタイズ。

VLAの役割： 対象の質感（箱/袋）に応じて「吸着」か「挟み込み」かを瞬時に自動選択。



柔軟ソフトハンド

特徴： 空気圧駆動のエラストマー・シリコンフィンガー。

用途： 農作物収穫（トマト/果実）、傷つきやすい食品のハンドリング。

VLAの役割： 視覚的に柔らかさや接触状態を推定し、最適な空気圧・把持力を調整。



高剛性爪+工具アタッチメント

特徴： 金属製頑丈アーム、高トルクモジュール、電動工具連動。

用途： 廃家電解体（リサイクル）、重工業部品処理。

VLAの役割： ネジ穴位置合わせ、コード切断、固定力制御などツール連携動作の自動化。

 VLAモデル 産業応用事例 比較まとめ

産業分野	プロジェクト/企業	主要VLAタスク	使用ハンド	導入効果・価値
自動車・精密製造	BMW × Figure AI	板金パーツ配置・治具レス補正	触覚統合5本指ハンド	ティーチング工数削減
物流・倉庫	Physical Intelligence (π0)	不定形物ピッキング・デパレタイズ	平行2指＋真空吸着	未登録品への即時対応
リサイクル処理	E-waste自動解体	電池・基板識別、工具分解・破断	高剛性爪＋電動工具連動	危険作業の完全自動化
農業・フードテック	自動収穫ロボット	熟度判定・障害物回避・繊細摘み取り	シリコン爪＋カッター	人手不足解消・高品質収穫
自動運転・モビリティ	Turing VLA Model	カメラ映像からの直接車両操作量生成	車両アクチュエータ直接制御	未知状況での意思決定向上

Image Sources



<https://mediapool.bmwgroup.com/cache/P9/202408/P90563506/P90563506-figure02-bmw-group-plant-spartanburg-2666px.jpg>
Source: www.press.bmwgroup.com



https://i.ytimg.com/vi/YyXCMhnb_IU/maxresdefault.jpg
Source: www.therobotreport.com



<https://www.theagilityeffect.com/app/uploads/2024/10/Desassemblage.jpg>
Source: www.theagilityeffect.com



<https://www.denso.com/jp/ja/driven-base/assets/img/project/artemy/mv-en.webp>
Source: www.denso.com



<https://www.e-consystems.com/blog/camera/wp-content/uploads/2025/11/What-is-The-Role-of-Embedded-Cameras-at-Different-Levels-of-Autonomous-Driving.jpg>
Source: www.e-consystems.com