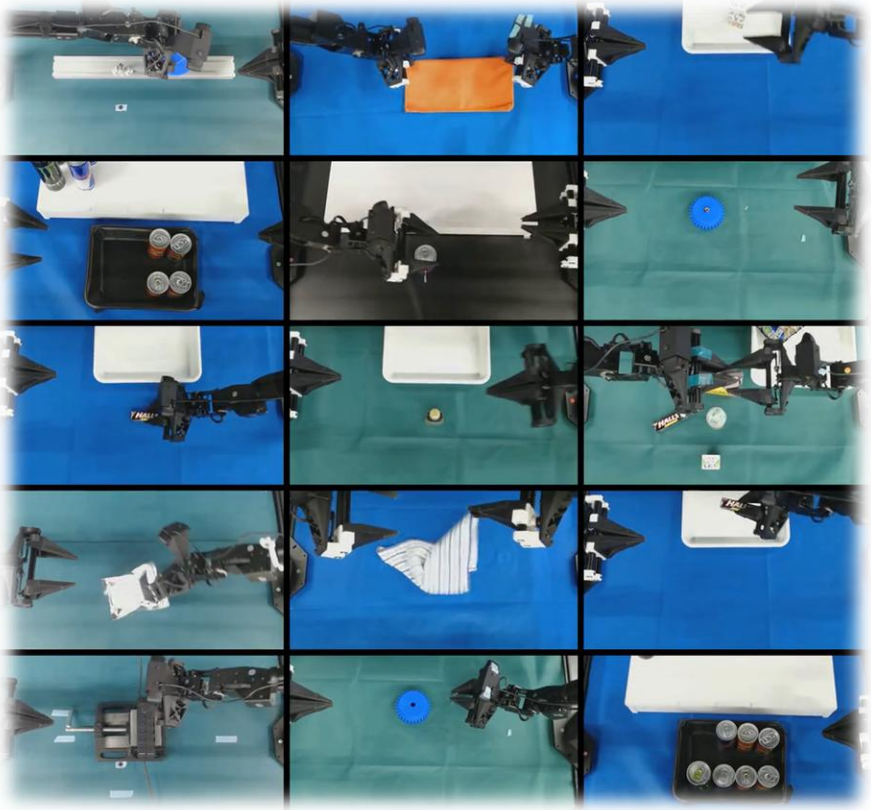
基盤モデル・データ収集研究事例紹介
AIST-Bimanual Manipulationデータセットの公開と取り組みについて

発表者：元田智大
産業技術総合研究所　人工知能研究センター　実体知能研究チーム

2025/11/07

目次

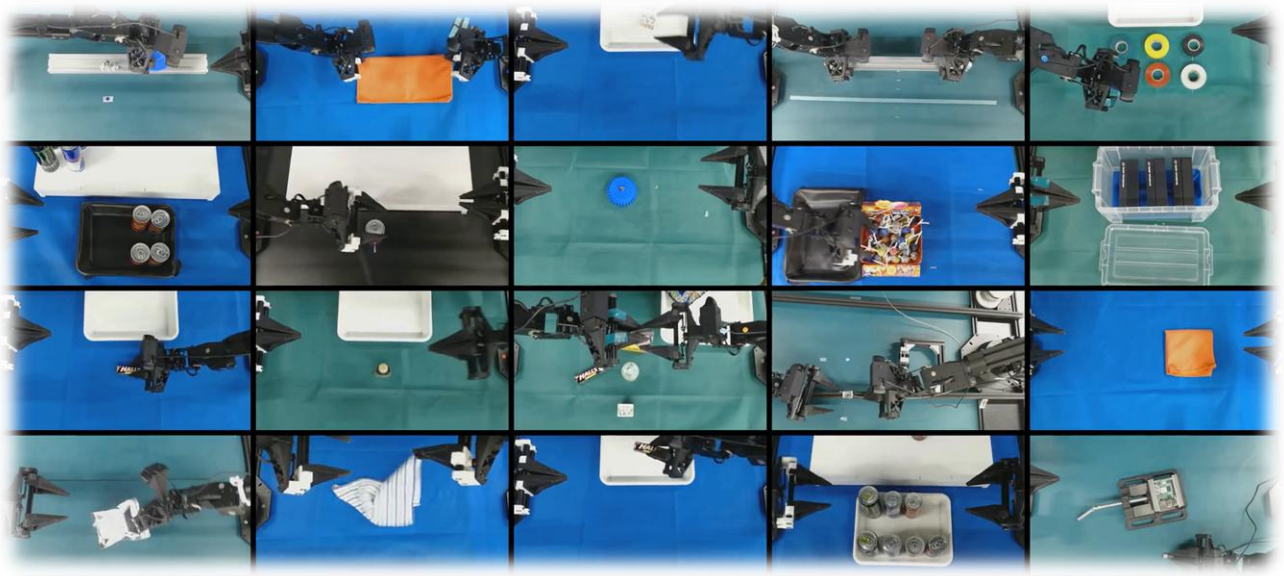
1. 研究背景
2. 「AIST-Bimanual Manipulation」Datasetの概要
   - システムセットアップ
   - データ収集とその過程
   - データの詳細情報について
3. 予備検証と今後の展望
4. まとめ



発表概要

- ロボット基盤モデルの発展に必須なデータの収集の取り組みについて紹介

- 我々は**双腕**を使って遠隔操作するための“**オープン**”ハードウェア「ALOHA」を用い，実際に10,000+件の**双腕**軌道データを収集し，ベンチマークや事前学習向けに公開



図：ウェブ上で公開中
https://aistairc.github.io/aist_bimanip_site/

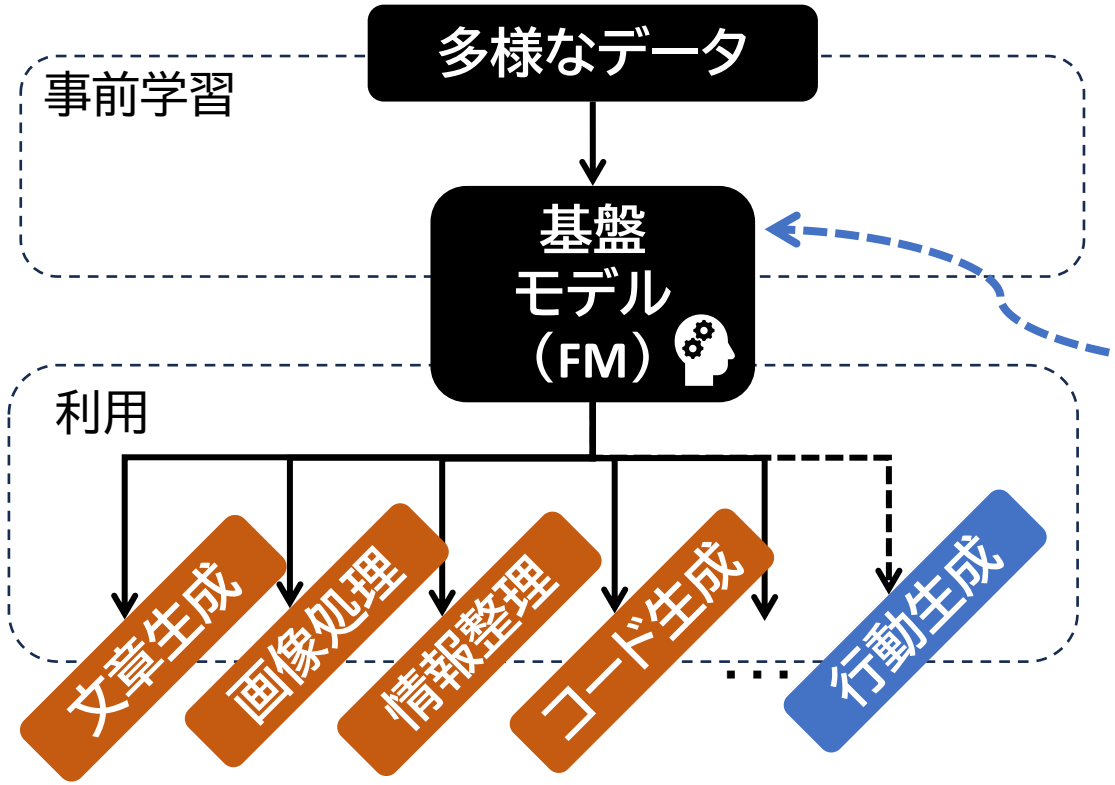
図：産総研ではロボット基盤モデル開発をサポートする
ソフトウェアとデータセット（本発表）を提供

研究背景：ロボットデータの特殊性

<<課題>>「環境との相互作用」を含むためwebからスクレイピングできない

- センサや行動，その結果の組み合わせが必要，現実世界特有の知見がある

人手や実機を用いたデータ収集そのものが重要なプロセスあり，インフラそのものである



図：理想的な基盤モデルの概要図

図：基盤モデルを拡張にはロボットデータが必須

（スケーラビリティの観点が重要）

研究背景：双腕データの重要性

なぜ双腕ロボットなのか？

- 人間の動作を模倣する
  人間の作業は両手を使うことが前提
  - 高精度かつ器用な複雑作業が可能
- 双腕ロボットの強み
  - 複雑な組立作業
  - 繊細な部品の取り扱い
  - 自動車・電子機器製造

- 将来的な展望：協働ロボットの進化
  - 人との安全な協働作業
  - 複数アームの統合による柔軟な作業対応

関連研究調査: 既存データセット（双腕）との比較

- 海外コミュニティや企業のレベルは驚異的
- 様々なロボット, センサ情報, 特筆すべきモダリティなどが含まれている
- 国内における事例は限定的（直近で, 日本の**AIロボット協会**の動向がある）

データセット	#Ep.（エピソード数）	時間（h）	特徴						タスク例
			オープン	実機	両腕	深度	マルチビュー	テキスト	
MIME [1]	8.6k		✓	✓	✓	✓	✓	✗	Grasp; push; stack; pour; open bottles
RT-1 [2]	130k		✓*	✓	✗	✗	✗	✓	Grasp; place; move, knock over; openclose
RH20T [3]	> 110k		✓	✓	✗	✓	✓	✓	Cut; plug; pour; fold
ALOHA [4]	< 1k		✓	✓	✓	✗	✓	✗	Unwrap candy; put on shoes
Mobile ALOHA [5]	< 1k		✓	✓	✓	✗	✓	✗	Push chairs; move bowls; wipe, push buttons
Droid [6]	92k		✓	✓	✗		✗	✓	Pick&place; put; clean up; cook
BRMData [7]	500		✓	✓	✓	✓	✓	✓	household bimanual-mobile
RDT-1B [8]	> 6k		✓	✓	✓	✓	✓	✓	diverse bimanual teleop
DAA [9]	224k		✓	✓	✓	✗	✓	✓	Pick up coins; open zippers, thread needles
GR00T N1 Humanoid Pre-Training Dataset [10]		88.4 h	✓	✓	✓	✗	✗	✓	pick&place
Gemini Robotics [11]		1k h	✗	✓	✓	—	✓	—	proprietary bimanual
π_0 [12]		10k h	✗	✓	✓	—	✓	✓	fold shirt; sort; pour lettuce; put container
TRI-Ramen-Real [13]	> 46k		✗	✓	✓	—	✓	✓	Put Kiwi in center of table; cut apple in slices
Agibot [14]	> 1M		✓	✓	✓	✓	✓	✓	pick&place; sort; pour lettuce; put container
Ours	10.2k		✓	✓	✓	✓	✓	✓	pick&place; hand over; fold towels; assemble gears

✓: 利用可, ✗: 利用不可, ✓*: 部分的に公開. “—”: 不記載（未公表）. k: 千, M: 百万. ≥: 以上, ≤: 以下, >: より大きい（超）, <: 未満. エピソード数／時間の列では、未公表の場合は空白としています。

[1] Sharma+, 2018
[2] Brohan+, 2022
[3] Fang+, 2024
[4] Zhao+, 2023
[5] Fu+, 2024
[6] Khazatsky+, 2024
[7] Zhang+, 2024
[8] Liu+, 2024
[9] Kim+, 2024
[10] Bjorck+, 2025
[11] Google, 2025
[12] Black+, 2024
[13] Barreiros+, 2025
[14] Bu+, 2025

関連研究調査: 既存データセット（双腕）との比較

時間・コスト・量を売りするのは, あまりよくないアプローチであるが,
①どのようにして品質を担保するのか？（★本発表に含まれる. しかし, 課題は残る）
②データの詳細を分析して“このデータによって何の貢献があるか”“何が提供できるか”

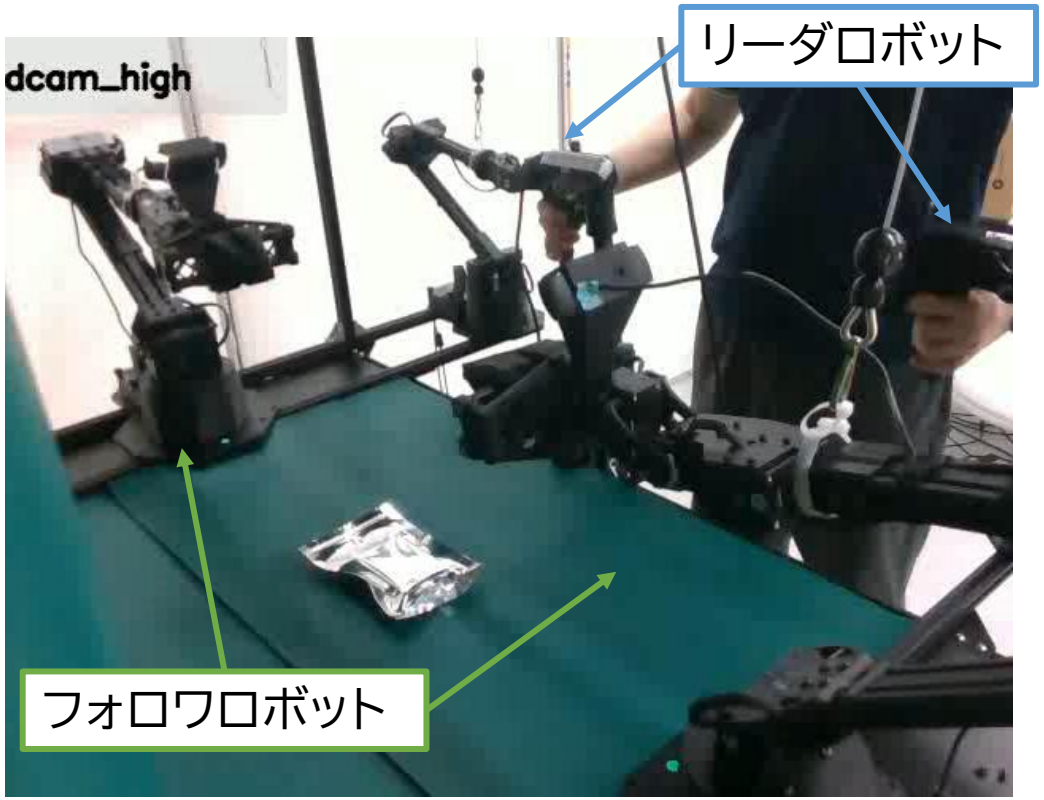
データセット	#Ep.（エピソード数）	時間（h）	特徴						タスク例
			オープン	実機	両腕	深度	マルチビュー	テキスト	
MIME [1]	8.6k		✓	✓	✓	✓	✓	✗	Grasp; push; stack; pour; open bottles
RT-1 [2]	130k		✓*	✓	✗	✗	✗	✓	Grasp; place; move, knock over; openclose
RH20T [3]	> 110k		✓	✓	✗	✓	✓	✓	Cut; plug; pour; fold
ALOHA [4]	< 1k		✓	✓	✓	✗	✓	✗	Unwrap candy; put on shoes
Mobile ALOHA [5]	< 1k		✓	✓	✓	✗	✓	✗	Push chairs; move bowls; wipe, push buttons
Droid [6]	92k		✓	✓	✗		✗	✓	Pick&place; put; clean up; cook
BRMData [7]	500		✓	✓	✓	✓	✓	✓	household bimanual-mobile
RDT-1B [8]	> 6k		✓	✓	✓	✓	✓	✓	diverse bimanual teleop
DAA [9]	224k		✓	✓	✓	✗	✓	✓	Pick up coins; open zippers, thread needles
GR00T N1 Humanoid Pre-Training Dataset [10]		88.4 h	✓	✓	✓	✗	✗	✓	pick&place
Gemini Robotics [11]		1k h	✗	✓	✓	—	✓	—	proprietary bimanual
π_0 [12]		10k h	✗	✓	✓	—	✓	✓	fold shirt; sort; pour lettuce; put container
TRI-Ramen-Real [13]	> 46k		✗	✓	✓	—	✓	✓	Put Kiwi in center of table; cut apple in slices
Agibot [14]	> 1M		✓	✓	✓	✓	✓	✓	pick&place; sort; pour lettuce; put container
Ours	10.2k		✓	✓	✓	✓	✓	✓	pick&place; hand over; fold towels; assemble gears

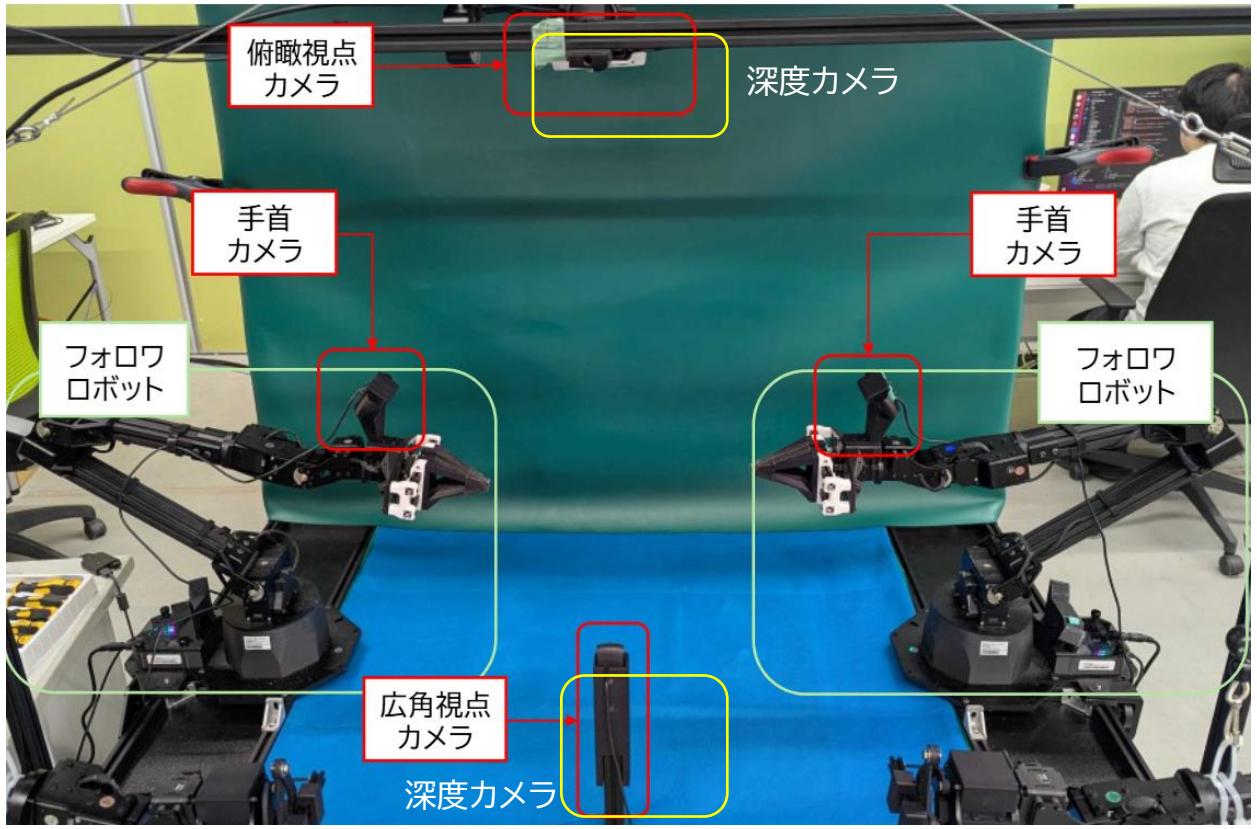
✓: 利用可, ✗: 利用不可, ✓*: 部分的に公開. “—”: 不記載（未公表）. k: 千, M: 百万. ≥: 以上, ≤: 以下, >: より大きい（超）, <: 未満. エピソード数／時間の列では、未公表の場合は空白としています。

[1] Sharma+, 2018
[2] Brohan+, 2022
[3] Fang+, 2024
[4] Zhao+, 2023
[5] Fu+, 2024
[6] Khazatsky+, 2024
[7] Zhang+, 2024
[8] Liu+, 2024
[9] Kim+, 2024
[10] Bjorck+, 2025
[11] Google, 2025
[12] Black+, 2024
[13] Barreiros+, 2025
[14] Bu+, 2025

データ収集環境

ALOHA (低コストかつ直感的操作に対応したリーダフォロワ) ※ALOHA 1
[Zhao+, RSS2023]





図：RGB-Dカメラ(2台)をオリジナルセットアップに対して追加

ALOHAを用いた模倣学習の検証

教師(エキスパート)データの収集

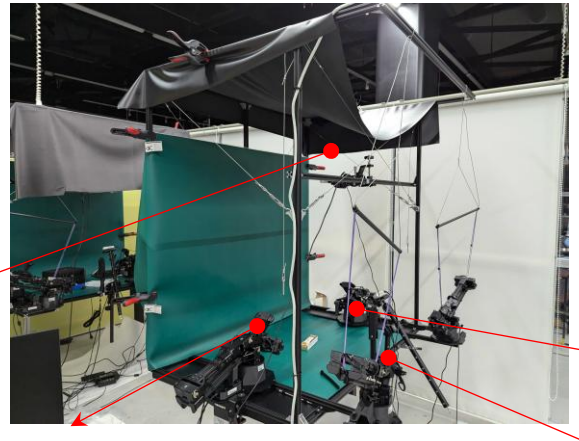
1000秒(約17分)の動作データ

X 50

学習済みモデルの実行(デプロイ)

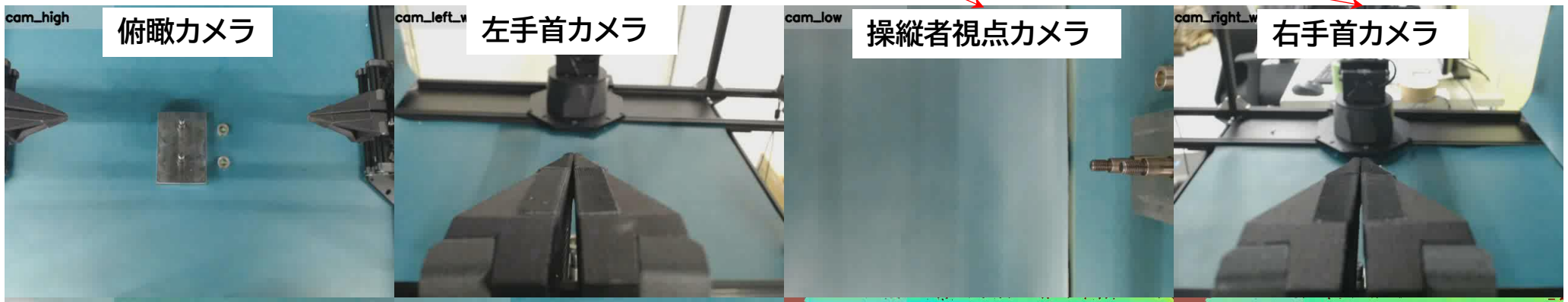
ACTを用いた事例
学習時間:約1時間30分)

データ収集環境

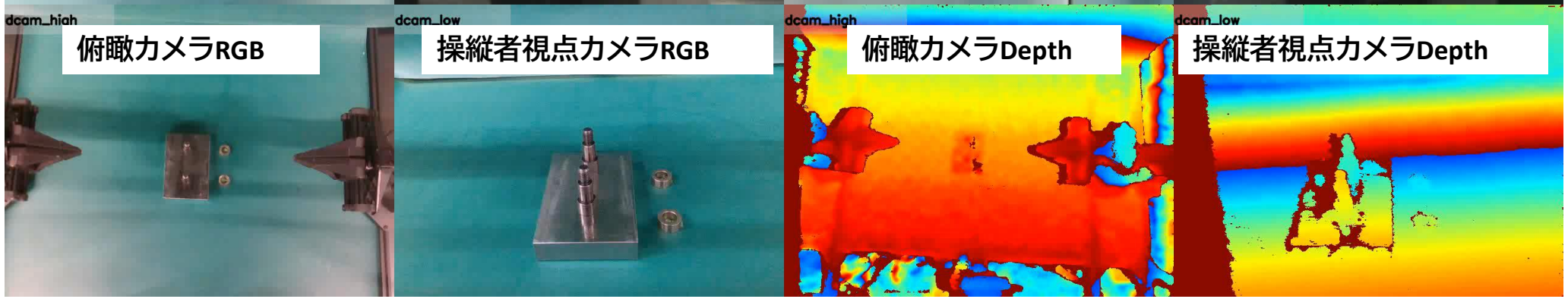


標準搭載：Logitech C922Pro x4
（追加）： Realsense D435i
（追加）： Realsense D415

標準搭載



（追加）

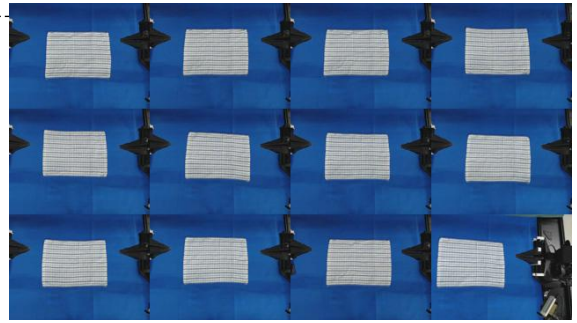


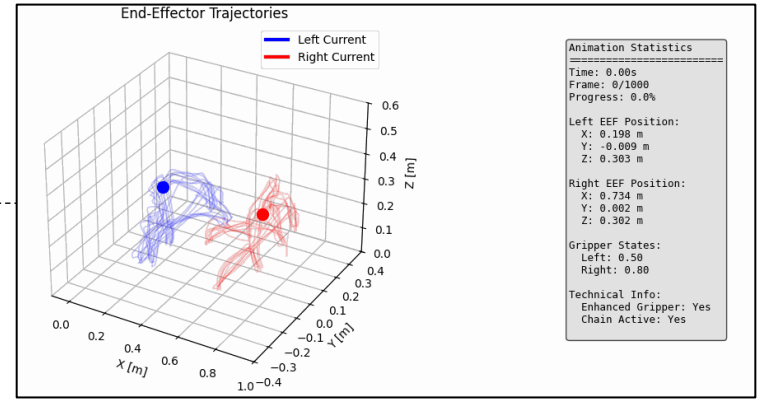
データ収集環境

グループ	備考 (次元数)
action	リーダ関節位置情報 (14,)
observations	
depth	
dcam_high	深度カメラ（480, 640,）
dcam_low	深度カメラ（480, 640,）
images	
cam_high	俯瞰視点カメラ（480, 640, 3）
cam_left_wrist	左手首カメラ（480, 640, 3）
cam_low	広角視点カメラ（480, 640, 3）
cam_right_wrist	右手首カメラ（480, 640, 3）
dcam_high	俯瞰視点カメラ RGB（480, 640, 3）
dcam_low	広角視点カメラ RGB（480, 640, 3）
qpos	フォロワ関節位置情報 (14,)
qvel	フォロワ関節速度情報 (14,)
text	
prompt	テキストプロンプト
text_embedding	テキストの埋め込みベクトル [7] (384,)

データ収集プロトコル

1. 環境設定
   作業領域の整備、カメラ位置・照明調整、自動露出の防止
2. タスク設定
   操作手順・初期状態（固定/ランダム）・成功条件の明確化
3. オペレータ訓練
   練習セッションの実施、操作精度・一貫性の向上
4. データ収集の実行
   複数回の実演、ACT事例に倣い50件程度の収集
5. データ品質検証
   可視化ツールによるチェック、必要に応じて再収集
6. データ後処理
   メタデータ付与、圧縮処理による共有可能な形式へ

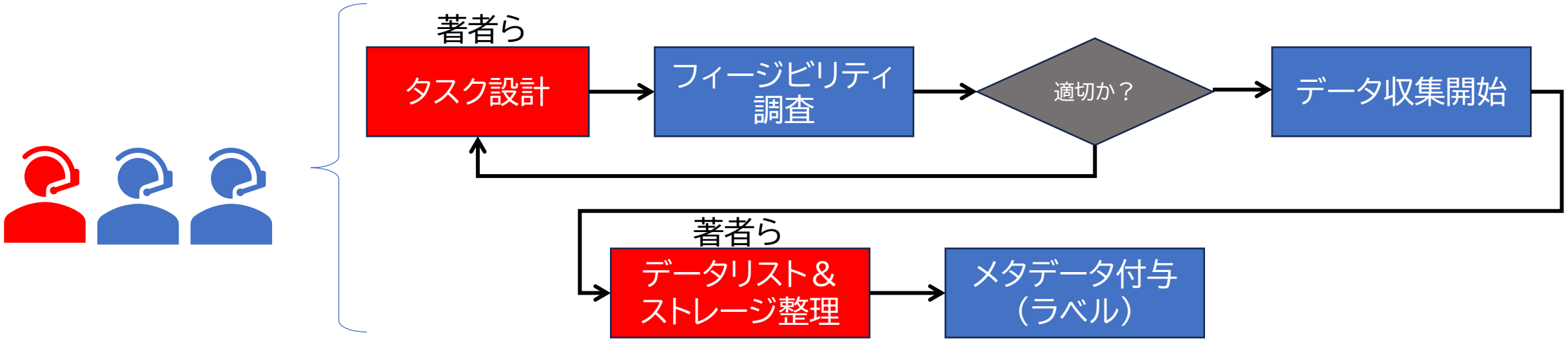




- テキスト情報の整備
- HDF5形式の保存
- 他の形式への変換を準備

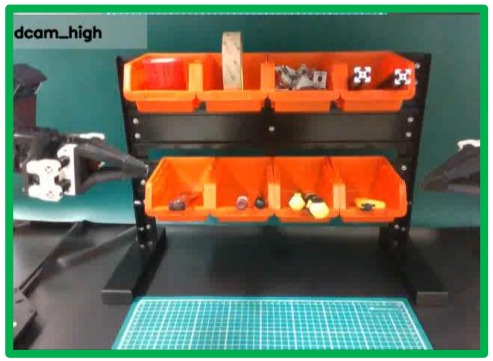
データ収集プロトコル

- “ALOHA”のフィージビリティを考慮してタスクを設計
- 各操作タスクは**50-100 エピソード**を標準として**2-3名の作業者**（≠ロボット専門家）



判断の基準：

- 小規模モデルのデモ
- 作業者視点の難易度（やりづらい, etc.）をヒアリングし決定



…

データ概要

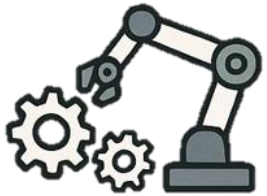
各データの識別名

テキスト（動作プロンプト）

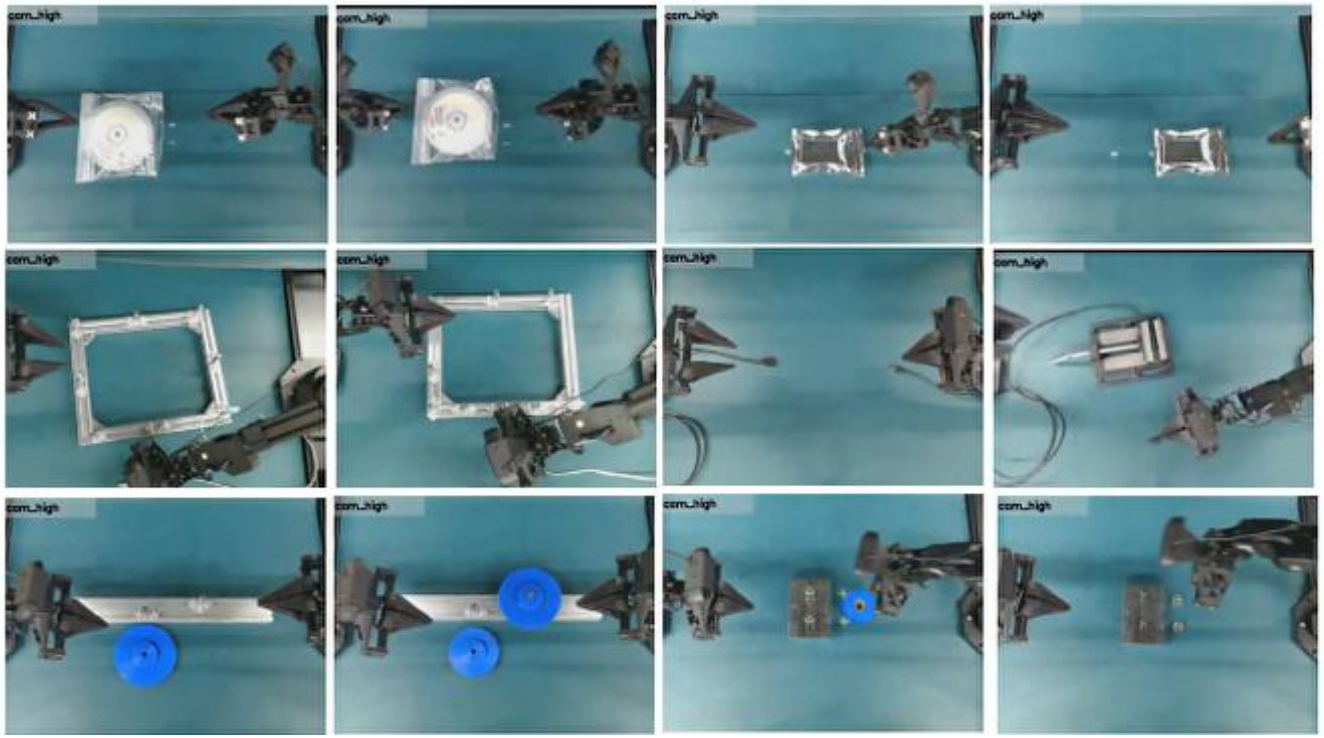
Task I	Task Name	Action Type
0001	bluebox_failed	Pass box from right arm to left arm
0004	bluebox_right_left	Pass box from right arm to left arm and left arm to rihgt arm
0008	gear_task/01_washer_len700	Put washer on shaft
0009	gear_task/02_gear_len500	Put big gear on shaft
0012	insert_usb_cable	Insert the male usb cable into the female usb cable with both arms
0013	insert_usb_cable_fix_parts	Insert the male usb cable into the female usb cable on vise-mounted base
0020	gear_task_003_01	Put big gear on shaft
0021	gear_task_003_02	Put medium gear on shaft
0024	cable_routing_8pin_cross	Pass through eight pins alternately
0025	cable_routing_cross_narrow	The ball passes through four closely spaced pins in an alternating fashion
0047	color_tape_grab_and_insert	Use the left and right arms to grab the tape and place it in the basket
0050	goods_grab_and_insert	Use the left and right arms to grab the item and put it in the basket.
0056	silver_pouch_001	Stand the silver pouch
0057	silver_pouch_002	Stand the silver pouch change the position
0059	silver_pouch_004	Stand the silver pouch six positions 2cm intervals
0060	clear_zipper_bag_001	Stand the clear_zipper_bag
0061	clear_zipper_bag_reel_001	Remove the reel from the clear zipper bag
0062	clear_zipper_bag_reel_002	Remove the reel from the clear zipper bag six positions 2cm intervals
0063	**receive_goods_and_add_to_basket**	The left arm grabs the product and passes it to the right arm, which then plac
0072	tape_placed_in_random_places_long	Grab randomly placed tapes and place them in a basket

図：各軌道データに対するメタデータ（サンプル）

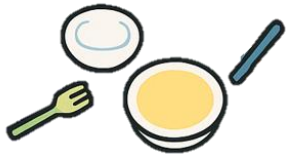
データ概要



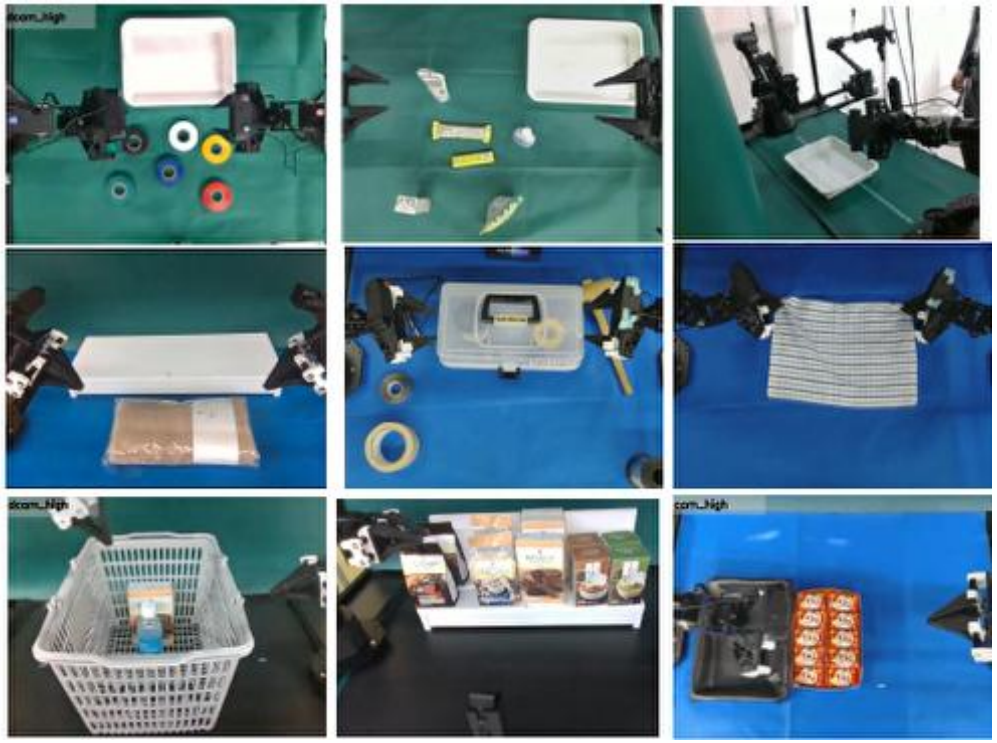
産業応用を想定したタスク



例）袋から出す，ケーブル配策，組み立て



日常タスク



例）ピック＆プレース，折り畳み，受け渡し

データ概要

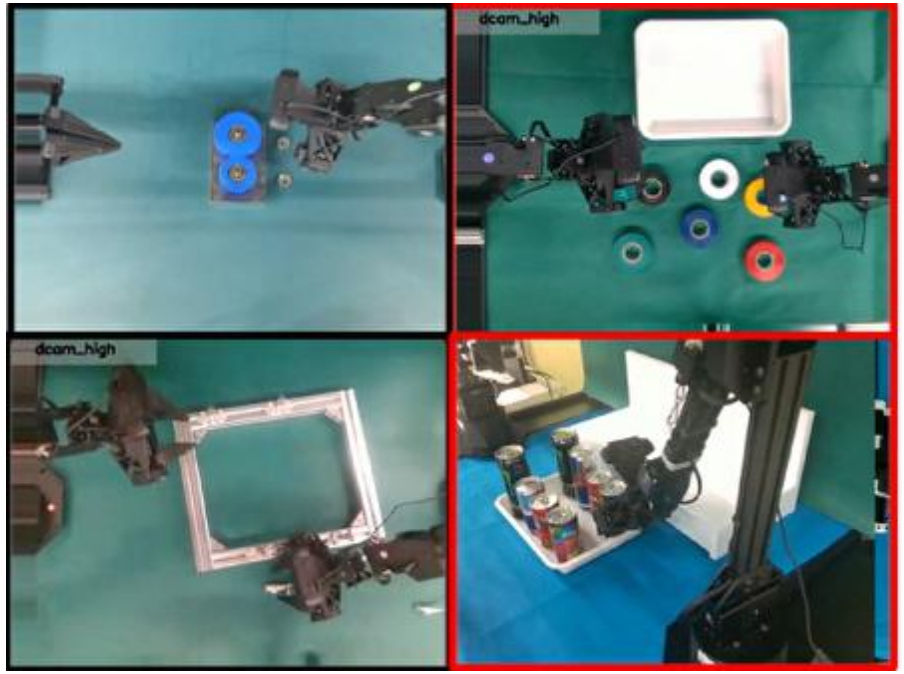
各操作タスクの種類に大別できると仮定.
→ それぞれが含まれることを目指す

独立動作
= 片手のみを利用

片手動作
＝左右が独立

左右非同期動作

左右同期動作



Put etc.　　Pick&Place　　Hand over　　Lift

データ概要

固定配置もしくはランダム配置のデータが含まれる
→ 目的(特定の動きを学習or多様性の学習)に応じて検証

配置や動作が
一定になるタスク

配置がランダム
複数の軌道例が
データに含まれる

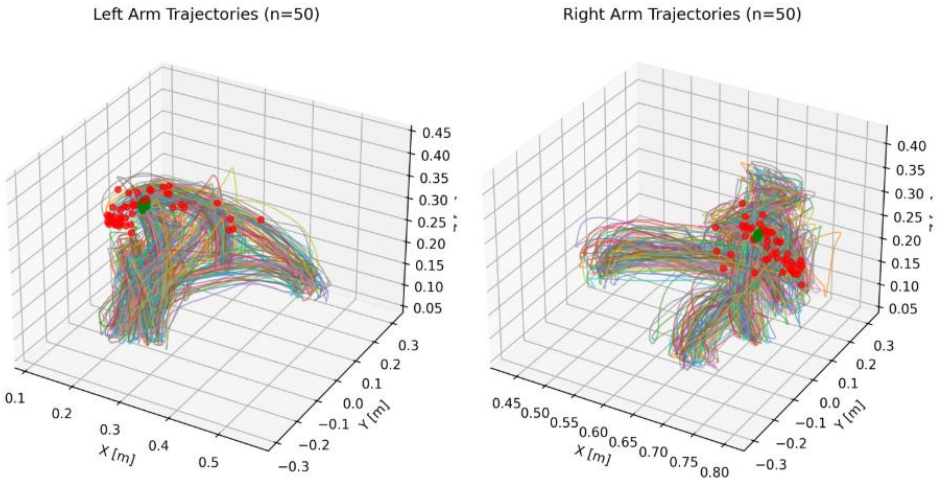


図:エンドエフェクタのxyzの軌道①

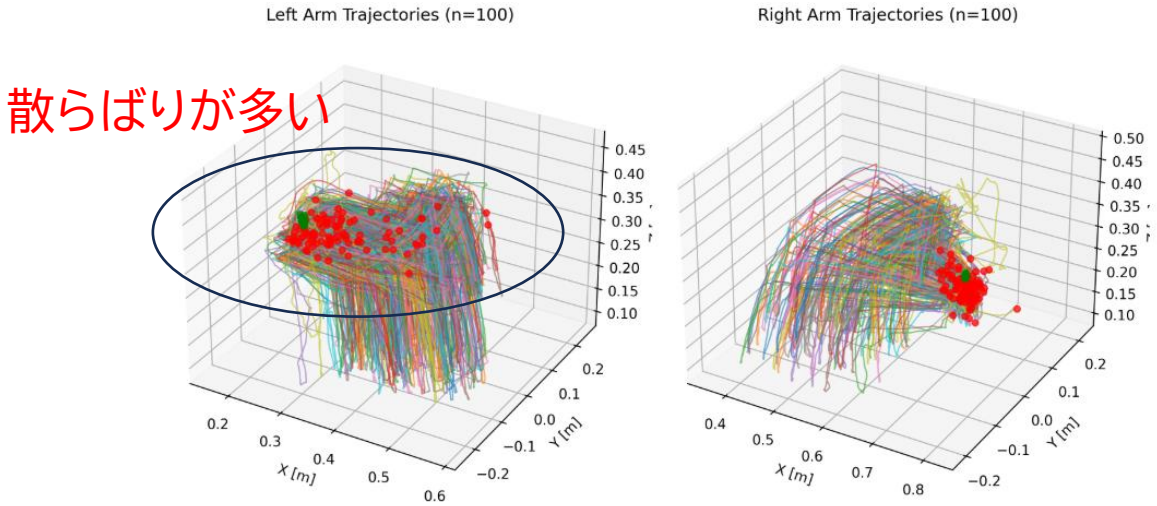


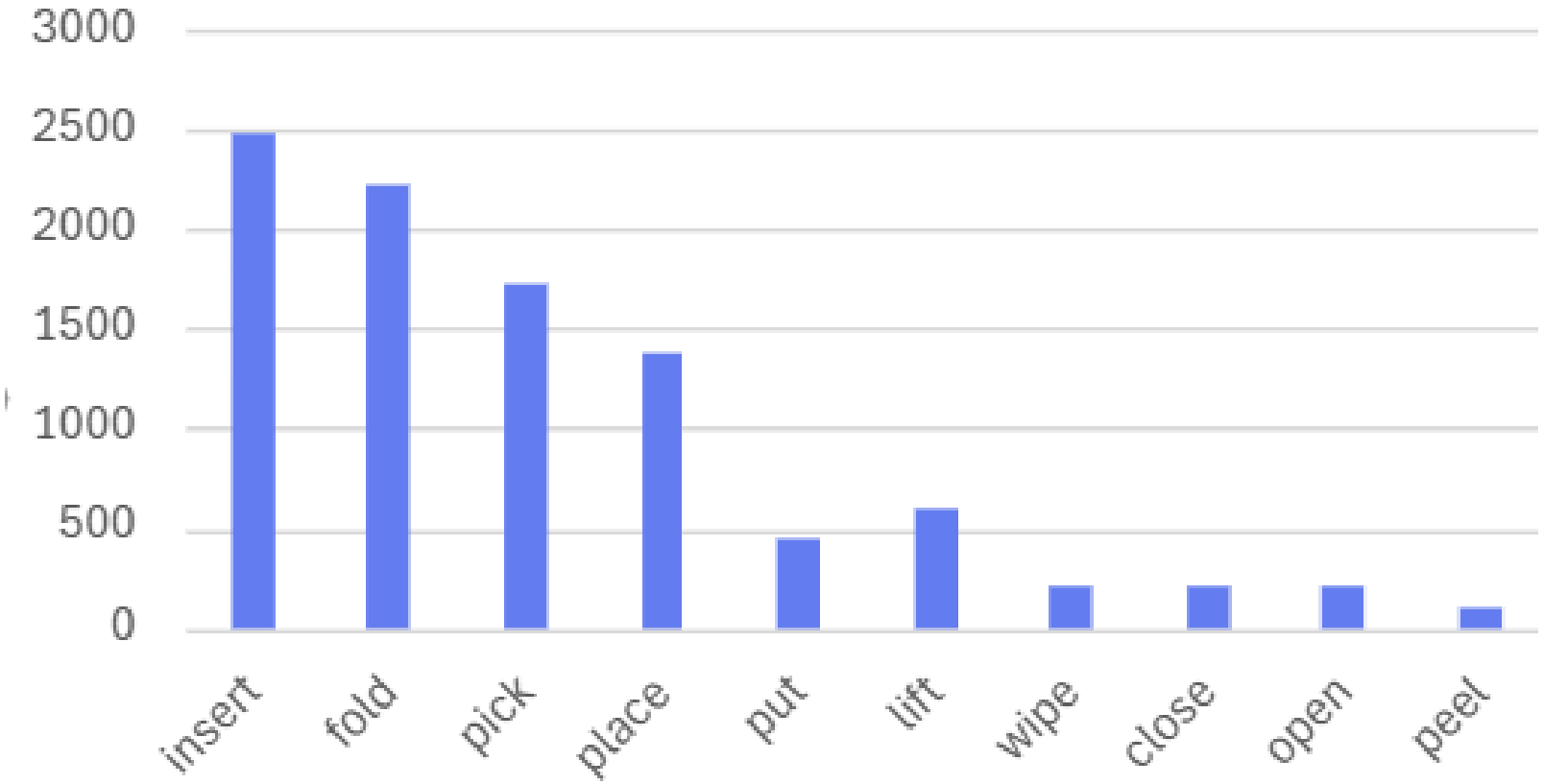
図:エンドエフェクタのxyzの軌道②

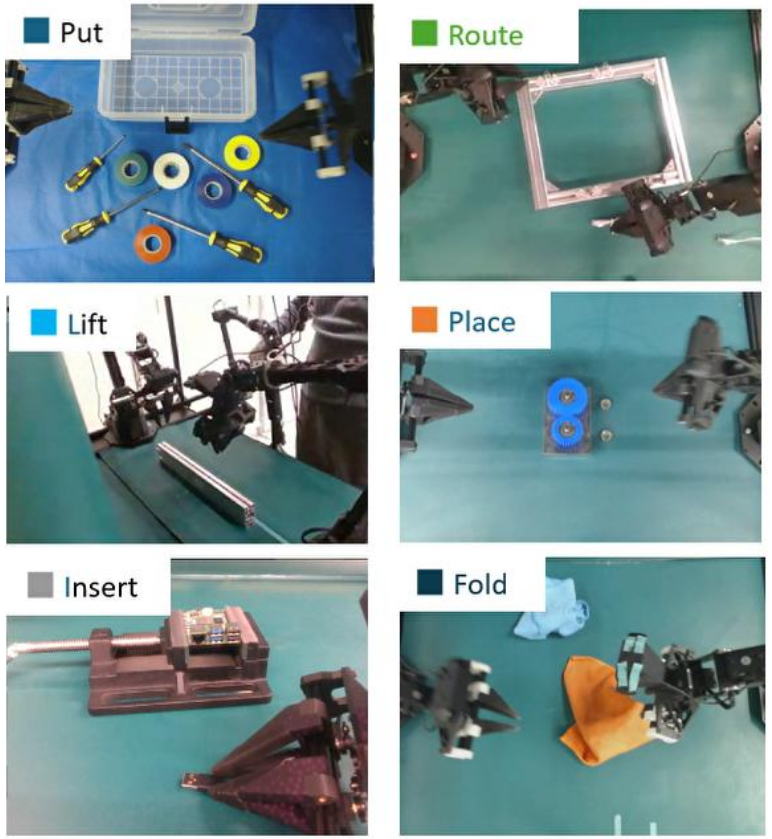
データ概要

各操作タスクは50-100 エピソードを標準として2-3名の作業者

データ概要

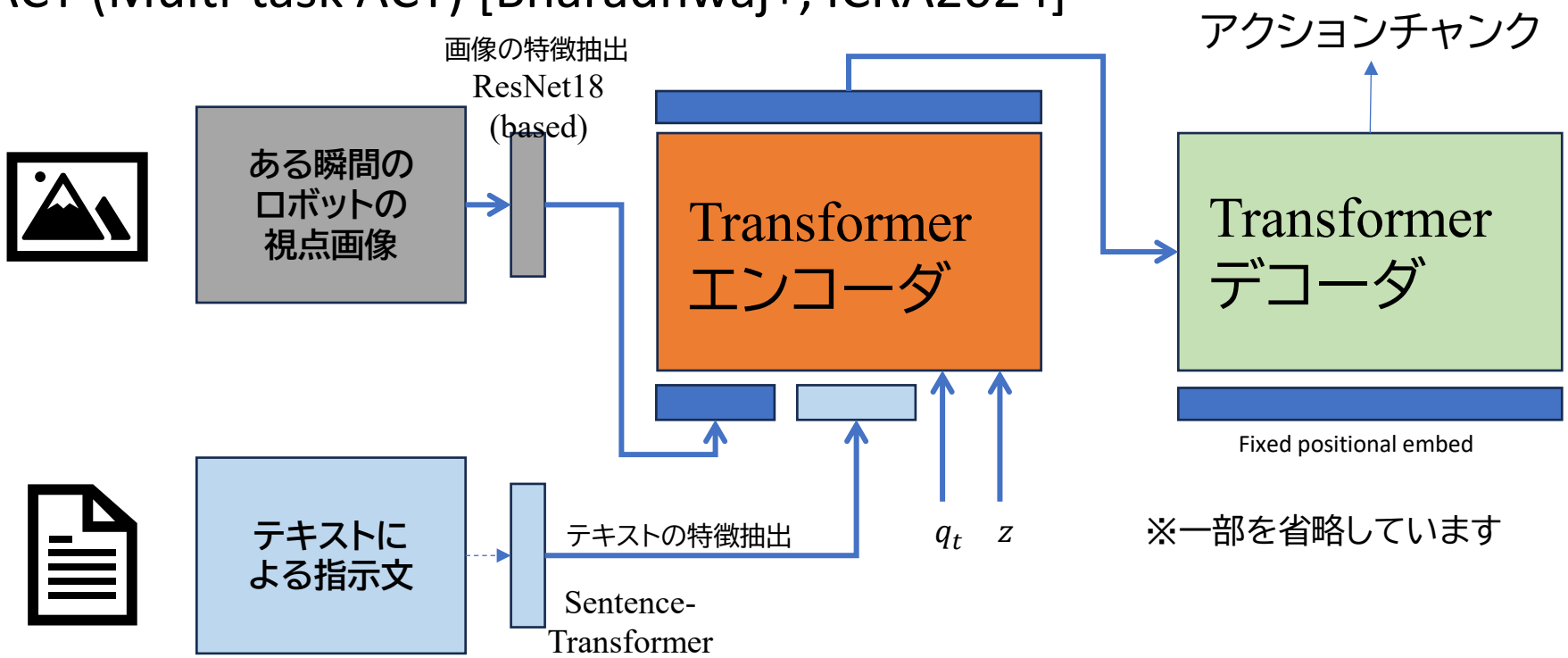
スキル


3000
2500
2000
1500
1000
500
0
insert
fold
pick
place
put
lift
wipe
close
open
peel


Put
Route
Lift
Place
Insert
Fold

予備実験

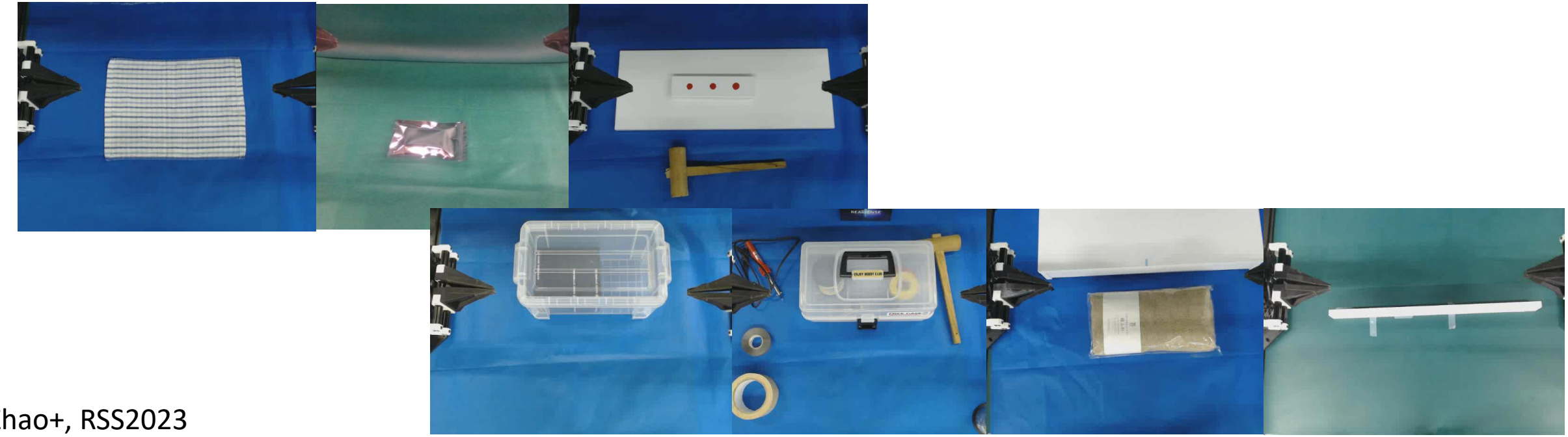
- 今回のデータセット（Ours）とALOHA（ロボット）を使用
- “どういうベンチマーク”や“開発ができるか”という事例を紹介
- この目的のために**小規模のVLAモデル（ACT, MT-ACT）**の検証
  - ACT (Action Chunking Transformer) [Zhao+, RSS2023]
  - MT-ACT (Multi-task ACT) [Bharadhwaj+, ICRA2024]



予備実験 - データセットの効果

ACT (Action Chunking Transformer)　※テキストなし　それぞれ50 エピソードを模倣

	Fold Towel			Hand-over Silver bag		Hand-over Hammer		Open the lid	Open the toolbox	Lift the towel and place it on the shelf		Lift the board with both hands	Ave.
	Fold	Pick	Place	Grab	Hand-over	Hand-over	Pat	Open	Open	Lift	Place	Lift	
ACT[1]	92	72	72	92	88	80	72	**100**	**100**	**100**	4	**100**	82.5



[1] Zhao+, RSS2023

予備実験 -データセットの効果

ACT (Action Chunking Transformer)　※テキストなし　それぞれ100 エピソードを模倣
→ 位置のランダム性が顕著である場合

Pick Tape

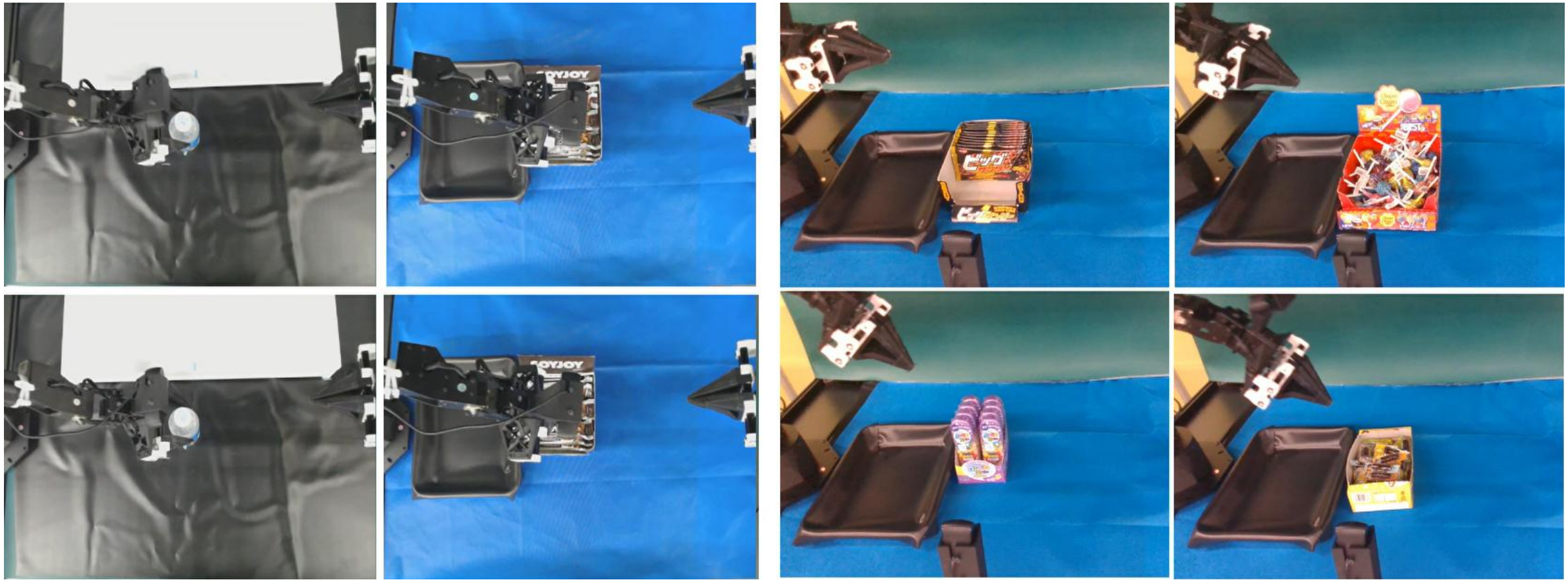
Pick Box

予備実験 - スケーラビリティ

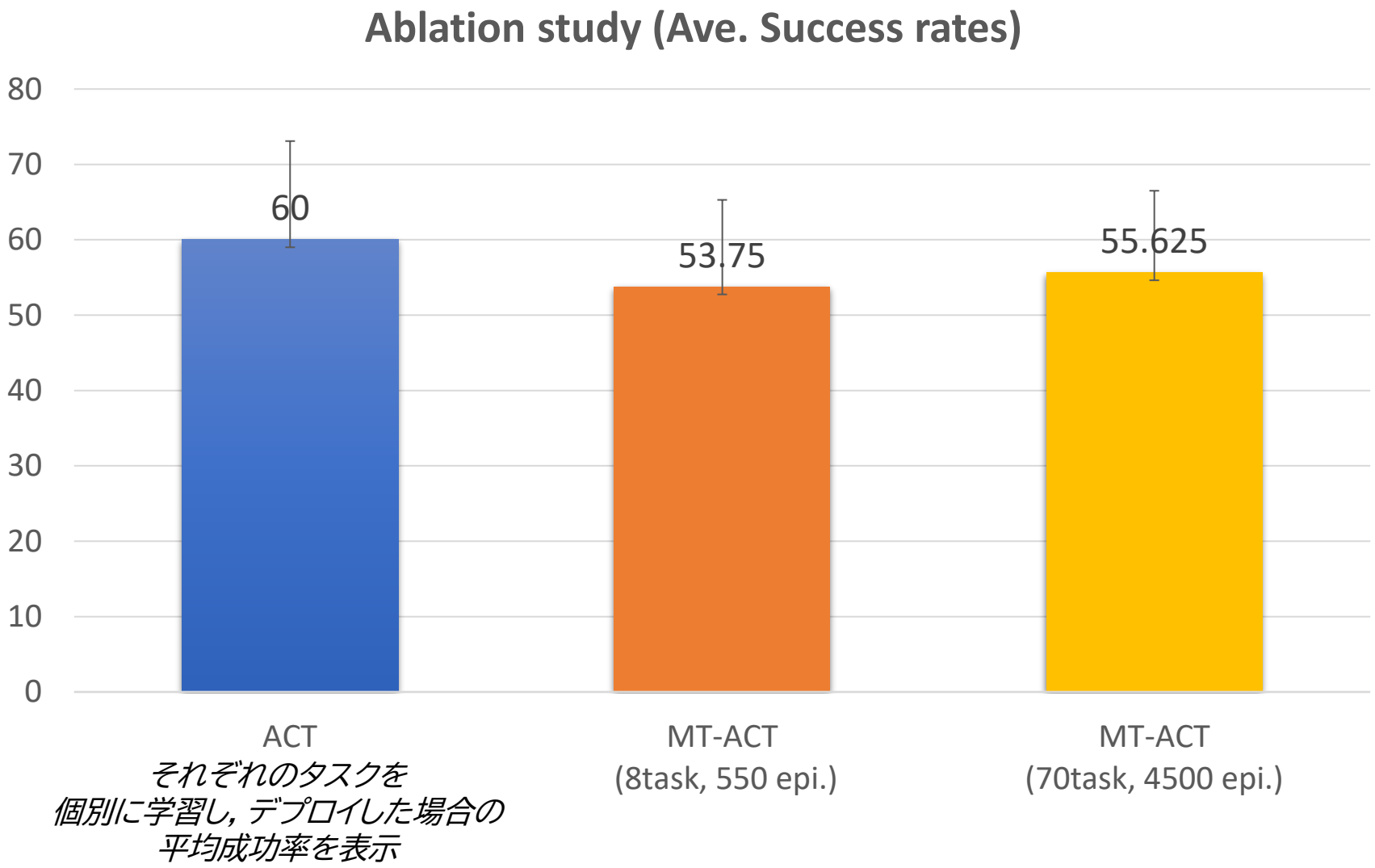
◆① 500 エピソード（8タスク）とALOHAの片腕（ViperX300S）を利用して学習＋検証
小規模のVLAモデル（MT-ACT）の検証　※テキスト文あり

予備実験 - スケーラビリティ

◆② **4500 エピソード（70タスク）**とALOHAの片腕（ViperX300S）を利用して学習＋検証
小規模のVLAモデル（MT-ACT）の検証　※テキスト文あり



予備実験 - スケーラビリティ



予備実験 - スケーラビリティ（補足）

それぞれ対応するエキスパートデータが含まれる場合において，
類似するPick＆Placeデータを拡張すると成功率の向上もある

議論

- **データバランスの最適化**
  - 各種類や分散の比率調整
  - 協調操作・同期双腕操作データの強化
    →双腕操作データにより特徴を持たせる
- **ベンチマーク足りうる対象物の選定**
  - 視覚(サイズ, 光沢)/ 複雑形状 etc.
- **（意図的に）失敗データの追加**
  - エラー回復・異常対応能力の向上
  - 実運用での信頼性強化
- **新規タスクドメインの拡張**
  - 複雑な認知的タスクの導入
  - 判断・推論を要する高次タスクで包括的学習環境を構築

まとめと今後

- 「**AIST-Bimanual Manipulation**」Datasetを紹介
  - 収集におけるプロトコル
  - データセットの概要
- 予備検証として小規模VLAモデルに適用した事例を提示
- 今後は, ロボット基盤モデル開発及び, 多様なフィジカルAIのためのデータセットとして拡張.
- データをオープンにすることで“ロボット学習の発展”の一助を期待