

バイオ実験作業の知の獲得とロボットへの転写

東京大学 生産技術研究所
佐藤 洋一

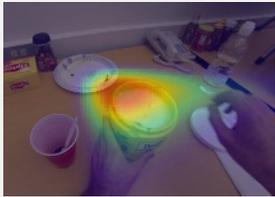
主たる共同研究者：光山統泰（産総研人工知能研究センター）
Kris Kitani（Robotics Institute, CMU）



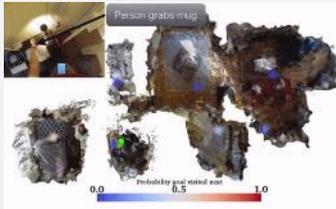
AIP加速課題研究の背景と狙い

CREST 集合視プロジェクト

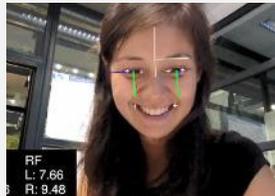
集合視によるグループ行動・注意の理解



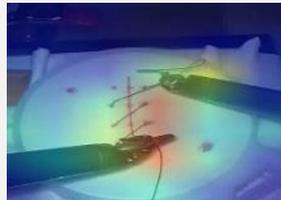
自己視点推定



行動予測



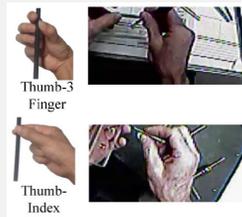
他者視点推定



スキルレベル推定



移動予測



動作認識



一人称手術映像閲覧支援

AIP加速課題プロジェクト

バイオ実験作業



実験技術者が専用ツールを用いて行う高度な作業

作業内容・作業技能（作業知）の獲得

バイオ実験ロボットへの転写



- 実験再現性の改善 への貢献
- 実験技術者の育成

各研究項目の狙い・アプローチ

1. バイオ実験作業映像データセット構築

- 一人称視点と複数外部視点からの観察
- 多様性の担保（実験プロトコル、実験技術者）
- 作業内容の多角的理解のための密なアノテーション



2. 実験作業認識技術の開発

頑健かつ高精度な動作認識

- 外乱や環境変化に頑健な動作認識モデル
- 少量データからの動作認識モデルの学習

頑健かつ高精度な手作業理解

- 環境変化に頑健な手抽出モデル
- 部分遮蔽に頑健なアクティブ物体検出

3. 実験作業技能モデリング技術の開発

作業スキルレベルの評価と可視化

- ツールを用いた作業スキルの定量評価・要因可視化

実験器具の3次元トラッキング

- 透明&金属物体を含む大規模データセットの構築
- 複雑シーン・部分遮蔽に対する頑健性
- 精度と速度の両立

4. 実験ロボットへの作業技能転写技術の開発

人のデモンストレーションに基づくロボット動作の獲得

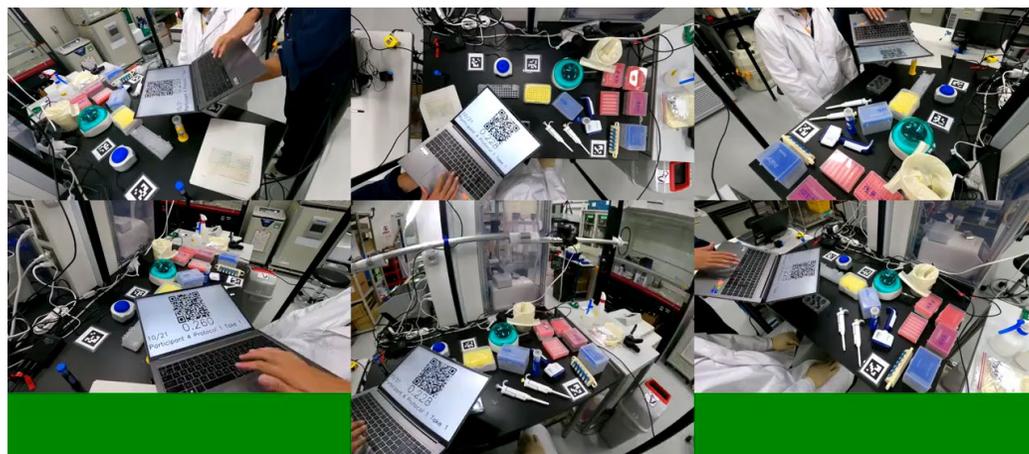
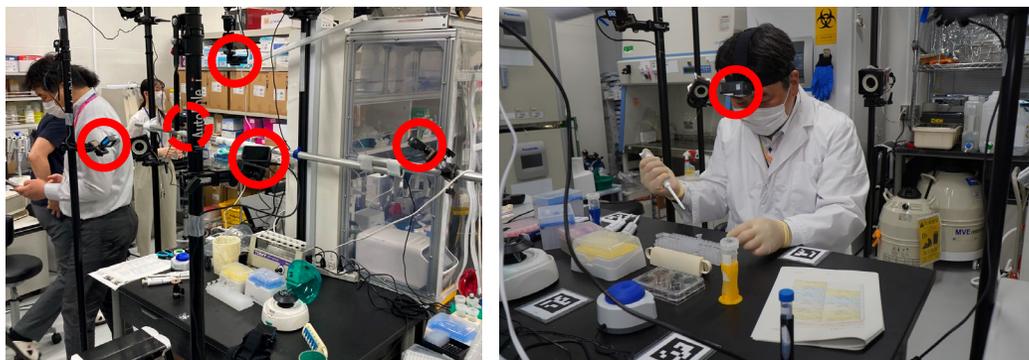
- 動作モデル転移における手とハンドのギャップの克服



バイオ実験映像データセット構築：映像収集

一人称視点映像と外部視点映像による記録

カメラは全て幾何キャリブレーション&時刻同期済み



実験プロトコル

培養細胞の回収と溶解、磁気ビーズによるDNA抽出、PCR実験、スピンドラムによるDNA抽出から7パターン

	培養細胞の回収と溶解
1	ウェルから培地を除去する
2	PBS 1mLを加える
3	プレートを手で軽くゆする
4	PBS 1mLを除去する
5	PBS 1mLを加える
6	プレートを手で軽くゆする
7	PBS 1mLを除去する
8	細胞溶解液を添加する
9	プレートを手で軽くゆする
10	細胞溶液を1.5mLチューブに移す
11	スピンドウン
12	上清を除去する

データセットのサイズ

- 実験者48名×5プロトコル×試行1~2回
- 総試行回数380、実験時間の合計約25時間（1試行平均4分）、映像長さの合計約150時間
- 実験者数、プロトコル数、実験時間の合計、カメラ視点数ともに既存のバイオ実験映像データセットBioVL2[西村+, 自然言語処理, 2021]を大幅に越え最大規模を達成

※ 2022年末までに収録完了予定

バイオ実験映像データセット構築：アノテーション

階層的かつ密なアノテーションの付与

各プロトコルの作業映像に対し

タスク、動作、物体の3階層でアノテーションを付与

実験プロトコル

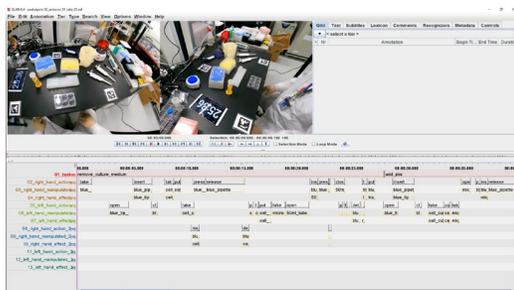
培養細胞の回収と溶解

タスク系列

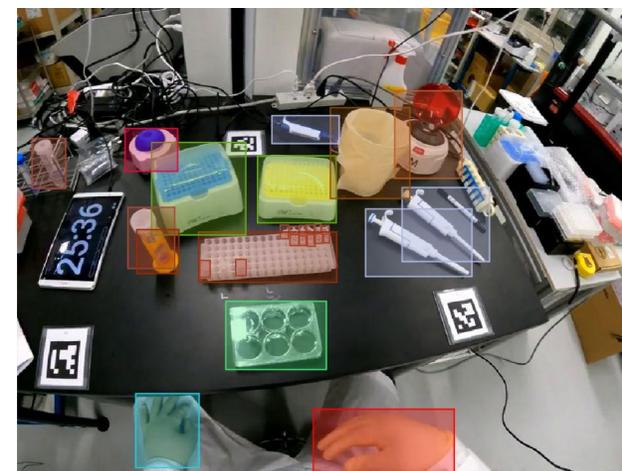
1	ウェルから培地を除去する
2	PBS 1mLを加える
3	プレートを手で軽くゆする
4	PBS 1mLを除去する
5	細胞溶解液を添加する
6	プレートを手で軽くゆする
7	細胞溶液を1.5mLチューブに移す
	⋮

動作系列

動作 1	Take pipette
2	Take tube
3	Open tip
4	Insert pipette into tube
	⋮



remove_culture_medium



take pipette table

- タスク (26クラス) のラベルと時間区間 【約6,600】
- 動作 (10クラス, 物体31クラス) のラベルと時間区間 【約10万】

手・物体の領域 【約15万】

各研究項目の狙い・アプローチ

1. バイオ実験作業映像データセット構築

- 一人称視点と複数外部視点からの観察
- 多様性の担保（実験プロトコル、実験技術者）
- 作業内容の多角的理解のための密なアノテーション



2. 実験作業認識技術の開発

頑健かつ高精度な動作認識

- 外乱や環境変化に頑健な動作認識モデル
- 少量データからの動作認識モデルの学習

頑健かつ高精度な手作業理解

- 環境変化に頑健な手抽出モデル
- 部分遮蔽に頑健なアクティブ物体検出

3. 実験作業技能モデリング技術の開発

作業スキルレベルの評価と可視化

- ツールを用いた作業スキルの定量評価・要因可視化

実験器具の3次元トラッキング

- 透明&金属物体を含む大規模データセットの構築
- 複雑シーン・部分遮蔽に対する頑健性
- 精度と速度の両立

4. 実験ロボットへの作業技能転写技術の開発

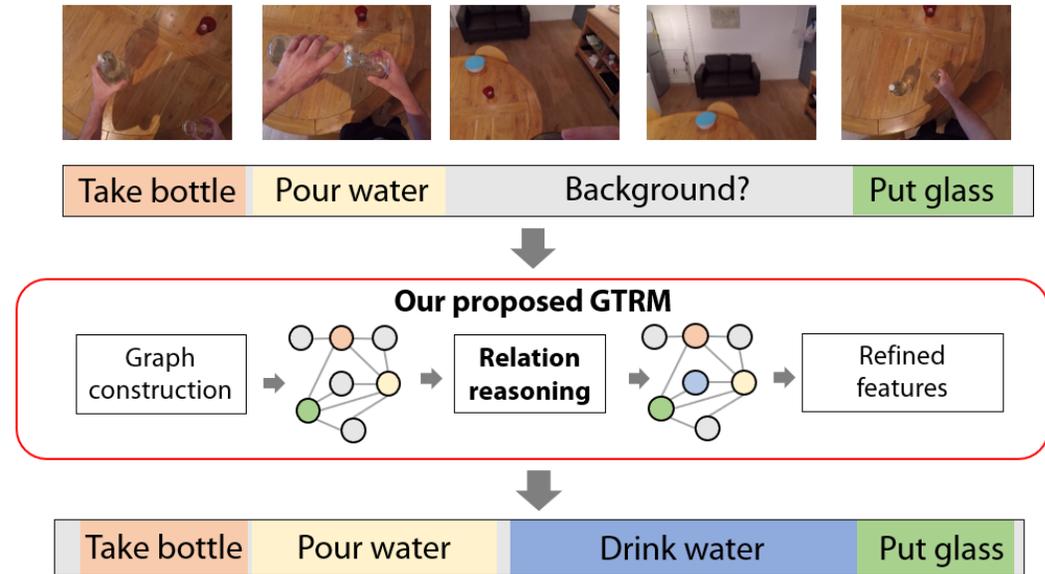
人のデモンストレーションに基づくロボット動作の獲得

- 動作モデル転移における手とハンドのギャップの克服



アクションセグメンテーションの高精度化

- **アクションセグメンテーション**（映像中の動作の分割と識別）は行動理解のための重要な要素技術
- 一人称視点映像では**カメラモーションや視野の制限による性能低下**が課題
- Graph Convolutional Networkで**前後動作間の関係性をモデリング**するというアイデアにより、**動作区間検出・動作カテゴリ識別ともに大幅な性能向上**を実現



大きなカメラモーションを伴う映像の例



EGTEA dataset, 2.5x playback speed



正解



ベースライン(m-GRU)



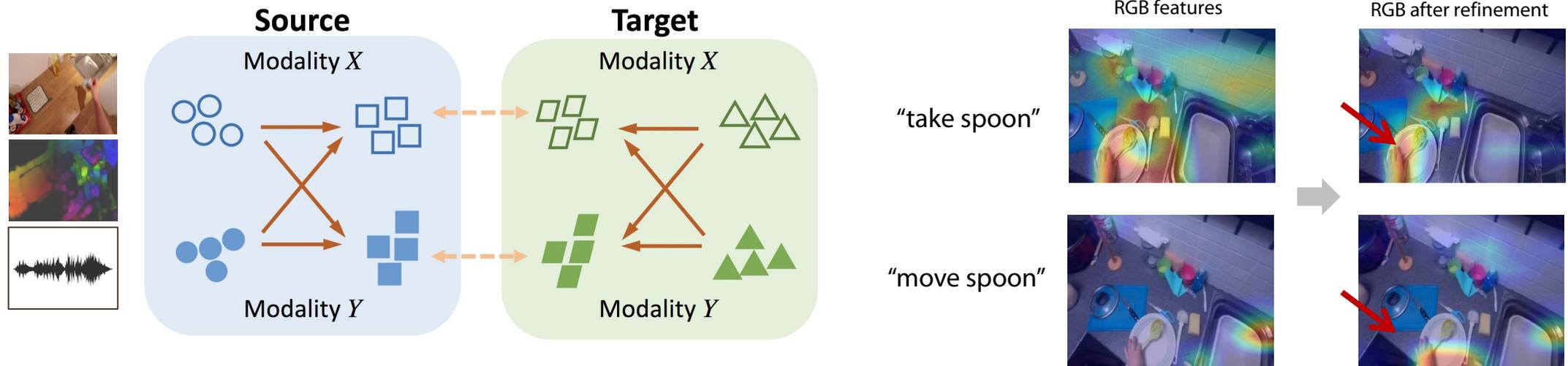
提案手法

アクション認識の頑健性向上

- アクション認識において、モデルの学習時とテスト時のドメインギャップによる性能低下が課題



- マルチモダリティ情報（RGB, フロー, 音）の相補性・整合性に基づき特徴量をリファインするというアイデアにより、アクション認識の教師無しドメイン適応でSOTA性能を実現



手領域抽出の頑健性向上：教師無しドメイン適応

- **手領域抽出**は手作業理解のための基本タスク
- 照明等の**撮影条件の変化**（ドメインギャップ）による**大幅な精度低下**が課題



学習ドメイン



学習済みモデル



テストドメイン

- **自己訓練によるドメイン適応**において疑似ラベルの不確かさをベイジアンCNNで見積もるというアイデアにより、**手領域抽出の教師無しドメイン適応でSOTA性能**を実現

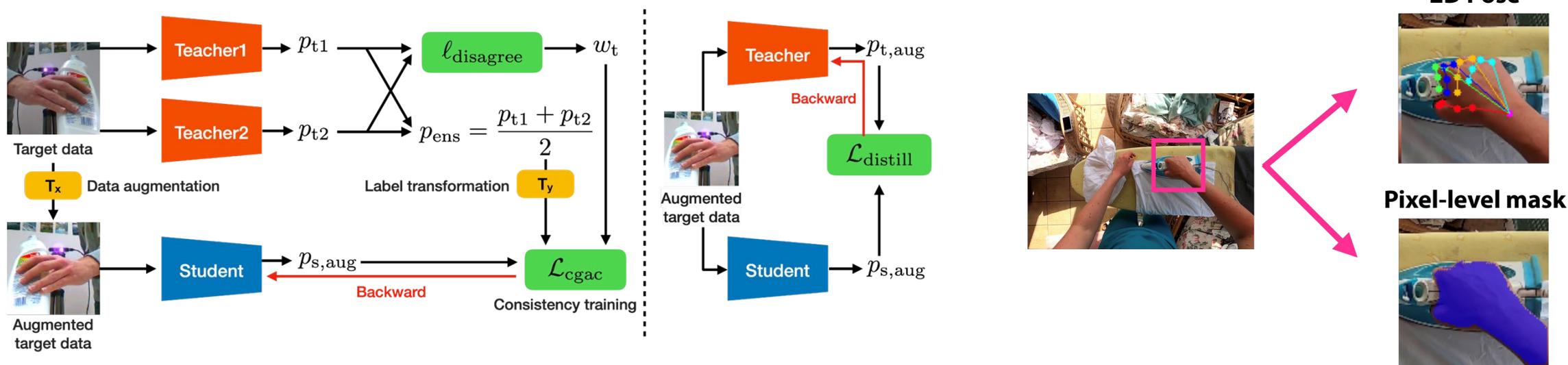


手領域抽出の頑健性向上：コンセンサス疑似ラベル

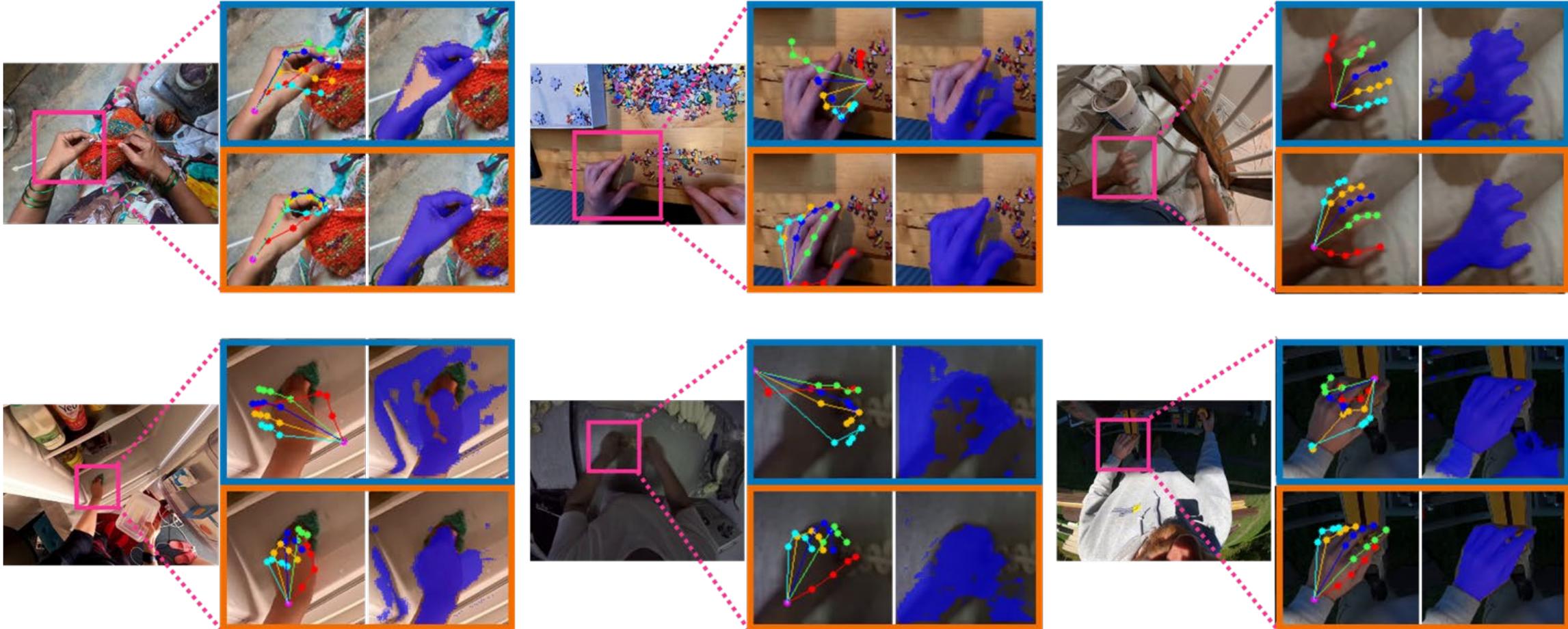
- 自己訓練によるドメイン適応において、**疑似ラベルの良さを複数モデルの推定の一致度合いにより見積もる（コンセンサス疑似ラベル※）**というアイデアにより**更なる性能向上**を実現

※ [Ohkawa+, IEEE Access, 2021], [特願2021-144956、PCT/JP2022/31009]

- 手領域抽出に加えて、**手指2D姿勢推定**を同時に実現



Qualitative Results in the Wild



— Baseline w/o adaptation — Our method

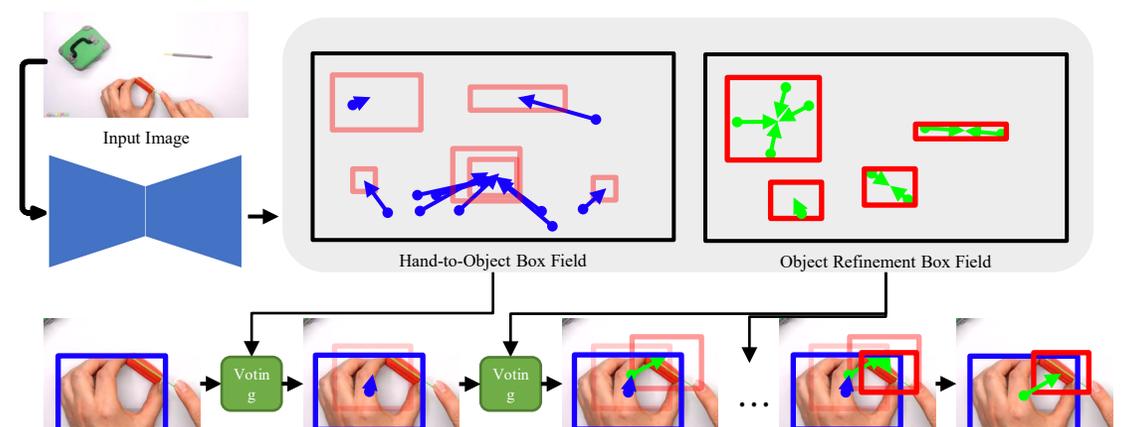
アクティブ物体検出の頑健性向上

- 手物体操作における**アクティブ物体検出**は手作業理解のための基本タスク
- **手による遮蔽に対する頑健性、小物体への対応**が課題

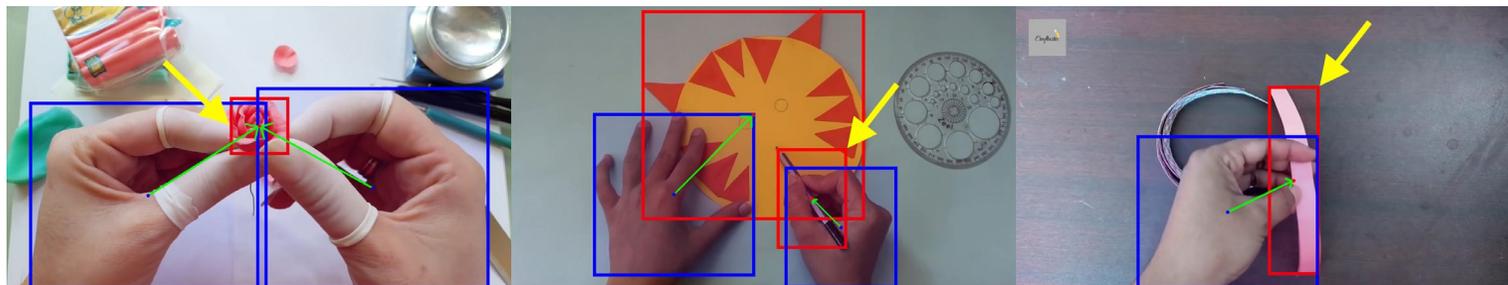


物体領域の検出と手から物体への対応の発見

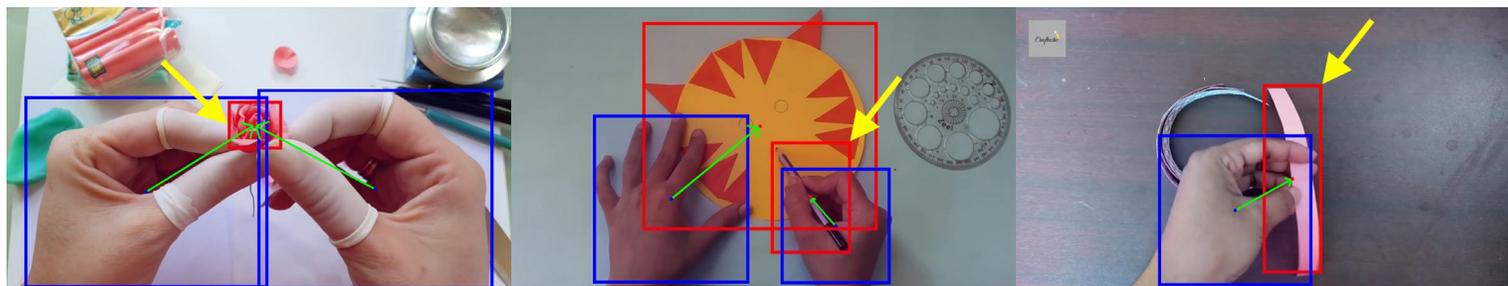
- アクティブ物体の**相対位置・サイズと手のアピアランスの関係**に着目し、**Relational Box Fieldに基づく逐次投票操作**により、アクティブ物体検出の大幅な性能向上を達成



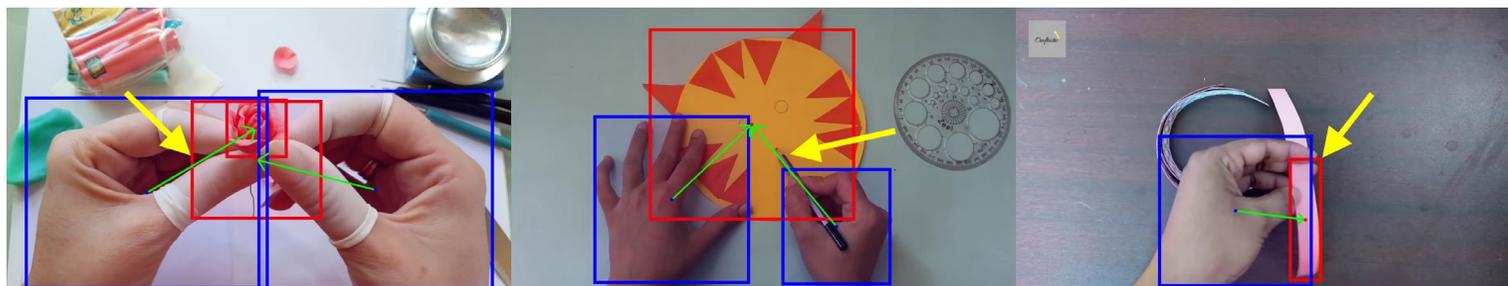
正解



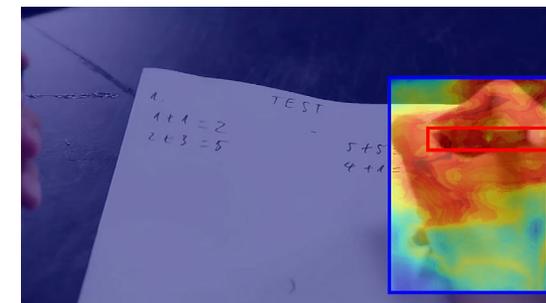
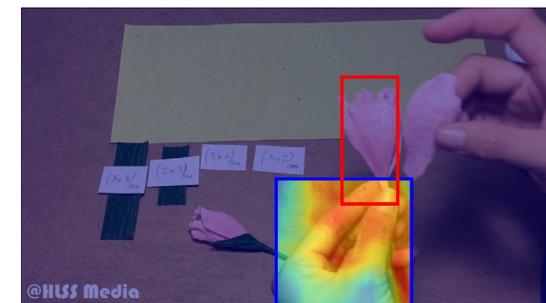
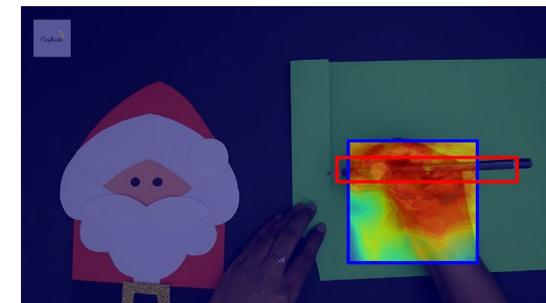
提案手法



HO Detector
[CVPR 2020]



アクティブオブジェクト検出の例



RBFによる投票重みの可視化

各研究項目の狙い・アプローチ

1. バイオ実験作業映像データセット構築

- 一人称視点と複数外部視点からの観察
- 多様性の担保（実験プロトコル、実験技術者）
- 作業内容の多角的理解のための密なアノテーション



2. 実験作業認識技術の開発

頑健かつ高精度な動作認識

- 外乱や環境変化に頑健な動作認識モデル
- 少量データからの動作認識モデルの学習

頑健かつ高精度な手作業理解

- 環境変化に頑健な手抽出モデル
- 部分遮蔽に頑健なアクティブ物体検出

3. 実験作業技能モデリング技術の開発

作業スキルレベルの評価と可視化

- ツールを用いた作業スキルの定量評価・要因可視化

実験器具の3次元トラッキング

- 透明&金属物体を含む大規模データセットの構築
- 複雑シーン・部分遮蔽に対する頑健性
- 精度と速度の両立

4. 実験ロボットへの作業技能転写技術の開発

人のデモンストレーションに基づくロボット動作の獲得

- 動作モデル転移における手とハンドのギャップの克服



作業スキルレベルの定量評価の高精度化

- 作業技能を捉えるためには**スキルレベルの定量評価**が重要
- 従来のスキルレベル推定では、手・器具・物体などの**要素間の関係**を**上手く捉えることが出来ない**という課題

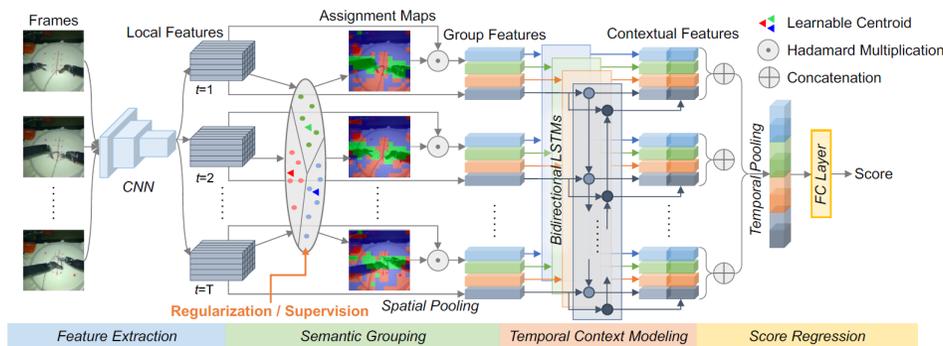
JIGSAWS surgical skill dataset



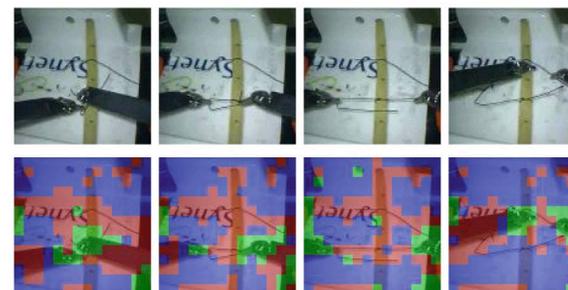
Low skilled (17/30)

High skilled (19/30)

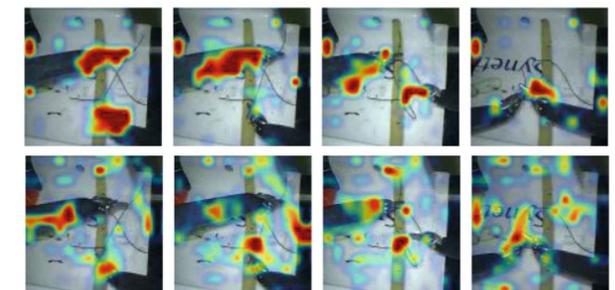
- 画像フレームを画一的に扱うのではなく、**意味的要素を発見し、それらの時空間的關係から作業スキルを捉える**というアイデアにより**SOTA手法を大幅に越える性能**を達成



提案手法 (VISA)の概要



セマンティックグルーピングの可視化



スキルレベルに寄与している箇所の可視化

提案手法

Baseline

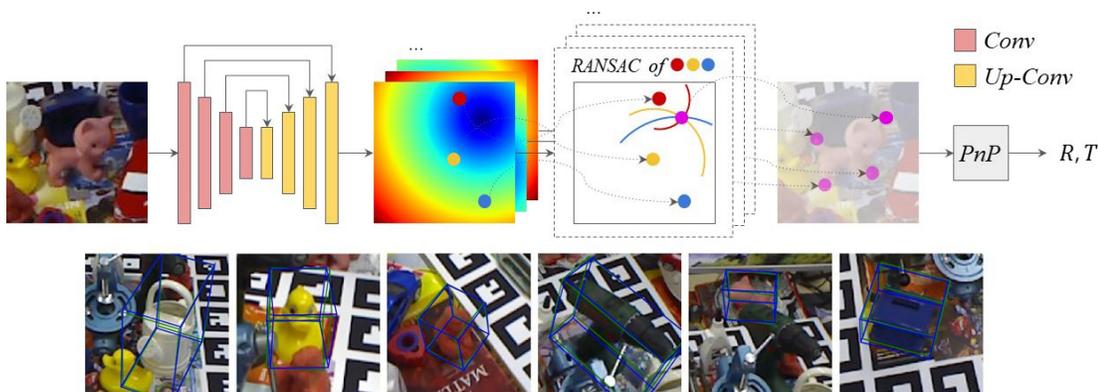
頑健かつ高速な物体 3 次元姿勢推定の実現

従来手法の課題

複雑背景や部分遮蔽による性能低下

解決のアプローチ

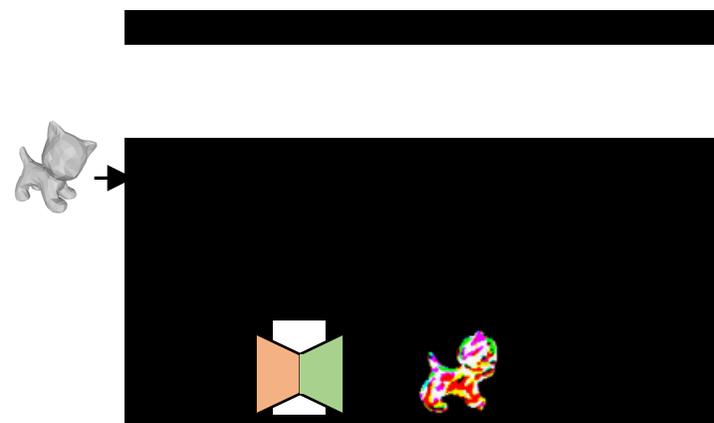
キーポイント対応による姿勢推定 (PnP) に対して、**Keypoint Distance Field表現に基づく投票**というアイデアにより、**部分遮蔽物体や透明物体の姿勢推定でSOTA性能を実現**



推定精度と処理速度の両立が困難

解決のアプローチ

- **特徴量をテクスチャとする物体モデルのレンダリング**とLM法による最適化により姿勢を逐次更新。
- SOTAの姿勢精度と**90+fpsの高速処理**を達成



各研究項目の狙い・アプローチ

1. バイオ実験作業映像データセット構築

- 一人称視点と複数外部視点からの観察
- 多様性の担保（実験プロトコル、実験技術者）
- 作業内容の多角的理解のための密なアノテーション



2. 実験作業認識技術の開発

頑健かつ高精度な動作認識

- 外乱や環境変化に頑健な動作認識モデル
- 少量データからの動作認識モデルの学習

頑健かつ高精度な手作業理解

- 環境変化に頑健な手抽出モデル
- 部分遮蔽に頑健なアクティブ物体検出

3. 実験作業技能モデリング技術の開発

作業スキルレベルの評価と可視化

- ツールを用いた作業スキルの定量評価・要因可視化

実験器具の3次元トラッキング

- 透明&金属物体を含む大規模データセットの構築
- 複雑シーン・部分遮蔽に対する頑健性
- 精度と速度の両立

4. 実験ロボットへの作業技能転写技術の開発

人のデモンストレーションに基づくロボット動作の獲得

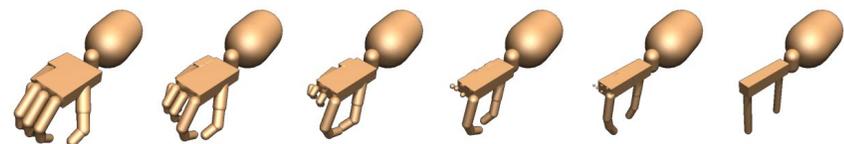
- 動作モデル転移における手とハンドのギャップの克服



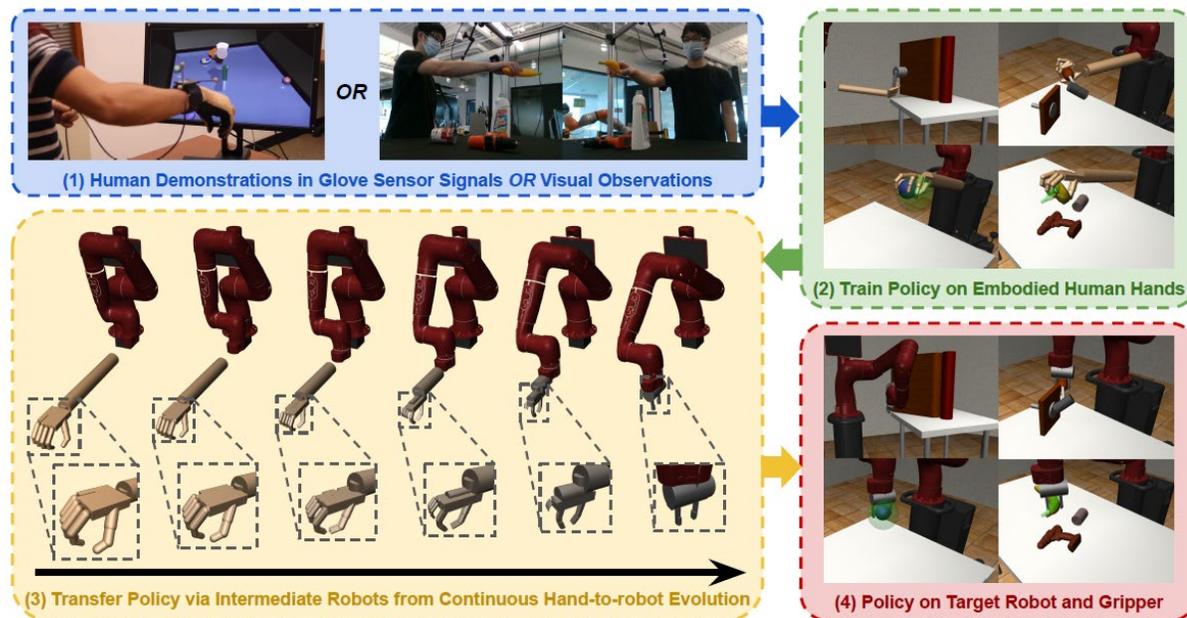
人のデモンストレーションに基づくロボット動作の獲得

課題：人の手からロボットハンドへ動作モデルの転移を実現するために、如何にして**手とハンドの構造の違い**を克服するか

手からハンドへの連続的進化に沿って動作モデルを徐々に転移させるというアイデアにより、動作の転移における**構造のギャップの問題を解決**

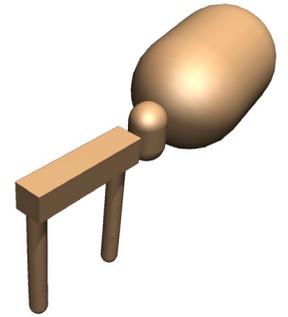
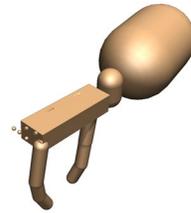
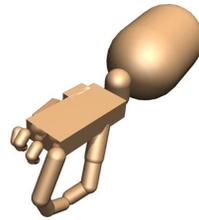
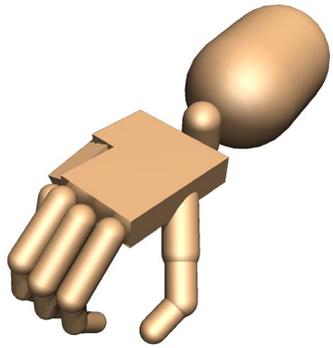


Inter-robot imitation learningの新パラダイム



Liu et al., REvolver: continuous evolutionary models for robot-to-robot policy transfer, ICML 2022 (long oral, top 2.1%)

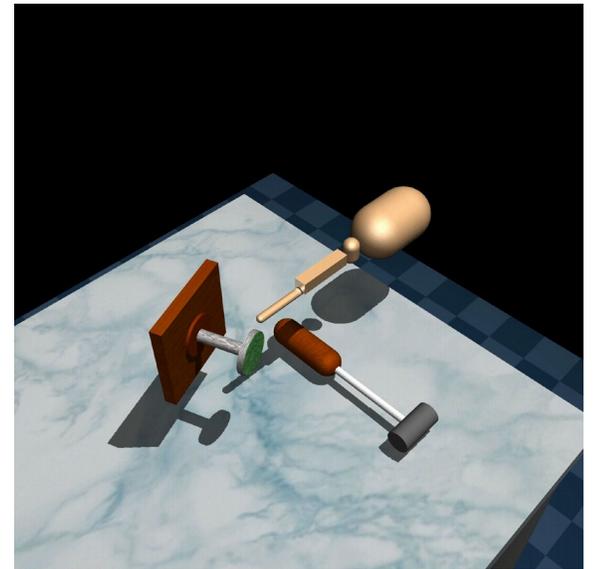
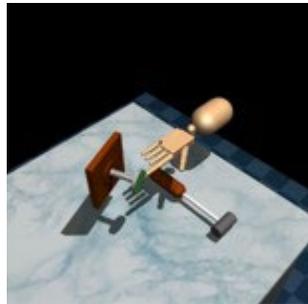
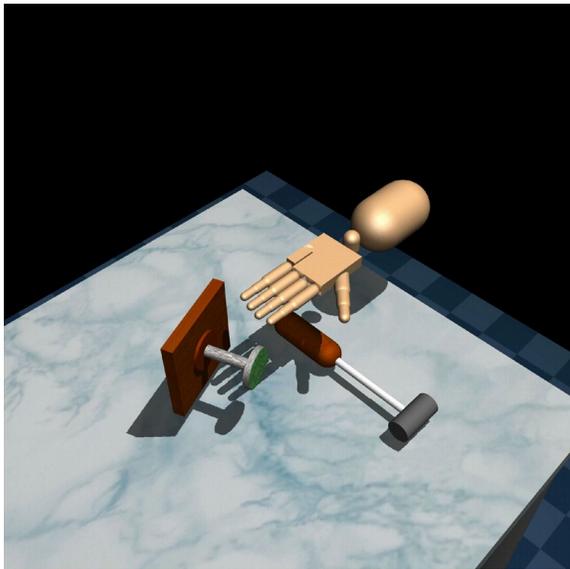
Liu et al., HERD: continuous human-to-robot evolution for learning from human demonstration, CoRL 2022



Robot at evolution
progress $\alpha = 0.0$



$\alpha = 1.0$



まとめ：実験作業知の獲得とロボットへの転写の実現に向けて

バイオ実験映像データセットの構築

- 多様性・データ量の両面で最大規模
- 作業内容の多角的理解のための密なアノテーション

作業認識技術の開発

- 動作間の関係性に基づく頑健かつ高精度なアクションセグメンテーション
- モダリティ間のインタラクションによる動作認識のドメイン適応
- 手作業理解のための頑健な手領域抽出とアクティブ物体検出

作業技能モデリング技術の開発

- 基本要素間の関係を捉えたスキルレベル推定と要因可視化

ロボットへの作業技能転写技術の開発

- 連続的進化に基づく手からロボットハンドへの動作モデルの転移

今後の展開

- 大規模言語モデルやVision & Languageモデルの活用、特定作業ドメインへの適応
- システム構築による概念検証