公開

研 究 終 了 報 告 書

「模倣 AI エージェントによる人物行動理解」

研究期間： 2020 年 11 月～2023 年 3 月
研 究 者： 大川 武彦

1. 研究のねらい

少子高齢化社会や後継者不足の問題は極めて深刻であり，国内では熟練ノウハウを持つ団塊世代の大量退職によるモノづくり産業の減衰が顕著となっている．持続的成長可能な社会の実現に向けて，デジタル技術を活用した熟練者技能の継承・伝承も検討されるべきである．熟練者の動作には，勘やコツなどの個人の感性に帰属する「暗黙知」の割合が多く，先端人工知能技術やロボティクスなどの最新研究を導入しても形式化することが難しい現状である．

この熟練者動作のモデリングとして，模倣アルゴリズムを利用して人の手動作系列の生成を行い，シミュレーション可能，またはロボット実行可能にする試みがなされている．しかし，このような模倣による人工知能やロボットの構成法は，このシステムの入力となる手操作情報の取得が困難であることから実験室環境等の制限された条件下でしか応用が進んでいない．

本研究では，調理，組み立て，生化学実験等の様々な人の作業映像から手動作系列を取得する技術を開発する．具体的には，実験室環境などの既存のラベル付けされた手操作認識データセットから学習したモデルを，それらの応用場面ごとに適応する戦略を考える．この新規映像における手操作認識の課題を解決することは，調理や製造，生化学実験を行うロボットの開発やその応用の促進へ貢献するものである．

2. 研究成果

(1)概要

模倣による人工知能やロボットの構成に向けて，様々な作業映像において人の手操作を認識することは重要な課題である．手操作に関する情報の中で手領域抽出，手姿勢推定，手物体検出等のタスクは，手がどのように物体を操作するかを詳細に表現することができる．

しかし，実世界の新しい映像データに手操作認識モデルを適用した場合のこれらのタスクの精度は未だ低く，どのようにこの新規映像にて高い性能の学習器を構成するかは非自明である．既存の手操作認識に関するデータセットは，手領域，手関節，手と物体位置に関するアノテーションの難易度の高さから環境の多様性が限定的な実験室で構築されている．このようなデータセットをもとに学習した手操作認識モデルは，新規映像において認識精度が悪化し，手操作に関する情報の自動取得に利用できない．

上記の課題を解決するために，本研究では新規映像データが与えられた際に，自己教師あり学習やデータ拡張を活用して，手操作認識モデルを改善する方法を提案する．それらの詳細について後段にて記述する．

また，本 ACT-X 研究を通じて国内外の研究機関との共同研究や対外交流，国際的な会合への参加が実現した．アメリカのカーネギーメロン大学へ訪問研究者として滞在して研究実績を積むことができ，オムロンサイニックエックス社と産学連携して研究を実施した．さらにチューリッヒ工科大学，Meta 社，Microsoft 社などの有名研究機関へ訪問して，私の専門分野とそ

の隣接分野の研究の状況，今後の社会実装への見通しなど議論することができた．

（2）詳細

A. **新規映像における手操作認識**

様々な作業映像にて手操作認識を行うために，学習モデルが未だ経験してない新規の映像データにおける推論精度を向上させる研究を実施した．特に，手領域抽出，手姿勢推定，手物体検出のタスクは詳細なレベルでの手操作の状態を表現するため，ユーザの行動理解やロボットによる模倣に向けて重要な課題である．

1. **手領域抽出タスク（IEEE Access 2021 採択[1]）**

手領域抽出は，画像から手のピクセル単位のラベルを予測するタスクであり，手位置の同定，指先の状態の解析等に利用できる．研究[1]では，新規映像におけるユーザの手の見えと背景の環境の相違を明示的に解消し，信頼できる形式で自己教師あり学習を行う手法を提案する（図 1）．第一に，データ間の前景・背景に関する見えの相違を解消するために，画像スタイル変換ネットワークを提案し，前景同士，背景同士の画像スタイルの差異を解消した．第二に，ラベルのない新規映像に擬似的なラベルを付与してモデルを再学習する擬似ラベリング手法を提案した．ここでは，2 体のネットワークから予測の確信度を推定して，信頼できる擬似ラベルのみを再学習に使用することで性能向上を図った．

本研究成果は，国際ジャーナル IEEE Access 2021 へ採択され，国内会議 MIRU2021 にてロングオーラル発表を行い，学生奨励賞を受賞した．さらに，本研究はオムロンサイニックエックス株式会社と共同研究を実施した成果である．

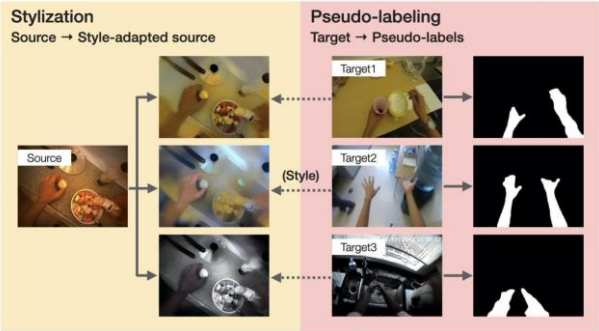

図 1: 新規映像における手領域抽出手法

2. **手領域抽出と手姿勢推定タスク（ECCV 2022 採択 & ワークショップ x2 発表[2]）**

[1]からさらに発展させ，新規映像において手姿勢推定と手領域抽出を同時に解く課題に取り組んだ．手姿勢推定は，手のキーポイントの座標を回帰するタスクであり，手領域より明示的に手の把持姿勢を推定するものである．しかし，手姿勢アノテーションは一層困難であり，利用可能なラベルありデータは実験室環境のものに限られる．研究[2]では，実験室データから構成された既存モデルを屋外などの非常に異なる撮影条件でうまく推論させることを目的とした（図2）．

本研究では，新規映像データ上で幾何的なデータ拡張前後における自己教師あり学

習手法を提案した．データ拡張前後の一貫性学習に予測の信頼度推定，Teacher-Student 学習機構を組み合わせることで安定して性能向上することを発見した．

本研究は，カーネギーメロン大学へ訪問研究員として滞在していた期間の成果であり，一流国際会議 ECCV 2022 に採択された．さらに，同会議から手操作認識に関するワークショップ，複数視点からの人物行動理解に関するワークショップに招待され，ポスター発表を行なった．

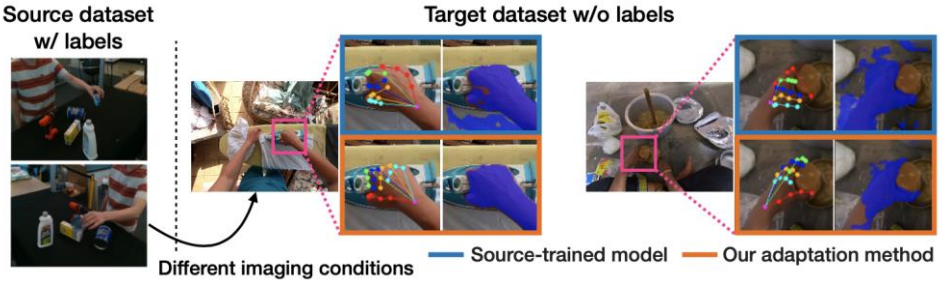

図 2: 新規映像における手姿勢推定と手領域抽出

3. 手物体検出タスク（ECCV 2022 ワークショップ発表[3]）

手を対象とするだけでなく操作している物体も同時に検出することは，手と物体の作用関係を明らかにする役割を持つ．新規映像において，2 枚の学習画像を重ね合わせてデータの偏りを緩和する Mixup と呼ばれるデータ拡張が有効であることが示されている．しかし，手・物体検出においては，2 枚の手作業中の画像を混合すると，特定の領域に手や物体が集中して，手・物体境界の識別能力が低下するなど，意図しないバイアスが発生する．研究[3]は，手・物体検出におけるこの意図しないバイアスを軽減しつつ，データ混合の効果を活用する Background Mixup と呼ばれるデータ拡張手法を提案する．図 3 のように，生化学実験等の小規模の学習データに，既存の大規模な背景データを混合することで，データの多様性を向上させ，手・物体検出器をより頑健にする．

本研究成果は，ECCV 2022 の手操作認識に関するワークショップにて採択され，ポスター発表を行った．

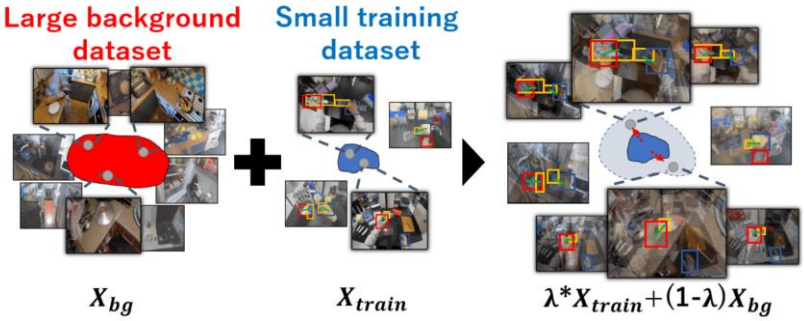

図 3: 手・物体検出における背景画像混合データ拡張

3. 今後の展開

A. コンテクスト情報を利用した手操作認識

本 ACT-X 研究で開発した手操作認識モデルは画像フレーム毎に推論するものであったが，特に一人称視点映像ではモーションブラーや手が視野外に消えてしまうなど単一フレームからの推論では，手操作の状態推定が難しいケースが多い．これを解決するための今後の展開として，映像のコンテクスト情報を利用して動作の意味や時間的な整合性を考慮した手操作認識が考えられる．また，これらの補助情報を元にした認識モデルの改善は，新規映像に対する適応においても重要な手がかりとなると考える．

将来的に一人称視点からの手操作認識の実用に向けて，上記のようなモデリングの可能性についてコミュニティを上げて議論する必要がある．そのため，これまでのように手操作認識にのみ注視するのではなく，今後数年間で手操作認識とコンテクストの関連に関する学会のワークショップやチャレンジを構築して，当テーマの探索を加速させる必要がある．

4. 自己評価

模倣人工知能やロボット構成に関わる重要な人物行動理解のタスクとして，手操作認識の理解に焦点を当て，国内外の著名な研究機関と共同で研究を進めて，一流国際会議やジャーナルへの採択等の業績を上げることができた．研究費については，これらの共同研究を行う際の旅費・滞在費に主に充てたため，本 ACT-X の支援なしではこれらの共同研究を行うことが出来なかった．本研究の社会実装を進めるために，企業訪問や産学連携により現在の実用状況と今後の展開について議論ができた．さらに，ACT-X 領域会議や国際学会，現地の研究所訪問を通して，隣接するロボット，AR/VR，グラフィクス，自然言語処理等の研究者と交流を深めて，自身の専門とその立ち位置を明確にすることが出来た．

5. 主な研究成果リスト

（1）代表的な論文（原著論文）発表

研究期間累積件数：4件

[1] Takehiko Ohkawa, Takuma Yagi, Atsushi Hashimoto, Yoshitaka Ushiku, and Yoichi Sato. Foreground-Aware Stylization and Consensus Pseudo-Labeling for Domain Adaptation of First-Person Hand Segmentation. IEEE Access, vol. 9, pp. 94644-94655, 2021.
新規映像における手領域抽出に関する研究．内容は 2.研究成果(2)詳細 A.1 に記載．
[2] Takehiko Ohkawa, Yu-Jhe Li, Qichen Fu, Ryosuke Furuta, Kris M. Kitani, and Yoichi Sato. Domain Adaptive Hand Keypoint and Pixel Localization in the Wild. In Proceedings of European Conference on Computer Vision (ECCV), 2022 (in press) and Two Invited Posters in European Conference on Computer Vision Workshops (ECCVW), 2022.
新規映像における手姿勢推定と手領域抽出に関する研究．内容は 2.研究成果(2)詳細 A.2 に記載．
[3] Koya Tango, Takehiko Ohkawa, Ryosuke Furuta, and Yoichi Sato. Background Mixup Data Augmentation for Hand and Object-in-Contact Detection. In European Conference on Computer Vision Workshops (ECCVW), 2022 (in press).

手物体検出における画像混合データ拡張に関する研究. 内容は 2.研究成果(2)詳細 A.3 に記載.

（2）特許出願

研究期間全出願件数：0 件（特許公開前のものも含む）

（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

- 2021 年 7 月 MIRU 学生奨励賞 （研究[1]に関する受賞）
- 2022 年 2 月 UTokyo-IIS Research Collaboration Initiative Award