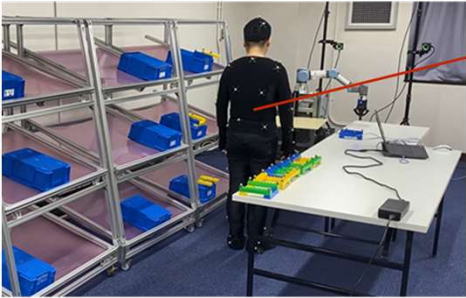


ABCI 3.0開発加速利用（2024年度）成果概要（公開用）

課題名：Video-LLMによる熟練者作業理解とロボット動作学習への転移	実施時期：2025.02~03 所属機関名：東京大学大学院 代表者氏名：Yusheng Wang
成果概要： 本研究は実際の生産現場の作業者の動画から、作業の内容、継続期間、動作の情報などを抽出し、ロボットによって再現することを目的とする。学習済のVideo-LLMは、特定領域である工場の作業をうまく理解できないため、ファインチューニングが必要。本研究はLoRAなどでファインチューニングを試す。 成果のポイント： 本研究は、実際の生産現場における作業者の作業映像をもとに、作業の種類や内容、各作業の継続時間、さらに細かな動作に関する情報を抽出し、それらの特徴をロボットに学習・再現させることを目的としている。これにより、現場で人が行っている複雑な作業をロボットに効率的に移行させるための基盤技術の確立を目指す。 成果として、まず、State-of-the-artモデルをhugging faceからダウンロードし、実装を行い、結果を確認した。既存モデルは大体なキャプションを生成することが可能。例えば、人間は左手を挙げたような基本的なこと。一方で、タイミング情報はだいぶ間違っている。そこで、ファインチューニングのニーズがあげられる。現在所有する作業動画の数はごくわずかで、ファインチューニングでも一定規模なデータが所要し、それを構築している。これが完成後、LoRAでファインチューニングを行う。  • 人間の作業内容： 左手でパーツを取る 両手で組み立てるなど • タイミング： 個の動作1分30秒～1分40秒まで継続 	
成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：	