様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ６ 月 １６ 日現在

機関番号：８２６２６
研究種目：研究活動スタート支援
研究期間：2022～2023
課題番号：２２Ｋ２１２９６

研究課題名（和文）映像とキャプション系列のマルチモーダル解析による物体状態認識

研究課題名（英文）Object State Recognition via Multi-Modal Analysis of Videos and Video Caption Sequences

研究代表者

八木　拓真（Yagi, Takuma）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・研究員

研究者番号：５０９６４２７７

交付決定額（研究期間全体）：（直接経費）　　2,200,000円

研究成果の概要（和文）：映像中に出現する物体の状態（例：卵が割れている、ゆでられている）を認識する計算モデルを開発した。物体状態の認識にあたっては映像と対応する物体状態に関する注釈が必要であるが、多様な物体状態の教師情報の収集はコストが高く現実的でない。そこで本研究では、インターネット映像中に含まれる説明文（実況）の情報に大規模言語モデル（LLM）を適用することで多様な物体状態に関する教師情報を自動で生成しモデルの学習を行う新たなフレームワークを提案した。

研究成果の学術的意義や社会的意義
従来人の行動やその周辺環境の理解にあたっては、人が何をしているか（行動）および何があるか（物体）の認識が主で、ある物体が人の行動の結果どのような状態になったかといったシーンの詳細に関する認識が十分に取り組まれていなかった。様々な物体状態を映像から自動で認識することで、例えばロボットが行動を意図した通りに実行できたかを実際に物体の状態が変化したかによって判定でき、より信頼性の高いタスク遂行が期待できる。また、LLMは任意の状態記述に対応できるため語彙の変更が容易で、ユーザの要求に合わせた認識結果を提供することも可能となる。

研究成果の概要（英文）：We developed a computational model that recognizes the states of objects that appear in a video (e.g., an egg is cracked or boiled). Recognizing object states requires annotations of the object states that correspond to the video, but collecting training information for various object states is costly and unrealistic. In this study, we proposed a new framework that automatically generates training information for various object states by applying large language models (LLM) to the information in the narrations included in Internet videos.

研究分野：コンピュータビジョン

キーワード：物体状態認識　大規模言語モデル　映像字幕からの学習

様　式　C－１９、F－１９－１、Z－１９（共通）

１．研究開始当初の背景

ロボットなどが物体操作を伴うタスクを実行する際、対象物体の状態を正しく認識し、ある行動がどのような変化をもたらすかを予測することはタスクの確実かつ安全な遂行に不可欠である。従来、動画像からの物体状態認識では事前に取りうる状態をコップが空/一杯などの離散変数として表現してきたが、状態空間を固定してしまうと、想定外の状態に対応できない。解釈可能かつより柔軟な表現として、動画像とそれを説明する文章（キャプション）との対応関係を学習する方法があるが、物体状態およびその変化を認識する文脈では十分に取り組まれてこなかった。

２．研究の目的

本研究では動画像中の物体の状態およびその変化を明示的に説明したキャプション系列（状態記述キャプション）から物体単位の状態認識を実現することを目的とする。具体的には、物体状態変化を含む動画像に対して出現物体の位置・状態およびその変化をもたらした行動や現象を説明するキャプションを新たに付与し、対象物体および周辺の見えの変化と対応づける学習を行うことで物体単位での特徴表現を獲得する。

３．研究の方法

当初計画では物体単位の状態記述キャプションをクラウドソーシングなどにより人手で収集することを予定していたが、研究期間中に急激に進歩のあった大規模言語モデル（LLM）を用いることによる解決を試みた。

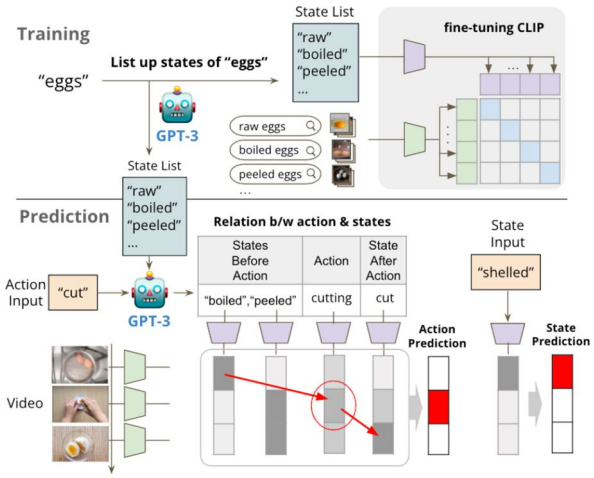

図　１ (1)の提案フレームワークの概要

(1) LLM を用いた状態ラベル自動列挙

従来の映像からの物体状態認識の研究では、事前に認識対象の状態カテゴリおよびそこで行われる行動情報が既知であると仮定されており、特に様々な状態語が考えられる物体状態について柔軟性が不足していた。そこで LLM に含まれる行動-状態間の関係性に関する世界知識を利用することで行動語からその行動の前/後に期待される物体状態語を複数列挙し、それらを既存の画像-言語モデルと組み合わせることにより、より正確に物体状態変化のタイミングを推定する方法を考案した（図 1）。

(2) 説明文付きの映像情報からの LLM を用いた情報抽出に基づく物体状態認識

(1)では映像中に単一回の物体状態変化が起こることを仮定していたが、実際の映像では対象とする物体が複数回様々な物体状態変化を経ることが一般的であり、行動と状態変化とが 1 対 1 で対応するとは限らない。そうした多様な状態の出現区間を推論するためには従来多量の教師ラベルが必要とされ、現実的には困難であった。

そこで本項では、音声実況などの説明を含む映像から LLM を用いて、生のナレーションをまず行動情報に変換し、それをその行動の結果現れる状態記述キャプションに変換し、最後に認識対象の状態ラベルの有無を推定するフレームワークを提案した（図 2）。ナレーション付きの映像の状態に関する説明の不足を補いつつ自動的に状態認識のための訓練ラベルを収集する方法を構築した。図 3 にりんごの調理映像に対する様々な物体状態の有無に関する予測例を示す。完全ではないものの、複数の状態が同時に出現する複雑な映像についても、それぞれの出現区間を柔軟に認識できることを示した。

４．研究成果

当初の予定とは異なる結果をえたものの、当初目標にて設定した映像からの柔軟な物体状態認識について、LLM を用いることによって期待通りの成果を得られたと考える。研究機関を通じて (1) ナレーション付きの映像からの状態カテゴリの自動列挙 (2) LLM を用いた状態認識器の学習 それぞれについて国内学会で発表済・発表予定であり、(2)については査読付国際会議に投稿中である。提案した LLM ベースの情報

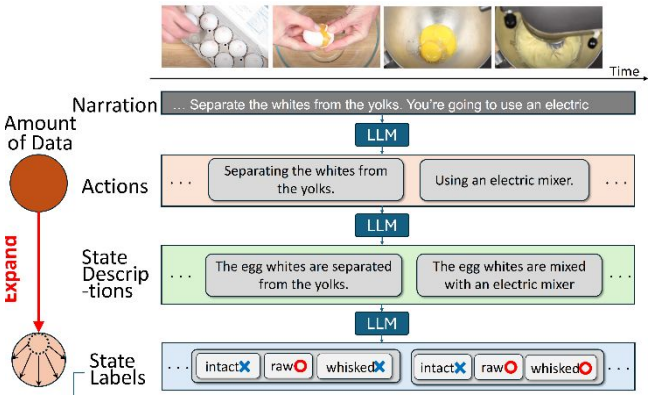

図　２ (2)の提案フレームワークの概要

変換フレームワークは物体状態認識に限らず、教師情報が豊富な属性（例：説明文中に含まれる行動情報）から教師情報が少ない属性へ知識を転移（図 2 左）することが有効なドメインについても適用可能であり、人手による教師情報付与を回避しながら映像からの各種属性認識を行える可能性を示した。
様々な物体状態を映像から自動で認識することで、例えばロボットが行動を意図した通りに実行できたかを実際に物体の状態が変化したかによって判定でき、より信頼性の高いタスク遂行が期待できる。また、LLM は任意の状態記述に対応できるため語彙の変更が容易で、ユーザの要求に合わせた認識結果を提供することも可能となる。

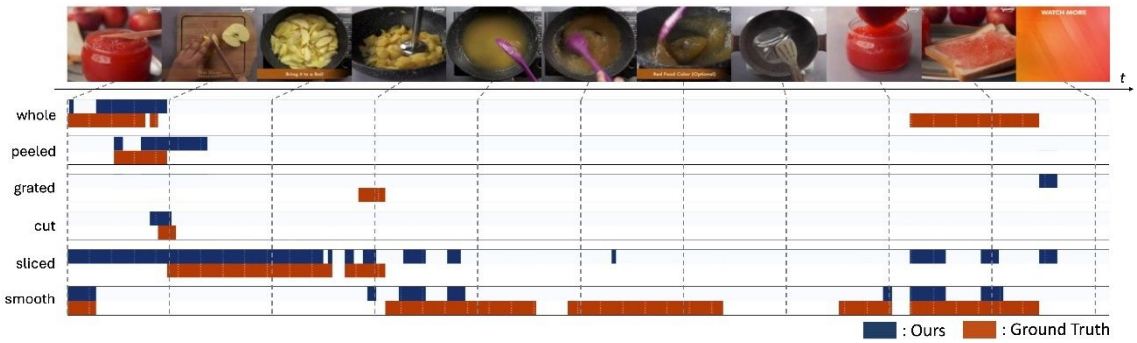

図 3 リンゴにおける(2) を用いて学習したモデルの物体状態認識の予測例（青：予測結果、赤：正解）

5．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計3件（うち招待講演　0件／うち国際学会　1件）

1．発表者名 舘野将寿、八木拓真、古田諒佑、佐藤洋一
2．発表標題 大規模言語モデルを用いた学習カテゴリの自動決定による映像からのオープン語彙物体状態認識
3．学会等名 第26回画像の認識・理解シンポジウム
4．発表年 2023年

1．発表者名 Masatoshi Tateno, Takuma Yagi, Ryosuke Furuta, Yoichi Sato
2．発表標題 Learning Object States from Actions via Large Language Models
3．学会等名 第27回画像の認識・理解シンポジウム
4．発表年 2024年

1．発表者名 Masatoshi Tateno, Takuma Yagi, Ryosuke Furuta, Yoichi Sato
2．発表標題 Learning Object States from Actions via Large Language Modelsa
3．学会等名 CVPR Workshop Learning from Procedural Videos and Language: What is Next?（国際学会）
4．発表年 2024年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

—

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関