COM Kitchensデータセットのご紹介

IDRユーザフォーラム2024

橋本敦史 / オムロン サイニックエックス株式会社 / PI

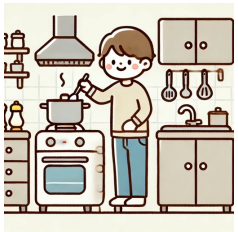
SINIC X

2024.12.13(Fri)

IDR

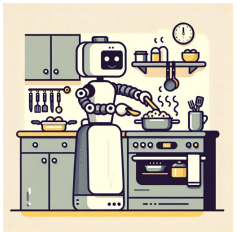
COM Kitchens Dataset: 取り組みの背景

手順+作業動画の理解は技能継承の重要な鍵



Human to Human

Human to Robot



固定視点映像は動画の主要なフォーマットだが未探索

- 最近のスマホは広い作業エリアを一台で撮影可能。
  - 数多くの潜在的なスマホアプリ応用が考えられる
- ウェブ動画や一人称視点映像とのドメインギャップ
  - 編集済み v.s. 未編集, 作業領域へのフォーカスあり v.s. なし, など。

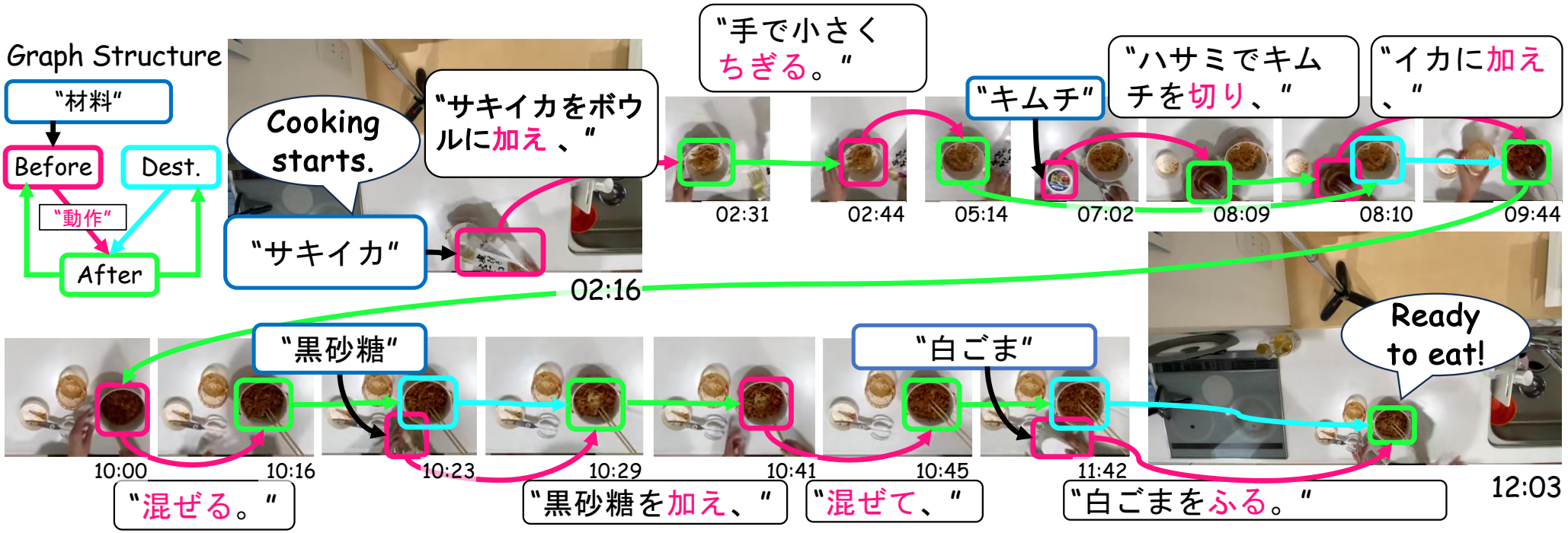
COM Kitchens Dataset: データ収集プロセス

- iPhone 11 Pro (リアカメラ)
  - Full HD (1920x1080)
  - 超広角モードで撮影
- Obtained 213 valid videos
  - **84 home kitchens**
  - **197 recipes**
  - **平均動画長: 16 minutes**
  - 訓練/検証用セット (80%) がIDRから公開されています.
    - 評価用セットはコンペ開催のために非公開のままとってあります。

他のデータセットとの主要な差異 → 撮影環境とタスク(recipe)の多様性

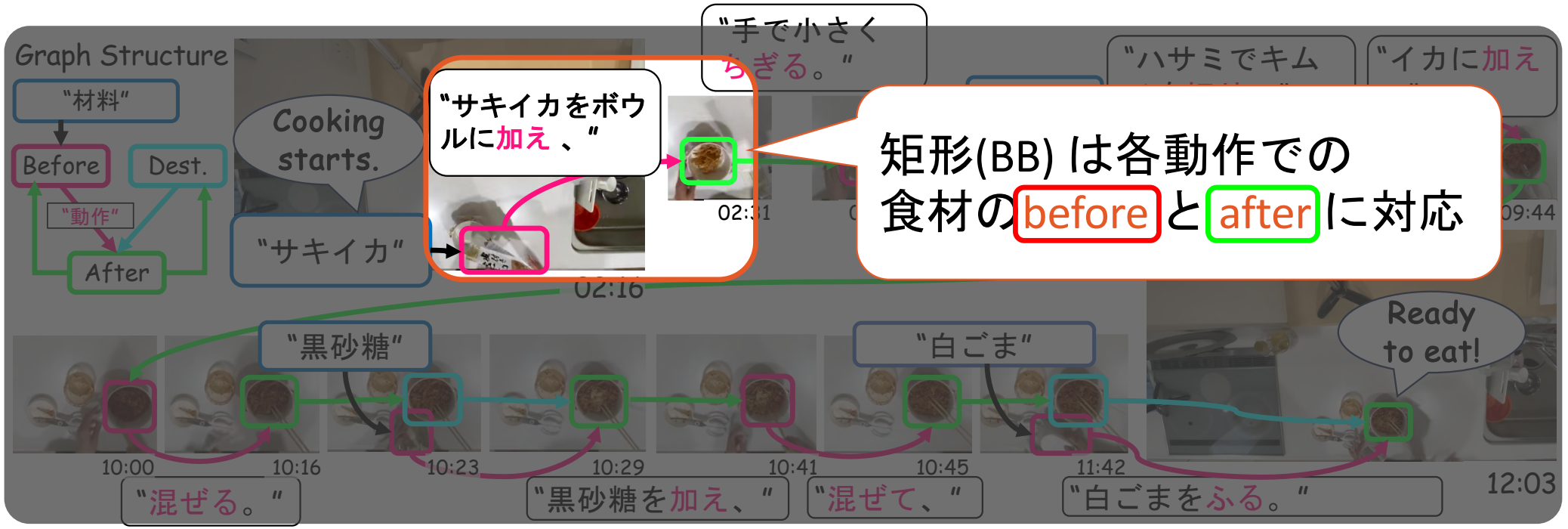
COM Kitchens Dataset: アノテーション

Visual Action Graph が 213本の動画のうち145本に付与済み (合計40時間分)



COM Kitchens Dataset: アノテーション

Visual Action Graph が 213本の動画のうち145本に付与済み (合計40時間分)



COM Kitchens Dataset: アノテーション

Visual Action Graph が 213本の動画のうち145本に付与済み (合計40時間分)

ベンチマーク課題1: オンラインレシピ検索

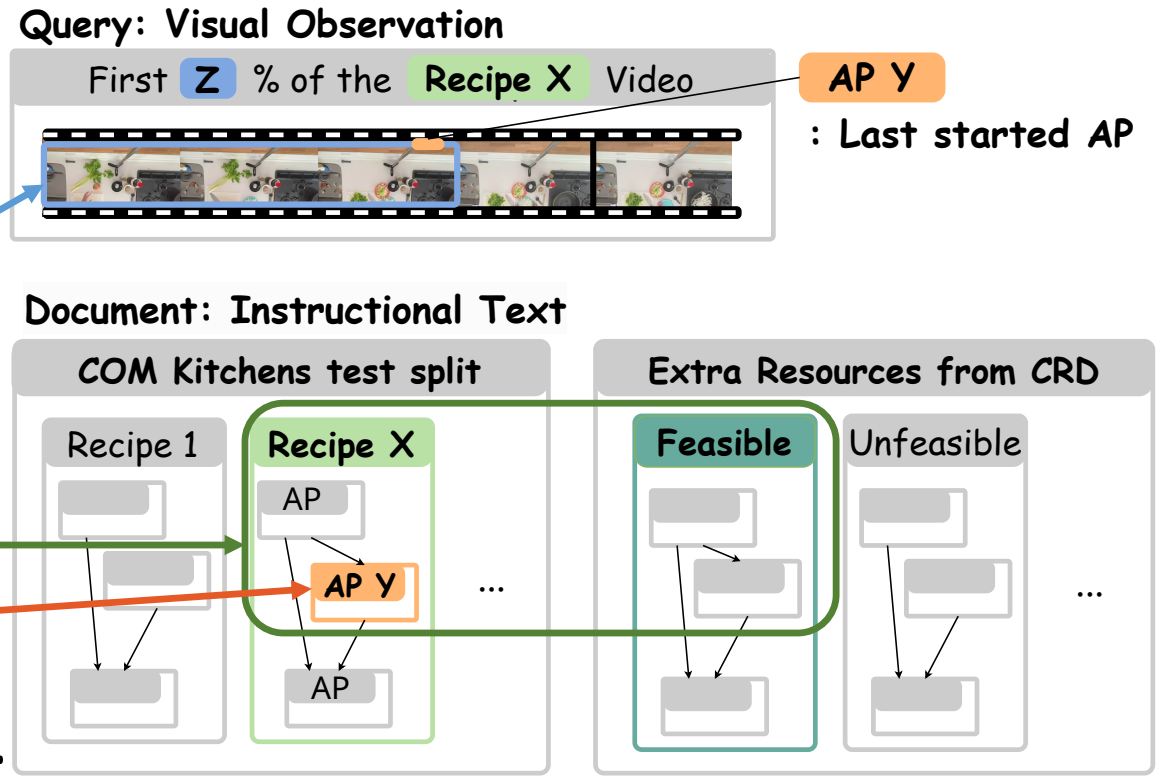
想定する応用例

1. オンラインレシピ推薦
2. 調理ナビゲーション

クエリ: 観測動画
- 作業途中までの動画(=online)

検索対象: 手順書 + 進捗

1. **移行可能レシピ**
2. **レシピ内の現在ステップ**



現行のSOTA手法ではほぼ解けない.

(詳細は論文を参照ください)

ベンチマーク課題2: Dense Video Captioning (DVC)

動機: ウェブ動画でのDVCとのドメインギャップを明確にすること.

入力: 映像全体 (= offline)
出力: 動作ごとの手順テキスト

Challenges

1. 多くの動作が繰り返し行われる
2. 作業対象物体へのフォーカスなし
3. 映像長 (Ours: 16分, YouCookII: 5分)

Model	FT	AG	SODA_c(↑)	CIDEr(↑)	METEOR(↑)
PDVC [49]	-	-	0.022	0.000	0.000
Vid2Seq [53]	-	-	0.017	0.066	0.010
Vid2Seq	✓	-	0.369	2.832	0.642
Vid2Seq	✓	RL	0.211	1.381	0.285
Vid2Seq	✓	AS	0.266	2.513	0.423
Vid2Seq	✓	RL+AS	0.581	6.195	1.142

Fine Tuning w/ COM Kitchens

Attention supervised by Visual Action Graphs (in two ways: RL&AS)

SOTA手法(3月当時)ではほぼ対応できず
(Visual Action Graphを学習に利用しても不十分)

低い自動評価指標のスコアはウェブ動画で事前学習された従来手法が有効でないことを示唆している

データセット利用の際は、下記文献を引用ください。

COM Kitchens:

An Unedited Overhead-view Video Dataset as a Vision-Language Benchmark

ECCV2024, Milan, Italy

Koki Maeda[*1,2], Tosho Hirasawa[*1,3], Atsushi Hashimoto[1],
Jun Harashima[4], Leszek Rybicki[4], Yusuke Fukasawa[4], Yoshitaka Ushiku[1]

* equally contributed
1. OMRON SINIC X Corp., 2. Tokyo Institute of Technology, 3. Tokyo Metropolitan Univ., 4. Cookpad Inc.

SINIC X cookpad