

動的グラフを用いた調理シーンの時空間解析

井阪正俊、伏見 卓恭 (東京工科大学コンピュータサイエンス学部)

概要

背景

近年、SNSに調理動画が活発に投稿されている。レシピ文による調理動画の検索が必要である。

目的

レシピ文を生成モデルで生成することで、レシピ検索の効率化を図りたい。

課題

多くの物体や動作が映っている調理動画では、レシピ文の生成が難しい。
また、動画処理は大量のリソースが必要である。

提案

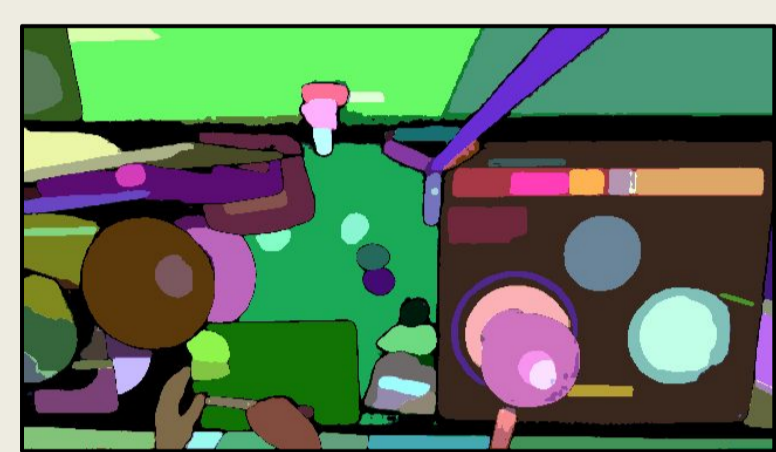
まずは、調理動画で行われている動作の分類を低リソースで行うことを目指す。

提案手法

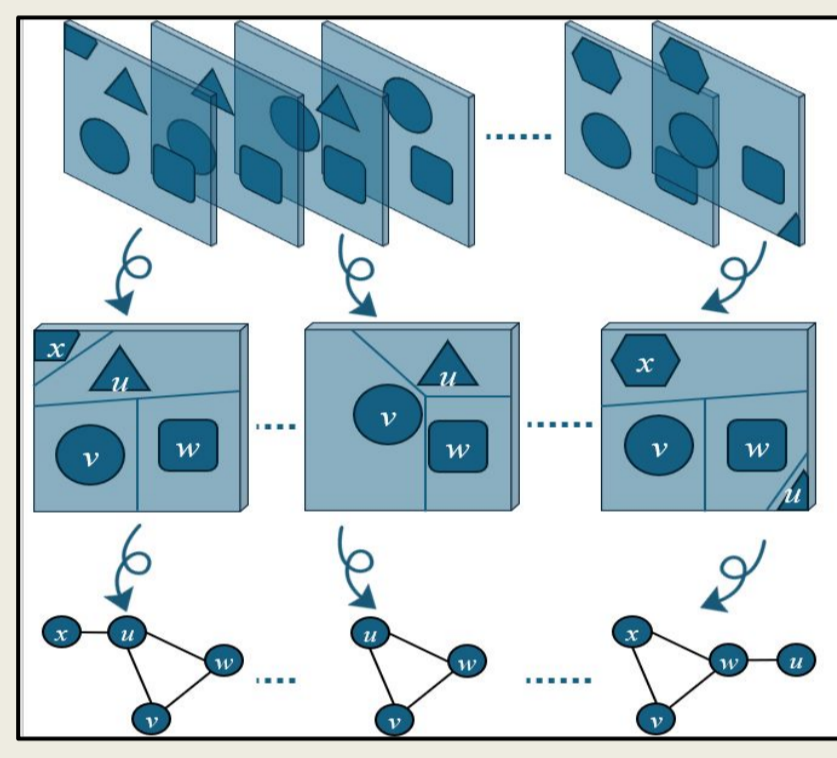
1. 動的グラフの作成

- 調理動画を30フレームごとにセグメンテーションする
- セグメンテーション領域をノードとしたグラフを作成する
- 隣接するフレームにおいて、同一なオブジェクトと判定されたノード同士を接続することで、動的グラフを作成する

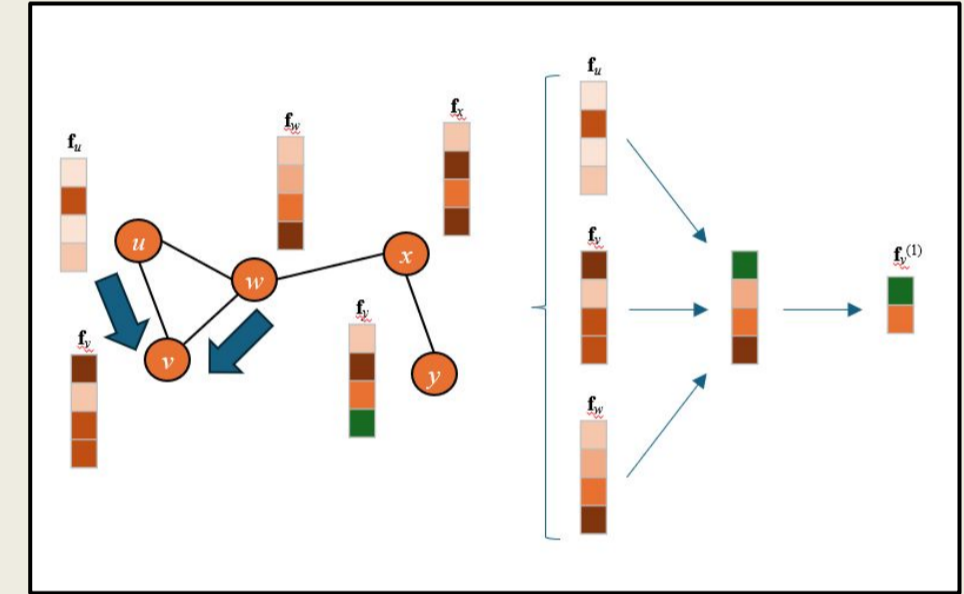
セグメンテーション後



作成手順



GCNの概要



2. 調理動作の分類

- 各フレームのノード間の関係をGCNで抽出する
- フレーム間で接続されたノード同士を時系列特徴としてTransformerに入力する
- 動画内の全特徴を平均して、全結合層で調理動作の分類を行う

実験設定

- 実験データ (COM-kitchensにおけるアクション動画)
 - 動画数: 2260件
 - フレーム数
 - 平均値: 1369, 最大値: 41136, 最小値: 6
 - 動作ラベルの分布
 - add(加える): 820, others(その他): 1440
 - 動的グラフにおける track_idの種類 (=ノード数)
 - 平均値: 81, 最大値: 239, 最小値: 39
- 比較手法
 - TimeSformerによるアクション分類
- 実験設定
 - 損失関数: CrossEntropy
 - 学習率: 3e-5
 - バッチサイズ: 16
 - エポック数: 40
- 評価指標
 - 各クラスにおける、適合率と再現率とF1値
 - 提案手法と比較手法における混同行列

結果と考察

提案手法における各クラスの混同行列

	予測:0	予測:1	合計(実際)
実際:0	114	50	164
実際:1	97	191	288
合計(予測)	211	244	452

考察

実験結果より、クラス0の再現率は提案手法の方が比較手法よりも0.19ほど高い。クラス0を正しく予測できたサンプル数は、提案手法の方が比較手法よりも30ほど高い。クラス1の適合率は、提案手法の方が比較手法よりも0.06ほど高い。以上のことから、提案手法は比較手法よりもクラス0の正しい予測が行え、クラス全体の予測を適切に行えたといえる。

比較手法における各クラスの混同行列

	予測:0	予測:1	合計(実際)
実際:0	84	80	164
実際:1	69	219	288
合計(予測)	153	299	452

適合率と再現率とF1値の表

クラス	クラス0(add)			クラス1(others)		
	適合率	再現率	F1値	適合率	再現率	F1値
提案手法	0.54	0.70	0.60	0.79	0.66	0.72
比較手法	0.55	0.51	0.53	0.73	0.76	0.75

今後の課題

- 二値分類におけるその他のクラスについて、より細かく階層化された分類を行う必要がある
- 物体の出現・消失や重なりに弱く、フレーム間で同一物体を安定して紐づけるのが難しい

謝辞

本研究では、国立情報学研究所のIDRデータセット提供サービスによりオムロンサイニックエクス株式会社から提供を受けた「OSX 調理映像データセット(COM-Kitchens)」を利用しました。心より感謝いたします。