

研究分野
Department複合知能メディア
Intelligent Media研究者
ResearcherZongshang Pang 中島悠太
Z. Pang Y. Nakashimaキーワード
Keyword動画、意味内容理解、検索、編集
Videos, Semantics, Retrieval, Editing応用分野
Application広告、エンターテインメント
Advertisement, Entertainment

研究開発段階

基礎

実用化準備

応用化

背景

動画共有サービスが一般化した現在、多くの人が動画の検索や閲覧、さらには動画自体の作成に関わる機会が増加しており、高度な動画の意味内容理解と、その結果を利用した効率的な動画の利活用が期待されています。

概要・特徴

大規模言語モデルや大規模マルチモーダルモデル等を利用し、セマンティクス(意味・意図・イベント)を中心に据えた動画理解。

技術内容

- **LLMベースの部分動画検索を再設計**: 検索対象の始まりと終わりの時刻を正確に予測しようとする既存手法に対して、提案手法は大規模言語モデルを活用して検索対象の意味に注目するアプローチを採用。

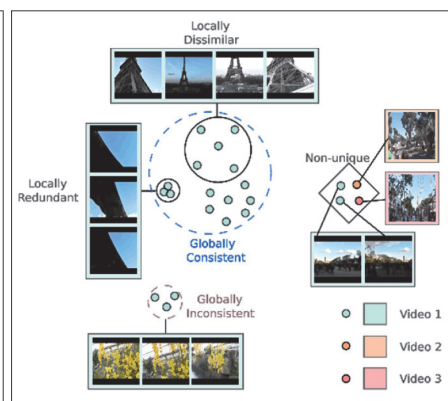
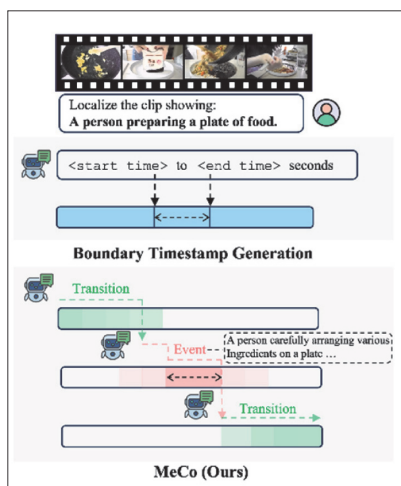
- **動画要約を対照学習の枠組みで直接定式化**: 真値から動画中の重要箇所を学習する既存手法に対して、動画内容の類似度やユニークさ等を教師なしの対照学習に自然に組み込むことができること、また結果として教師なしの動画要約を高精度化できることを示唆。

社会への影響・期待される効果

動画を見た目の動きではなく意図・イベント・要点として扱えるようにすることで、一般の長尺動画でも検索(いつ何が起きたか)や要約(どこが重要か)を低コストに高精度化しています。この成果は、例えば動画による政治広告のような場面では操作的編集の可視化によるメディア・リテラシー向上や公正な意思決定支援を促すなどの高度な動画内容理解の応用をはじめとして、広告・エンターテインメント・教育・アーカイブ運用・コンテンツ管理/モデレーション・業務効率化など幅広い領域で、動画の利活用を自動化、半自動化により後押しする効果が期待されます。

【論文 Paper】

[1] Z. Pang, M. Otani, Y. Nakashima, "Measure Twice, Cut Once: Grasping Video Structures and Event Semantics with LLMs for Video Temporal Localization," arXiv: 2503.09027, 2025. [2] H. Liu, Y. Nakashima, N. Babaguchi, "PALADIN: Understanding Video Intentions in Political Advertisement Videos," Proc. WACV 2025. [3] Z. Pang, Y. Nakashima, M. Otani, "Contrastive Losses Are Natural Criteria for Unsupervised Video Summarization," Proc. WACV 2023.



↑ 教師なしの動画要約手法の概要。
← LLMによる部分動画検索の既存手法(上)と提案手法(下)。