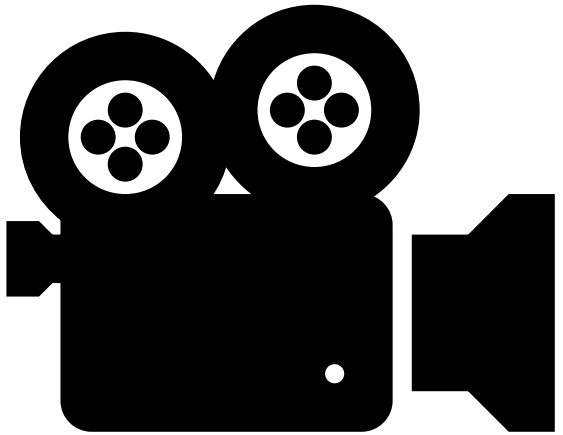
2021 ICCV OCTOBER 11-17 VIRTUAL

On the hidden treasure of dialog in video question answering



Deniz Engin, François Schnitzler,
Ngoc Q. K. Duong, Yannis Avrithis

interdigital

Inria

Introduction

Goal

- Knowledge-based video question answering on TV shows

Recent works

- Rely on human-generated external sources

Our approach

- Substituting human-generated sources by automatically generated knowledge

Motivation

Episode

Input scene in the episode A

Leonard: Come on. Is that really necessary?
Sheldon: Leonard, I believe it is. This is trash talk. Trash talk is a traditional component in all sporting events.

Scene elsewhere in the episode A

Kripke: Word around the plasma lab is you built a robot?
Leonard: Yes, we did, Kripke.
Sheldon: His name is Monte.

What did the guys name their robot?

A) Killer Robot

B) Terminator

C) Monte

D) Crippler

Solution: Dialog Summarization

Input scene in the episode A

Leonard: Come on. Is that really necessary?
Sheldon: Leonard, I believe it is. This is trash talk. Trash talk is a traditional component in all sporting events.

Kripke is going to name his robot Scrap Metal. Sheldon and Leonard are going to defeat Kripke's robot because theirs is better in design and execution.

Scene elsewhere in the episode A

Kripke: Word around the plasma lab is you built a robot?
Leonard: Yes, we did, Kripke.
Sheldon: **His name is Monte.**

Leonard and Raj have built a robot called Monte.
Kripke is going to enter him in the Southern California Robot Fighting League Round Robin Invitational.

Episode A

... Leonard and Raj have **built a robot called Monte.**
Kripke is going to enter him in the Southern California Robot Fighting League Round Robin Invitational. ...

Episode A – Scene 1

?

What did the guys name their robot?

A) Killer Robot B) Terminator
C) Monte D) Crippler

Inputs sources of the proposed method

Video Description

Rely on character and action recognition, place classification, object relation detection
ROLL [Garcia et al., ECCV 2020]

Dialog

Dialog Summarization

Scene Dialog Summary

Concatenating
scene dialog summaries

Episode Dialog Summary

The overview of the proposed method

Video Description

Scene Dialog Summary

Episode Dialog Summary

Video Stream

QA

Scene Dialog Summary Stream

QA

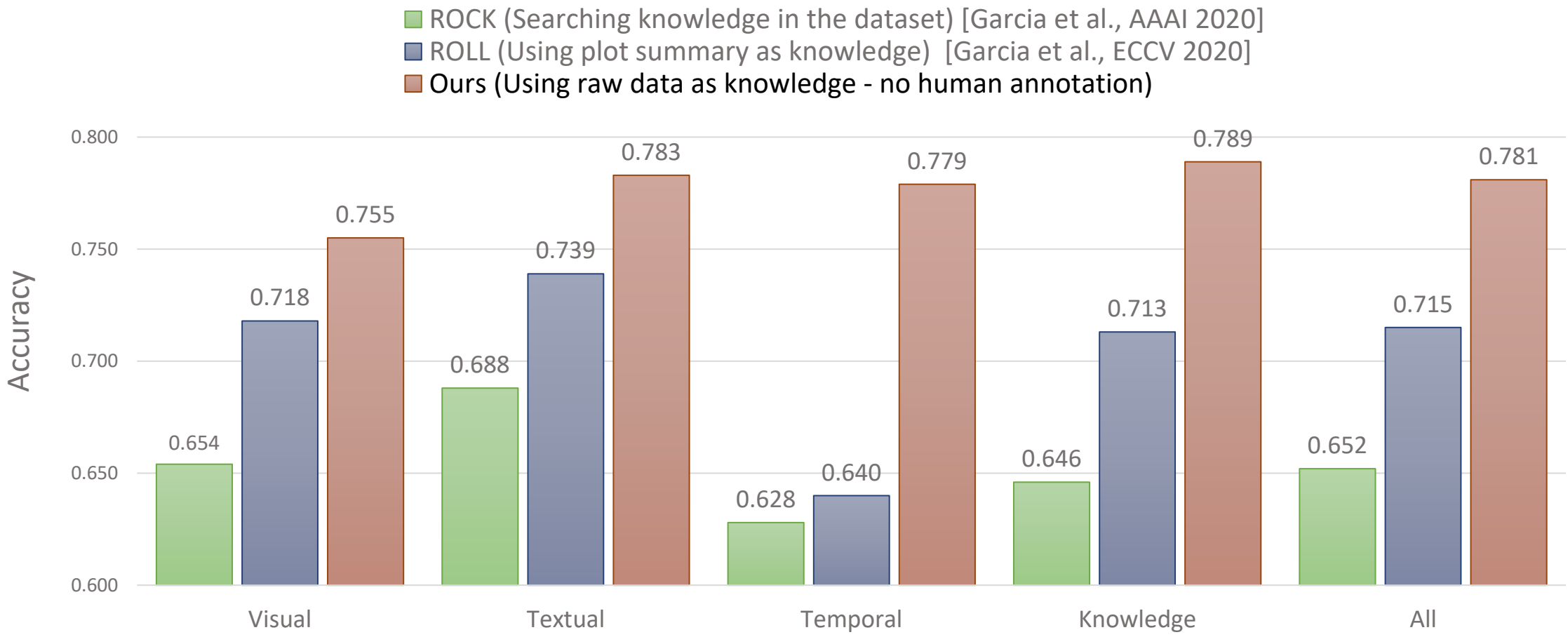
Episode Dialog Summary Stream

QA

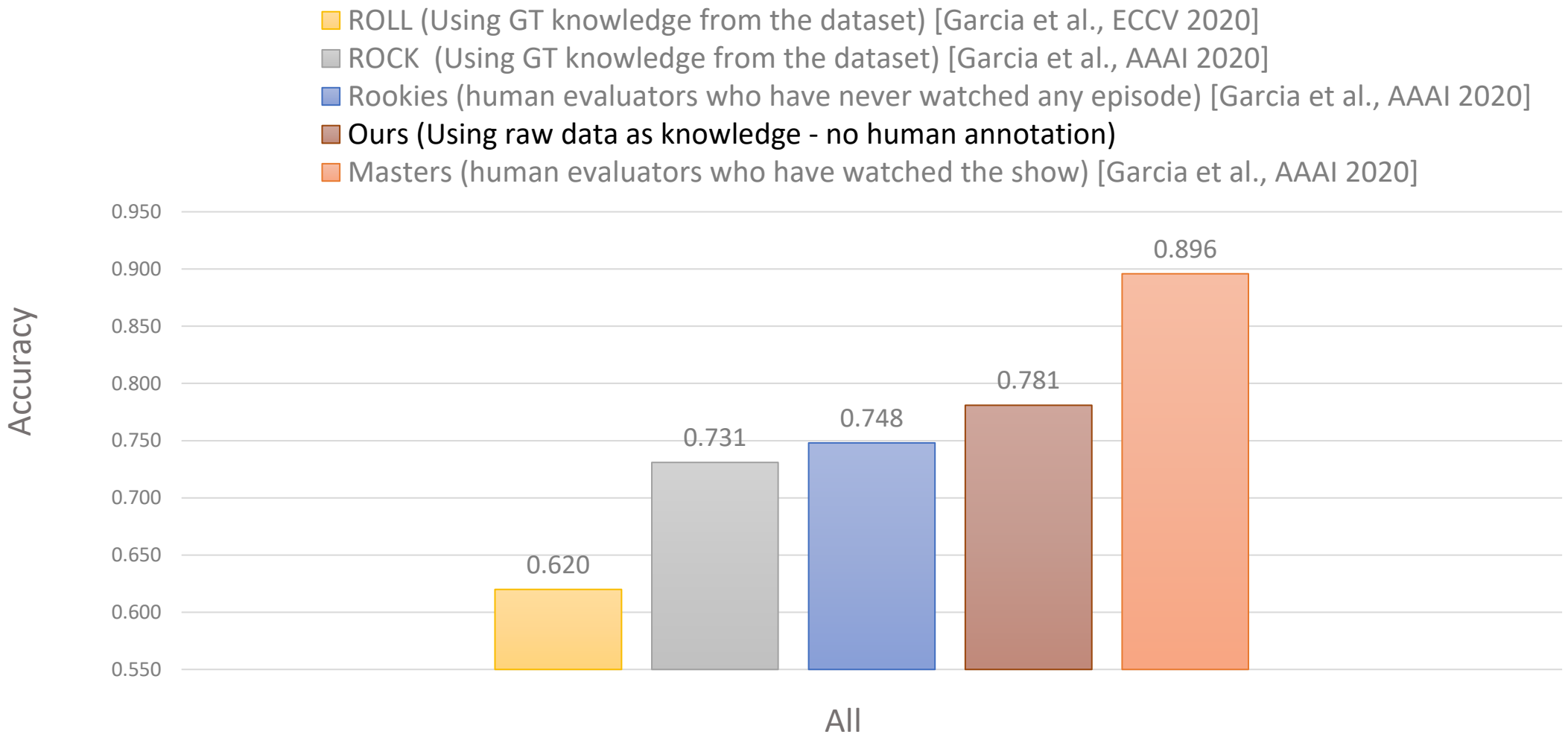
Fusion

Answer

Our method outperforms state-of-the-art on KnowIT VQA



Our method outperforms rookie human evaluators

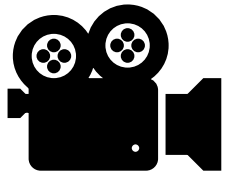


Contributions

- Building a knowledge-base VideoQA system without extra human annotation
- Applying dialog summarization to VideoQA
- Introducing a weakly-supervised soft temporal attention approach for localization
- Proposing a simple multimodal fusion mechanism

2021 ICCV OCTOBER 11-17 VIRTUAL

On the hidden treasure of dialog in video question answering



Deniz Engin, François Schnitzler,
Ngoc Q. K. Duong, Yannis Avrithis

Project Page

interdigital

Inria