



DCU



DCU ADAPT at TRECVID 2022

Deep Video Understanding Challenge

A. Potyagalova, Gareth J. F. Jones
Dublin City University, Ireland

Overview

- Introduction
- Approach
- Scene-level track
- Movie-level track
- Results and Conclusions
- Questions

The background features two large, overlapping geometric shapes. On the left, a teal-colored shape extends from the top-left corner towards the bottom-right. On the right, a light gray shape extends from the top-right corner towards the bottom-left. The two shapes meet at a diagonal line that divides the page. The word "Introduction" is centered in the white space between these shapes.

Introduction

Introduction

▶ Approach

- ▶ Scene-level track solution includes recognition of actions and looking for the closest term.
- ▶ Movie-level track approach focused on building a knowledge graph based on the full-text description of the movie.

▶ **Results** Achieved reasonably good results (11.00% accuracy) for the scene-level track, but results for the movie-level track were not good enough for final submission.

▶ **Ways to improve** Process the knowledge graph for movies to clarify the relations between entities; perform a more detailed analysis of the timeline of the actions for video segments

The background features two large, overlapping geometric shapes. On the left, a teal-colored shape extends from the top-left corner towards the bottom-right. On the right, a light gray shape extends from the top-right corner towards the bottom-left. The two shapes meet at a diagonal line that divides the page. The word "Approach" is centered in the white space between these two shapes.

Approach

Approach

General scheme

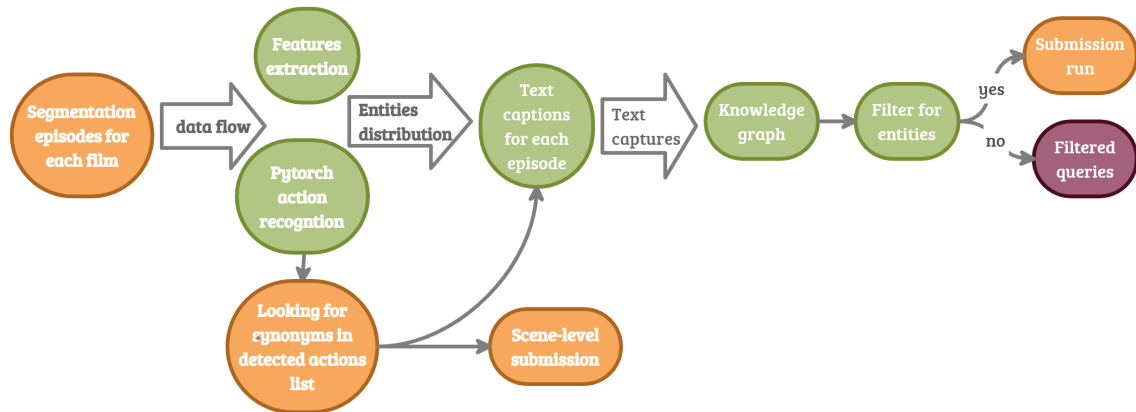


Figure 1: General workflow of the proposed approach

Approach

General scheme

Algorithm for movie processing:

- ▶ Recognize the actions in the short video episodes
- ▶ Detect the entities in the movie segments using the reference images
- ▶ Generate the text capture for segments by combining the results of previous stages
- ▶ Create the piece of text describing the whole movie; forms the base of the knowledge graph
- ▶ Build the knowledge graph for the whole movie

Approach

Scene-level

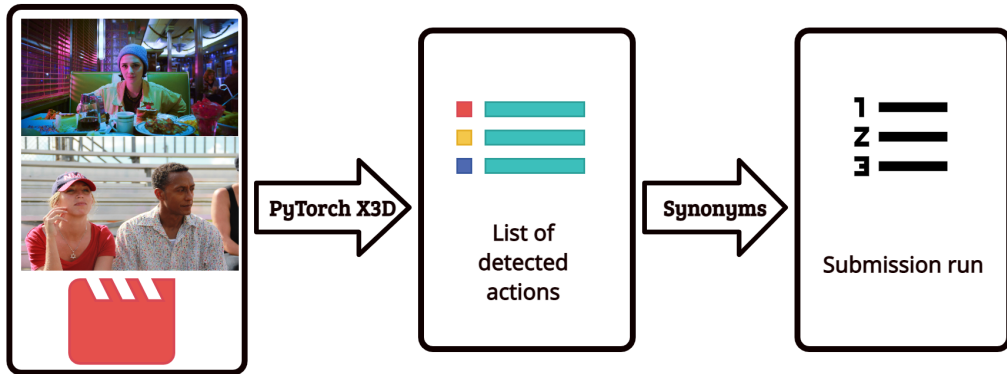


Figure 2: General description of scene-level solution

Approach

Movie-level



Figure 3: General description of movie-level solution

The background features two large, overlapping geometric shapes. On the left, a teal-colored shape starts from the top-left corner and extends diagonally towards the bottom-right. On the right, a light gray shape starts from the top-right corner and extends diagonally towards the bottom-left. The two shapes meet in the center, creating a white triangular area where the text is located.

Scene-level track

Scene-level track

Action recognition

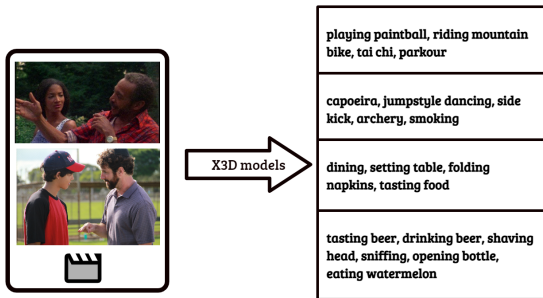


Figure 4: Recognition of actions for video segments

- ▶ Process movie segments
- ▶ Create several variants of lists using different X3D models

Scene-level track

Search for the synonyms

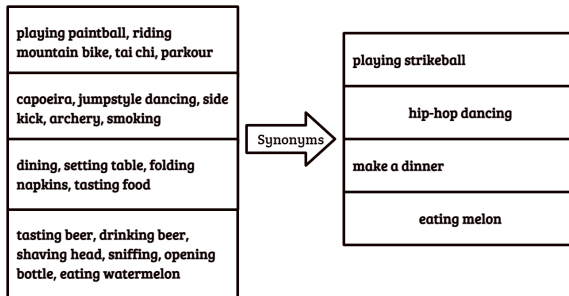


Figure 5: Search for coincidences and synonyms

- ▶ Process lists using the Gensim library
- ▶ Submit detected coincidences or synonyms

The background features two large, overlapping geometric shapes. On the left, a teal-colored shape extends from the bottom-left corner towards the center. On the right, a light gray shape extends from the bottom-right corner towards the center. The two shapes meet at a diagonal line that divides the page. The text 'Movie-level track' is centered in the white space between these two shapes.

Movie-level track

Movie-level track

Caption generation

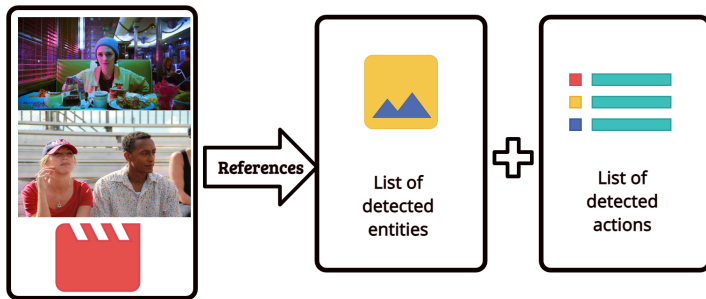


Figure 6: Building the capture using the entities and previously recognized actions

- ▶ Detect the required entities in the movie segments
- ▶ Using the results from the scene-level actions lists, create the text caption for each movie episode

Movie-level track

Knowledge graph

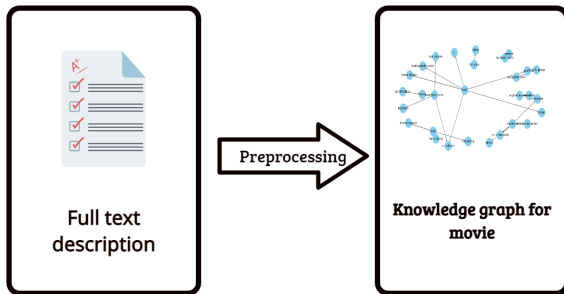


Figure 7: Building the knowledge graph based on the text descriptions

- ▶ Preprocess the text descriptions
- ▶ Build the knowledge graph
- ▶ Preprocess the knowledge graph to clarify relations between entities (unfinished)

The slide features a white background with two large, overlapping geometric shapes. On the left, a teal triangle points downwards and to the right. On the right, a light gray triangle points upwards and to the left. The two triangles meet at a diagonal line that divides the slide.

Results and Conclusions

Results and Conclusions

Results and scores

Table 1: TRECVID-2022 results

Scene-level task results		
Team	Points	Percentage
Columbia_1	11.35	15.8%
Adapt	11.00	15.3%
Columbia_2	8.35	11.6%
WHU_NERCMS_1	8.00	11.1%
WHU_NERCMS_2	2.25	3.1%

Results and Conclusions

Results and scores

Table 2: Detailed scene-level results

Scene-level results for each movie		
Movie	Points	Average
Calloused hands	1.00	0.083
Chained for life	1.00	0.083
Liberty kid	3.00	0.25
Like me	3.00	0.25
Little rock	0.0	0.0
Losing ground	3.00	0.25

Results and Conclusions

Future plans

- ▶ **Text descriptions** Prepare more accurate and detailed text descriptions
- ▶ **Action recognition** Create the timeline for the episodes with accurate correspondence between action and timestamp
- ▶ **Knowledge graph** Add the extra parameters to the text parsing algorithm

The background features two large, overlapping geometric shapes. On the left, a teal-colored shape extends from the bottom-left corner towards the center. On the right, a light gray shape extends from the bottom-right corner towards the center. The two shapes meet at a diagonal line that divides the page. The word "Questions" is centered in the white space between these two shapes.

Questions