

NII-UIT at TRECVID 2020: Video Summarization Task

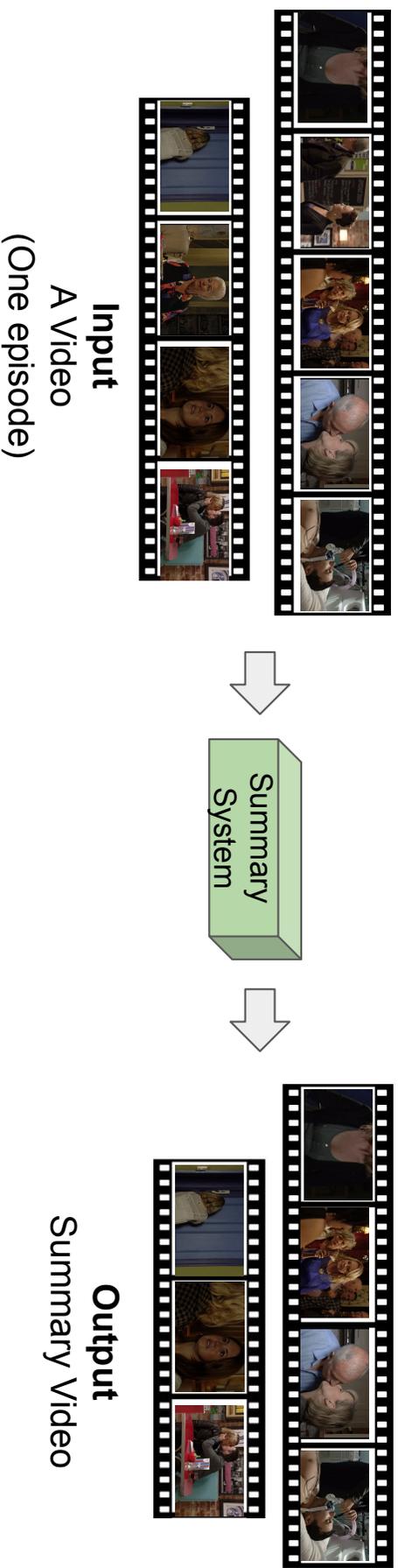
Dung-Minh Nguyen^{1,2}, Hung-Quoc Vo^{1,2}, Tien-Van Do^{1,2}, Thinh-Le-Gia Pham^{1,2},
Tri-Le-Minh Vo^{1,2}, Thua-Ngoc Nguyen^{1,2}, Vinh-Tiep Nguyen^{1,2}, Thanh-Duc Ngo^{1,2},
Duy-Dinh Le^{1,2}, Zheng Wang³ and Shin'ichi Satoh³

¹University of Information Technology, Ho Chi Minh City, Vietnam

²Vietnam National University, Ho Chi Minh City, Vietnam

³National Institute of Informatics, Tokyo, Japan

Introduction - Video Summarization Task

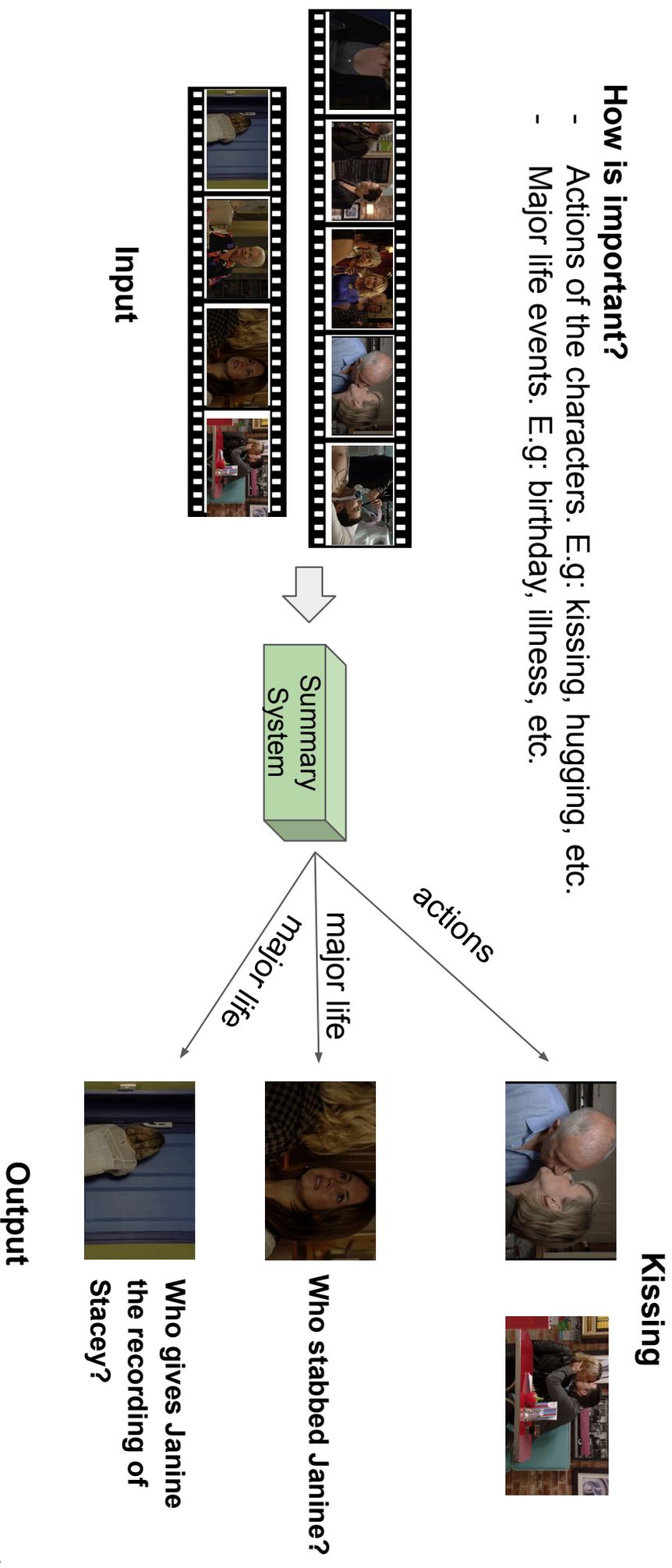


- create a short skim of the original video
- keep the most **important content**

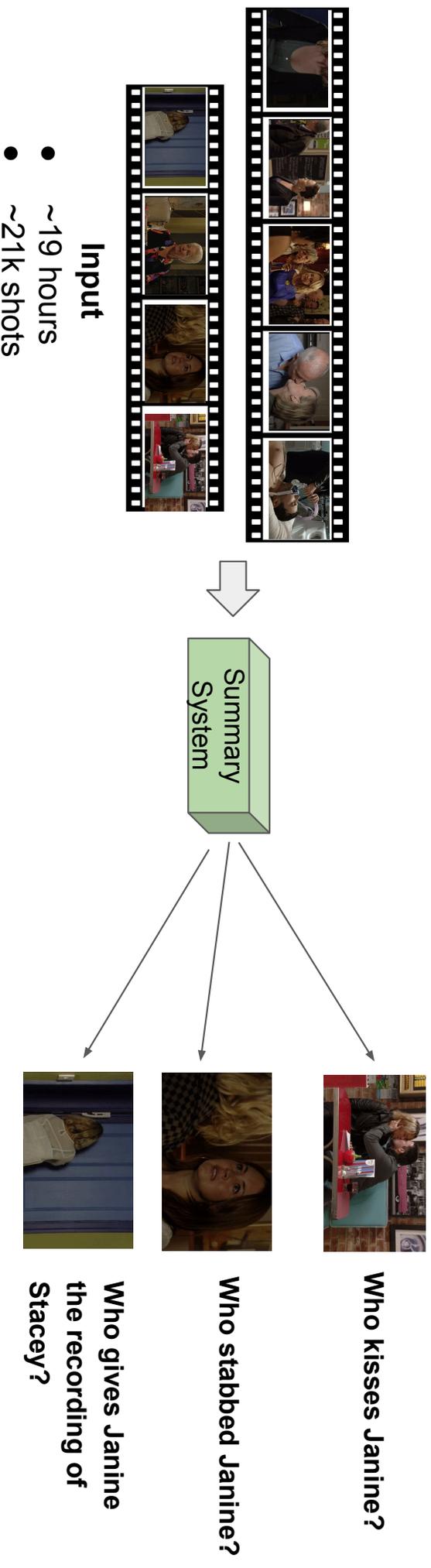
Introduction - Video Summarization Task

How is important?

- Actions of the characters. E.g: kissing, hugging, etc.
- Major life events. E.g: birthday, illness, etc.



Introduction - TRECVID 2020: VSUM Task



- Ratio: 5-20/21K shots → 0.02% - 0.10%
- Summary **major life event** of each character in input video

Introduction - TRECVID 2020: VSUM Task

- #Videos: 11
- #Shots: 21,212
- #Frames: 1,714,091
- Duration : 68,563.64 (~ 19 hrs)
- 3 People



Janine

Ryan

Stacey

Character	Janine	Ryan	Stacey
Start Shot #	shot175_1	shot175_1	shot175_1
End Shot #	shot185_1736	shot185_1736	shot185_1736
Images	Images	Images	Images
Max # Shots Run 1	5	5	5
Max Summary Length Run 1	150 seconds	150 seconds	150 seconds
Max # Shots Run 2	10	10	10
Max Summary Length Run 2	300 seconds	300 seconds	300 seconds
Max # Shots Run 3	15	15	15
Max Summary Length Run 3	450 seconds	450 seconds	450 seconds
Max # Shots Run 4	20	20	20
Max Summary Length Run 4	600 seconds	600 seconds	600 seconds

Specifies of VSUM task 2020

Challenges

- The ratio of summary is *too short* (0.02-0.1%)
- How to define what importance is? → *we do not know the queries in advance*



Our summary for Ryan



GT for the Ryan query 1:
**How does Janine attempt to kill
Ryan in the hospital?**

VSUM Queries

- 6 **WHAT**-questions: About actions of characters do
- 1 **HOW**-question
 - How does Janine attempt to kill Ryan while in the hospital? (Both Janine and Ryan)
- 6 **WHO**-questions: The person regarding the characters
- 1 **WHERE**-question:
 - Where is Ryan trapped?
- Some query regarding 2 characters:
 - Who gives Janine the recording of Stacey?

→ *We need to find out the person giving the recording of Stacey for Janine. How do we know whose the recording? - Difficult*

Janine Query



- Q1 - What is causing Ryan to be sick in bed?
- Q2 - How does Janine attempt to kill Ryan while in the hospital?
- Q3 - What happens when Janine attempts to play recording of Stacey?
- Q4 - Who stabbed Janine?
- Q5 - Who gives Janine the recording of Stacey?

Ryan Query



Q1 - How does Janine attempt to kill Ryan in the hospital?

Q2 - What does Ryan do when Janine is lying in the hospital?

Q3 - Where is Ryan trapped?

Q4 - What does Ryan tell Phil he can do for him?

Q5 - Who is Ryan with when going to put his name on the babies birth cert?

Stacey Query



- Q1 - Who climbs up the roof to talk Stacey out of jumping off?
- Q2 - What does Stacey reveal when in a cell with Janine, Kat, and Pat?
- Q3 - What does Stacey admit to her mum in bedroom when mum is upset?
- Q4 - Who confronts Stacey in restroom where Stacey finally admits to killing Archie?
- Q5 - Who calls to Stacey's door to tell her to get her stuff and go after Stacey's mum had called the police?

Groundtruth of Janine Query

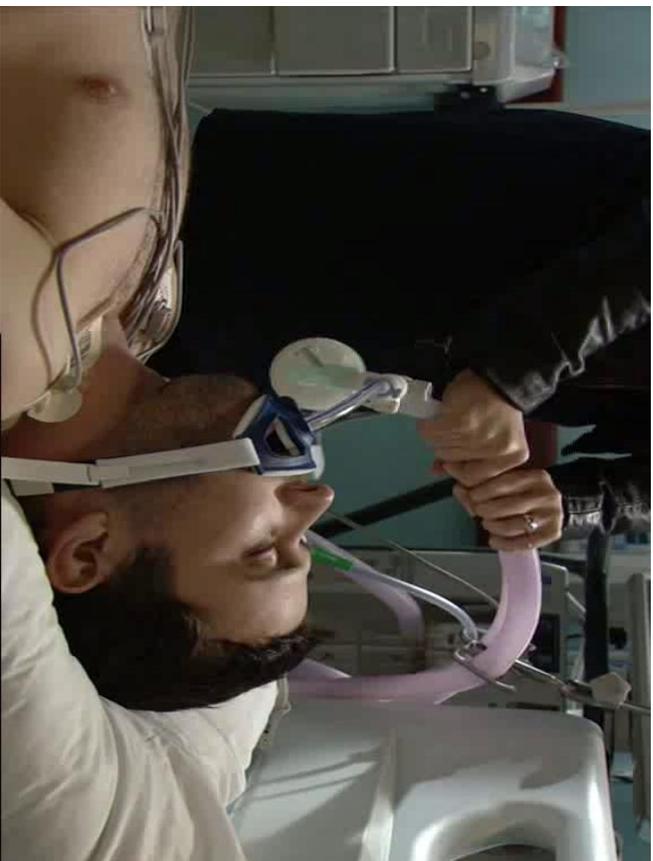
Q1 - What is causing Ryan to be sick in bed?



Janine gave Ryan a poison drink

Groundtruth of Janine Query

Q2 - How does Janine attempt to kill Ryan while in the hospital?



She tried to pull out his breathing tube

Groundtruth of Janine Query

Q3 - What happens when Janine attempts to play recording of Stacey?



Janine played the wrong recording

Groundtruth of Janine Query

Q4 - Who stabbed Janine?



Stacey stabbed Janine



Major life event of Janine but not that of Stacey.

Groundtruth of Janine Query

Q5 - Who gives Janine the recording of Stacey?



The girl

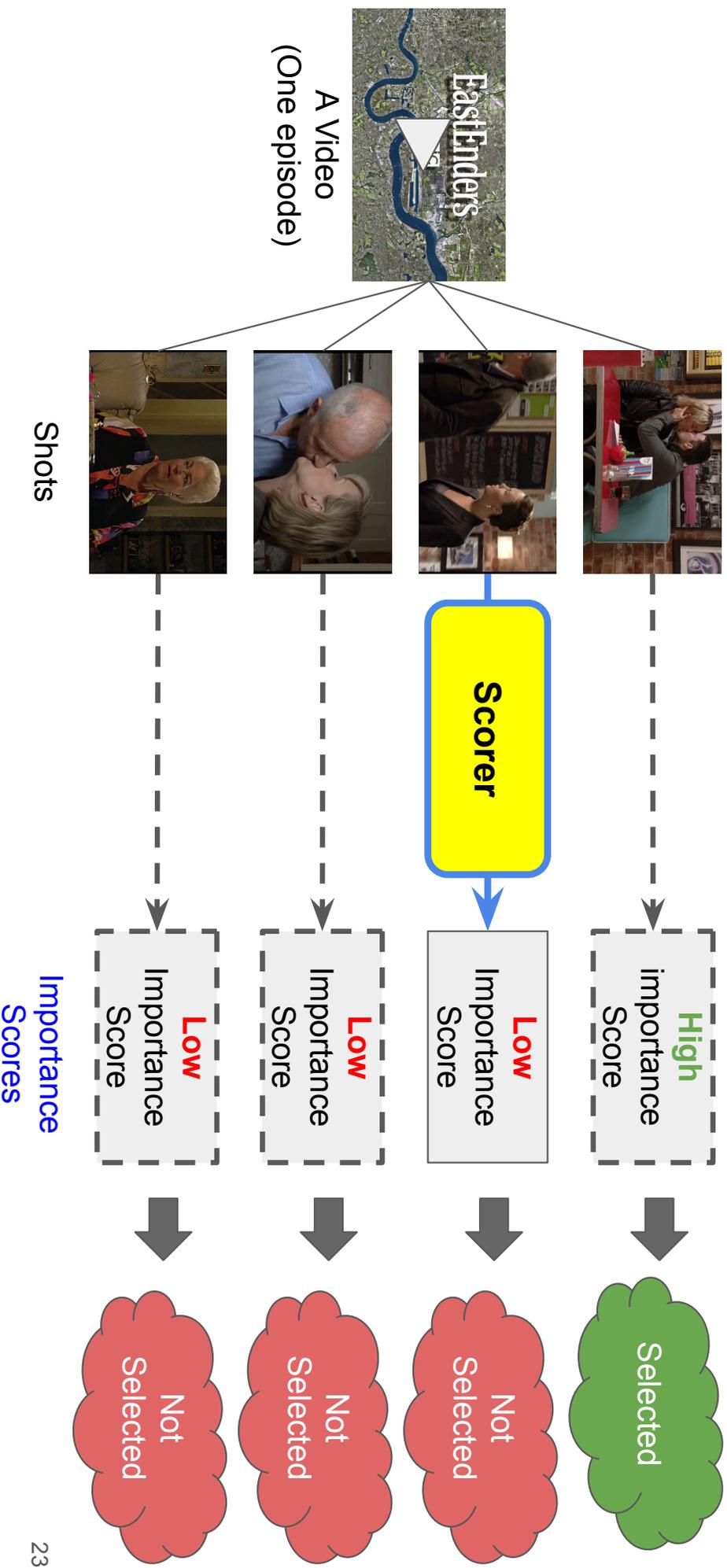
System Overview

The most recent approaches include 3 phases:

- Segmentation
- Score
- Selection

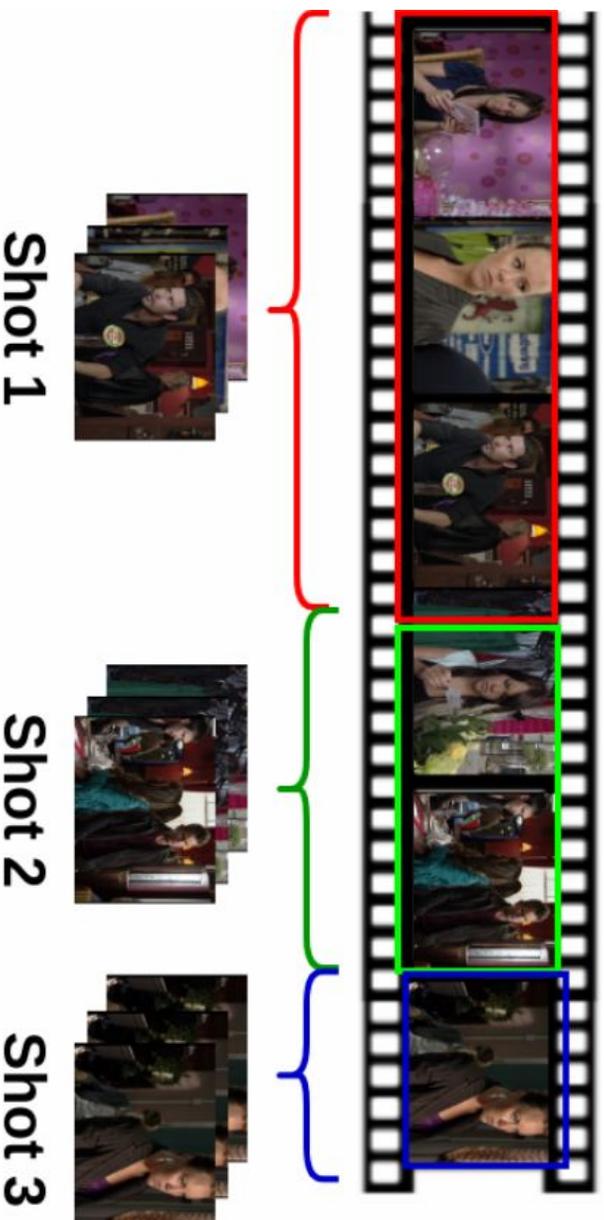


Proposed System - NII-UIT VSUM 2020 Pipeline



Proposed System - Segmentation Module

- Keyframes are extracted at *5fps*
- BBC EastEnders shot



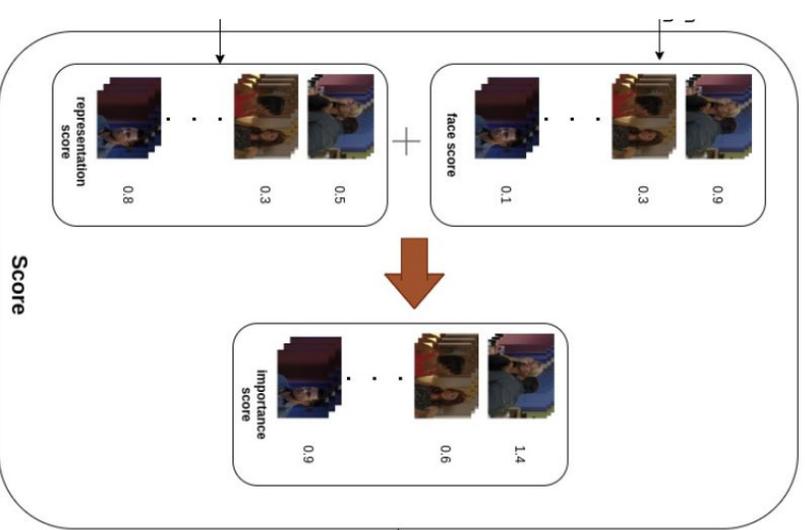
Proposed System - Scoring Module

$$\text{importance_score} = w_1 * \text{person_score} + w_2 * \text{representation_score}$$

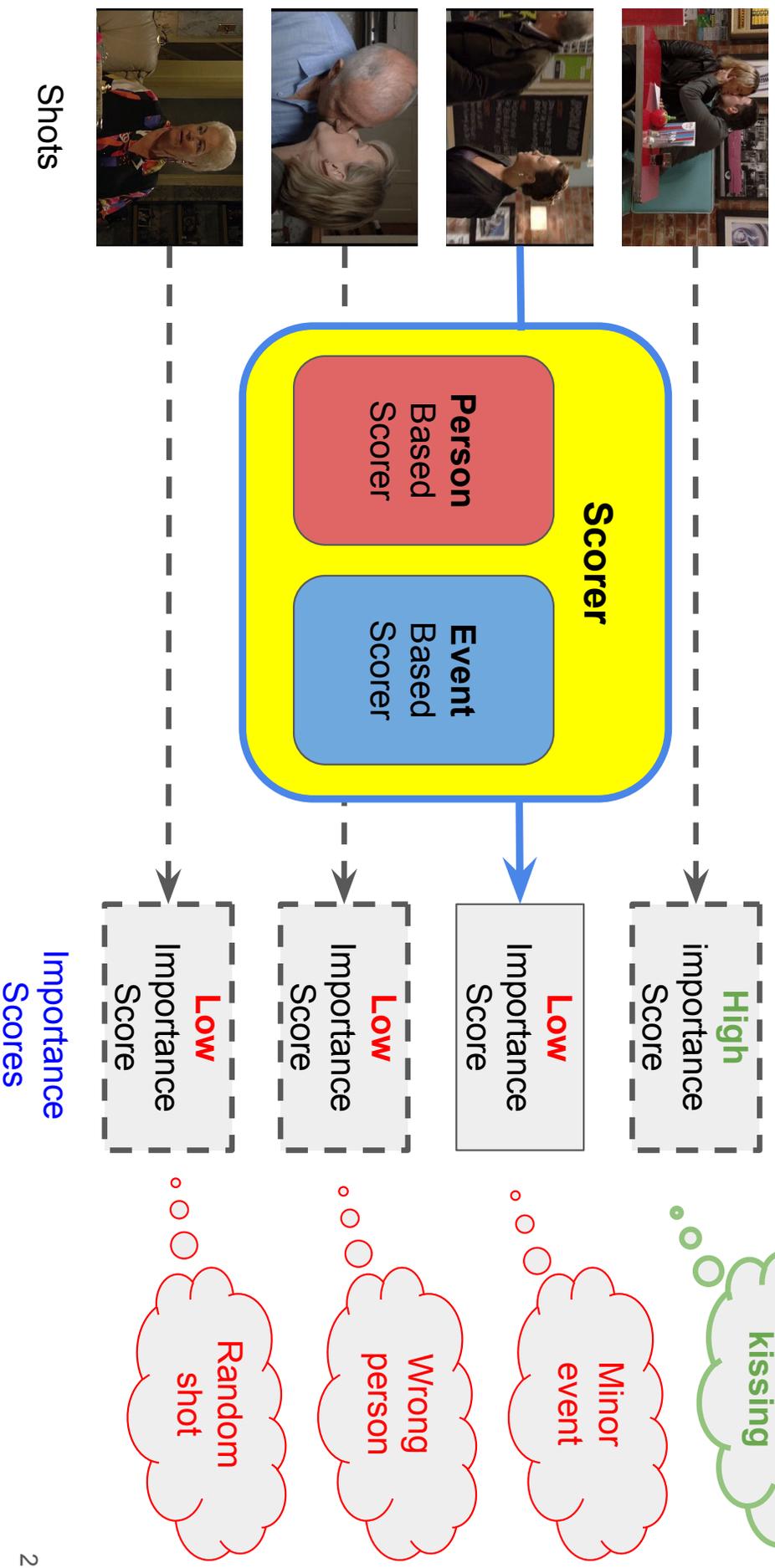
person_score: To get the desired character.

representation_score: To get the major life events of character.

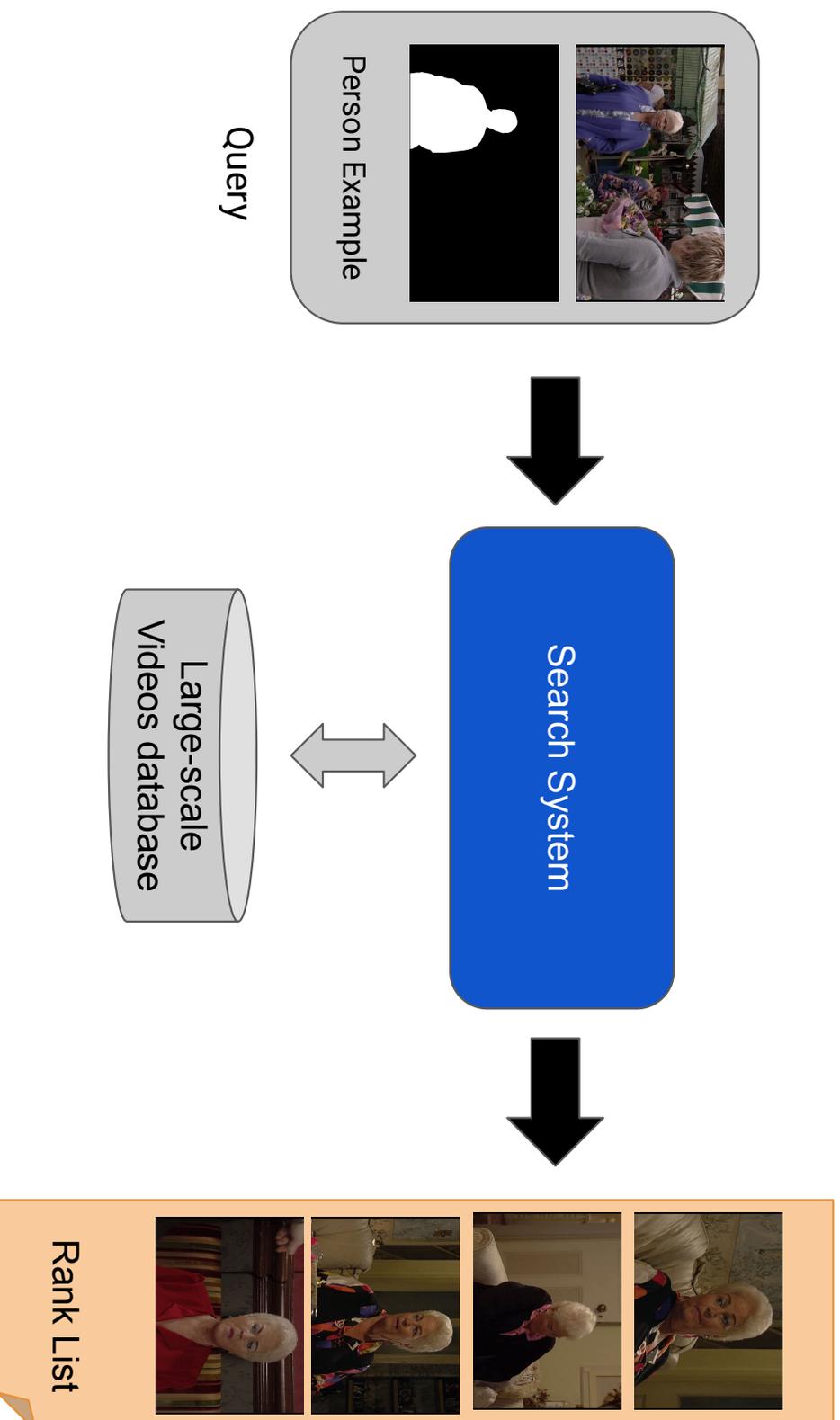
w1, w2: weight of two scores



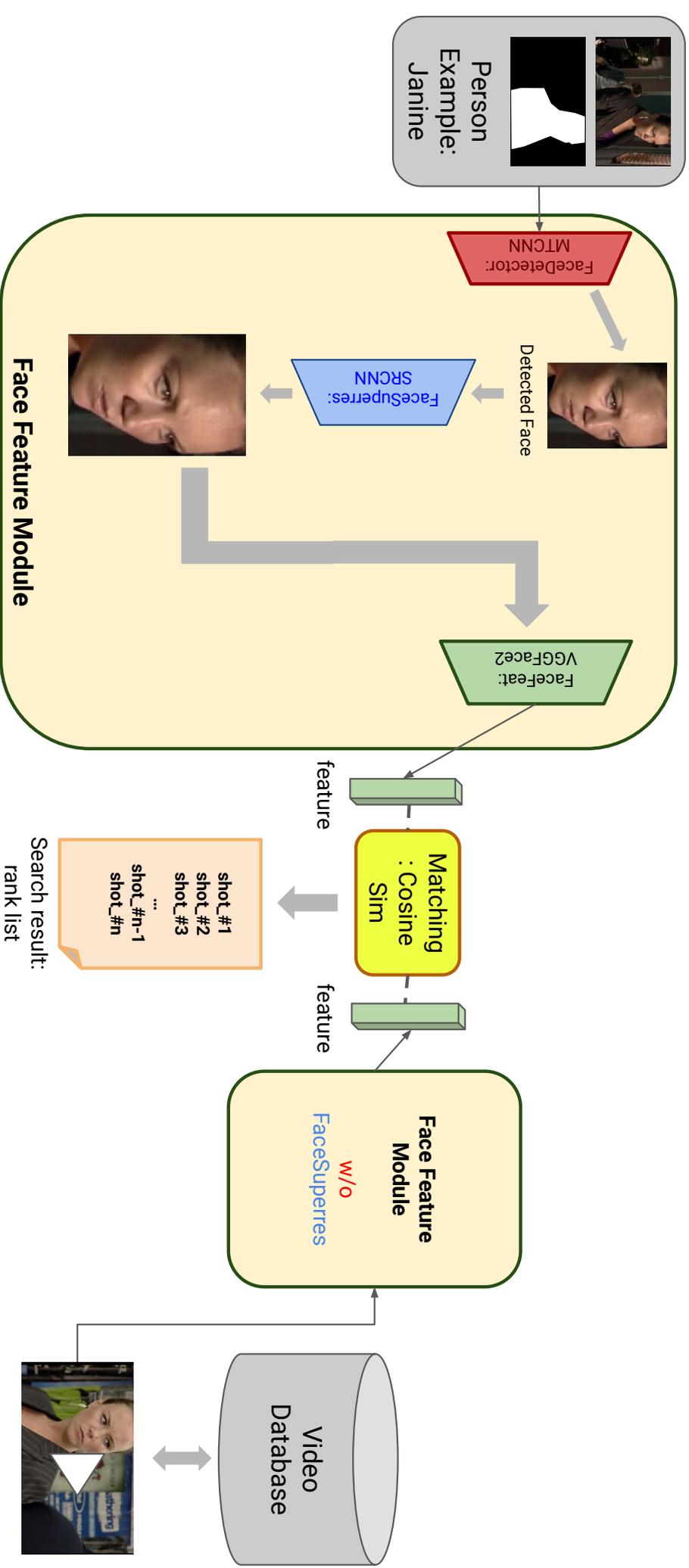
Proposed System - Scoring Module



Proposed System - Scoring Module (Person)

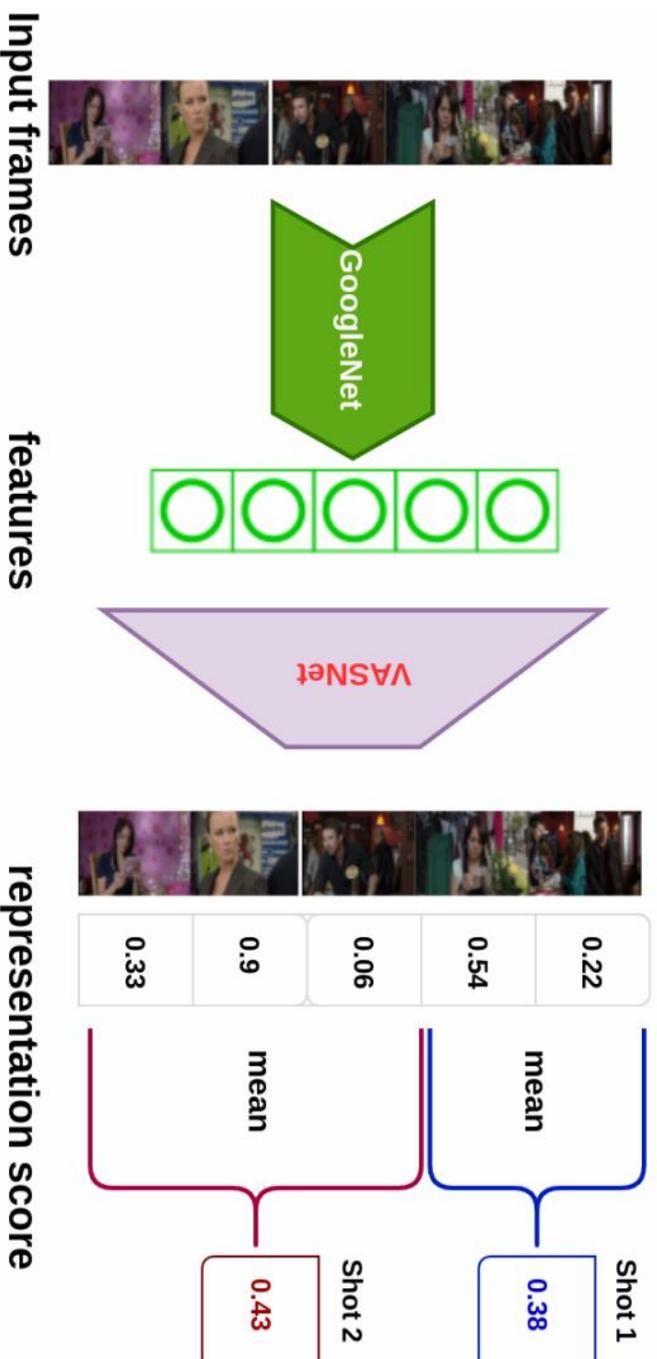


Proposed System - Scoring Module (Person)



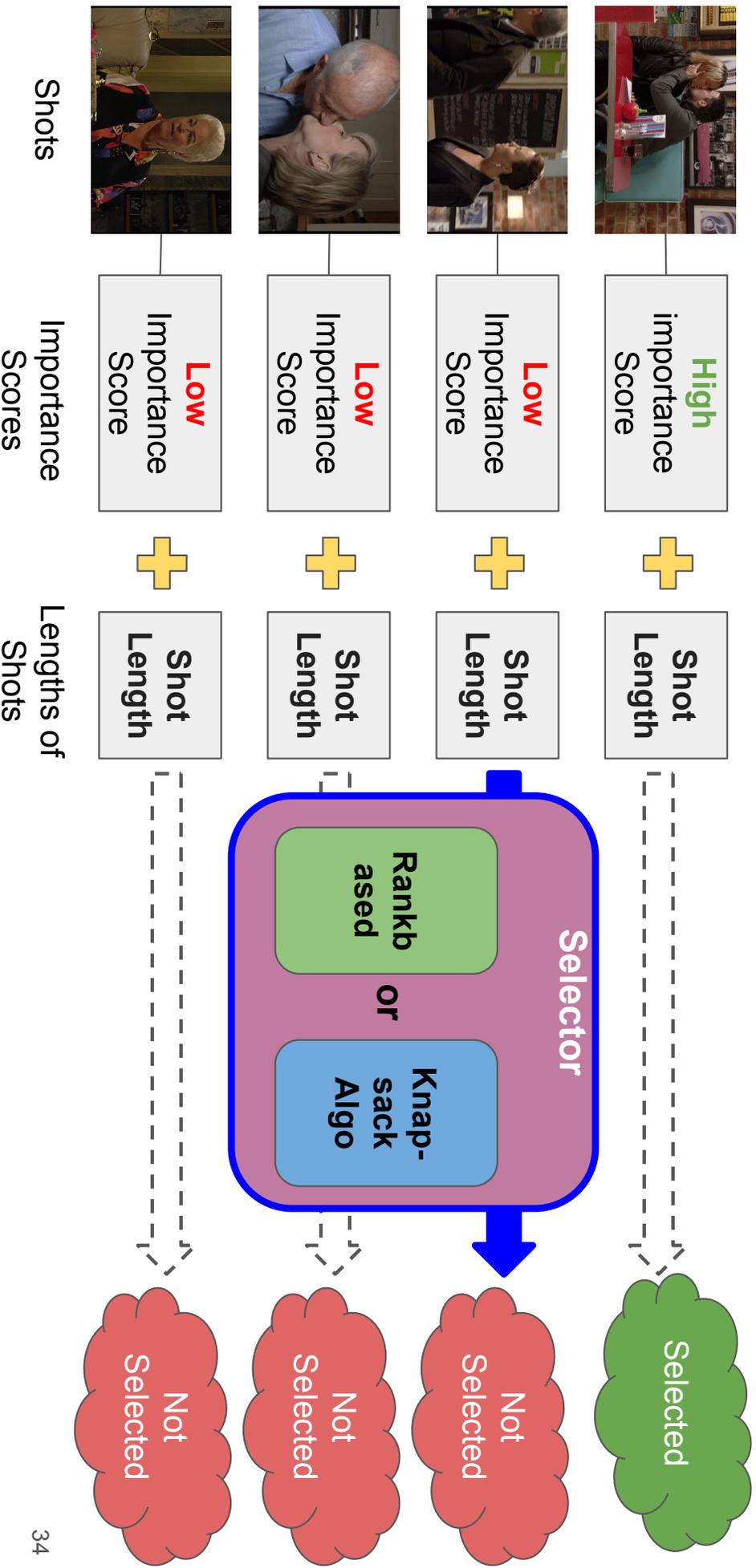
Proposed System - Scoring Module (Representation)

- VASNet[1] is an approach to sequence to sequence transformation for video summarization based on soft, self-attention mechanism.
→ Using VASNet to calculate the importance score.



[1] Fajtl, et al. Summarizing videos with attention. In ACCV2018 AIU2018 workshop

Proposed System - Selection Module



Proposed System - Selection Module

Why using Knapsack algorithm for selection?

→ Recently common approaches[1,2] for video summarization is using Knapsack algorithm for selecting shot in the summary and get high performances.

[1] Fajtl, et al. Summarizing videos with attention. In ACCV2018 AIU2018 workshop

[2] Kaiyang Zhou, Yu Qiao, Tao Xiang. Deep Reinforcement Learning for Unsupervised Video Summarization with Diversity-Representativeness Reward. In AAAI 2018

Proposed System - Selection Module (Knapsack)



Shot 1

- Score: 0.25
- Length: 2



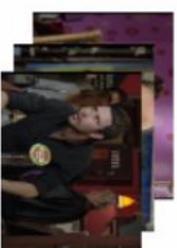
Shot 2

- Score: 0.4
- Length: 55

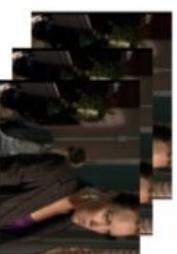


Shot 3

- Score: 0.3
- Length: 45



Shot 1



Shot 3

Selected shots



Final Summary

Proposed System - Selection Module (Ranking)

- Ranking method:
 - Filter 1000 shots with highest person score
 - Take n shots with highest VASNet score for summary
 - Concatenate selected shot into summary final.
- To compare two selection methods:
 - Apply Ranking for Run 1&3
 - Apply Knapsack for Run 2&4

Experimental Result

Criteria for evaluation:

- Content of queries
- Tempo
- Contextuality
- Redundancy

Team_Run	Percentage
NII_UT_1	9%
NII_UT_2	8%
NII_UT_3	8%
NII_UT_4	6%

- → Do not capture any major life events of characters.

Experimental Result

Q5: Who gives Janine the recording of Stacey?



Run1: Janine



Query 5 GT

- Our submit usually focuses on shots having **one** character. While many ground-truth shots could contain **more than one** character.
- Query 5 is actions giving a recording.
→ Our importance definition does not work

Experimental Result

Knapsack



Ranking



The tempo, contextuality, and redundancy of generated video summaries by Ranking is better than that by Knapsack.

→ The summary using Knapsack is usually only composed of short segments

Conclusion

- Proposed a simple approach to generate the video summary for VSUM task.
 - Using INS system to find out shots of character.
 - Using VASNet to get the importance shots.
 - Finally, create the summary video with Knapsack or Ranking.
- Our system failed to take major life events because:
 - VASNet network focuses on visual information and gets highlight and obvious main person in shots. While the main person in shots having major life events is not usually highlight.