AI-Based Soccer Game Summarization: From Video Highlights to Dynamic Text Summaries

Sushant Gautam (MSISE020)

In collaboration with **simulamet**, Norway

Supervisor:
Er. Dinesh Baniya Kshatri

Department of Electronics and Computer Engineering

Institute of Engineering, Thapathali Campus
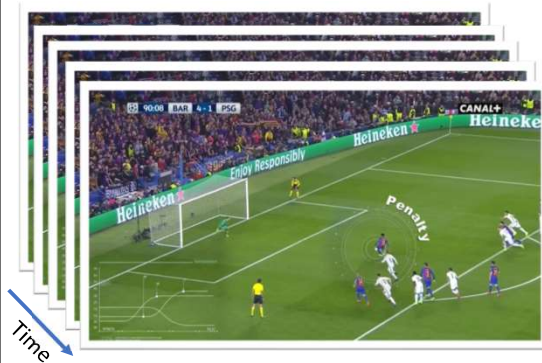
September 18, 2022

1

Presentation Outline

2

2

Motivation

Time

• Soccer: popular sport

- **One-line(Tweet):**

Aalesund lost to Mjøndalen 1-2.

- **Multi-lines(Video Description):**

The game ended with a score of 1-2. Aalesund was awarded a penalty in the second half, which they scored to level the game...

- **News article:**

Aalesund and Mjøndalen played to a 1-2 score in a Norwegian Eliteserien matchup on Sunday. Aalesund took an early lead in the game, but Mjøndalen responded with two goals to take a lead. Aalesund then scored on a penalty kick in the second half to even the score...

How to obtain updates on missed matches?

9/17/2022

3

3

Background

• A **soccer broadcast pipeline** can consist of

- Event detection and classification

- Highlight clip generation, thumbnail selection

- Statistics generation

• Overall **game summarization**:

- Last stage of the pipeline

- Video or text summary of the game

- Subjective nature of viewer expectations result in multiple interpretations

- Research in this area is limited (esp. text summaries)

9/17/2022

4

4

Problem Statement

- **Game summarization** in soccer broadcast pipeline

- Currently dominated by **manual** efforts

- No fully **automated** dynamic end-to-end system available

- Emphasize **automation, configurability** and **multimodal input**

5

Objectives of Thesis

- To design **automated summarization pipeline**
  - *Dynamic:* variable length summary texts
  - *Multimodal input:* game audio, commentary and text metadata
- To **validate** the pipeline
  - *Objective:* system performance and real-time operation
  - *Subjective:* user studies, feedback from consumers of summaries

6

Scope of Thesis

- **Capabilities**:

  - English language text summaries for association football (soccer)

  - User defined constraints (length, etc)

- **Limitations**:

  - Only rely on existing multimedia from game

  - Only rely on existing metadata (not create own annotations)

  - Only text summaries

7

Originality of Thesis

- **Automated** end-to-end pipeline

  - Use state-of-the-art ML algorithms (e.g., transformer based few-shot learning)

- **Dynamic** configuration

  - Variable length output

- **Multimodal** input: both audio and text metadata as input

- New **dataset**

  - HOST dataset along with the existing public SoccerNet dataset

  - includes soccer game multimedia and metadata

8

Applications

- **Sports Companies**
  - Integrate pipeline in commercial broadcast soccer production pipeline
- **Journalism**
  - Use pipeline for automatic sports news generation
- **Soccer Clubs**
  - Use pipeline for social media automation, content creation

- **Sports Websites**
  - Use pipeline for generating automatic news tickers in portal
- **Web Crawlers**
  - Use summary from the pipeline for video and audio indexing and SEO

9

9

Literature Review – [1]
(Automated Soccer Video Production)

- [1]: Challenges of AI-based video production for soccer (2022)

  - State-of-the-art approaches are far from being adequate

  - Enhancement operations are needed after an event has been detected

- [6]: Motion entropy features for video segmentation (2008)

- [5]: Temporal Segment Net for video action recognition (2018)

- [7]: 3D CNN for action/event detection (2020)

10

10

Literature Review – [2]
(Natural Language Generation)

- [2]: Template-based data-to-text system for soccer (2018)

  - Multilingual Emotional Football Corpus (MeMo FC)

- [8]: Subjective summarization via social media interactions (2014)

- [9]: News articles for election results (2017)

- [10]: Grammatical rules around meteorology (2003)

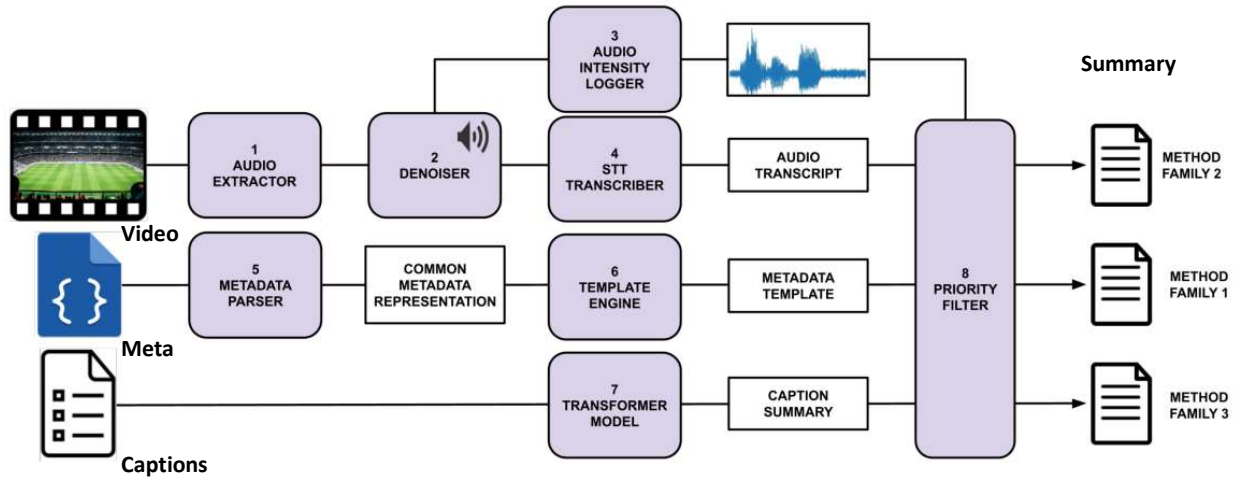# Literature Review – [3]
(Summarization)

- [11]: Clustering technique & hybrid TF-IDF for tweet summary (2011)

- [12]: Exemplary tweets using graph techniques (2012)

- [2]: multi-view multi-modal summarization (2022)

  - Uses image and texts on tweets

- [4]: Sports news from live commentary and knowledgebase (2022)

  - Transformer based approach

Methodology - [1]
(System Block Diagram)

13

Methodology - [2]
(Sample: Naive Template)

```
('action_corner', 'team'): "{d[team][value]} got corner.",
('action_end_of_game',): "Game ended.",
('action_end_phase', 'phase'): "{d[phase][value]} ended.",
('action_free_kick', 'offending_player'): "{d[offending_player][value]} got free kick.",
('action_free_kick', 'offending_player', 'team'): "{d[team][value]} was awarded free kick because of {d[offending_player][value]}.",
('action_free_kick', 'team'): "{d[team][value]} got free kick.",
('action_goal', 'after_set_piece', 'scorer', 'shot_type', 'team'): "{d[scorer][value]} scored a goal for {d[team][value]} with {d[shot_t
('action_goal', 'assist_by', 'scorer', 'shot_type', 'team'): "{d[scorer][value]} scored a goal for {d[team][value]} with assistance from
('action_goal', 'scorer', 'shot_type', 'team'): "{d[scorer][value]} from {d[team][value]} scored a goal.",
('action_offside', 'team'): "{d[team][value]} got offside.",
('action_penalty', 'team'): "{d[team][value]} got penalty.",
('action_red_card', 'player', 'team'): "{d[player][value]} from {d[team][value]} got red card.",
('action_shot', 'player', 'shot_type', 'team'): "{d[player][value]} from {d[team][value]} got {d[shot_type][value]} shot.",
('action_shot', 'shot_type', 'team'): "{d[team][value]} got {d[shot_type][value]} shot.",
('action_start_phase', 'phase'): "{d[phase][value]} started.",
('action_substitution', 'player_in', 'player_out', 'team'): "{d[player_in][value]} replaced {d[player_out][value]} in {d[team][value]}."
('action_yellow_card', 'player', 'team'): "{d[player][value]} from {d[team][value]} got red card.",
('action_yellow_card', 'team'): "{d[team][value]} got yellow card.",
```

14

Methodology - [3]
(Template Engine)
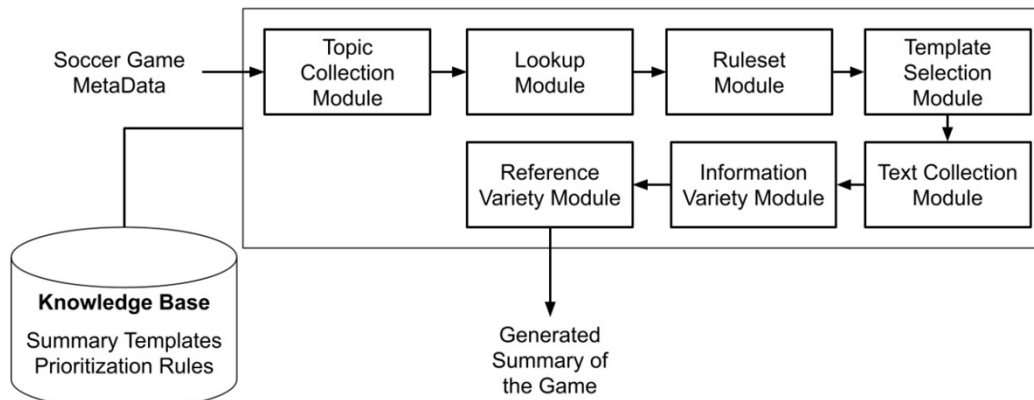
15

15

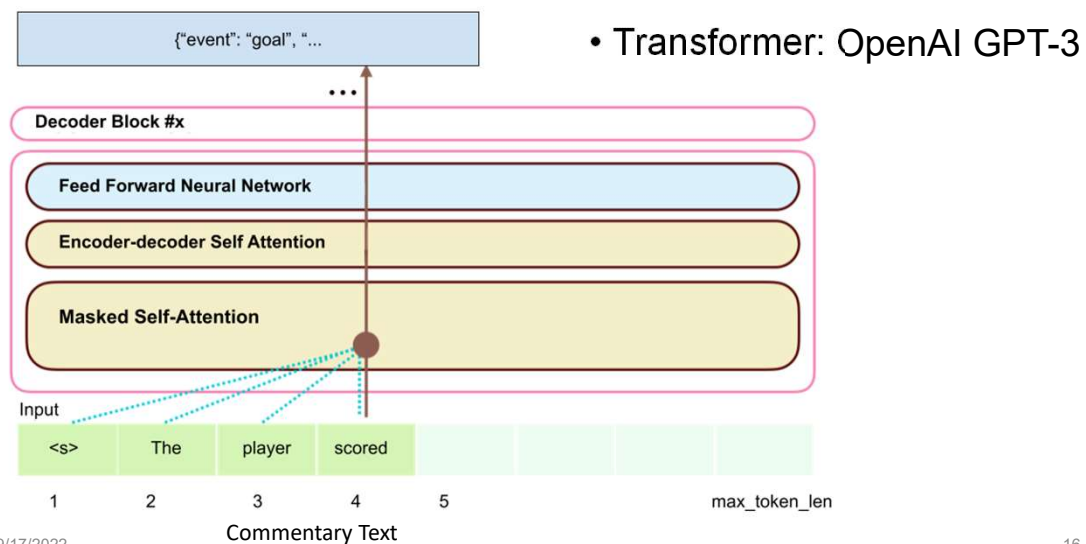# Methodology - [4]
(Meta Data Generation)

• Transformer: OpenAI GPT-3

16

16

8

Methodology - [5]
(Training Longformer)

1st half started. Aalesund was awarded free kick because of Shuaibu Lalle Ibrahim. Tonny Brochmann Christiansen from Mjøndalen hit standard shot. Jørgen Hatlehol from Aalesund hit standard shot. Mjøndalen was awarded free kick because of Vetle Fiskerstrand. Aalesund was awarded free kick because of Markus Lund Nakkim. Aalesund was awarded free kick because of Stian Semb Aasmundsen. Stian Semb Aasmundsen from Mjøndalen got red card. Aalesund was awarded free kick because of Tonny Brochmann Christiansen. Aalesund was awarded free kick because of Alexander Betten Hansen. Lars Olden Larsen from Mjøndalen got red card. Parfait Bizoza from Aalesund hit standard shot. Vetle Fiskerstrand from Aalesund hit standard shot. Aalesund got corner.
…
Game ended.

Longformer Model
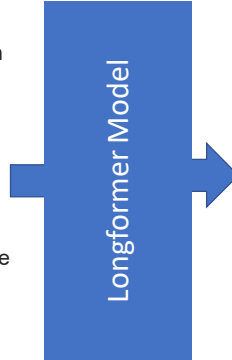
Aalesund and Mjøndalen played to a 2-2 draw. Aalesund took the lead in the first half through goals from Parfait Bizoza and Vetle Winger Dragsnes. Mjøndalen equalized in the second half through goals from Shuaibu Lalle Ibrahim and Sondre Liseth. The game was marred by two red cards, one for each team. Aalesund's Parfait Bizoza was sent off in the second half for a second yellow card, and Mjøndalen's Sigurd Hauso Haugen was sent off in the second half for a second yellow card.

<Template Generated From Meta Constraints>	Training Dataset	**<Expected Target Summary>**

9/17/2022   17

17

Methodology - [6]
(Longformer-LED: Architecture)

```
LEDModel(
 (shared): Embedding(50265, 768, padding_idx=1)
 (encoder): LEDEncoder(
  (embed_tokens): Embedding(50265, 768, padding_idx=1)
  (embed_positions): LEDLearnedPositionalEmbedding(16384, 768)
  (layers): ModuleList(
   (6x): LEDEncoderLayer(
    (self_attn): LEDEncoderAttention(
         (longformer_self_attn): LEDEncoderSelfAttention(
         (query): Linear(in=768, out=768, bias=True)
         (key): Linear(in=768, out=768, bias=True)
         (value): Linear(in=768, out=768, bias=True)
         (query_global): Linear(in=768, out=768, bias=True)
         (key_global): Linear(in=768, out=768, bias=True)
         (value_global): Linear(in=768, out=768, bias=True)
         )
     (output): Linear(in=768, out=768, bias=True) )
    (self_attn_layer_norm): LN((768,), eps=1e-05, eAf=True)
    (activation_fn): GELUActivation()
    (fc1): Linear(in=768, out=3072, bias=True)
    (fc2): Linear(in=3072, out=768, bias=True)
    (final_layer_norm): LN((768,), eps=1e-05, eAf=True)
   )
  )
  (LN_embedding): LN((768,), eps=1e-05, eAf=True)
 )
```

```
 (decoder): LEDDecoder(
  (embed_tokens): Embedding(50265, 768, padding_idx=1)
  (embed_positions): LEDLearnedPositionalEmbedding(1024, 768)
  (layers): ModuleList(

   (6x): LEDDecoderLayer(
    (self_attn): LEDDecoderAttention(
         (k_proj): Linear(in=768, out=768, bias=True)
         (v_proj): Linear(in=768, out=768, bias=True)
         (q_proj): Linear(in=768, out=768, bias=True)
         (out_proj): Linear(in=768, out=768, bias=True)
         )
    (activation_fn): GELUActivation()
    (self_attn_layer_norm): LN((768,), eps=1e-05, eAf=True)
    (encoder_attn): LEDDecoderAttention(
     (k_proj): Linear(in=768, out=768, bias=True)
     (v_proj): Linear(in=768, out=768, bias=True)
     (q_proj): Linear(in=768, out=768, bias=True)
     (out_proj): Linear(in=768, out=768, bias=True)
     )
    (encoder_attn_layer_norm): LN((768,), eps=1e-05, eAf=True)
    (fc1): Linear(in=768, out=3072, bias=True)
    (fc2): Linear(in=3072, out=768, bias=True)
    (final_layer_norm): LN((768,), eps=1e-05, eAf=True)
   ) )
  (LN_embedding): LN((768,), eps=1e-05, eAf=True)
 ))
```
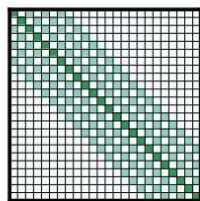
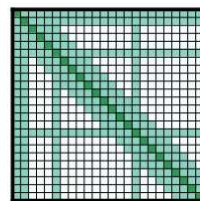9/17/2022   18

18

9

Methodology - [7]
(Longformer Attention)
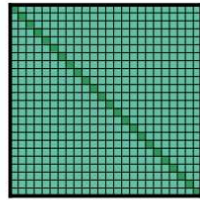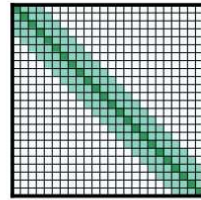
(a) Full n^2 attention   (b) Sliding window attention

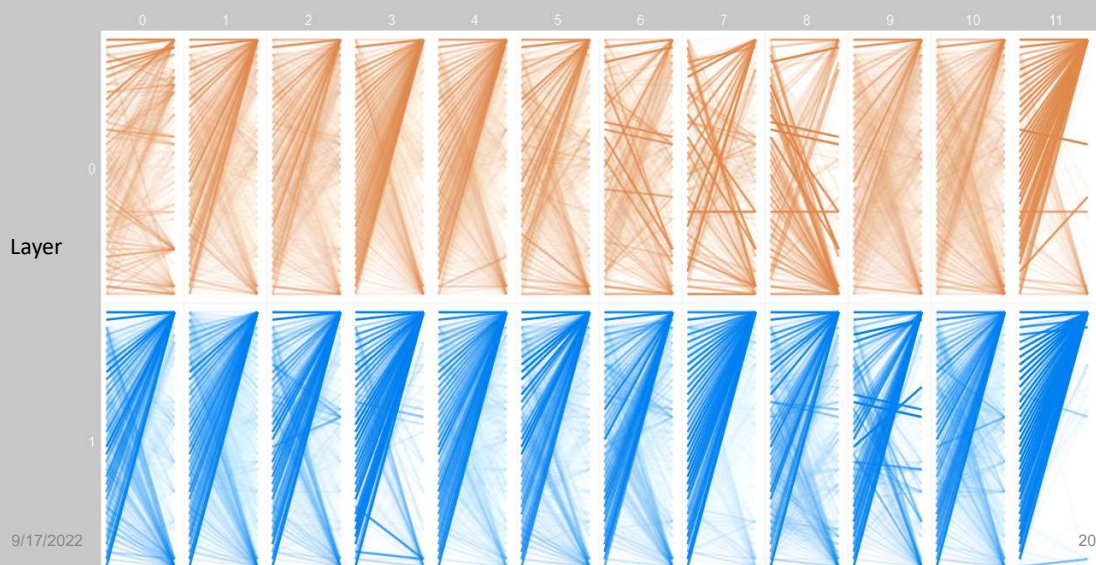c) Dilated sliding window   (d) Global+sliding window

19

19

Methodology - [8]
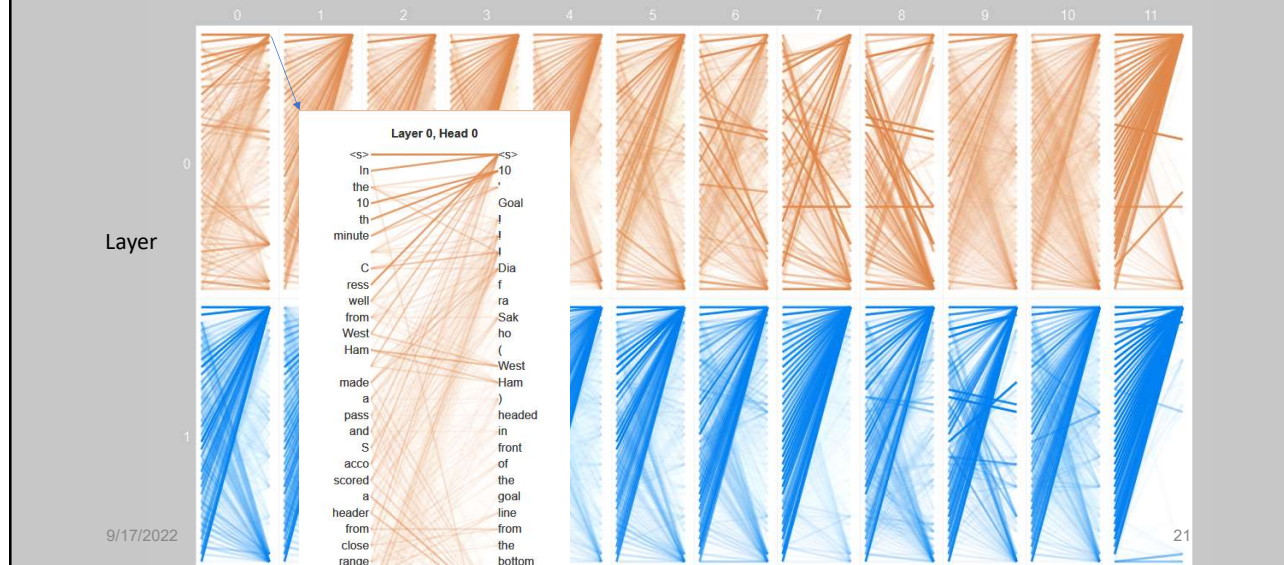## (Longformer Attention)

12 Attention Heads

Layer

20

20

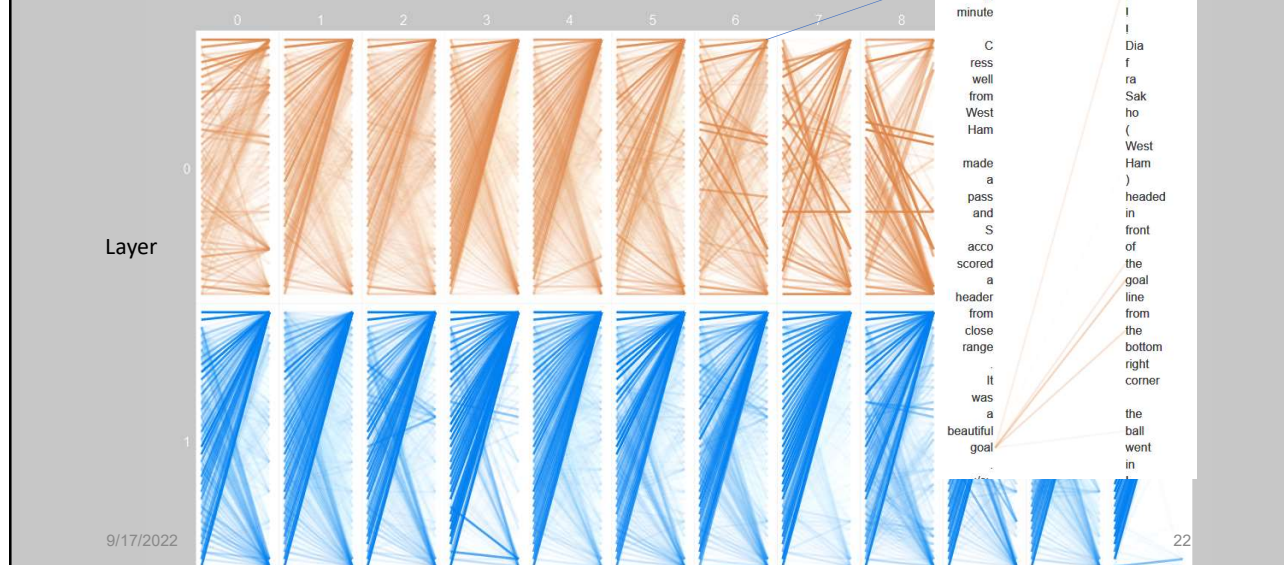Methodology - [9]
(Longformer Attention)
12 Attention Heads

21

Methodology - [10]
(Longformer Attention)
12 Attention Heads

22
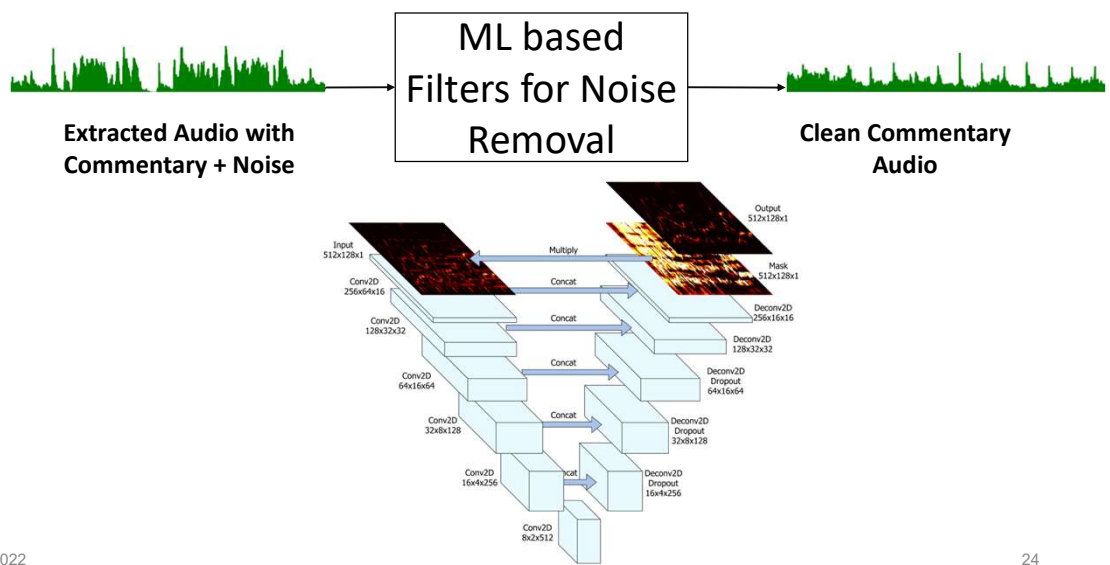

Methodology - [11]
(Length Control)

• Variance in the model output

• Solution:

  • Feed length requirement to the input

  • As natural language input to model

  • Eg: Prepend: "summarize to 512 : "

23

23

Methodology - [12]
(Noise Reduction)

Extracted Audio with Commentary + Noise → ML based Filters for Noise Removal → **Clean Commentary Audio**

24

24

Methodology - [13]
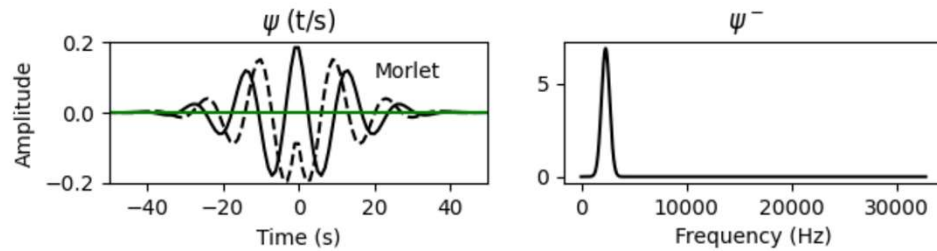## (Wavelet Analysis)

Mother Morlet Wavelet

25

Methodology - [14]
(Instrumentations)

- Software: FFMPEG, Spleeter, Transformers Python Library

- OpenAI API for finetuning the GPT-3 model

- A100 GPUs for Fine-tuning Longformer

- Django server for a web-based based GUI

- Server to handle task queue for STT jobs

- IBM Watson API for speech-to-text for commentary audio

26

Methodology - [15]
(Datasets)

```
{
    '<timestamp>',
    '{
        "team":
            {
                "id"    : <team-id>,
                "type"  : "team",
                "value" : "<team-name>"
            },
        "action": "<yellow/red> card",
        "player":
            {
                "id"    : <player-id>,
                "type"  : "player",
                "value" : "<player-name>"
            }
    }'
}
```

```
{
    "gameTime" : "<half Number>-<time from start of half>",
    "label"    : "<action type>",
    "position" : "<time in ms from beginning of game>",
    "team"     : "<home/away/not applicable>",
    "visibility" : "<visible/not shown>"
}
```
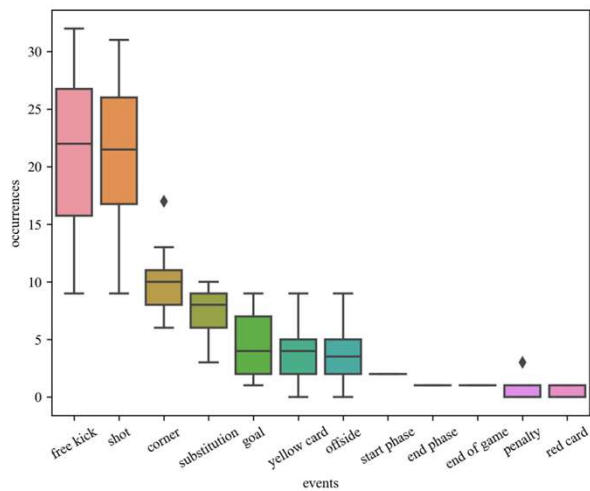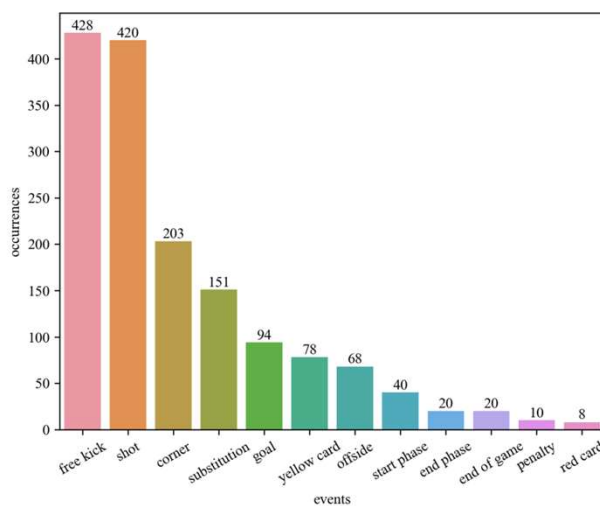
HOST Dataset
20+ Videos

SoccerNet Dataset
500 Videos

9/17/2022

27

27

Methodology - [18]
(Distribution of Actions in HOST Dataset)

9/17/2022

28

28

14

Methodology - [19]
(Datasets: SportsSum)

Caption: {
  "id": "109",
  "s1": "0",
  "s2": "3",
  "s": "",
  "t": "91'",
  "m": "The match ended with Hamburg 0-3 Bayern Munich, thank you for watching Sina Football Live."
 },

News: After the big rotation in the German Cup in midweek, Heynckes returned to the main line-up, Ribery returned, . . .

SportsSum Dataset: Scope and Length of news is large.
Original: 5000+ Commentary and News pairs in Chinese
Contribution: Translation of Commentary and News

29

Methodology - [20]
(Datasets: SportsSum)

Caption: {
  "id": "109",
  "s1": "0",
  "s2": "3",
  "s": "",
  "t": "91'",
  "m": "The match ended with Hamburg 0-3 Bayern Munich, thank you for watching Sina Football Live.",
  "meta": {
  "event": "GameEnded",
   "info": "Hamburg 0-3 Bayern Munich"
  }
 },

SportsSum Dataset: Scope and Length of news is large.
Original: 5000+ Commentary and News pairs in Chinese
Contribution: Meta Extraction using GPT3

30

Methodology - [21]
(Datasets: K-SportsSum)

```
[
  {
    "_id": "239",
    "commentary": [
        [
        "Bayer Leverkusen made a direct pass to the left side of the frontcourt!",
        "1st Half 7'",
        "0-0"
        ], . . . ...
                   ],
      "news": "After the first half of the game, Yatai turned the main back. In the first minute, Yatai
launched a quick attack from the right side. . .. "
```

SportsSum Dataset: Scope and Length of news is limited.
Original: 7000+ Commentary and News pairs in Chinese
Contribution: Translation of Commentary and News

31

Results – [1]
(Method Families)

ID	Input	Family	Denoising	Priority Filtering
1.1	M	Naive	✗	✗
1.2	M+A	metadata	✗	✓
1.3	M+A	template	✓	✓
2.1	A		✗	✗
2.2	A	Naive	✓	✗
2.3	A	STT	✗	✓
2.4	A		✓	✓
3.1	C		✗	✗
3.2	C+A	Transformer model	✗	✓
3.3	C+A		✓	✓

32

Results – [2]
(Pipeline Output from Method 1.2)

- Ball out of play happened in 3rd minutes. Corner from Swansea in 4th minutes. Shots on target from Swansea in 4th minutes. Ball out of play happened in 9th minutes. Ball out of play happened in 19th minutes. Clearance from Manchester United in 21st minutes. Goal from Manchester United in 27th minutes. Goal from Swansea in 29th minutes. Kick-off from Manchester United in 30th minutes. Ball out of play happened in 31st minutes. Ball out of play happened in 32nd minutes. Throw-in from Swansea in 34th minutes. Ball out of play happened in 34th minutes. Foul from Manchester United in 37th minutes. Offside from Swansea in 41st minutes. Yellow card from Manchester United in 44th minutes. Substitution from Manchester United in 0th minutes. Offside from Swansea in 2nd minutes. Yellow card from Manchester United in 5th minutes. Corner from Manchester United in 9th minutes. Substitution from Manchester United in 25th minutes. Ball out of play happened in 30th minutes. Offside from Manchester United in 35th minutes. Goal from Swansea in 54th minutes. Substitution from Swansea in 57th minutes. Substitution from Manchester United in 66th minutes. Ball out of play happened in 67th minutes. Yellow card from Manchester United in 77th minutes. Shots on target from Swansea in 80th minutes. Yellow card from Swansea in 84th minutes. Yellow card from Manchester United in 85th minutes. Foul from Swansea in 90th minutes. Yellow card from Swansea in 90th minutes. Substitution from Swansea in 95th minutes.

33

Results – [3]
(Pipeline Output from Method 2.4)

- In the area he loves today . And he is happening since this month yes united states has changed the since looking for a little bit of revenge for diamond shadow play before to remember any second support to strike you so talking auctions to full structural . Not so cheap three gigs . The smart finish . So we have to say . One minute's second this between . Last season a really successful time as well and wanted to take it but he feels that this is his home and gary monkey for lengthy discussions . These try to find party mean area . If so just caught the wrong side of his mom . S-delivery is trust . Quite readjustment is needed . Yesterday we want the feet in the last nineteen ames . Situation behind the solution . The past has been plenty of sign for manchester market they've never panicked in these situations . How it's john joselbe . Immediately apologises i think he tried to actually . On the banks that he may not get the goals . The swansea city now push everybody back all hands to the pump . Just to stop us to quite rise the correctly .

34

Results – [4]
## (Pipeline Output from Method 3.1)

- After the opening Jeonbuk Hyundai took the lead with its first shot. In the 12th minute Eninho made a cross from the right and Han Kyo-won headed the ball into the lower right corner from the front of the penalty area. In halftime stoppage time Jeonsang Hyundai took a right corner kick from the left and Eninho's left-footed shot was blocked by the opponent's defender. In half-time stoppage Je Jeonba Hyundai took another right corner and Jeonbaruk Hyundai's left foot shot from outside the opponent's penalty area was saved by the opposing goalkeeper. After the first half of the game was over the two sides changed sides and fought again and the situation remained the same. After halftime the game remained in a stalemate with the two teams fighting fiercely and neither was able to create any scoring opportunities. The two sides finally shook hands and fought successively with a 1-1 draw at home.

35

Results – [5]
(Training Loss for Method 3.1)
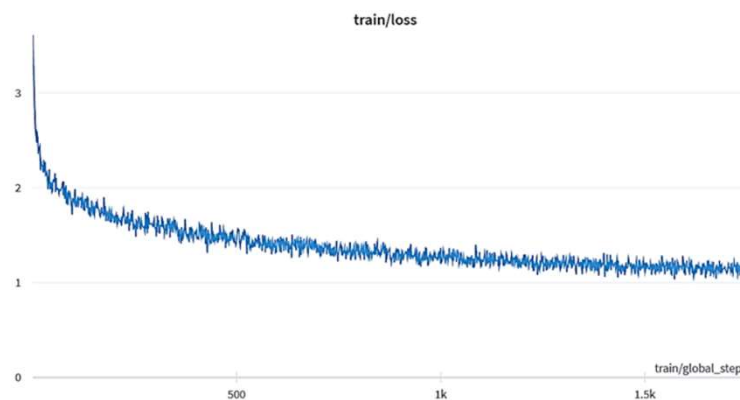
$$loss = -\sum_{k \in C} \{t_{ck} \log g_{ck} + (1 - t_{ck}) \log(1 - g_{ck})\}$$

36

Results – [6]
(Evaluation)

Method	Dataset	ROUGE-1	ROUGE-2	ROUGE-L
1.1	SoccerNet	0.08	0.00	0.08
1.2	SoccerNet	0.13	0.01	0.09
2.1	SoccerNet	0.26	0.06	0.10
2.4	SoccerNet	0.29	0.04	0.11
3.1	K-SportsSum	0.52	0.27	0.31

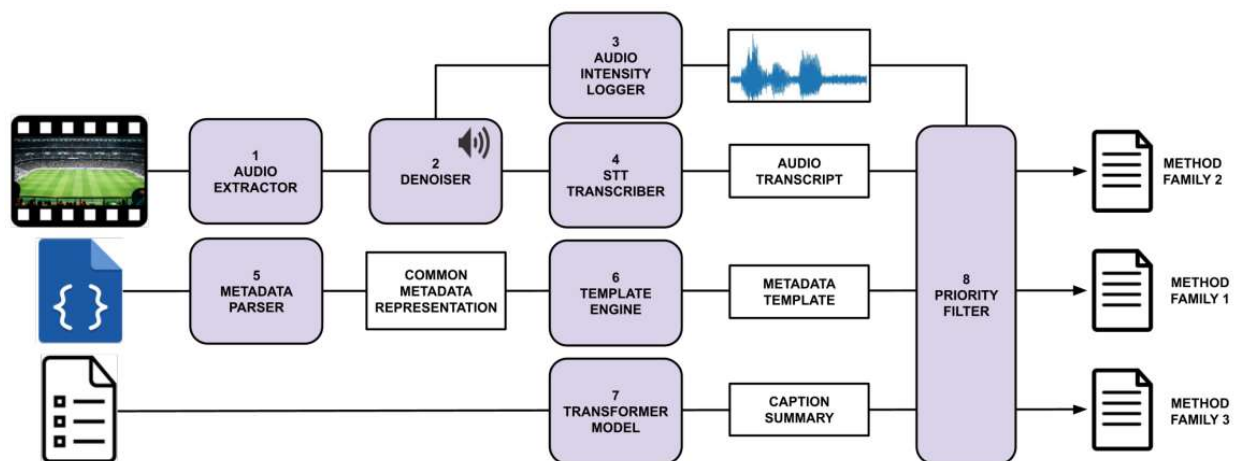**Preliminary results for methods 1.1, 1.2, 2.1, 2.4, and 3.1 in terms of ROUGE1, ROUGE-2, and ROUGE-L metrics.**

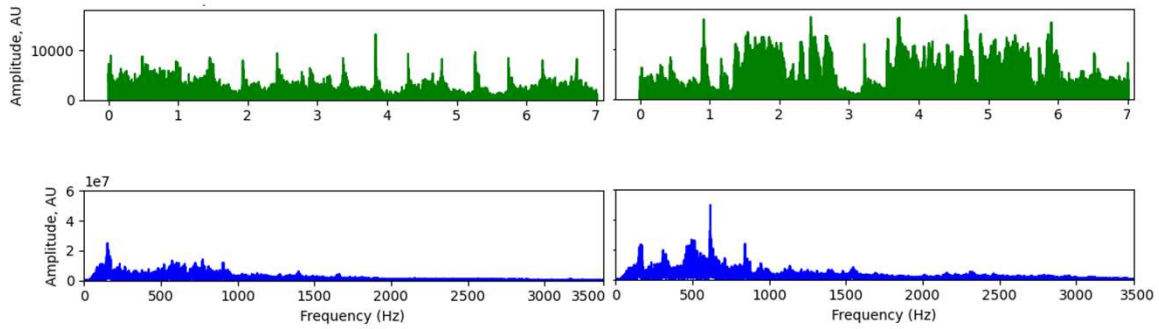9/17/2022   37

37

Results – [7]
(Framework)

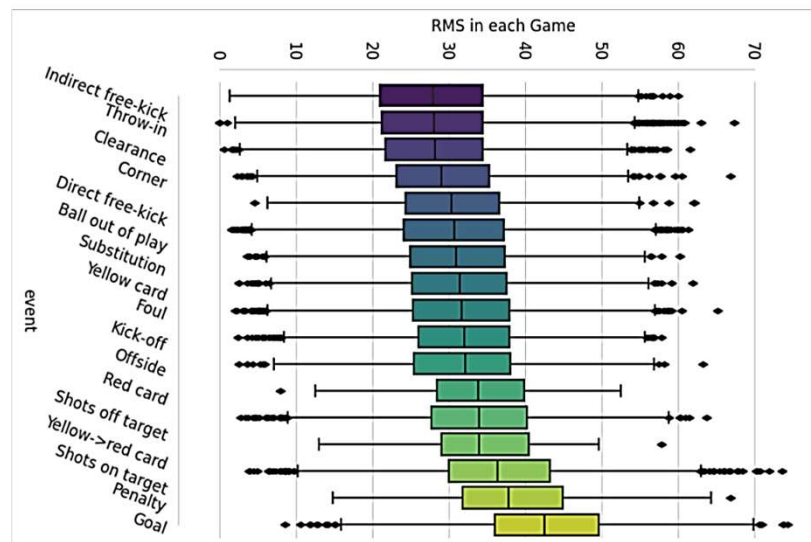9/17/2022   38

38

Results – [8]
(Noise-free STT)

Background Noise   **Commentary**

39

Results – [9]

Figure: The distribution of audio RMS of different types of events.

40

Results – [10]
(Wavelet Analysis)
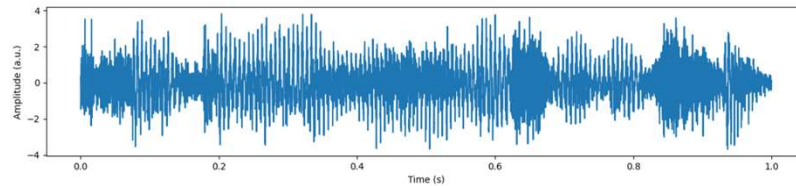

Figure: 1-second time audio series after a goal event
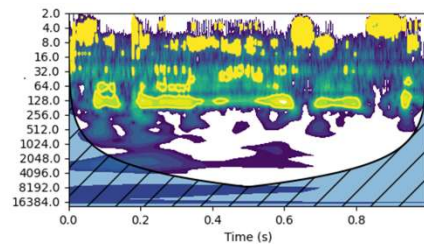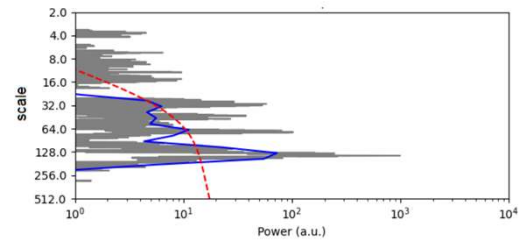

Figure: Wavelet Power Spectrum

Figure: Global Wavelet Spectrum
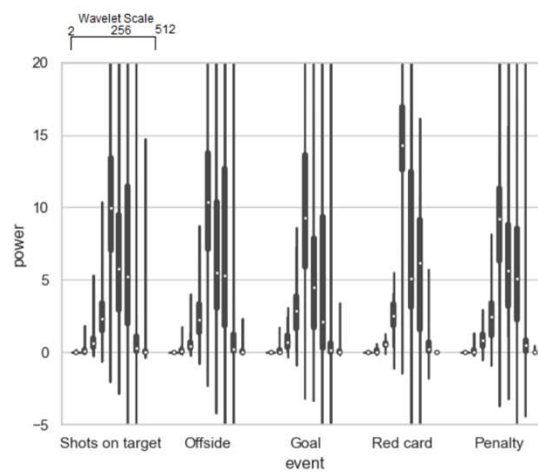
41

Results – [11]
(Wavelet Analysis)

Figure: Power distribution across Wavelet scales

42

Results – [12]
(Length Control)

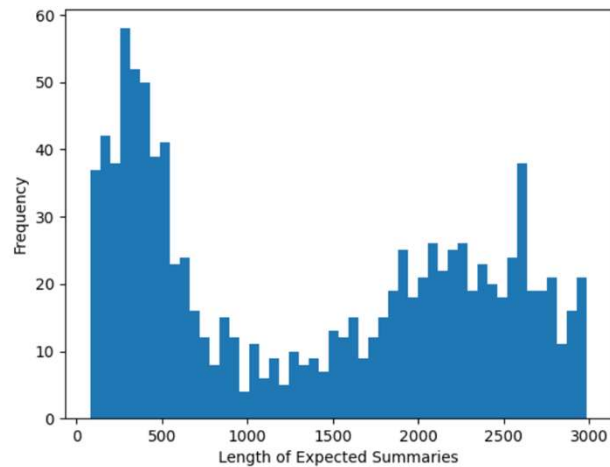Figure: Distribution of length of summary sampled from validation set.

9/17/2022   43

43

Results – [13]
(Length Control)

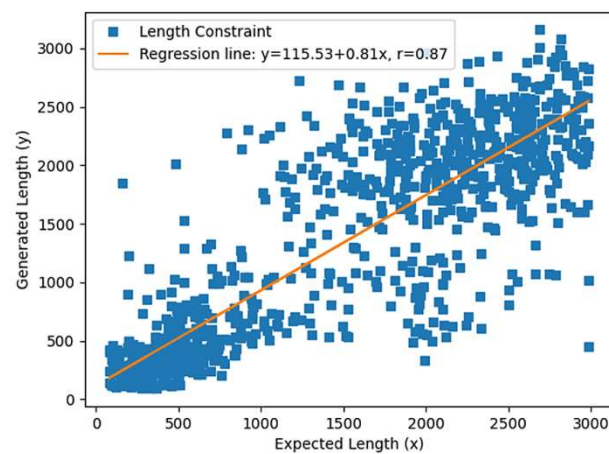Figure: Expected Vs. Generated Length of Summary

9/17/2022   44

44

Discussion and Analysis [1]

- Method 1.1:
  - **Input**: Meta-data
  - Generated form naive template engine
  - Complete information of game.
  - Very long: a sentence for each event in meta.
  - Sentence structures are very similar and boring for readers

45

45

Discussion and Analysis [2]

- Method 1.2:
  - **Input**: Meta-data and RMS audio level
  - **Priority filters** as per audio levels in the game,
  - Shorter summary and contains key events in the game
  - Sentence structures still very similar and boring to read

46

46

Discussion and Analysis [3]

- Method 2.1:

  - **Input**: Game Audio

  - Contains STT results on the whole audio

  - Extremely lengthy; difficult to put into context.

  - Problems due to:

    - Quality of audio in dataset

    - Accuracy of STT

    - Efficiency of denoising module

47

47

Discussion and Analysis [4]

- Method 2.4:

  - **Input**: Game Audio and RMS audio level

  - **Priority filters** as per audio levels in the game

  - Summary are shorter

  - Contents events with high RMS energy audio

  - Texts still very difficult to put into context

48

48

Discussion and Analysis [5]

- Method 3.1:

  - **Input**: Meta, Caption from Static Template Engine

  - Short summary

  - Contains the key events in the game.

  - Output structure like the training data

49

49

Discussion and Analysis [6]

- Longformer (Method 3) demonstrates the potential

  - Fixed-length maximum input, but very large:16K Tokens

  - ML based methods able to adapt

    - Based on output pattern in the dataset

    - Based on their training and configuration

50

50

Future Works

- Pipeline Improvements

- Effectiveness of STT, Priority Filter Evaluation

- More evaluation metrics

- Subjective human evaluation

- More data curation

9/17/2022   51

51

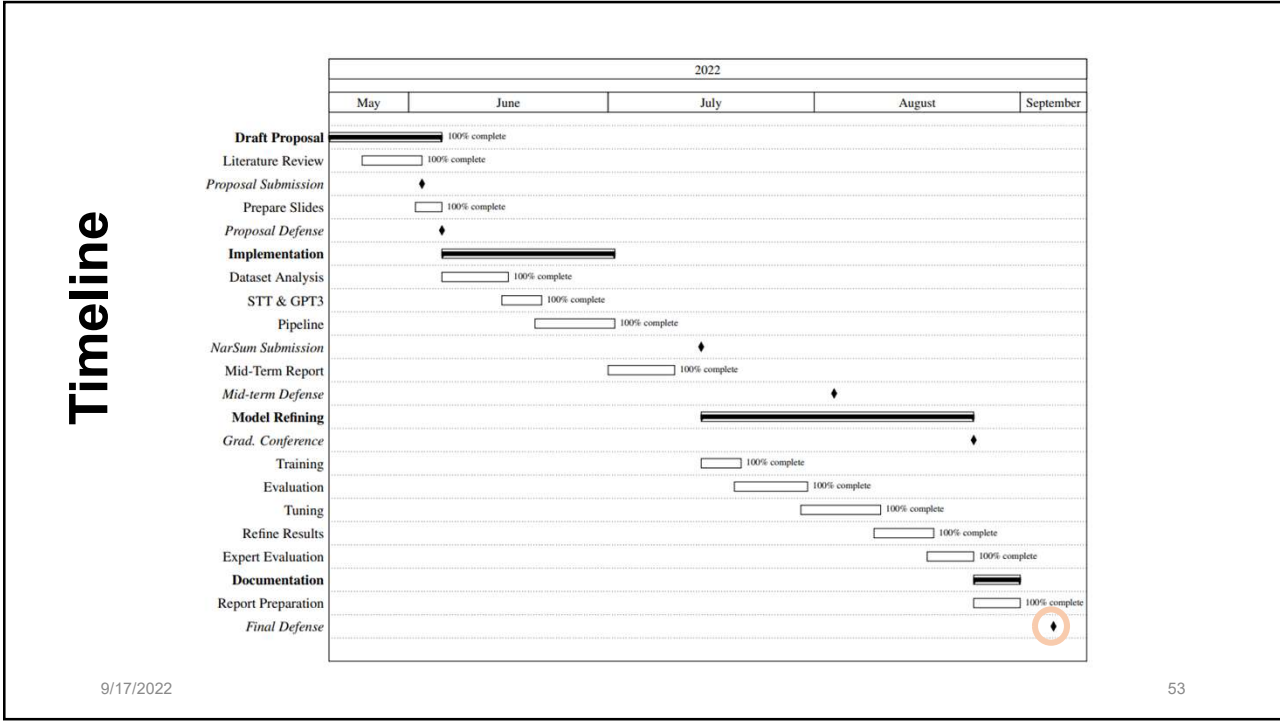# Conclusion

- Explored the domain of automatic game summarization
    - Particularly focusing on end-to-end method

- Utilize multi-model data as input to the system

- Explored the importance of audio intensity on event filtering

- Extended existing datasets from summarization viewpoint

9/17/2022   52

52

Timeline

53

References - [1]

[1]   Midoglu, C., Hicks, S. A., Thambawita, V., Kupka, T., & Halvorsen, P. (2022). MMSys'22 Grand Challenge on AI-based Video Production for Soccer. arXiv, 2202.01031. Retrieved from https://arxiv.org/abs/2202.01031v1

[2]   Aloufi, S., & Saddik, A. E. (2022). MMSUM Digital Twins: A Multi-view Multi-modality Summarization Framework for Sporting Events. ACM Trans. Multimedia Comput. Commun. Appl., 18(1), 1–25. doi: 10.1145/3462777

[3]   Van der Lee, C., Krahmer, E., & Wubben, S. (2017). PASS: A Dutch data-to-text system for soccer, targeted towards specific audiences. ACL Anthology, 95–104. doi: 10.18653/v1/W17-3513

[4]   Wang, J., Li, Z., Zhang, T., Zheng, D., Qu, J., Liu, A., ...Chen, Z. (2022). Knowledge Enhanced Sports Game Summarization. WSDM '22: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. Association for Computing Machinery. doi: 10.1145/3488560.3498405

54

References - [2]

[5]　　Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X., & Van Gool, L. (2018). Temporal Segment Networks for Action Recognition in Videos. IEEE Trans. Pattern Anal. Mach. Intell., 41(11), 2740–2755. doi: 10.1109/TPAMI.2018.2868668

[6]　　Chen, C.-Y., Wang, J.-C., Wang, J.-F., & Hu, Y.-H. (2008). Motion Entropy Feature and Its Applications to Event-Based Segmentation of Sports Video. EURASIP J. Adv. Signal Process., 2008(1), 1–8. doi: 10.1155/2008/460913

[7]　　Rongved, O. A. N., Hicks, S. A., Thambawita, V., Stensland, H. K., Zouganeli, E., Johansen, D., ...Halvorsen, P. (2020). Real-Time Detection of Events in Soccer Videos using 3D Convolutional Neural Networks. 2020 IEEE International Symposium on Multimedia (ISM). IEEE. doi: 10.1109/ISM.2020.00030

[8]　　Corney, D. P., Martin, C., & Göker, A. (2014, April). Two Sides to Every Story: Subjective Event Summarization of Sports Events using Twitter. In *SoMuS@ ICMR*.

References - [3]

[9]　　Leppänen, L., Munezero, M., Granroth-Wilding, M., & Toivonen, H. (2017, September). Data-driven news generation for automated journalism. In *Proceedings of the 10th International Conference on Natural Language Generation* (pp. 188-197).

[10]　　Sripada, S., Reiter, E., & Davy, I. (2003). SumTime-Mousam: Configurable marine weather forecast generator. *Expert Update*, *6*(3), 4-10.

[11]　　Inouye, D., & Kalita, J. K. (2011, October). Comparing twitter summarization algorithms for multiple post summaries. In *2011 IEEE Third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing* (pp. 298-306). IEEE.

[12]　　Nichols, J., Mahmud, J., & Drews, C. (2012, February). Summarizing sporting events using twitter. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces* (pp. 189-198)

Thank You
For Your Attention

1/7/2022

57

57